Title
From Transposable Elements to DNA Methylation: The Role of Genome Regulation in the Evolution of the Jellyfish Medusa

Author
Zhang, Xinhui

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

From Transposable Elements to DNA Methylation: The Role of Genome Regulation in the Evolution of the Jellyfish Medusa

A dissertation submitted in partial satisfaction
of the requirements for the degree Doctor of Philosophy
in Biology

by

Xinhui Zhang

2022

ABSTRACT OF THE DISSERTATION

From Transposable Elements to DNA Methylation:
The Role of Genome Regulation
in the Evolution of the Jellyfish Medusa

by

Xinhui Zhang
Doctor of Philosophy in Biology
University of California, Los Angeles, 2022
Professor David K. Jacobs, Chair

One phenomenon that has perplexed biologists for decades is the lack of correlation between genome size and organismal complexity, aka the 'C-value paradox'. It is now generally accepted that the paradox is caused by repetitive, non-coding sequences such as transposable elements. These 'selfish' or 'parasitic' sequences have profound impacts on the host genome. For instance, DNA methylation is thought to have originated as a defense mechanism against these genome parasites. As one of the most conserved epigenetic modifications in eukaryotes, DNA methylation plays important regulatory roles in animals. However, the evolution of transposable elements and DNA methylation have rarely been studied in non-bilaterian animals. This dissertation explores genome

regulation in Cnidaria, the sister group of Bilateria, and its potential role in the evolution of complex life histories.

Chapter 1 studies the distribution of DNA methylation across Cnidaria using published genomes and transcriptomes of over 70 species. By analyzing a proxy of DNA methylation, I show this epigenetic modification is prevalent in Cnidaria with instances of loss in a group of endoparasitic species. I also show cnidarian DNA methylation shares many similarities with bilaterian invertebrates, supporting the hypothesis that gene body methylation is the ancestral state in Eumetazoa. Additionally, cnidarians with complex life cycles tend to have heavier methylation, and both gene body methylation and repeat methylation increase with genome sizes.

Chapter 2 examines DNA methylation patterns across the life history of a scyphozoan jellyfish, *Aurelia coerulea*, via stage-specific whole-genome bisulfite sequencing. This work characterizes the distribution of DNA methylation on the *Aurelia* genome, the correlation between methylation and expression levels, and highlights that sparsely-methylated genes tend to be expressed dynamically. Moreover, many genes are differentially methylated across life history stages; however, methylation changes do not predict expression changes. This study indicates that the regulatory role of gene body methylation lies more in stabilizing expression, a notion emerging as the primary function of DNA methylation in animals.

Chapter 3 studies transposable elements (TEs) in the *Aurelia* genus. TEs in 12 species are annotated and quantified using whole genome sequencing. One clade shows signatures of ancient TE expansion events, indicating the TEs have undergone different evolutionary histories. Based on the phylogeny within the genus, these expansion events predate the closure of the isthmus of panama which separated the two clades of *Aurelia*.

Additionally, the genome size difference between *Aurelia aurita* and *Aurelia coerulea* is unlikely due to activities of known TEs; however, *Aurelia* genomes harbor many unique TEs that would be worth further exploring.

Overall, this dissertation research shows that Cnidaria is one of the most fascinating systems to study the evolution of genome regulation. As the first animals with real tissue layers, neuro-sensory systems, and a wide variety of different life histories, Cnidaria can shed light on how the genome 'dark matter' plays a role in the evolution of animal complexity.

The dissertation of Xinhui Zhang is approved.

Allen G. Collins

Volker Hartenstein

Jingyi Li

Matteo Pellegrini

David K. Jacobs, Committee Chair

University of California, Los Angeles

2022

To my mom Liu Limin and dad Zhang Juncheng.

They taught me to have faith that there is a solution to every problem,

and to have a sense of humor when there isn't.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

LIST OF TABLES

CHAPTER 1

CHAPTER 3

ACKNOWLEDGMENTS

I cannot imagine finishing this journey without all the amazing people in my life. I am immensely grateful for their presence, and only wish that I will pay it forward by helping others in the future.

I would like to thank my advisor, David Jacobs, for his guidance and support throughout the years. I hope I have learned well his scientific rigor and integrity, attention to details, hardworking spirits, and the pure drive to do good science and good work. I am also grateful that he has indulged my curiosity, and always encouraged me to just try when we did not know if things would work out. Even though I am leaving academia, I will cherish and carry on these qualities that I picked up from him.

I want to thank my lab mates, Benjamin Hà, Hayden Speck, and Rachel Turba. Not only do I admire their brilliant minds, but also I am grateful that they always believed in me when I doubted myself, and they were always able to make the most bleak days more bearable. They always went above and beyond to be helpful even when it must have been hard to find the mental capacity. It is a privilege to have worked with them.

I also want to thank Dr. Michelle Rensel, Dr. Mitra Hooshmand, Dr. Erin Brown, and Dr. Katie Dixie for their mentoring. As the inspirational and kind women academics that they are, they helped me navigate when I was most lost. I also thank Sorel Fitz-Gibbon, Clément Goubert, Fides Lay, and Marco Morselli for graciously teaching me bioinformatics and sequencing.

I feel very lucky that I have met Diane Gu, Yuqing Hou, and Richard Mendez, for they are such caring and fun human beings, and not biologists. They each gives me new perspectives and helps me become a more well-rounded person. My scuba dive buddies,

Ashely Arnold, Ben Callahan, Rachel Fang, Tomas Geurerro, Norbert Lee, Xia Li, Abe Rodrigues, and many others, offered so many fun moments and helped me stay sane during the craziest times.

Additionally, I am thankful for Lana, the best kitty in the world. She was born in the same year that I started graduate school, so I would like to think that she and I have been growing together.

I would not have been where I am now if it were not for my parents, who did not want me to do a PhD because they wanted me to be happy. I obviously did not listen. Nevertheless, they are always supportive of my dreams and did everything they could to help me become the best version of myself. There are not enough words in the world for me to express my love and gratitude for them.

VITA

EDUCATION

2015 B.S., Biology
The Ohio State University
Columbus, OH

PUBLICATIONS

Robin AN, Denton KK, Horna Lowell ES, Dulay T, Ebrahimi S, Johnson GC, Mai D, O'Fallon S, Philson CS, Speck HP, **Zhang X**, Nonacs P. Major Evolutionary Transitions and the Roles of Facilitation and Information in Ecosystem Transformations. *Frontiers in Ecology and Evolution*. 2021:866.

Zhang X, Jacobs D. A Broad Survey of Gene Body and Repeat Methylation in Cnidaria Reveals a Complex Evolutionary History. *Genome Biology and Evolution*. 2022 Feb;14(2):evab284.

SELECTED AWARDS AND HONORS

2014 Summer Research Fellowship
The Ohio State University

2016 Sigma Xi Grants-in-Aid of Research
UCLA

2020 EEB Annual Research Symposium 3rd Place in Graduate Presentations
UCLA

2021 Alan Alda Center for Communicating Science Training Award
UCLA

GBE

A Broad Survey of Gene Body and Repeat Methylation in Cnidaria Reveals a Complex Evolutionary History

Xinhui Zhang and David Jacobs *

Department of Ecology and Evolutionary Biology, University of California, Los Angeles, USA

*Corresponding author: E-mail: djacobs@ucla.edu.

Accepted: 18 December 2021

Abstract

DNA methylation, an important component of eukaryotic epigenetics, varies in pattern and function across Metazoa. Notably, bilaterian vertebrates and invertebrates differ dramatically in gene body methylation (GbM). Using the frequency of cytosine-phospho-guanines (CpGs), which are lost through mutation when methylated, we report the first broad survey of DNA methylation in Cnidaria, the ancient sister group to Bilateria. We find that: 1) GbM differentially relates to expression categories as it does in most bilaterian invertebrates, but distributions of GbM are less discretely bimodal. 2) Cnidarians generally have lower CpG frequencies on gene bodies than bilaterian invertebrates potentially suggesting a compensatory mechanism to replace CpG lost to mutation in Bilateria that is lacking in Cnidaria. 3) GbM patterns show some consistency within taxonomic groups such as the Scleractinian corals; however, GbM patterns variation across a range of taxonomic ranks in Cnidaria suggests active evolutionary change in GbM within Cnidaria. 4) Some but not all GbM variation is associated with life history change and genome expansion, whereas GbM loss is evident in endoparasitic cnidarians. 5) Cnidarian repetitive elements are less methylated than gene bodies, and methylation of both correlate with genome repeat content. 6) These observations reinforce claims that GbM evolved in stem Metazoa. Thus, this work supports overlap between DNA methylation processes in Cnidaria and Bilateria, provides a framework to compare methylation within and between Cnidaria and Bilateria, and demonstrates the previously unknown rapid evolution of cnidarian methylation.

Key words: Cnidaria, DNA methylation, $CpG_{o/e}$, parasitism, gene bodies, repetitive elements.

Significance

DNA methylation shows dramatically different patterns in bilaterian vertebrates and invertebrates, yet the origin of these differences is unclear. We present the first comprehensive survey of DNA methylation in Cnidaria, the sister group to Bilateria, and show that gene body methylation is likely the ancestral state for Eumetazoa and that several factors likely contribute to the complex evolutionary history of DNA methylation in Cnidaria. This work shows that DNA methylation is highly conserved in Cnidaria, and provides an important piece of the puzzle of the evolution of DNA methylation in animals.

Introduction

In Metazoa, cytosines in cytosine-phospho-guanine dinucleotides (CpG sites) are the predominant target of methylation. However, the overall level and pattern of such methylation varies greatly across Metazoa. Methylated cytosine is hypermutable converting to thymine at a high rate. In this work we use the resulting depletion of CpG sites as a proxy for DNA methylation, examining 76 species across Cnidaria. Although bisulfite sequencing remains the gold standard in DNA methylation research, it would require high-cost de novo sequencing in these 76 taxa to achieve this wide range of sampling (Dimond and Roberts 2016; Aliaga et al. 2019). Thus, our application of this proxy provides a critical avenue forward for evolutionary study, as methylation sequencing outside of vertebrates is limited and CpG depletion can be measured using genomic and transcriptomic data accumulated for other purposes.

Cnidaria, the sister group of Bilateria, contains a suite of ancient taxa with substantial variation in life history, symbiosis, coloniality, and parasitism. Thus, Cnidaria provides an interesting system to understand the evolution of DNA methylation across the Metazoa. For the rest of the introduction, we will review: 1) the patterns of DNA methylation in different Metazoan taxa, 2) the two main gene families involved in DNA methylation machinery in Metazoa—DNA methyltransferases (DNMTs) and methyl-CpG-binding domain proteins (MBDs), and 3) the measurement of CpG depletion.

CpG methylation on the gene bodies (transcription units) and repetitive elements vary greatly across Metazoa. In most bilaterian invertebrates examined, methylation is more concentrated at gene bodies than transposable elements (TEs) (Suzuki et al. 2007; Feng et al. 2010; Zemach et al. 2010) (fig. 1). Vertebrates, on the other hand, consistently methylate their genomes globally regardless of TE content (Zemach et al. 2010). Although defense against genome parasites has been proposed as the ancestral function of DNA methylation in eukaryotes (Bestor 1990), the varied patterns observed in vertebrates and bilaterian invertebrate suggest a complicated evolutionary history in Metazoa, and Sarda et al. (2012) among others argued that gene body methylation (GbM) was the ancestral DNA methylation pattern in animals. A recent study on sponge methylation showed heavy methylation on TEs as well as gene bodies, comparable to vertebrates, suggesting either convergence or multiple loss of this global methylation pattern (de Mendoza et al. 2019).

In bilaterian invertebrates, methylated and unmethylated gene bodies are typically found at approximately similar frequency, a pattern usually described as "bimodal." Notably, genes with different modes of methylation status have been found to differ in functional enrichment, expression levels, and plasticity. Genes with high GbM tend to serve housekeeping functions with high and stable expression. Those with low GbM tend to function in stress response and developmental regulation, and exhibit relatively low, more plastic and more tissue or condition specific expression (Suzuki et al. 2007; Zemach et al. 2010; Glastad et al. 2014; Dimond and Roberts 2016).

GbM is functionally important as exemplified by studies in the honey bee (Kucharski et al. 2008), the Pacific oyster (Riviere et al. 2013), and the moon jellyfish (Fuchs et al. 2014). GbM is thought to repress spurious transcription initiation, which could allow for more efficient transcription elongation. This inference is consistent with the correlation between GbM and expression level and plasticity noted above. More direct evidence comes from a study in the anemone *Aiptasia* where genes with heavy GbM showed less intragenic promoter activity (Li et al. 2018). Aside from repressing intragenic promoters, Flores et al. (2012) proposed that GbM regulates alternative splicing. All these observations suggest the functional importance of GbM in certain circumstances, but confirmation of specific functions, are few and the generality of these functions across taxa is not well understood.

Interestingly, a number of Metazoan species, including the nematode *Caenorhabditis elegans*, and the fruit fly *Drosophila melanogaster* lack cytosine methylation (Wenzel et al. 2011; Takayama et al. 2014). Loss of cytosine methylation potentially relates to particular developmental modes, such as mosaic development where cell fates are determined early on, or shortened life cycles associated with transient larval resources (Bestor 1990; Canestro et al. 2007). Such methylation loss often correlates with drastic genome compaction such as occurs in parasitic animals. Understanding the context and consequences of these losses is important for understanding the evolution of global methylation patterns.

DNMTs, a shared attribute of crown group eukaryotes, mediate methylation of CpG sites, among which DNMT1and 3 are specific to Metazoa (Zemach et al. 2010). DNMT1 maintains DNA methylation and restores symmetric methylation of hemimethylated CpGs after DNA duplication, whereas DNMT3 methylates previously unmethylated CpG sites (Holz-Schietinger et al. 2011). DNMT1 and DNMT3 both have undergone frequent duplications and losses across Metazoa, and the loss of DNMTs typically concur with loss of methylation (Tweedie et al. 1997; Field et al. 2004; Goll et al. 2006; Yi and Goodisman 2009; Wenzel et al. 2011).

Although DNMTs are the writers of DNA methylation, this information is read by MBDs. Invertebrate genomes commonly possess only one gene in the MBD family, MBD2/3, whereas the gene family has expanded in mammals (Hendrich and Tweedie 2003; Cramer et al. 2017). This expansion is evolutionarily coincident with the increase of global methylation documented more broadly in the vertebrates (Hendrich and Tweedie 2003).

Methylated cytosines spontaneously deaminate to uracils more readily than unmethylated cytosines (Duncan and Miller 1980). Due to this hypermutability, sequences that are historically highly methylated in the germline over evolutionary time become depleted in CpG. CpG depletion has been shown to correlate with direct measures of somatic DNA methylation using methods such as bisulfite sequencing, making it a reliable first order approximation of DNA methylation (Bock and Lengauer 2008; Elango and Yi 2008; Xiang et al. 2010; Park et al. 2011; Sarda et al. 2012; Keller et al. 2016). Thus, depletion of CpG in a genomic region, often calculated as the ratio of observed to expected CpG ($CpG_{o/e}$) as discussed in Materials and Methods, provides a proxy metric for methylation in the absence of bisulfite sequencing data. Consequently, a number of studies have used CpG depletion to understand phylogenetic aspects of DNA methylation in animals, yet application of this approach outside of Bilateria is minimal (Yi and Goodisman 2009; Sarda et al. 2012; Dixon et al. 2014; Aliaga et al. 2019).

Cnidaria, the sister group to Bilateria, encompasses species with diverse morphology, life history, coloniality, and

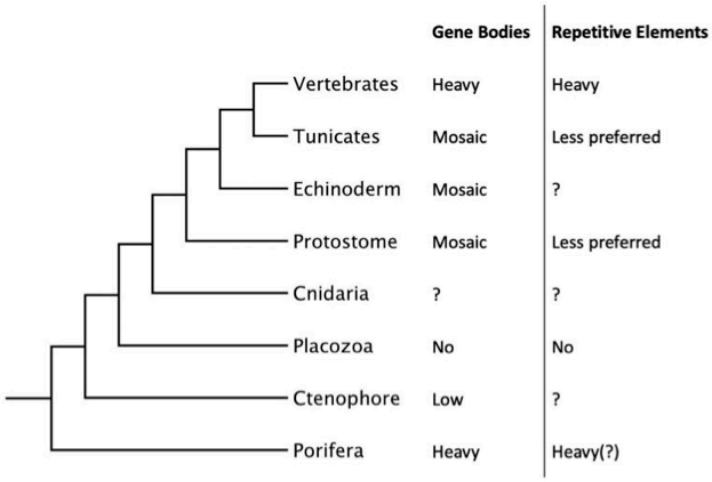


Fig. 1.—Methylation on the gene bodies and repetitive elements vary greatly across the tree of animals (Gavery and Roberts 2010; Hunt et al. 2010; Zemach et al. 2010; Glastad et al. 2011; Nanty et al. 2011; Sarda et al. 2012; Falckenhayn et al. 2013; Dabe et al. 2015).

skeletonization; while feeding modes include parasitism, predation, and photosymbiosis. Therefore, not only are Cnidaria interesting in comparison to Bilateria, they also merit comparative investigation among themselves due to this diversity, antiquity, and as an ancient parallel evolutionary radiation to Bilateria. Yet, previous studies on Cnidarian DNA methylation have been limited to the class Anthozoa including *Nematostella vectensis, Aiptasia*, and several coral species, (Zemach et al. 2010; Dixon et al. 2014, 2016; Dimond and Roberts 2016; Li et al. 2018), and the parasitic clade Myxozoa (Kyger et al. 2020). We survey the extent and distribution of DNA methylation in Cnidaria, to better understand how the variation of methylation relates to their diverse biology, and to inform reconstruction of the state of DNA methylation in the last common ancestor of Eumetazoa.

Results

Forty-one of the 76 species examined in this study show an average gene body $CpG_{o/e}$ lower than 0.75 (table 1) consistent with substantial methylation (Aliaga et al. 2019), suggesting prevalent GbM in Cnidaria. The average $CpG_{o/e}$ of gene bodies varies greatly across the phylum, spanning from 0.58 in the jellyfish *Cassiopea xamachana* to 1.04 in the parasite *Myxobolus cerebralis* (table 1 and fig. 2). Only 11 of the 41 species with substantial gene body CpG depletion showed discrete bimodal distribution patterns of $CpG_{o/e}$ based on analysis by Notos, a kernel density estimation (KDE) method, in contrast to most bilaterian invertebrates (fig. 3 and supplementary fig. S1, Supplementary Material online; Bulla et al. 2018). Additionally, we assessed the distribution patterns using Gaussian mixture models, which classified most species as multimodal due to the complex nature of the data. We will address these differences in results further in Discussion. Clustering of frequencies of $CpG_{o/e}$ distribution across all gene bodies revealed differing patterns of methylation (fig. 4). Many, but not all, clades show similar patterns, including Alcyonacea (soft corals), Scleractinia (hard corals), Myxozoa (parasites), and Scyphozoa (true jellies) and Cubozoa (box jellies).

DNMT1, DNMT3, and MBDs were found in most species examined, with six exceptions. None of the DNMTs and MBDs was recovered in the myxozoan parasites *Myxobolus cerebralis, Myxobolus squamalis, Henneguya salminicola, and Thelohanellus kitauei*. DNMTs were also not found in the anthozoan *Eunicella verrucosa*, and MBDs were not found in the hydrozoan *Aegina citrea* (table 1 and fig. 2). All four myxozoan species mentioned above show little to no sign of CpG depletion on the gene bodies, indicating a lack of methylation consistent with absence of the enzymes. DNMTs were initially found in *Myxobolus pendula*, but our phylogenetic analysis suggests they are more closely related to DNMTs found in the zebrafish than the other Cnidarians sampled here, suggesting the transcripts most likely originated from the fish host (supplementary fig. S2, Supplementary Material online). This possibility of contamination could also explain the different pattern of gene body $CpG_{o/e}$ distribution in *Myxobolus pendula* than the other myxozoans (figs. 2 and 4). *Eunicella verrucosa* and *Aegina citrea* also show elevated levels of gene body $CpG_{o/e}$. However, several additional species also have high gene body $CpG_{o/e}$ despite the presence in their genomes of DNMTs and MBDs. For instance, the hydrozoan *Ectopleura larynx* provides an example of implicit limited methylation

Table 1
Average Gene Body $CpG_{o/e}$ and Repeat $CpG_{o/e}$ of Species in This Study

Class	Ordinal Group	Species	Mean Gene Body $CpG_{o/e}$	Mean Repeat $CpG_{o/e}$
Anthozoa	Actiniaria ③	*Actinia tenebrosa*	0.69	0.90
		Aiptasia pallida	0.69	0.98
		Anthopleura elegantissima	0.68	
		Edwardsiella lineata	0.758	
		Nematostella vectensis	0.77	
	Alcyonacea ①	*Acanthogorgia aspera*	0.83	
		Briareum asbestinum	0.79	
		Clavularia sp.	0.82	
		Corallium rubrum	0.84	
		Dendronephthya gigantea	0.84	1.03
		Eleutherobia rubra	0.88	
		Eunicella cavolinii	0.85	
		Eunicella verrucosa[a]	0.84	
		Gorgonia ventalina	0.75	
		Leptogorgia sarmentosa	0.85	
		Xenia sp.	0.79	
	Corallimorpharia ④	*Corynactis australis*	0.81	
		Rhodactis indosinensis	0.71	
		Ricordea yuma	0.72	
	Helioporacea ②	*Heliopora coerulea*	0.66	
	Scleractinia ⑤	*Acropora digitefera*	0.70	0.90
		Acropora millepora	0.72	0.91
		Ctenactis echinata	0.64	
		Favia lizardensis	0.66	
		Lobactis scutaria	0.67	
		Madracis auretenra	0.75	
		Montastraea cavernosa	0.68	
		Orbicella faveolata	0.73	0.93
		Platygyra carnosus	0.72	
		Pocillopora damicornis	0.70	0.87
		Seriatopora hystrix	0.68	
		Stylophora pistillata	0.70	0.92
Hydrozoa	Aplanulata ⑯	*Ectopleura larynx*	0.97	
		Hydra oligactis	0.75	
		Hydra viridissima	0.64	
		Hydra vulgaris	0.63	1.03
	Capitata ⑳	*Porpita porpita*	0.67	
		Velella velella	0.80	
	Filifera III ⑰	*Hydractinia polyclina*	0.82	
		Hydractinia symbiolongicarpus	0.81	
		Podocoryna carnea	0.63	
	Filifera IV ⑳	*Turritopsis* sp.	0.83	
	Leptothecata ㉑	Clytia hemisphaerica	0.81	
		Dynamena pumila	0.83	
	Limnomedusae ⑭	*Craspedacusta sowerbyi*	0.90	
	Narcomedusae ⑮	*Aegina citrea*[b]	0.85	
	Siphonophorae ⑧	*Abylopsis tetragona*	0.72	
		Agalma elegans	0.76	
		Craseoa lathetica	0.74	
		Nanomia bijuga	0.83	
		Physalia physalis	0.76	

(continued)

Table 1 Continued

Class	Ordinal Group	Species	Mean Gene Body $CpG_{o/e}$	Mean Repeat $CpG_{o/e}$
Cubozoa	Carybdeida ⑨	*Alatina alata*	0.75	
		Copula sivickisi	0.71	
		Morbakka virulenta	0.65	0.66
		Tripedalia cystophora	0.66	
	Chirodropoda ⑩	*Chironex fleckeri*	0.67	
		Chironex yamaguchii	0.65	
Scyphozoa	Coronatae ⑪	*Atolla vanhoeffeni*	0.71	
	Rhizostomeae ⑬	*Cassiopea xamachana*	0.58	0.65
		Nemopilema nomurai	0.69	
		Rhopilema esculentum	0.66	
		Stomolophus meleagris	0.70	
	Semaeostomeae ⑮	*Aurelia aurita*	0.69	
		Aurelia coerulia	0.649	0.74
		Chrysaora fuscescens	0.67	
Staurozoa	Stauromedusae ⑧	*Calvadosia cruxmelitensis*	0.78	
		Craterolophus convolvulus	0.83	
		Haliclystus auricula	0.74	
		Haliclystus sanjuanensis	0.73	
		Lucernaria quadricornis	0.91	
Myxozoa	Bivalvulida ⑦	*Henneguya salminicola*[c]	1.01	
		Myxobolus cerebralis[c]	1.04	
		Myxobolus pendula[c]	0.82	
		Myxobolus squamalis[c]	1.00	0.62
		Thelohanellus kitauei[c]	0.96	1.07
Polypodiozoa	Polypodiidea ⑥	*Polypodum hydriforme*	0.68	

NOTE.—Circled numbers correspond to figures 2 and 4.
[a]Species in which DNMTs are absent.
[b]Species in which MBDs are absent.
[c]Species in which both DNMTs and MBDs are absent.

despite the presence of the requisite genes for methylation (table 1 and fig. 2).

To compare $CpG_{o/e}$ of conserved genes to nonconserved genes, ten representative species with high-quality transcriptomes were selected covering the major cnidarian groups. In nine of the ten species, orthologous genes shared by the ten selected species were found to have significantly lower $CpG_{o/e}$ than nonorthologous genes, with the exception of *Calvadosia* (table 2). In eight of the ten species, orthologous genes are enriched in those with lower $CpG_{o/e}$ (supplementary table S2, Supplementary Material online) with *Calvadosia and Clytia* being the two exceptions, both of which have overall average gene body $CpG_{o/e}$ above 0.75 suggesting minimal methylation (table 1). These observations indicate that more conserved genes have higher methylation, consistent with previous findings in other taxa (Sarda et al. 2012). Such comparison of conserved genes and nonconserved genes is likely less biologically meaningful in species with low methylation globally such as *Calvadosia* and *Clytia*. For each species pair, genes that are low in $CpG_{o/e}$ (high in methylation) in both species are overrepresented in pairwise orthologs (supplementary table S3, Supplementary Material online). Additionally, differentially expressed genes in *Aurelia coerulia* reported by Gold et al. (2019) have significantly higher $CpG_{o/e}$ than nondifferentially expressed genes ($P < 2.2e\text{-}16$; Welch two sample *t*-test).

$CpG_{o/e}$ of the repeats proved higher than that of gene bodies in 14 out of the 15 species surveyed for both, indicating that repeats are generally less preferred sites of methylation in Cnidaria; an intracellular parasite *Myxobolus squamalis* was the exception (fig. 5). Both gene body and repeat $CpG_{o/e}$ negatively correlate with repeat content (fig. 5*A* and *B*) (Gene body: $t = -3.4576$, $P = 0.003008$; Repeat: $t = -3.6855$, $P = 0.003591$; Pearson correlation) and positively correlate with each other (fig. 5*C*) ($t = 5.0099$, $P = 0.0003963$, Pearson correlation). The outliers *Hydra vulgaris* and *Myxobolus squamalis* were excluded from the statistical tests, as will be addressed in Discussion.

Discussion

The LCA of Eumetazoa Likely Preferentially Methylated Gene Bodies Relative to Repetitive Elements

Sarda et al. (2012) argued that gene body methylation is the ancestral methylation mode in Metazoa. Consistent with previous studies on *Nematostella* and *Acropora* (Zemach et al.

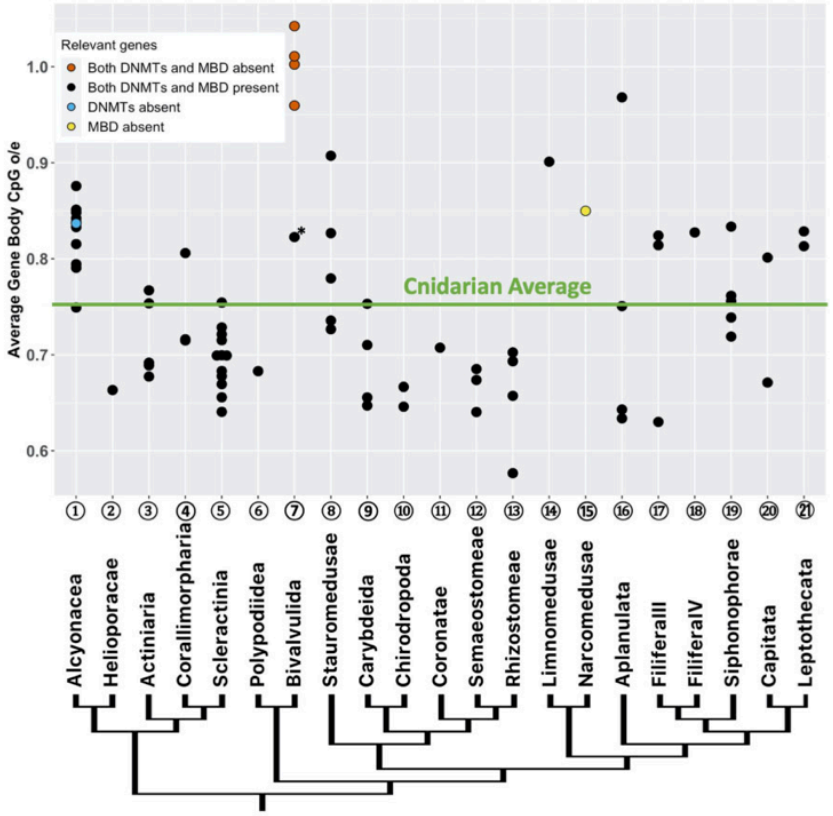


Fig. 2.—Average gene body $CpG_{o/e}$ of each cnidarian species organized in ordinal rank groups. Green horizontal line indicates the average among all species. Phylogenetic relationship between the major cnidarian groups rendered from Kayal et al. (2018) and Bentlage and Collins (2021). Circled numbers correspond to figure 4 and table 1. The species marked by the asterisk is *Myxobolus pendula*, where the relative low average gene body $CpG_{o/e}$ compared with the others in the class is possibly due to contamination from the fish host, which is discussed in detail in Results.

2010; Dixon et al. 2014), our results suggest pervasive higher GbM than repeat methylation across cnidaria. This study, combined with similar data for most nonvertebrate Bilateria, further supports GbM as the ancestral condition in Cnidaria, Bilateria, and the Eumetazoa.

Patterns of GbM in Cnidaria: Lack of Discrete Bimodality

Most species in our study did not exhibit the classic bimodal distribution of GbM found in bilaterian invertebrates where the two classes of genes significantly differ in terms of methylation status, functional enrichment, and expression plasticity (Zemach et al. 2010; Glastad et al. 2014; Dimond and Roberts 2016). However, our results show that conserved genes which are presumably expressed constitutively have higher methylation (table 2). Thus, cnidarian GbM appears to play a similar role in regulation of gene expression to that observed in bilaterian invertebrates, with the caveat that this pattern appears less pronounced or discretely bimodal in Cnidaria.

In addition to Notos, a KDE-based method, we further assessed the GbM distribution patterns using Gaussian mixture models, and models with three or more components best fit the GbM data for the vast majority of the species in question (fig. 3 and supplementary fig. S1, Supplementary Material online) (Fraley and Raftery 2003). Gaussian mixture models have previously identified three or more components in *Acropora* and some bilaterian animals as well (Dixon et al. 2016). The different results from Notos and Gaussian mixture models highlight the complexity of the data, which is further explored via the clustering analysis (fig. 4). From the perspective of these data, the default criteria set by Notos appear too stringent. In contrast, Gaussian mixture models are less interpretable due to excessive sensitivity. We note that assessing the complex distribution patterns of GbM using parameters previously applied to another group, in this case Bilateria, provides an arbitrary and incomplete picture of the variation.

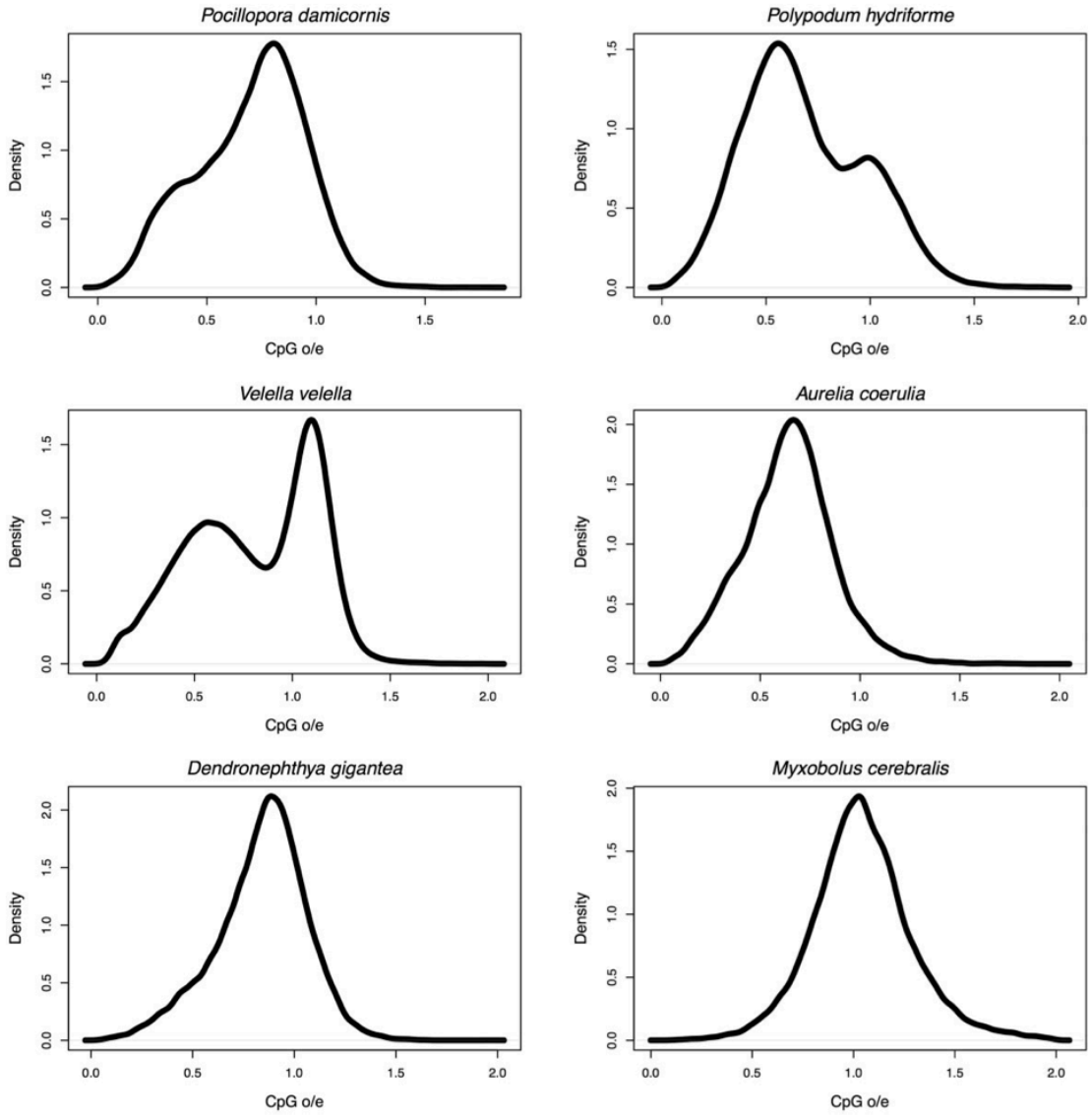


FIG. 3.—Density plots of gene body $CpG_{o/e}$ of selected species as examples. The discrete bimodality typically found in bilaterian invertebrates occurs only in 11 of the 41 cnidarian species where $CpG_{o/e}$ levels indicate substantial methylation. Among the six examples shown, *Polypodum and Velella* qualified as bimodal by Notos (and all taxa were classified as multimodal by mclust due to its sensitive nature). To further explore and group these patterns we conduct a clustering exercise (fig. 4), and a comparison of methylation in conserved and less conserved genes (table 2). *Pocillopora* and *Dendronephthya* are anthozoan corals; *Velella* is a hydrozoan, "by-the-wind sailor"; *Polypodum* belongs to the group Polypodiozoa, a sturgeon parasite, and *Myxobolus* is a member of Myxozoa which parasitizes bony fish; *Aurelia* is a scyphozoan jellyfish, "moon jelly."

Nevertheless, GbM is partially correlated with gene conservation in Cnidaria in a broad suite of species including those that are not diagnosed as formally bimodal using Notos (table 2). Thus, we do find evidence of similar processes operating in Bilateria and Cnidaria relative to methylation and gene expression and function. However, our results also highlight substantial differences between Bilateria and Cnidaria. Thus, caution is needed when choosing methods to describe distribution patterns of gene body methylation across previously unexamined groups.

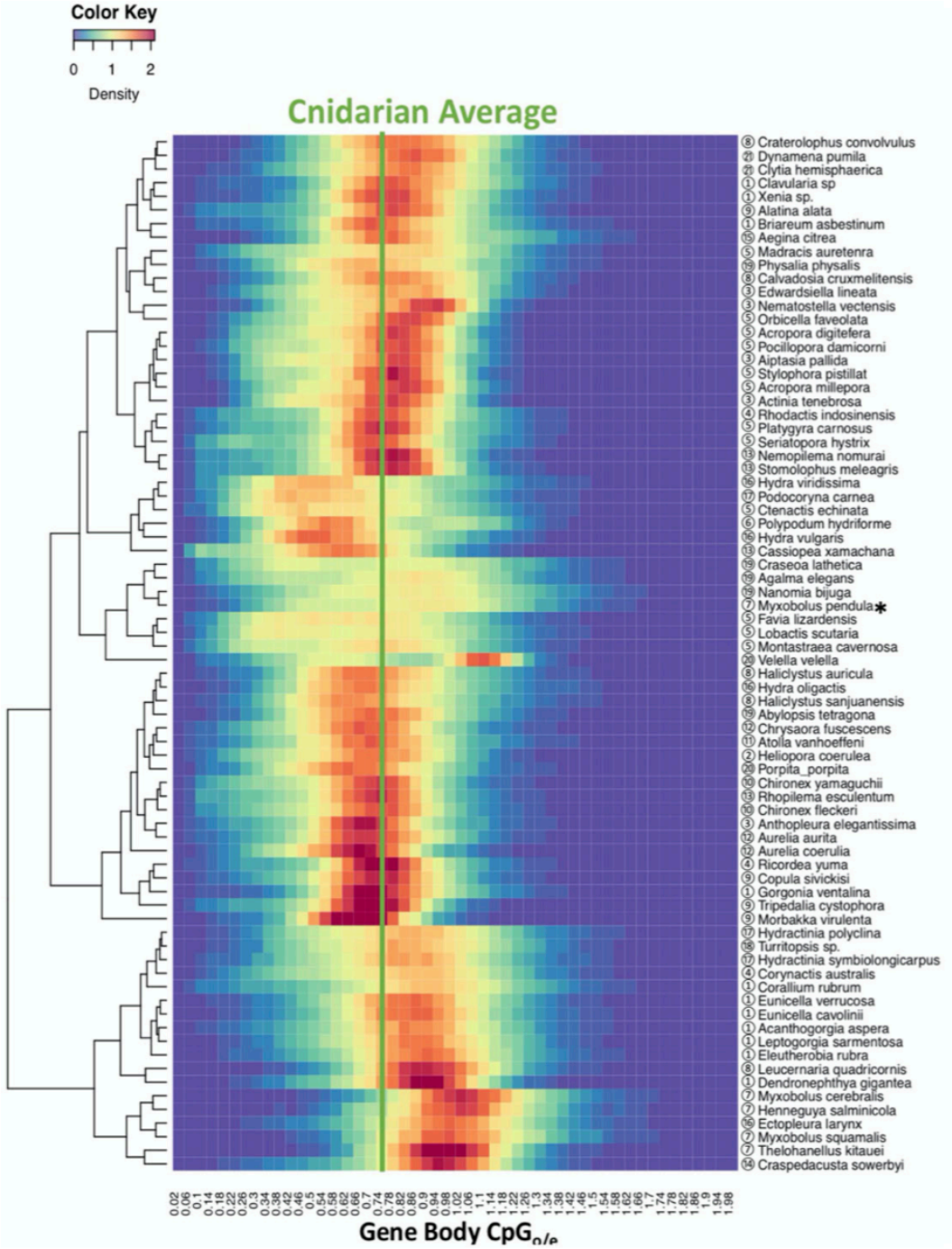


FIG. 4.—Heatmap of gene body $CpG_{o/e}$ distribution patterns in Cnidaria. Circled numbers next to the species names correspond to figure 2 and table 1. The species marked by the asterisk is *Myxobolus pendula*, where the distribution pattern might have been affected by possible contamination from the fish host, which is discussed in detail in Results.

Table 2
$CpG_{o/e}$ of Ten-Way Orthologous Genes versus Nonorthologous Genes in Each Species

Species	Mean $CpG_{o/e}$ of Nonorthologous Genes	Mean $CpG_{o/e}$ of Ten-Way Orthologous Genes	*t* stat.	*P* value
Acropora digitifera	0.7093909	0.6666281	15.794	2.2e-16
Aiptasia pallida	0.7152435	0.6313672	27.393	2.2e-16
Alatina alata	0.7597698	0.7381270	11.037	2.2e-16
Aurelia coerulia	0.6475442	0.6267479	6.8606	7.135e-12
Calvadosia cruxmelitensis	0.7769339	0.8065653	−15.144	2.2e-16
Clytia hemisphaerica	0.8164449	0.8091184	3.2878	0.001012
Hydra vulgaris	0.6437276	0.6156787	8.1367	4.315e-16
Morbakka virulenta	0.6508604	0.6392557	5.6923	1.271e-08
Nematostella vectensis	0.7860412	0.7203910	25.841	2.2e-16
Physalia physalis	0.7628115	0.6970781	16.648	2.2e-16

NOTE.—With the exception of *Calvadosia*, orthologous genes have higher methylation compared with non-ten-way orthologous genes in each species.

Patterns of GbM in Cnidaria: Lack of $CpG_{o/e}$ Higher Than 1

In addition to less-pronounced bimodality, most cnidarian species have very few genes with $CpG_{o/e}$ higher than 1 (fig. 4 and supplementary fig. S1, Supplementary Material online). This contrasts with bilaterian invertebrates where modes well above 1 are often observed (e.g., Park et al. 2011). These values above 1 may suggest general compensatory mechanisms that restore CpGs on gene bodies in bilaterian invertebrates that do not exist in Cnidaria. The presence of compensatory mechanisms could also explain the stronger bimodality observed in invertebrate Bilateria. There are a number of DNA repair mechanisms that could play a role in this, however of particular interest are the DNA repair functions exhibited by some MBD proteins in conjunction with other molecules (Watanabe et al. 2003; Wu et al. 2003). Further work in this area could prove revealing regarding the broad evolution of mechanisms controlling GbM and its role in epigenesis.

Patterns of GbM in Cnidaria: Complex Patterns Partially Reflecting Taxonomy and Life History

GbM varies dramatically across cnidarian taxa indicating a complex evolutionary history (figs. 2 and 4). This is not surprising given the antiquity of Cnidaria as a whole; they are sister to Bilateria with a Precambrian origin, and the separation of the major cnidarian groups occurred over 500 Ma (Khalturin et al. 2019). By sampling widely in Cnidaria, we show that the GbM pattern is quite complex (figs. 2 and 4) and is not consistent within phyla or classes. However, multiple orders within particular class level taxa show consistent GbM, which we discuss later. In contrast, in insects, a single class, there is relatively consistent GbM across ordinal rank taxa with a trend toward less methylation in the advanced (Holometabolous) insects (Provataris et al. 2018). However, it is important to note that the insects are a much younger group that radiated in the late Paleozoic, thus comparison between orders in the two groups may not be appropriate.

In our data, GbM shows some relationship to life history complexity. Cubozoa and Scyphozoa, both have complex life cycles that include a medusa phase, and both have higher levels of GbM relative to the other groups (Cubozoa and Scyphozoa are groups ⑨–⑬; fig. 2). Previous work in Scyphozoan jellyfish *Aurelia aurita* showed that disrupting DNMT1 halts strobilation, the transition from the sessile polyp stage to free-swimming medusa stage, suggesting that cytosine methylation is functionally important in regulating life history transitions (Fuchs et al. 2014). On the other hand, Myxozoa which includes endoparasitic species with reduced complexity in life history (as summarized in Kent et al. [2001]) have high $CpG_{o/e}$ suggesting the absence of GbM. In contrast, the sister group to Myxozoa, *Polypodum hydriforme* shows one of the lowest gene body $CpG_{o/e}$ indicating high methylation. It is a parasitic species, but with an elaborate life history including a free-living stolon stage (Raikova 1994; Kayal et al. 2018).

On the other hand, the correlation between life history complexity and GbM is less clear in Hydrozoa, where dramatic differences occur even within families. The two taxa studied from the Capitata taxon are in the family Porpitidae, the by-the-wind sailors which have an unusual mode of life as colonies of medusae function as surface vessels. Despite the similarity in this mode of life and familial relationship, the species studied have very different GbM. Within the Aplanulata group (which includes the well-known freshwater genus *Hydra* and lacks a planula phase), the three *Hydra* species samples have high GbM, whereas the other Aplanulata genus sampled, *Ectopleura*, has low GbM. Similarly, within the Filifera III group (Bentlage and Collins 2021), both members of *Hydractinia* examined have low GbM, whereas that of *Podocoryne* is much higher.

In examining evolution of methylation implied by these patterns it is important to consider the antiquity and diversity of Hydrozoa. The modern species in the genus Hydra diverged 100–200 Ma in the Mesozoic prior to the K/T extinction (Schwentner and Bosch 2015; Wong et al. 2019), whereas

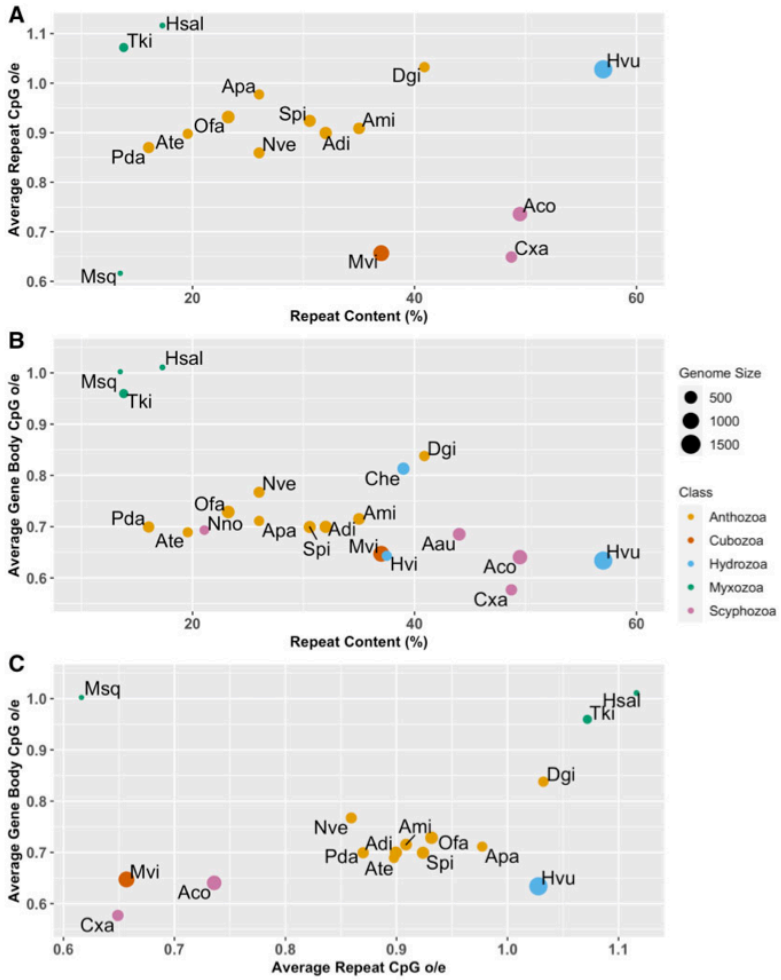


Fig. 5.—Both repeat and gene body $CpG_{o/e}$ correlate with repeat content. (*A*) The average $CpG_{o/e}$ of repeats is negatively correlated with the repeat content of the genome, with two notable exceptions (*Hydra vulgaris* and *Myxobolus squamalis*). (*B*) The average gene body $CpG_{o/e}$ is negatively correlated with repeat content. (*C*) Gene body $CpG_{o/e}$ and repeat $CpG_{o/e}$ are positively correalted with each other while gene body $CpG_{o/e}$ are generally lower than repeat $CpG_{o/e}$, indicating gene bodies are preferred targets of methylation. Genome sizes in Mb. Aau, *Aurelia aurita*; Aco, *Aurelia coerulia*; Adi, *Acropora digitefera*; Ami, *Acropora millepora*; Apa, *Aiptasia pallida*; Ate, *Actinia tenebrosa*; Che, *Clytia hemisphera*; Cxa, *Cassiopea xamachana*; Dgi, *Dendronephthya gigantea*; Hsal, *Henneguya salminicola*; Hvu, *Hydra vulgaris*; Msq, *Myxobolus squamalis*; Mvi, *Morbakka virulenta*; Nve, *Nematostella vectensis*; Ofa, *Orbicella faveolata*; Pda, *Pocillopora damicornis*; Spi, *Stylophora pistillata*; Tki, *Thelohanellus kitauei*.

a derived hydrozoan fossil from 520 Myr in the Cambrian suggests that the hydrozoan crown group is at least that old (Muscente et al. 2016; Song et al. 2021). In addition, Hydrozoa is the most diverse cnidarian group and higher order relationships within the hydrozoa are only now coming into focus (Bentlage and Collins 2021). Recent examination of the better fossilized Anthozoa suggests a still greater antiquity for the cnidarian radiation as a whole (McFadden et al. 2021). Our results suggest that DNA methylation has gone through dynamic changes in the evolutionary history of Cnidaria.

Others have noted that life history complexity and DNA methylation are not always correlated. For example, sponges have surprisingly high levels of GbM even though they are seemingly among the simplest forms of animals (de Mendoza et al. 2019). Thus, our results suggest that life history complexity appears associated with DNA methylation in cnidarians, but it does not appear to be the only factor. More detailed studies of GbM in Medusozoa would help illuminate the role of DNA methylation in the life history evolution across the group.

Patterns of GbM in Cnidaria: Absence of GbM in Parasitic Species

We present several cases of absence of CpG methylation in Myxozoa (fig. 2), a clade comprised parasitic species. Kyger et al. (2020) showed the absence of cytosine methylation in *Henneguya salminicola* and one additional Myxozoan species using bisulfite sequencing, and our result suggests that both DNMT1 and DNMT3 are absent in the entire clade. Aside from the reduction of life history complexity, this might also be due to the massively reduced genome sizes of myxozoans (Chang et al. 2015). Lechner et al. (2013) presented a positive correlation between genome size and DNA methylation across a range of metazoan species. However, in other contexts small genome size does not automatically lead to loss of methylation. *Plasmodium falciparum* maintains methylation, albeit in a different nucleotide context, in a much smaller genome than myxozoans (Ponts et al. 2013). Therefore, care must be taken in strong inferences of complete loss of methylation in small myxozoan genomes. Moreover, since Myxozoa is a diverse group of parasites inhabiting many teleost lineages, wider sampling might yield further insights.

Our results also support low GbM across the species of Alcyonacea analyzed. Consistent with low methylation, DNMTs were not recovered in the transcriptome of one of the species (*Eunicella verrucosa)*, suggesting either the expression of the enzymes are extremely low, or they have been lost in the genome (fig. 2).

Methylation on Repetitive Elements Correlates with Repeat Content

Previously Zemach et al. (2010) showed that *Nematostella* had less methylation on repeats than gene bodies using bisulfite sequencing. Our results add to this and are consistent with studies in other invertebrate phyla (Feng et al. 2010; Xiang et al. 2010; de Mendoza et al. 2019).

The correlation between overall methylation levels and repeat content has been unclear, although it is well documented that metazoan genome size and overall methylation level are positively correlated (Lechner et al. 2013; Zhou et al. 2020). For example, the pufferfish *Tetraodon nigroviridis* has very low repeat content, yet it still exhibits the typical vertebrate hypermethylation pattern (Zemach et al. 2010); the silk moth *Bombyx mori* has a transposon-rich genome, yet shows very low level of overall DNA methylation, as well as low GbM (Xiang et al. 2010). In the sponge *Amphimedon queenslandica* the genome is hypermethylated as in vertebrates, with interspersed repeats making up 35% of the genome. These repeats are heavily methylated approaching the vertebrate condition (de Mendoza et al. 2019). In this study, we found significant correlation between repeat methylation and gene body methylation, and both are positively correlated with repeat content (fig. 5). Our results suggest that global methylation, GbM, and repeat methylation are correlated with repeat content in Cnidaria, however, there are two exceptions worth discussing.

The *Hydra vulgaris* genome is drastically bigger than typical cnidarian genomes as the result of the invasion and expansion of a single class of TEs (Wong et al. 2019), and it has one the lowest gene body $CpG_{o/e}$ (fig. 2), indicating heavy methylation on the gene bodies. Yet the repeats showed no signs of CpG methylation (fig. 5). This may relate to the young age and rapid expansion of repeats as young repeats tend to be less methylated (Lechner et al. 2013). Alternatively, conversion of methylated C to T in recently methylated repeats may not have caught up with current methylation yielding a minimal departure from expected CpG frequency. Both of these could be tested by experimentally measuring DNA methylation on the repeats by methods such as bisulfite sequencing. Lastly, *Hydra* could have activated other mechanisms to defend against the TEs that do not involve CpG methylation (de Mendoza et al. 2019).

We observe an additional exception to the positive correlation between repeat content and methylation on the various genomic elements in *Myxobolus squamalis*, a myxozoan parasite which had no signs of methylation on the gene bodies but significant methylation on the repeats (fig. 5). Preferential methylation of repeats over gene bodies is opposite the results from a wide range of invertebrates (Zemach and Zilberman 2010; Sarda et al. 2012) as well as the cnidarians reported here. It is even more surprising as no DNMT or MBD was found in the transcriptome or the genome of *Myxobolus squamalis* (table 1). It would also be interesting to conduct a bisulfite sequencing study on *Myxobolus squamalis* to verify that there is indeed cytosine methylation on the repeats. $CpG_{o/e}$ has been shown to correlate well with experimentally determined methylation, thus *Myxobolus squamalis* provides quite an interesting system to follow up on regarding both $CpG_{o/e}$ as a reliable proxy for CpG methylation given the methylation of repeats where DNMTs are not currently observed.

To our knowledge, the relationship between repeat content and DNA methylation on various elements including TEs and gene bodies has not been previously studied with the consistent sampling across a large group comparable to our study. Nevertheless, previous work with fewer taxa convey important related information; bisulfite sequencing in several fungal species revealed a positive correlation between repeat content, genome-wide methylation, and TE methylation similar to our results (Hosseini et al. 2020). Comparative studies investigating such relationships require well-annotated genomes in closely related species. Several cnidarian genera such as *Acropora* in Anthozoa and *Aurelia* in Scyphozoa (Dawson and Jacobs 2001) appear poised for exploration in this fashion. Future studies could take advantage of these resources to examine evolution of methylation on TEs and GbM in more closely related cnidarian taxon sets.

From this comprehensive study, we show that DNA methylation patterns have some lineage specificity, but are often subject to significant evolutionary change within current taxon sampling. We report for the first time that genes with $CpG_{o/e}$ above 1 are rare in most cnidarians. This suggests the possibility of unknown compensatory mechanisms that restore CpG sites in bilaterian invertebrates that are absent in Cnidaria. In addition, the distribution of GbM across genes is less discretely bimodal in cnidarian taxa than it is in bilaterian invertebrates; nevertheless, we observe differences in GbM between conserved and differentially expressed genes. Thus, cnidarian GbM appears to also be involved in regulating gene expression. This builds on previous observations in two cnidarian taxa, *Nematostella* and *Acropora*, where GbM differs between stable and dynamically expressed genes (Dixon et al. 2014; Dimond and Roberts 2016). Intermediate levels of GbM appear likely to have been present in the last common ancestor of Eumetazoa. However, the incomplete agreement on the placement of Ctenophore and Placozoa in the Metazoa, and the limited methylation in these taxa makes it more difficult to be certain that the dramatically high methylation in sponges and vertebrates is a result of convergent evolution (Albalat 2008; Lechner et al. 2013; Dabe et al. 2015). Changes in GbM occur within several cnidarian orders, with parasitic taxa exhibiting dramatic methylation loss. We also note that scyphozoan and cubozoan jellyfish exhibit increased methylation and increased life history complexity, and increased repeat content is also associated with increased DNA methylation. However, neither life history changes nor evolution of repeat content appear sufficient to explain the patterns across the species studied. This work reveals a dynamic evolutionary history of DNA methylation among cnidarian groups and even within closely related clades. Our effort also points to several examples where bisulfite sequencing of a limited number of taxa would be revealing.

Materials and Methods

Measurement of CpG Depletion

CpG depletion is measured as the ratio of observed to expected CpG. This was calculated as $P_{CpG}/(P_C \times P_G)$, where P_{CpG}, P_C, and P_G are the frequencies of CpGs, cytosines, and guanines respectively (Elango et al. 2009). Loss of CpG due to DNA methylation is associated with conversion to TpG (Duncan and Miller 1980). Thus, to confirm that CpG depletion is a result of methylation, we also measured the enrichment of TpG, where $TpG_{o/e}$ was calculated in the same way.

Data Sources

In total, this study analyzed data from 76 species spanning across seven classes and 21 ordinal groups. Data sources used are summarized in supplementary table S1, Supplementary Material online. All transcriptomes had BUSCO scores higher than 50% For taxa with both an available annotated genome and transcriptome, the transcriptome is used for gene body $CpG_{o/e}$ as it was previously shown that measurement of CpG depletion using coding sequences and cDNAs better reflect experimentally determined methylation levels (Aliaga et al. 2019). Predicated gene models and transcripts that are shorter than 300 bp were excluded from subsequent analyses.

About ten species that have both transcriptomes and annotated genomes were selected for analyses of repetitive elements. Repetitive elements were identified de novo with RepeatScout and RepeatMasker (Smit and Hubley 2015; Smit et al. 2015). Repetitive elements longer than 50 bp and occur more than ten times in the genome were used for subsequent analyses. Repeats that overlap with predicted gene bodies were excluded. For all of these ten species, gene body CpG analysis was done using the transcriptomes, effectively excluding repeats that might reside in introns.

Statistical Analyses of Gene Body $CpG_{o/e}$ Distribution Modality

To estimate the number of components in the density distribution of gene body $CpG_{o/e}$ in each species, we employed 1) the *Notos* tool, a KDE-based approach (Bulla et al. 2018), and 2) a model-based clustering approach using the mclust package (Fraley and Raftery 2003) followed by Bayesian information criteria to measure the fit of each model. Both were conducted in R (www.r-project.org).

Search for DNMTs and MBDs

The presence of DNMT1, DNMT3, and MBDs in 76 species were screened via TBlastN searches and verified through reciprocal BLAST against the NCBI nonredundant nucleotide database. The queries used for DNMT1 are XP_012557244.1 (*Hydra vulgaris*) and XP_020612302.1 (*Orbicella faveolata*); the queries used for DNMT3 are XP_012561137.1 (*Hydra vulgaris*) and XP_015756999.1 (*Acropora digitifera*); the queries used for MBDs are XP_020895950.1 (*Aiptasia pallida*) and XP_020604760.1 (*Orbicella faveolata*).

To determine whether the DNMTs found in the parasite *Myxobolus pendula* are indeed from the cnidarian or from the fish host, we conducted phylogenetic analyses. DNMT1 and DNMT3 protein sequences were obtained using Transdecoder (https://github.com/TransDecoder/TransDecoder/wiki, last accessed February 4, 2022). The sequences were aligned using MUSCLE (Edgar 2004) and then manually curated. Phylogenetic trees were reconstructed using RAxML v8.2.12 (Stamatakis 2014).

Ortholog $CpG_{o/e}$ Analyses in Selected Species

Orthologs between *Acropora digitifera* (Anthozoa), *Aiptasia pallida* (Anthozoa), *Alatina alata* (Cubozoa), *Aurelia coerulia*

(Scyphozoa), *Calvadosia cruxmelitensis* (Staurozoa), *Clytia hemisphaerica* (Hydrozoa), *Hydra vulgaris* (Hydrozoa), *Morbakka virulenta* (Cubozoa), *Nematostella vectensis* (Anthozoa), and *Physalia physalis* (Hydrozoa) were determined using Orthofinder version 2.4.0 (Emms and Kelly 2015, 2019) with Diamond protein alignment (Buchfink et al. 2015). These species were selected for their high-quality transcriptomes to yield a sufficient number of orthologs for our analyses. For *Alatina*, *Calvadosia*, *Clytia*, and *Physalia*, transcripts were first translated into protein sequences using Transdecoder version 5.3.0 with default settings (https://github.com/TransDecoder/TransDecoder/wiki, last accessed February 4, 2022).

For each species, $CpG_{o/e}$ of ten-way orthologs was compared with genes that are not shared by all ten species. For each pair of species, $CpG_{o/e}$ of single orthologs in each species is compared.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Dr Joseph Ryan from the Whitney Laboratory for Marine Bioscience at the University of Florida for graciously directing us to the numerous cnidarian transcriptomic resources. We also would like to express our appreciation to Dr Matteo Pellegrini's group at UCLA for general guidance during our initial exploration in this new field. We additionally thank members of the Jacobs laboratory, Benjamin Ha, Rachel Turba, and Hayden Speck for feedback on the manuscript, and Qingru Sue Xu for her help with conducting the BLAST searches.

Data Availability

Computational codes used in this study are available on Github: https://github.com/pzhang312/Cnidaria-CpG-Comp.

Literature Cited

Albalat R. 2008. Evolution of DNA-methylation machinery: DNA methyltransferases and methyl-DNA binding proteins in the amphioxus *Branchiostoma floridae*. Dev Genes Evol. 218(11–12):691–701.

Aliaga B, Bulla I, Mouahid G, Duval D, Grunau C. 2019. Universality of the DNA methylation codes in Eucaryotes. Sci Rep. 9:1–11.

Bentlage B, Collins AG. 2021. Tackling the phylogenetic conundrum of Hydroidolina (Cnidaria: Medusozoa: Hydrozoa) by assessing competing tree topologies with targeted high-throughput sequencing. PeerJ. 9:e12104.

Bestor TH. 1990. DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes. Philos Trans R Soc Lond B Biol Sci. 326(1235):179–187.

Bock C, Lengauer T. 2008. Computational epigenetics. Bioinformatics 24(1):1–10.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 12(1):59–60.

Bulla I, et al. 2018. Notos – a galaxy tool to analyze CpN observed expected ratios for inferring DNA methylation types. BMC Bioinformatics 19(1):105.

Canestro C, Yokoi H, Postlethwait JH. 2007. Evolutionary developmental biology and genomics. Nat Rev Genet. 8(12):932–942.

Chang ES, et al. 2015. Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. Proc Natl Acad Sci U S A. 112(48):14912–14917.

Cramer JM, et al. 2017. Methylation specific targeting of a chromatin remodeling complex from sponges to humans. Sci Rep. 7:1–15.

Dabe EC, Sanford RS, Kohn AB, Bobkova Y, Moroz LL. 2015. DNA methylation in basal metazoans: insights from ctenophores. Integr Comp Biol. 55(6):1096–1110.

Dawson MN, Jacobs DK. 2001. Molecular evidence for cryptic species of *Aurelia aurita* (Cnidaria, Scyphozoa). Biol Bull. 200(1):92–96.

Dimond JL, Roberts SB. 2016. Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change. Mol Ecol. 25(8):1895–1904.

Dixon GB, Bay LK, Matz MV. 2014. Bimodal signatures of germline methylation are linked with gene expression plasticity in the coral *Acropora millepora*. BMC Genomics 15:1109–1111.

Dixon GB, Bay LK, Matz MV. 2016. Evolutionary consequences of DNA methylation in a basal metazoan. Mol Biol Evol. 33(9):2285–2293.

Duncan BK, Miller JH. 1980. Mutagenic deamination of cytosine residues in DNA. Nature 287(5782):560–561.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Elango N, Hunt BG, Goodisman MA, Soojin VY. 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. Proc Natl Acad Sci U S A. 106(27):11206–11211.

Elango N, Yi SV. 2008. DNA methylation and structural and functional bimodality of vertebrate promoters. Mol Biol Evol. 25(8):1602–1608.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16:157.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):1–14.

Falckenhayn C, et al. 2013. Characterization of genome methylation patterns in the desert locust *Schistocerca gregaria*. J Exp Biol. 216(Pt 8):1423–1429.

Feng S, et al. 2010. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A. 107(19):8689–8694.

Field LM, Lyko F, Mandrioli M, Prantera G. 2004. DNA methylation in insects. Insect Mol Biol. 13(2):109–115.

Flores K, et al. 2012. Genome-wide association between DNA methylation and alternative splicing in an invertebrate. BMC Genomics 13(1):480–489.

Fraley C, Raftery AE. 2003. Enhanced software for model-based clustering, density estimation, and discriminant analysis: MCLUST. J Classification. 20(2):263–286.

Fuchs B, et al. 2014. Regulation of polyp-to-jellyfish transition in *Aurelia aurita*. Curr Biol. 24(3):263–273.

Gavery MR, Roberts SB. 2010. DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). BMC Genomics 11:483–489.

Glastad KM, Hunt BG, Goodisman MA. 2014. Evolutionary insights into DNA methylation in insects. Curr Opin Insect Sci. 1:25–30.

Glastad KM, Hunt BG, Yi SV, Goodisman MAD. 2011. DNA methylation in insects: on the brink of the epigenomic era. Insect Mol Biol. 20(5):553–565.
Gold DA, et al. 2019. The genome of the jellyfish *Aurelia* and the evolution of animal complexity. Nat Ecol Evol. 3(1):96–104.
Goll MG, et al. 2006. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 311(5759):395–398.
Hendrich B, Tweedie S. 2003. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet. 19(5):269–277.
Holz-Schietinger C, Matje DM, Harrison MF, Reich NO. 2011. Oligomerization of DNMT3A controls the mechanism of de novo DNA methylation. J Biol Chem. 286(48):41479–41488.
Hosseini S, Meunier C, Nguyen D, Reimegård J, Johannesson H. 2020. Comparative analysis of genome-wide DNA methylation in Neurospora. Epigenetics 15(9):972–987.
Hunt BG, Brisson JA, Yi SV, Goodisman MA. 2010. Functional conservation of DNA methylation in the pea aphid and the honeybee. Genome Biol Evol. 2:719–728.
Kayal E, et al. 2018. Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. BMC Evol Biol. 18(1):68.
Keller TE, Han P, Yi SV. 2016. Evolutionary transition of promoter and gene body DNA methylation across invertebrate–vertebrate boundary. Mol Biol Evol. 33(4):1019–1028.
Kent ML, et al. 2001. Recent advances in our knowledge of the Myxozoa. J Eukaryot Microbiol. 48(4):395–413.
Khalturin K, et al. 2019. Medusozoan genomes inform the evolution of the jellyfish body plan. Nat Ecol Evol. 3(5):811–822.
Kucharski R, Maleszka J, Foret S, Maleszka R. 2008. Nutritional control of reproductive status in honeybees via DNA methylation. Science 319(5871):1827–1830.
Kyger R, et al. 2020. Myxosporea (Myxozoa, Cnidaria) lack DNA cytosine methylation. Mol Biol Evol. 38(2):393–404.
Lechner M, et al. 2013. The correlation of genome size and DNA methylation rate in metazoans. Theory Biosci. 132(1):47–60.
Li Y, et al. 2018. DNA methylation regulates transcriptional homeostasis of algal endosymbiosis in the coral model Aiptasia. Sci Adv. 4(8):eaat2142.
McFadden CS, et al. 2021. Phylogenomics, Origin, and Diversification of Anthozoans (Phylum Cnidaria). Syst Biol. 70(4):635–647.
de Mendoza A, et al. 2019. Convergent evolution of a vertebrate-like methylome in a marine sponge. Nat Ecol Evol. 3(10):1464–1473.
Muscente AD, Allmon WD, Xiao S. 2016. The hydroid fossil record and analytical techniques for assessing the affinities of putative hydrozoans and possible hemichordates. Palaeontology 59(1):71–87.
Nanty L, et al. 2011. Comparative methylomics reveals gene-body H3K36me3 in Drosophila predicts DNA methylation and CpG landscapes in other invertebrates. Genome Res. 21(11):1841–1850.
Park J, et al. 2011. Comparative analyses of DNA methylation and sequence evolution using Nasonia genomes. Mol Biol Evol. 28(12):3345–3354.
Ponts N, et al. 2013. Genome-wide mapping of DNA methylation in the human malaria parasite plasmodium falciparum. Cell Host Microbe. 14(6):696–706.
Provataris P, Meusemann K, Niehuis O, Grath S, Misof B. 2018. Signatures of DNA methylation across insects suggest reduced DNA methylation levels in Holometabola. Genome Biol Evol. 10(4):1185–1197.
Raikova EV. 1994. Life cycle, cytology, and morphology of *Polypodium hydriforme*, a coelenterate parasite of the eggs of acipenseriform fishes. J Parasitol. 80(1):1–22.
Riviere G, et al. 2013. DNA methylation is crucial for the early development in the oyster *C. gigas*. Mar Biotechnol. 15(6):739–753.
Sarda S, Zeng J, Hunt BG, Yi SV. 2012. The evolution of invertebrate gene body methylation. Mol Biol Evol. 29(8):1907–1916.
Schwentner M, Bosch TC. 2015. Revisiting the age, evolutionary history and species level diversity of the genus *Hydra* (Cnidaria: Hydrozoa). Mol Phylogenet Evol. 91:41–55.
Smit AFA, Hubley R. 2015. RepeatModeler Open-1.0.2008–2015. Available from: http://www.repeatmasker.org. Accessed February 4, 2022.
Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0.2013–2015. Available from: http://www.repeatmasker.org. Accessed February 4, 2022.
Song X, et al. 2021. Advanced Cambrian hydroid fossils (Cnidaria: Hydrozoa) extend the medusozoan evolutionary history. Proc Biol Sci. 288(1944):20202939.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.
Suzuki MM, Kerr ARW, De Sousa D, Bird A. 2007. CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res. 17(5):625–631.
Takayama S, et al. 2014. Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity. Genome Res. 24(5):821–830.
Tweedie S, Charlton J, Clark V, Bird A. 1997. Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol Cell Biol. 17(3):1469–1475.
Watanabe S, et al. 2003. Methylated DNA-binding domain 1 and methylpurine-DNA glycosylase link transcriptional repression and DNA repair in chromatin. Proc Natl Acad Sci U S A. 100(22):12859–12864.
Wenzel D, Palladino F, Jedrusik-Bode M. 2011. Epigenetics in *C. elegans*: facts and challenges. Genesis 49(8):647–661.
Wong WY, et al. 2019. Expansion of a single transposable element family is associated with genome-size increase and radiation in the genus *Hydra*. Proc Natl Acad Sci U S A. 116(46):22915–22917.
Wu P, et al. 2003. Mismatch repair in methylated DNA. J Biol Chem. 278(7):5285–5291.
Xiang H, et al. 2010. Single base–resolution methylome of the silkworm reveals a sparse epigenomic map. Nat Biotechnol. 28(5):516–520.
Yi SV, Goodisman MAD. 2009. Computational approaches for understanding the evolution of DNA methylation in animals. Epigenetics 4(8):551–556.
Zapata F, et al. 2015. Phylogenomic analyses support traditional relationships within Cnidaria. PLoS One 10(10):e0139068.
Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328(5980):916–919.
Zemach A, Zilberman D. 2010. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. Curr Biol. 20(17):R780–R785.
Zhou W, Liang G, Molloy PL, Jones PA. 2020. DNA methylation enables transposable element-driven genome expansion. Proc Natl Acad Sci U S A. 117(32):19359–19366.

Associate editor: Cristina Vieira

Supplemental Materials

Table S1. Genomic/transcriptomic resources used in this study. Gray shaded species indicate predicted gene models used for gene body $CpG_{o/e}$ analyses instead of transcripts.

Class	Order	Species	Source for GbM analyses	Source for repeat methylation analyses
Anthozoa	Actiniaria	*Actinia tenebrosa*	GCF_009602425.1	GCF_009602425.1
		Aiptasia pallida	Zapata *et al.* 2015	GCF_001417965.1
		Anthopleura elegantissima	GBXJ00000000; GBYC00000000	
		Edwardsiella lineata	Stefanik *et al.* 2014	
		Nematostella vectensis	GCA_000209225.1	GCA_000209225.1
	Alcyonacea	*Acanthogorgia aspera*	GETB00000000	
		Briareum asbestinum	GHBD00000000	
		Clavularia sp.	GHAW00000000	
		Corallium rubrum	SRR1552944	
		Dendronephthya gigantea	GCF_004324835.1	GCF_004324835
		Eleutherobia rubra	GHFI00000000	
		Eunicella cavolinii	SRR1324943	
		Eunicella verrucosa	SRR1324944	
		Gorgonia ventalina	SRR935083	
		Leptogorgia sarmentosa	SRR1324968	
		Xenia sp.	GHBC00000000	
	Corallimorpharia	*Corynactis australis*	GELM00000000	
		Rhodactis indosinensis	GELO00000000	
		Ricordea yuma	GELN00000000	
	Helioporacea	*Heliopora coerulea*	GFVH00000000	
	Scleractinia	*Acropora digitefera*	GCF_000222465.1	GCF_000222465.1
		Acropora millepora	GCF_004143615.1	GCF_004143615.1
		Ctenactis echinata	Okubo *et al.* 2016	

		Favia lizardensis	Okubo *et al.* 2016	
		Lobactis scutaria	SRR2300562	
		Madracis auretenra	Kayal *et al.* 2018	
		Montastraea cavernosa	SRR2306543	
		Orbicella faveolata	GCF_002042975.1	GCF_002042975.1
		Platygyra carnosus	SRR402974-5	
		Pocillopora damicornis	GCF_003704095.1	
		Seriatopora hystrix	SRR2300678	
		Stylophora pistillata	GCF_002571385.1	
Hydrozoa	Anthoathecata	*Ectopleura larynx*	Zapata *et al.* 2015	
		Hydra oligactis	SRR040466-9	
		Hydra viridissima	SRR040470-3	
		Hydra vulgaris	GCF_000004095.1	GCF_000004095.1
		Hydractinia polyclina	SRR923509	
		Hydractinia symbiolongicarpus	Zapata et al. 2015	
		Podocoryna carnea	Zapata et al. 2015	
		Porpita porpita	GHBA00000000	
		Turritopsis sp	Hasegawa *et al.* 2016	
		Velella velella	GHAZ00000000	
	Leptothecata	*Clytia hemisphaerica*	HAMU00000000	
		Dynamena pumila	GHMC00000000	
	Limnomedusae	*Craspedacusta sowerbyi*	SRR923472	
	Narcomedusae	*Aegina citrea*	Zapata *et al.* 2015	
	Siphonophorae	*Abylopsis tetragona*	Zapata *et al.* 2015	
		Agalma elegans	Zapata *et al.* 2015	
		Craseoa lathetica	Zapata *et al.* 2015	
		Nanomia bijuga	Zapata *et al.* 2015	
		Physalia physalis	Zapata *et al.* 2015	

		Alatina alata	Zapata *et al.* 2015	
		Copula sivickisi	GHBG00000000	
		Morbakka virulenta	GHAF00000000	GCA_003991215.1
	Carybdeida	*Tripedalia cystophora*	GHAQ00000000	
		Chironex fleckeri	SRR1819888	
Cubozoa	Chirodropoda	*Chironex yamaguchii*	GHAX00000000	
	Coronatae	*Atolla vanhoeffeni*	Zapata *et al.* 2015	
		Cassiopea xamachana	Kayal *et al.* 2018	GCA_900291935.1
		Nemopilema nomurai	GHAR00000000	
		Rhopilema esculentum	GEMS00000000	
	Rhizostomeae	*Stomolophus meleagris*	SRR1168418	
		Aurelia aurita	Brekhman *et al.* 2015	
		Aurelia coerulea	Gold et al., 2019	Gold et al., 2019
Scyphozoa	Semaeostomeae	*Chrysaora fuscescens*	SRR3180892	
		Calvadosia cruxmelitensis	Kayal *et al.* 2018	
		Craterolophus convolvulus	Kayal *et al.* 2018	
		Haliclystus auricula	HAHA00000000	
		Haliclystus sanjuanensis	Kayal *et al.* 2018	
Staurozoa	Stauromedusae	*Lucernaria quadricornis*	Kayal *et al.* 2018	
		Henneguya salminicola	GCA_009887335.1	
		Myxobolus cerebralis	GBKL00000000	
		Myxobolus pendula	SRR2472984;SRR2472987;SRR2472989	
		Myxobolus squamalis	GHBR00000000	GCA_010108815.1
Myxozoa	Bivalvulida	*Thelohanellus kitauei*	GCA_000827895.1	GCA_000827895.1
Polypodiozoa	Polypodiidea	*Polypodum hydriforme*	Kayal *et al.* 2018	

Table S2. Genes that have orthologs in at least one other species are tested for enrichment in high or low methylation. A gene is categorized as 'high' methylation if the $CpG_{o/e}$ is less than the median, and vice versa.

Acropora digitifera

	Low	High
Orthologous	12586	13471
Non-orthologous	7425	6541

X-squared = 3730.9, df = 3, p-value < 2.2e-16

Aiptasia pallida

	Low	High
Orthologous	11268	12281
Non-orthologous	3422	2409

X-squared = 10825, df = 3, p-value < 2.2e-16

Alatina alata

	Low	High
Orthologous	10616	13939
Non-orthologous	50872	47549

X-squared = 46824, df = 3, p-value < 2.2e-16

Aurelia coerulea

	Low	High
Orthologous	8277	9202
Non-orthologous	3250	2325

X-squared = 6295.1, df = 3, p-value < 2.2e-16

Calvadosia cruxmelitensis

	Low	High

Orthologous	14189	11822
Non-orthologous	36579	38946

X-squared = 24366, df = 3, p-value < 2.2e-16

Clytia hemisphaerica

	Low	High
Orthologous	16353	16024
Non-orthologous	36441	36770

X-squared = 15796, df = 3, p-value < 2.2e-16

Hydra vulgaris

	Low	High
Orthologous	6707	7070
Non-orthologous	4953	4590

X-squared = 15796, df = 3, p-value < 2.2e-16

Morbakka virulenta

	Low	High
Orthologous	10522	10791
Non-orthologous	3864	3595

X-squared = 6680.9, df = 3, p-value < 2.2e-16

Nematostella vectensis

	Low	High
Orthologous	13739	15012
Non-orthologous	6066	4792

X-squared = 8246.8, df = 3, p-value < 2.2e-16

Physalia physalis

	Low	High

Orthologous	10163	12618
Non-orthologous	20568	18113

X-squared = 4505.5, df = 3, p-value < 2.2e-16

Table S3. Single orthologs of various methylation status between pairs of species. In each species, the $CpG_{o/e}$ of each gene is compared to the species median, and is categorized as 'high' if the $CpG_{o/e}$ is less than the median, and vice versa. The rows are the methylation status in the first species, the columns are the methylation status in the second species.

Aurelia coerulea (Scyphozoa) *vs. Nematostella vectensis* (Anthozoa)

X-squared = 949.95, df = 3, p-value < 2.2e-16

	Low	High
Low	275	534
High	146	1046

Aurelia coerulea (Scyphozoa) *vs. Hydra vulgaris* (Hydrozoa)

X-squared = 90.261, df = 3, p-value < 2.2e-16

	Low	High
Low	514	449
High	614	753

Nematostella vectensis (Anthozoa) *vs. Hydra vulgaris* (Hydrozoa)

X-squared = 1140, df = 3, p-value < 2.2e-16

	Low	High
Low	278	143
High	969	1119

Acropora digitifera (Anthozoa) *vs. Aurelia coerulea* (Scyphozoa)

X-squared = 523.46, df = 3, p-value < 2.2e-16

	Low	High
Low	343	526
High	300	968

Acropora digitifera (Anthozoa) *vs. Nematostella vectensis* (Anthozoa)

X-squared = 1637, df = 3, p-value < 2.2e-16

	Low	High
Low	647	260
High	500	1740

Acropora digitifera (Anthozoa) *vs. Hydra vulgaris* (Hydrozoa)

X-squared = 595.96, df = 3, p-value < 2.2e-16

	Low	High
Low	423	894
High	308	1069

Alatina alata (Cubozoa) *vs. Aurelia coerulea* (Scyphozoa)

X-squared = 243.84, df = 3, p-value < 2.2e-16

	Low	High
Low	563	437
High	503	934

Alatina alata (Cubozoa) *vs. Nematostella vectensis* (Anthozoa)

X-squared = 1155.5, df = 3, p-value < 2.2e-16

	Low	High

Low	267	629
High	220	1255

Alatina alata (Cubozoa) *vs. Hydra vulgaris* (Hydrozoa)

X-squared = 299.26, df = 3, p-value < 2.2e-16

	Low	High
Low	554	424
High	743	1027

Alatina alata (Cubozoa) *vs. Acropora digitifera* (Anthozoa)

X-squared = 716.96, df = 3, p-value < 2.2e-16

	Low	High
Low	344	575
High	401	1189

Clytia hemisphaerica (Hydrozoa) *vs. Aurelia coerulea* (Scyphozoa)

X-squared = 73.083, df = 3, p-value = 9.329e-16

	Low	High
Low	352	281
High	207	408

Clytia hemisphaerica (Hydrozoa) *vs. Nematostella vectensis* (Anthozoa)

X-squared = 446.22, df = 3, p-value < 2.2e-16

	Low	High
Low	186	404
High	109	593

Clytia hemisphaerica (Hydrozoa) *vs. Hydra vulgaris* (Hydrozoa)

X-squared = 52.6, df = 3, p-value = 2.231e-11

	Low	High
Low	433	320
High	341	500

Clytia hemisphaerica (Hydrozoa) *vs. Acropora digitifera* (Anthozoa)

X-squared = 218.96, df = 3, p-value < 2.2e-16

	Low	High
Low	239	395
High	193	540

Clytia hemisphaerica (Hydrozoa) *vs. Alatina alata* (Cubozoa)

X-squared = 63.043, df = 3, p-value = 1.315e-13

	Low	High
Low	366	387
High	282	500

Physalia physalis (Hydrozoa) *vs. Aurelia coerulea* (Scyphozoa)

X-squared = 54.11, df = 3, p-value = 1.063e-11

	Low	High
Low	313	293
High	310	462

Physalia physalis (Hydrozoa) *vs. Nematostella vectensis* (Anthozoa)

X-squared = 562.62, df = 3, p-value < 2.2e-16

	Low	High
Low	150	433
High	131	665

Physalia physalis (Hydrozoa) *vs. Hydra vulgaris* (Hydrozoa)

X-squared = 123.3, df = 3, p-value< 2.2e-16

	Low	High
Low	486	300
High	461	628

Physalia physalis (Hydrozoa) *vs. Acropora digitifera* (Anthozoa)

X-squared = 317.6, df = 3, p-value < 2.2e-16

	Low	High
Low	221	376
High	225	640

Physalia physalis (Hydrozoa) *vs. Alatina alata* (Cubozoa)

X-squared = 113.57, df = 3, p-value = 1.315e-13

	Low	High
Low	358	373
High	307	611

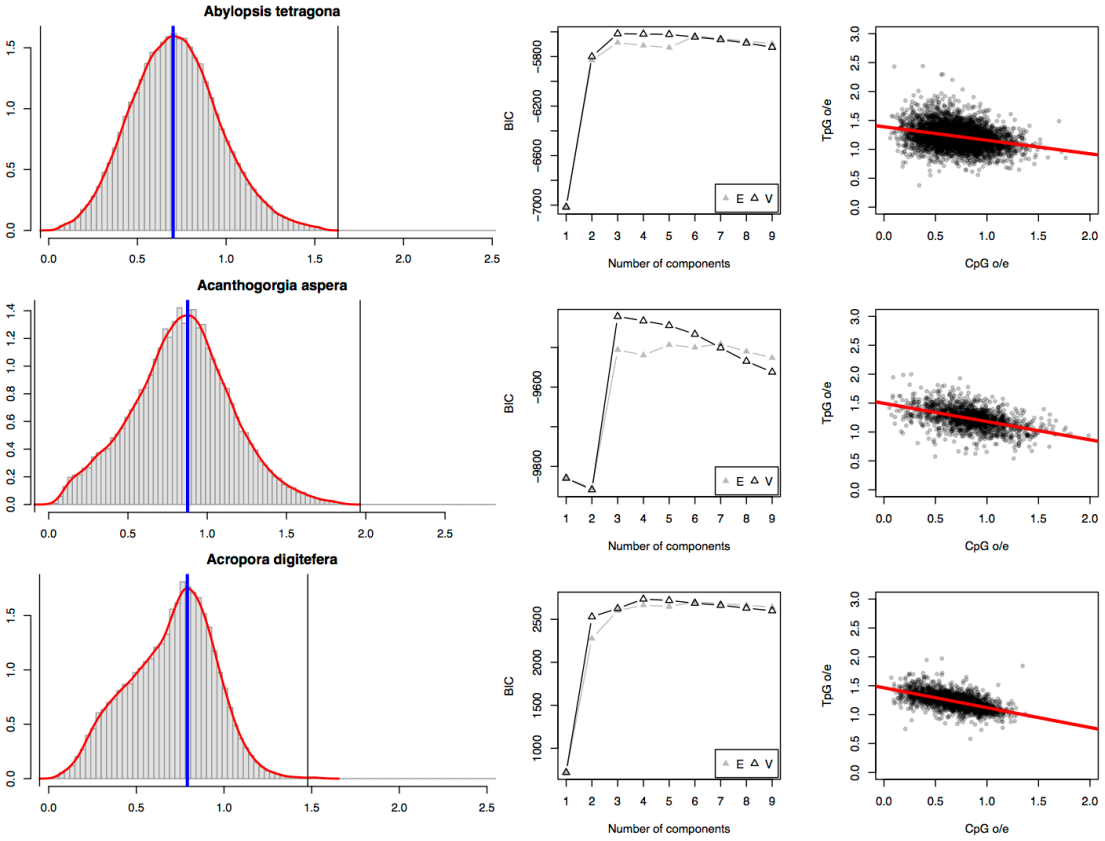
Physalia physalis (Hydrozoa) *vs. Clytia hemisphaerica* (Hydrozoa)

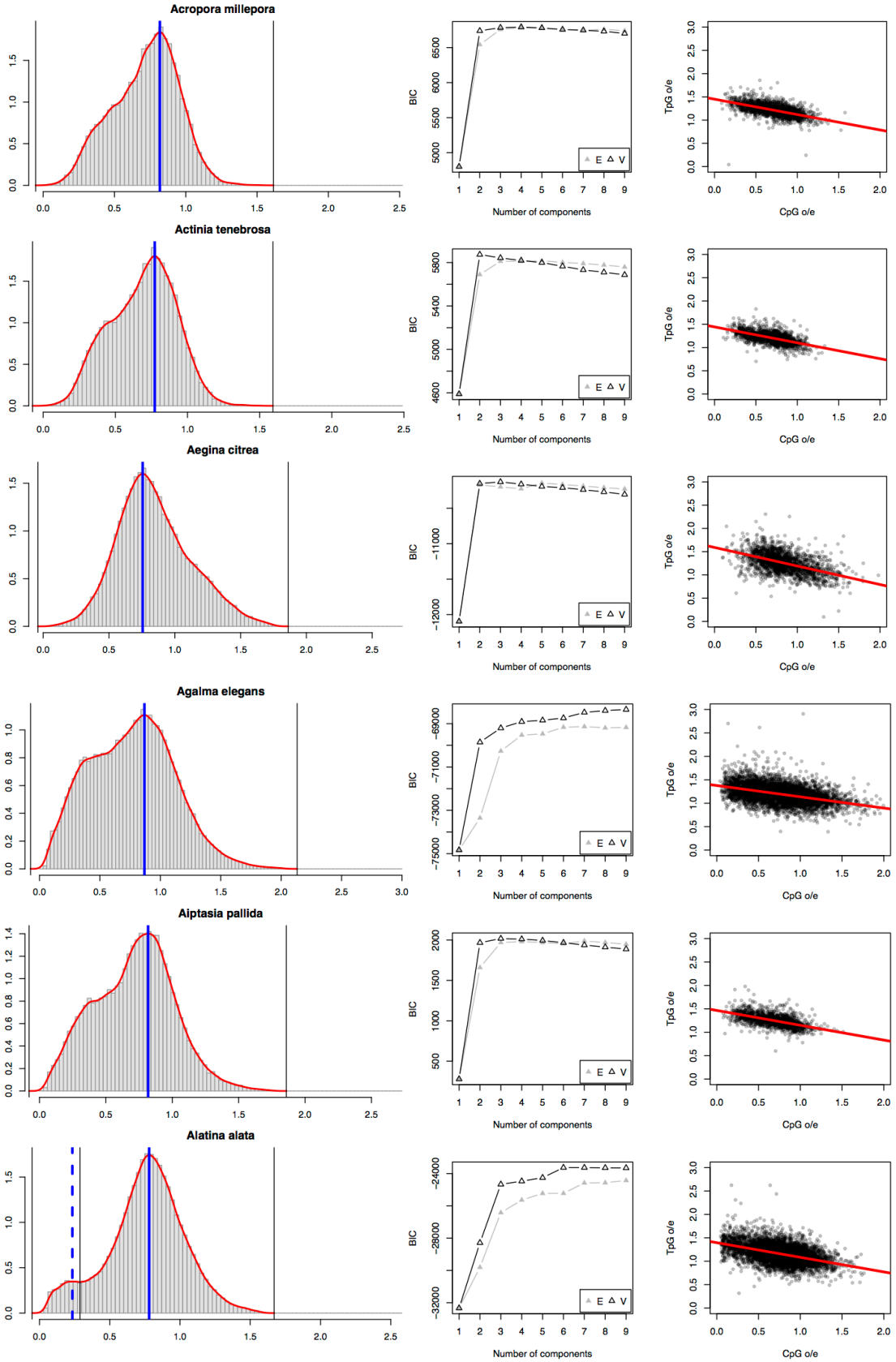
X-squared = 42.297, df = 3, p-value = 2.854e-09

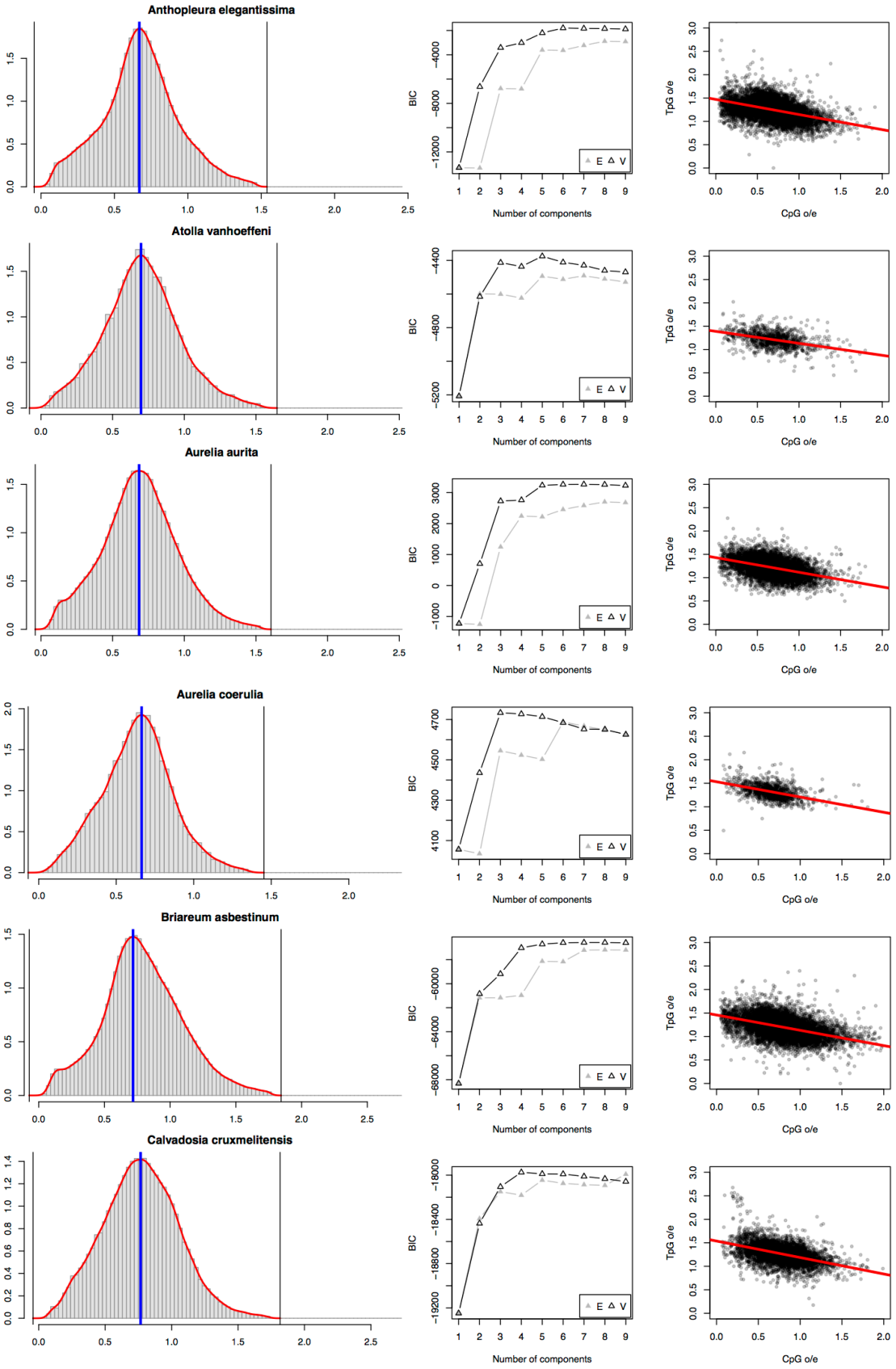
	Low	High
Low	327	242

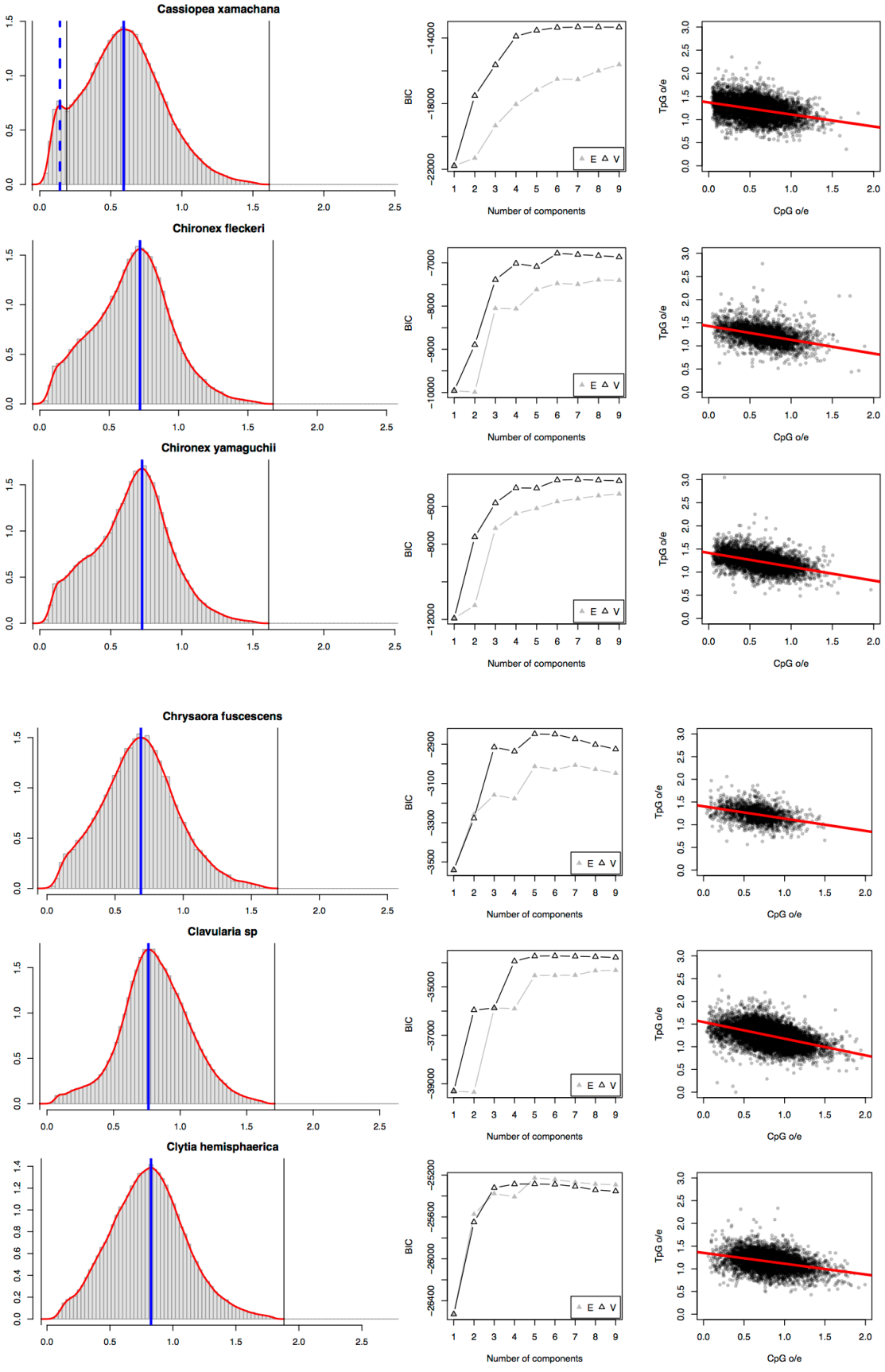
High	281	397

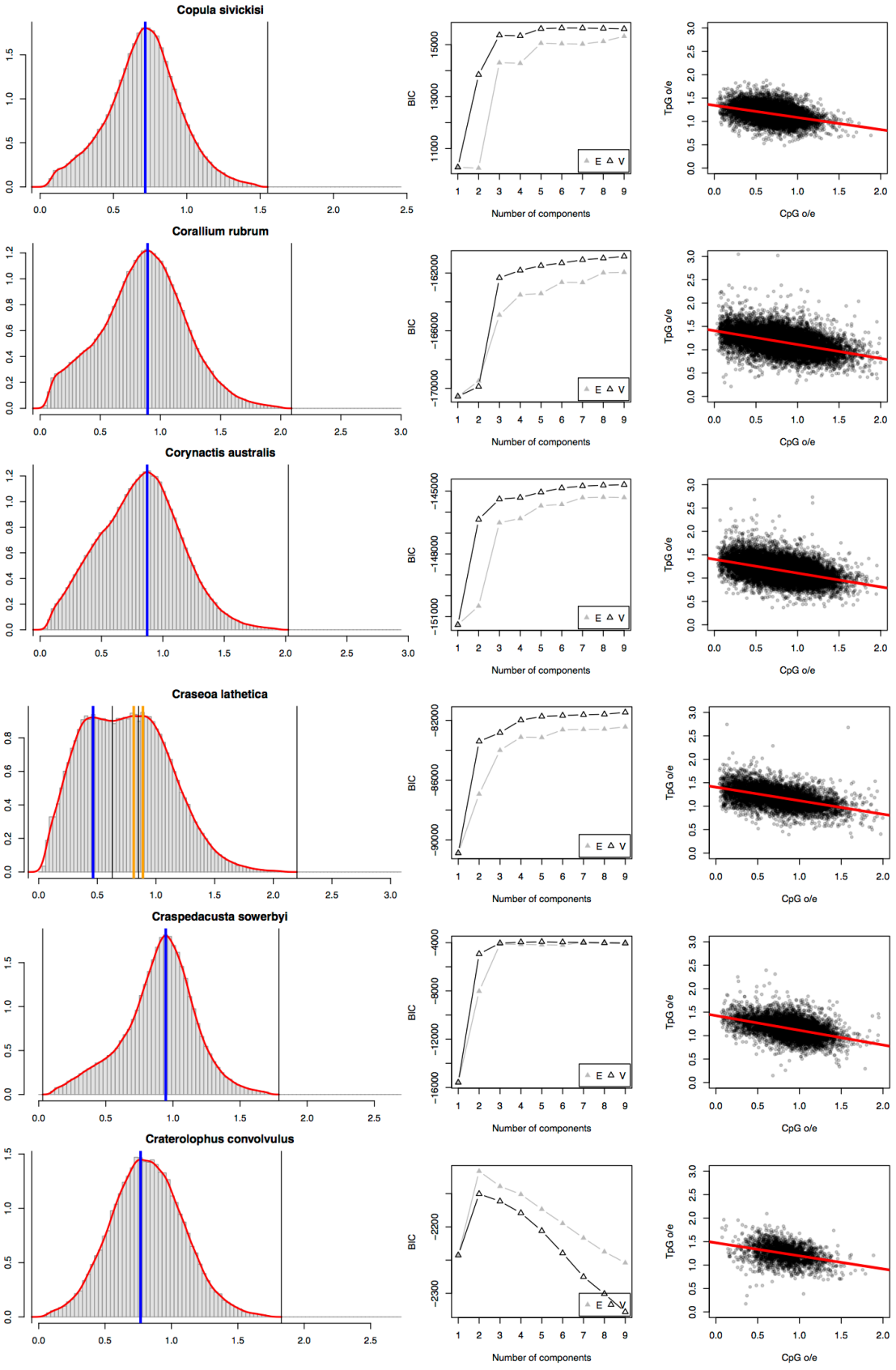
Figure S1. Density distribution of gene body $CpG_{o/e}$ of the 76 species in this study. For each species, the left panel shows the density distribution and mode detection by Notos. The blue or orange vertical lines denote the peaks of modes detected. 11 species are detected to be bimodal. The middle panel shows BIC scores of Gaussian mixture models. Most species are best described by mixture models with three or more components. The right panel shows correlation between $CpG_{o/e}$ and $TpG_{o/e}$ to verify that the depletion of CpG is indeed due to cytosine methylation.

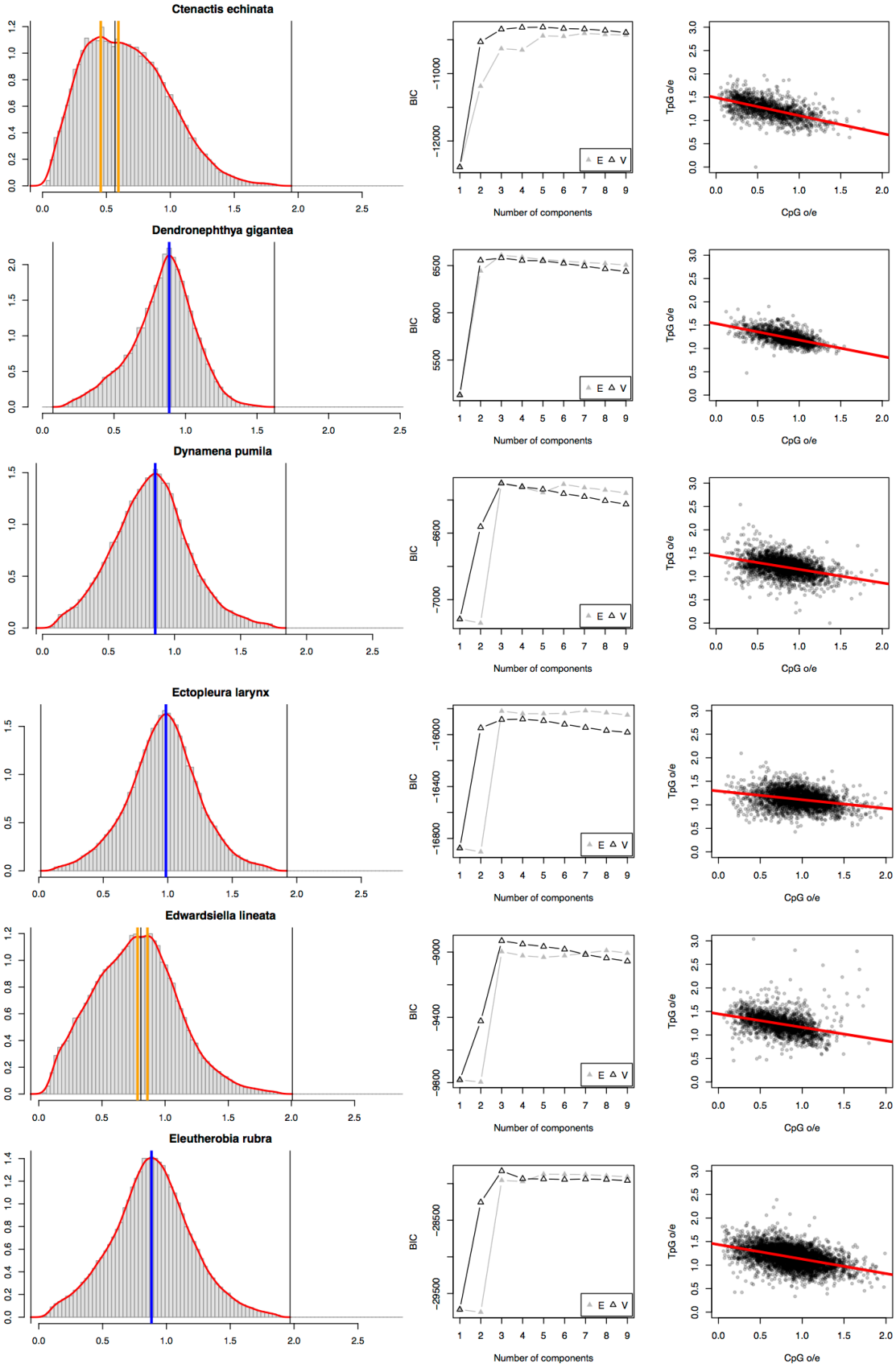


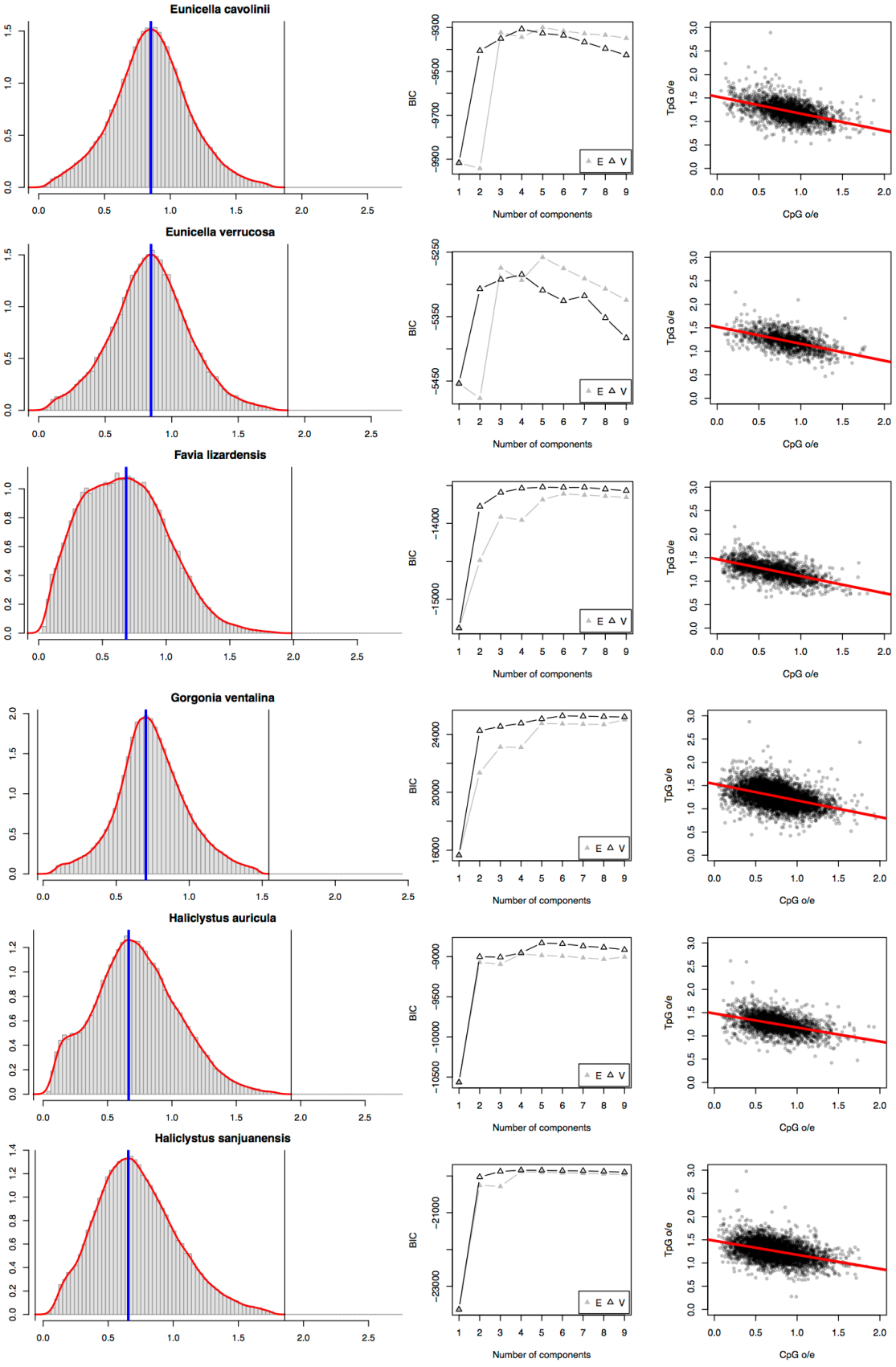

Acropora millepora
Actinia tenebrosa
Aegina citrea
Agalma elegans
Aiptasia pallida
Alatina alata
BIC
Number of components
E V
TpG o/e
CpG o/e

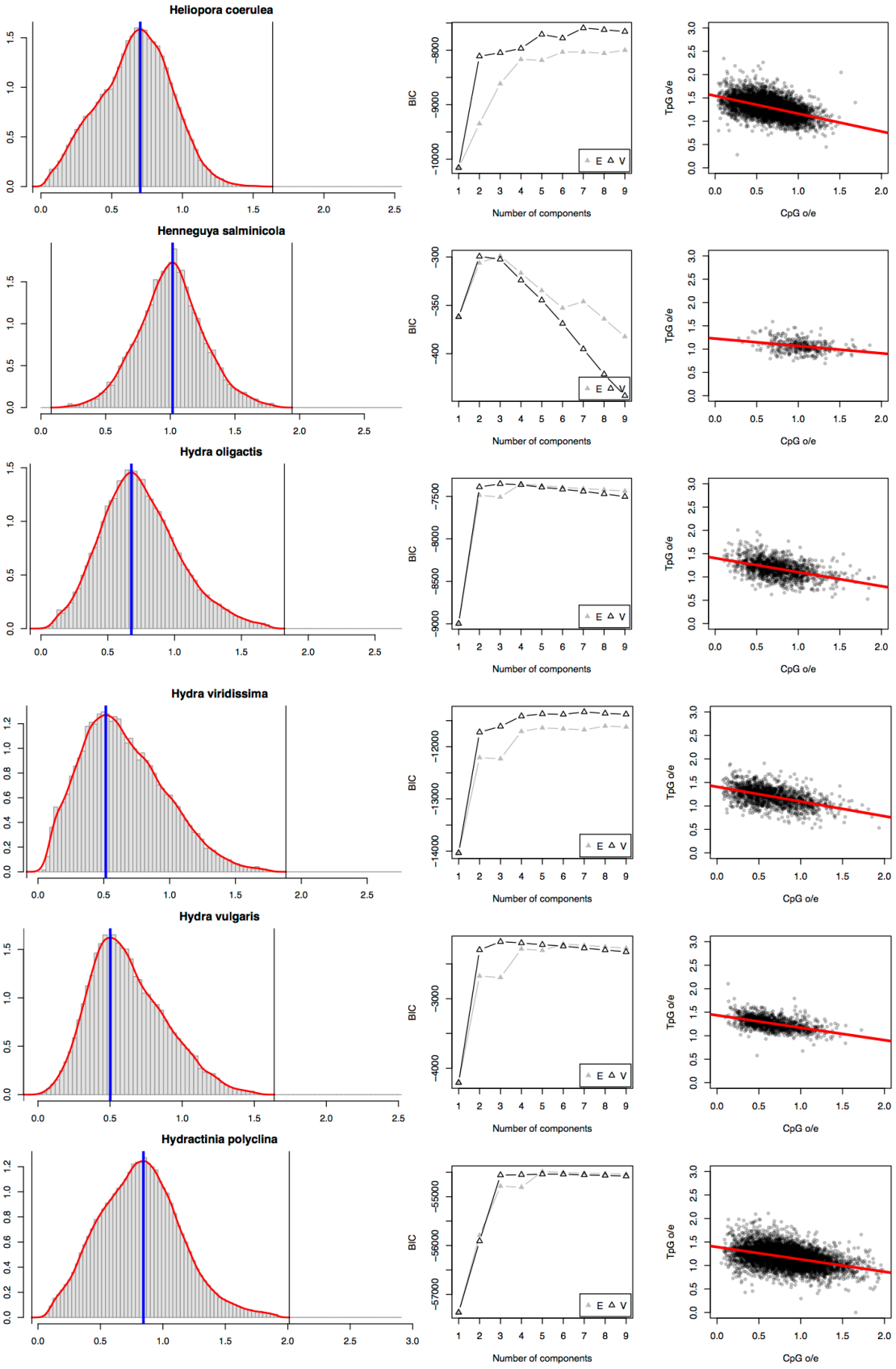

Anthopleura elegantissima
Atolla vanhoeffeni
Aurelia aurita
Aurelia coerulia
Briareum asbestinum
Calvadosia cruxmelitensis
BIC
Number of components
E V
TpG o/e
CpG o/e

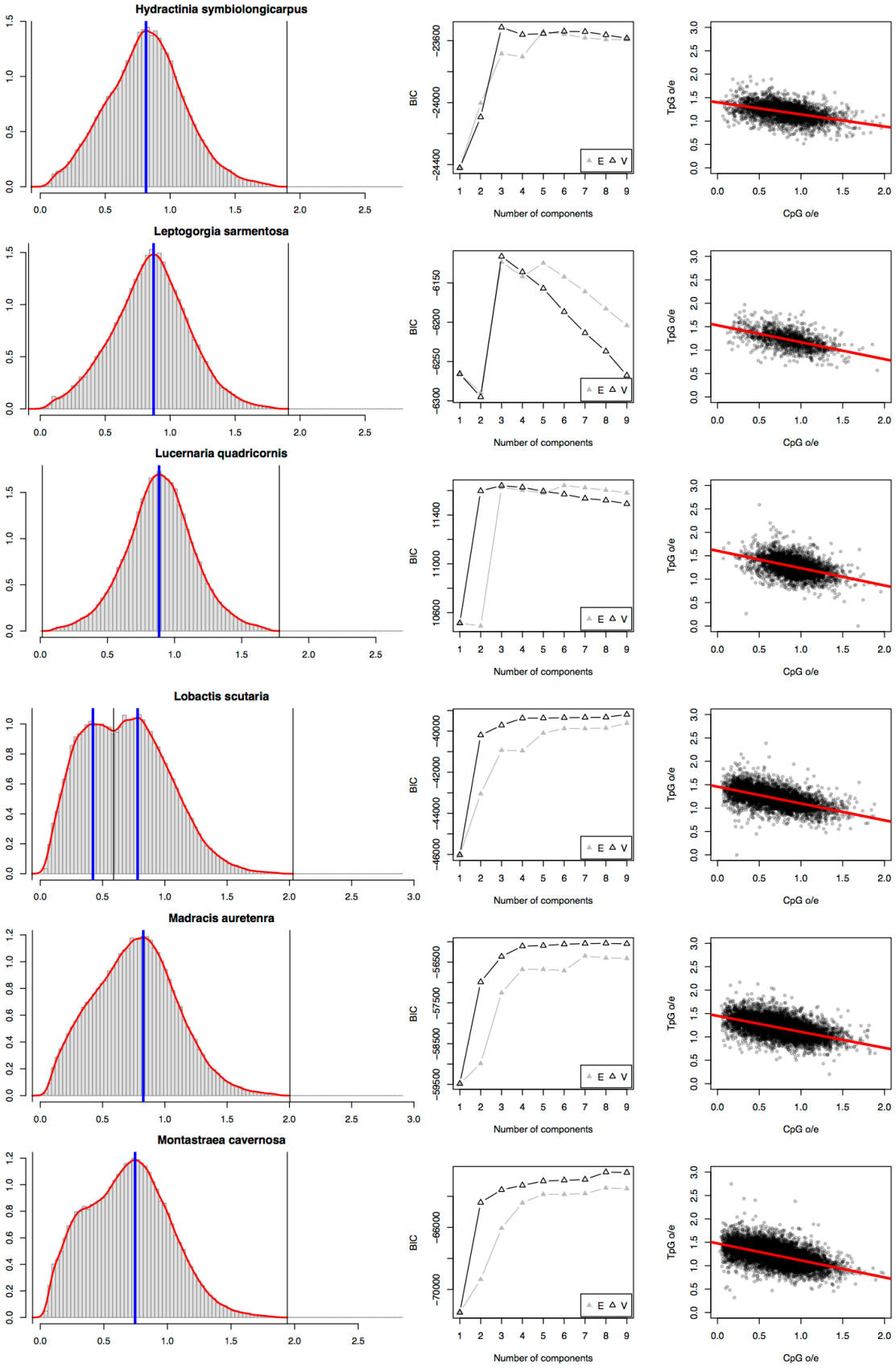

Cassiopea xamachana
Chironex fleckeri
Chironex yamaguchii
Chrysaora fuscescens
Clavularia sp
Clytia hemisphaerica
BIC
Number of components
E V
TpG o/e
CpG o/e

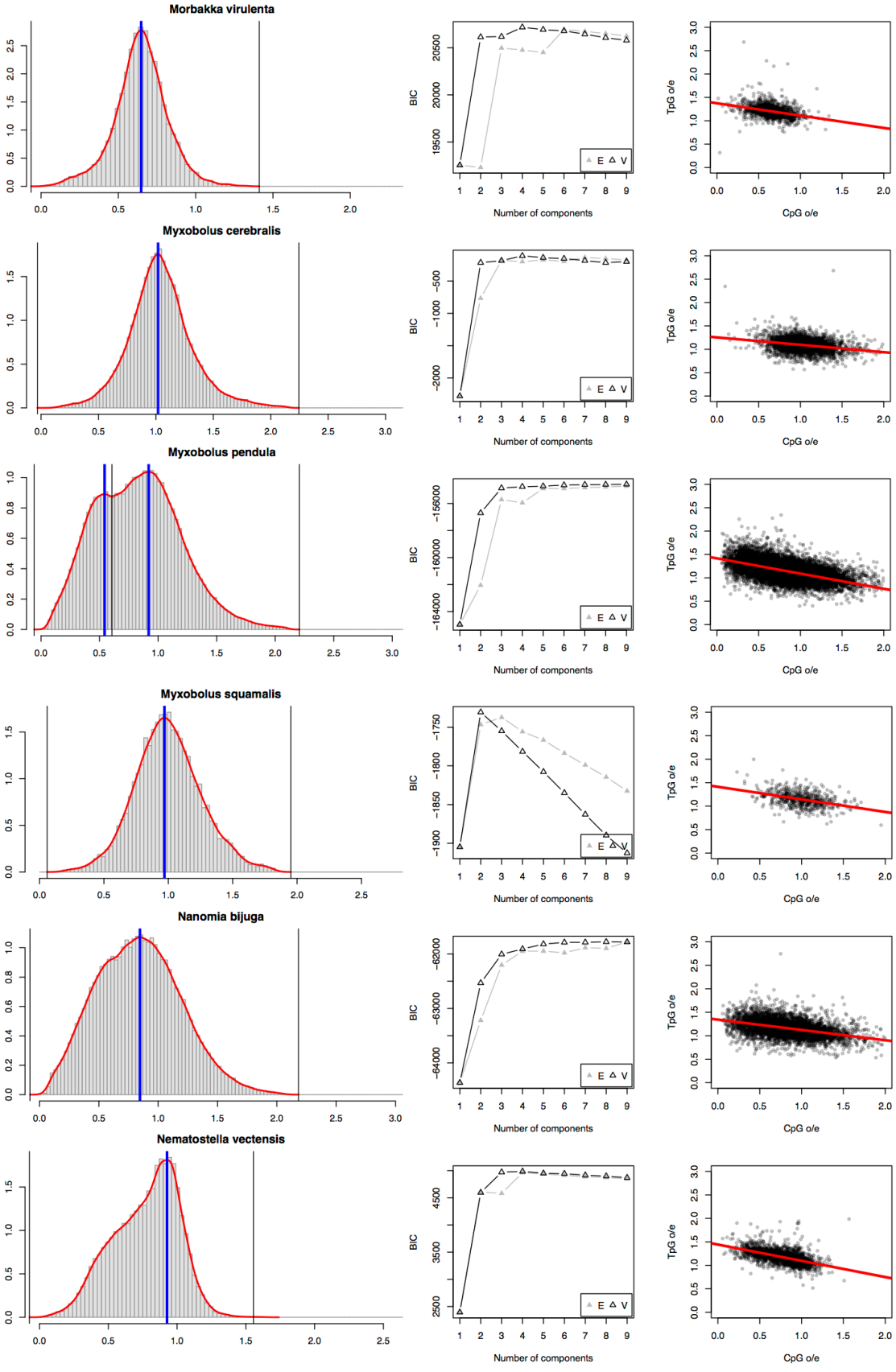

Copula sivickisi
Corallium rubrum
Corynactis australis
Craseoa lathetica
Craspedacusta sowerbyi
Craterolophus convolvulus
BIC
Number of components
E V
TpG o/e
CpG o/e

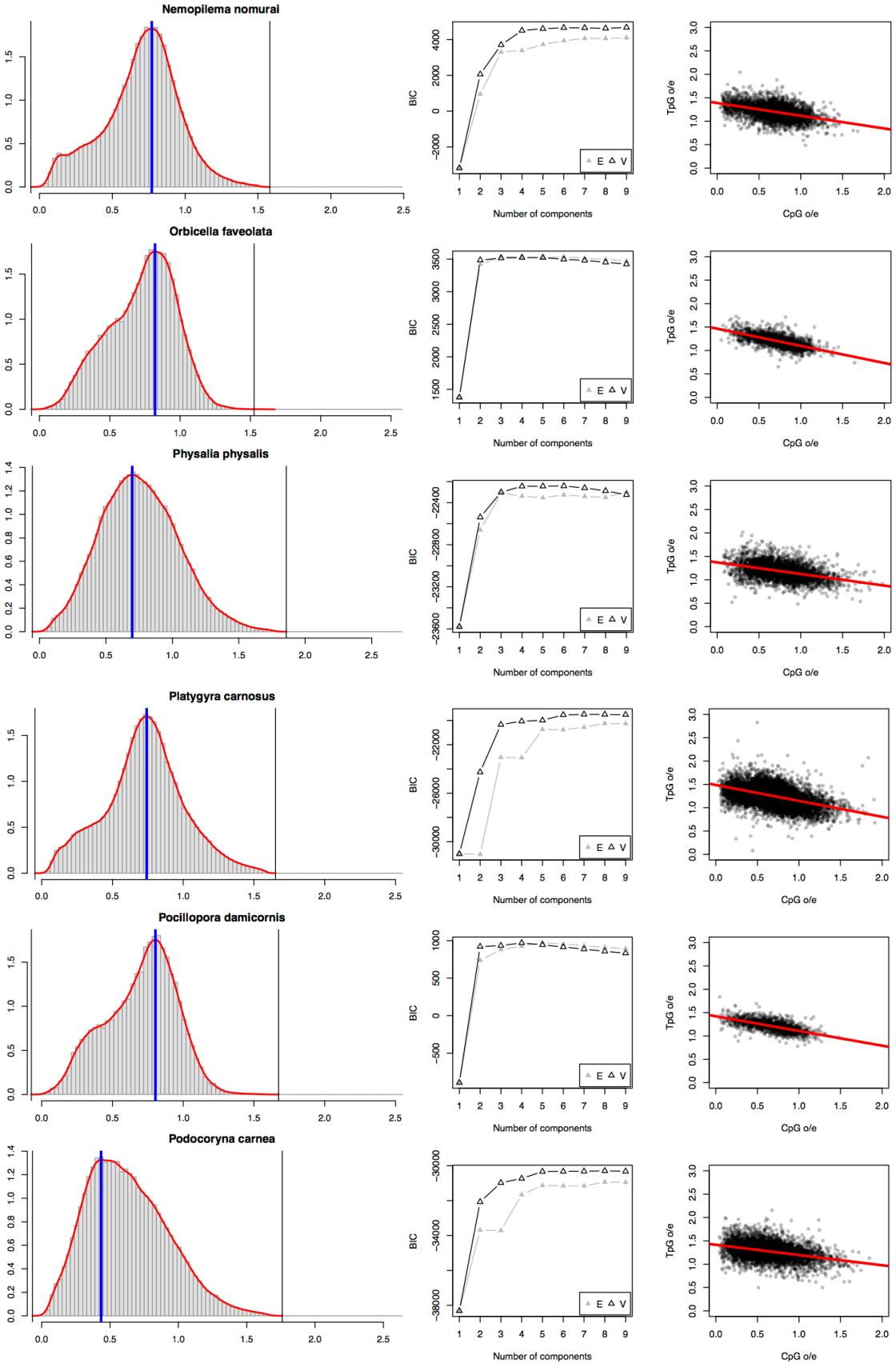

Ctenactis echinata
BIC
Number of components
E V
TpG o/e
CpG o/e
Dendronephthya gigantea
BIC
Number of components
E V
TpG o/e
CpG o/e
Dynamena pumila
BIC
Number of components
E V
TpG o/e
CpG o/e
Ectopleura larynx
BIC
Number of components
E V
TpG o/e
CpG o/e
Edwardsiella lineata
BIC
Number of components
E V
TpG o/e
CpG o/e
Eleutherobia rubra
BIC
Number of components
E V
TpG o/e
CpG o/e

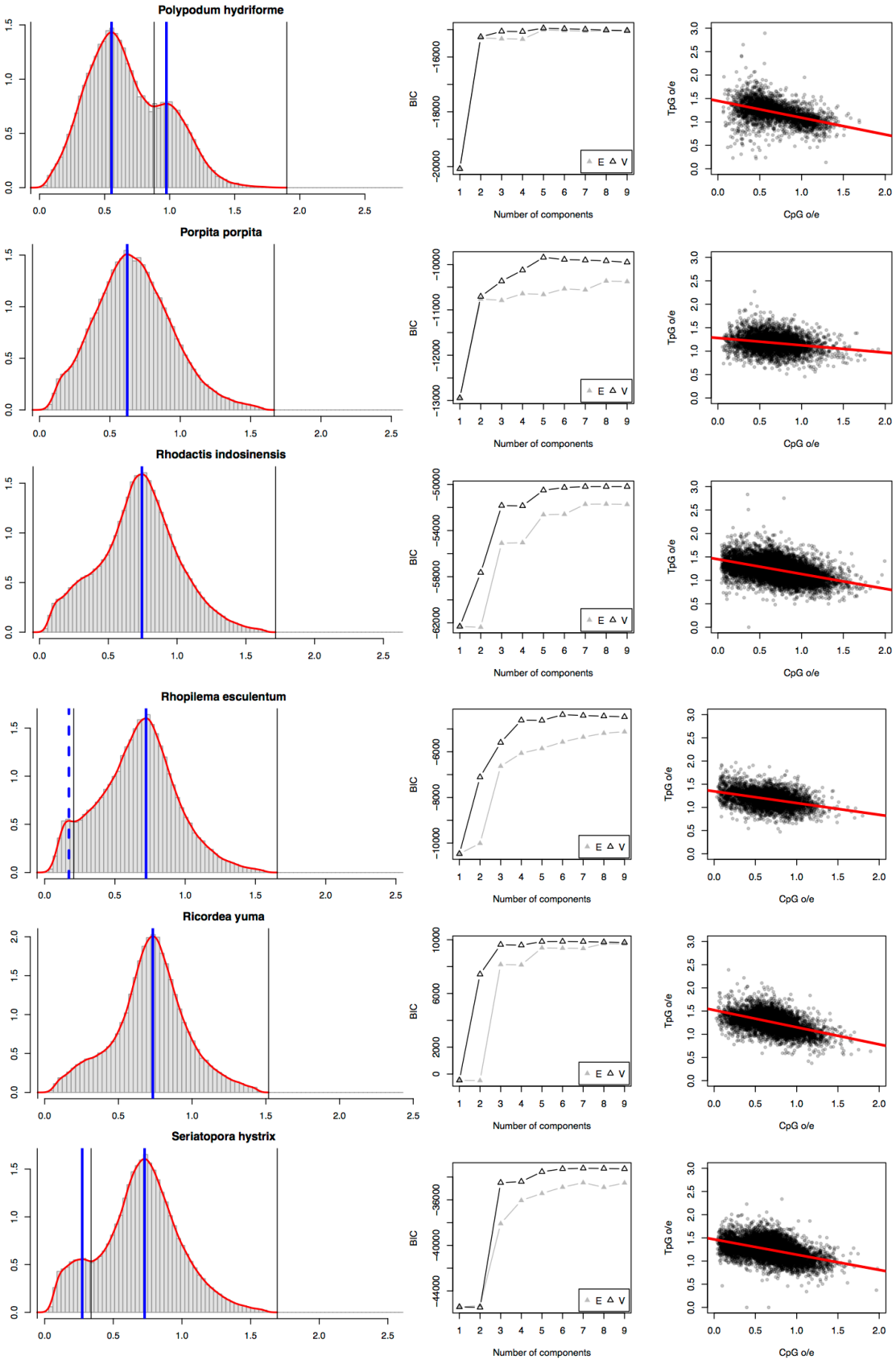

Eunicella cavolinii
Eunicella verrucosa
Favia lizardensis
Gorgonia ventalina
Haliclystus auricula
Haliclystus sanjuanensis
BIC
Number of components
E V
TpG o/e
CpG o/e

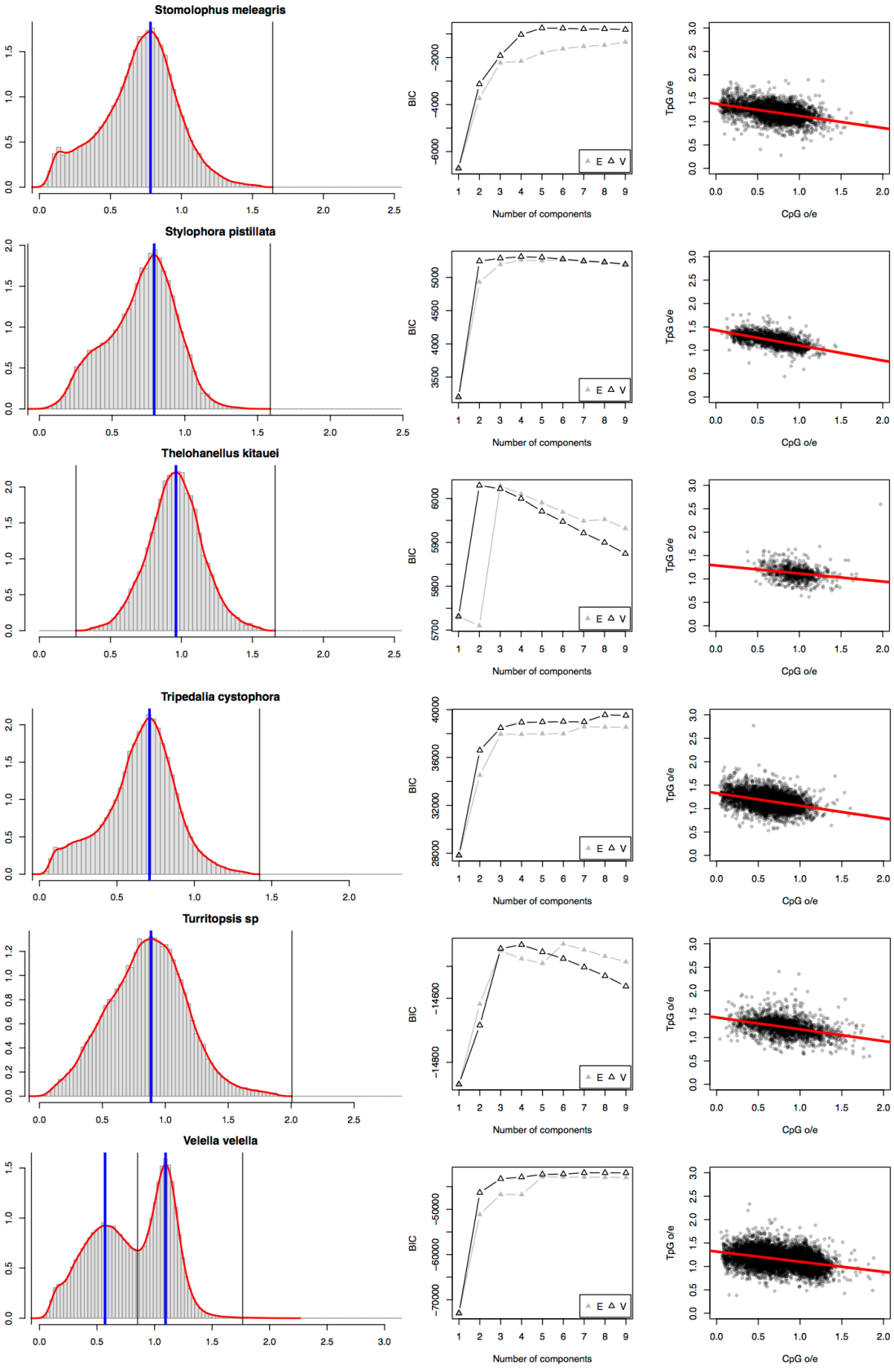

Heliopora coerulea
Henneguya salminicola
Hydra oligactis
Hydra viridissima
Hydra vulgaris
Hydractinia polyclina
BIC
Number of components
E V
TpG o/e
CpG o/e

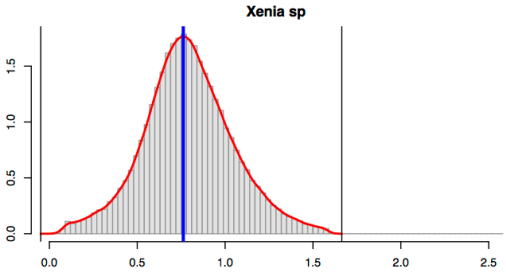

Hydractinia symbiolongicarpus
Leptogorgia sarmentosa
Lucernaria quadricornis
Lobactis scutaria
Madracis auretenra
Montastraea cavernosa
BIC
Number of components
E V
TpG o/e
CpG o/e

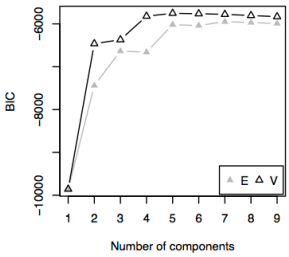

Morbakka virulenta
Myxobolus cerebralis
Myxobolus pendula
Myxobolus squamalis
Nanomia bijuga
Nematostella vectensis
BIC
Number of components
E V
TpG o/e
CpG o/e


Nemopilema nomurai
Orbicella faveolata
Physalia physalis
Platygyra carnosus
Pocillopora damicornis
Podocoryna carnea
BIC
Number of components
E V
TpG o/e
CpG o/e


Polypodum hydriforme
Porpita porpita
Rhodactis indosinensis
Rhopilema esculentum
Ricordea yuma
Seriatopora hystrix
BIC
Number of components
E V
TpG o/e
CpG o/e


Stomolophus meleagris
Stylophora pistillata
Thelohanellus kitauei
Tripedalia cystophora
Turritopsis sp
Velella velella
BIC
Number of components
E V
TpG o/e
CpG o/e


Xenia sp
0.0
0.5
1.0
1.5
2.0
2.5


BIC
−6000
−8000
−10000
1
2
3
4
5
6
7
8
9
Number of components
E
V

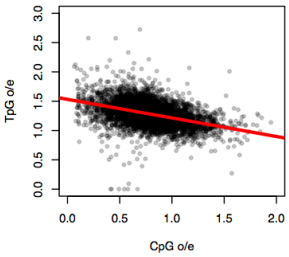
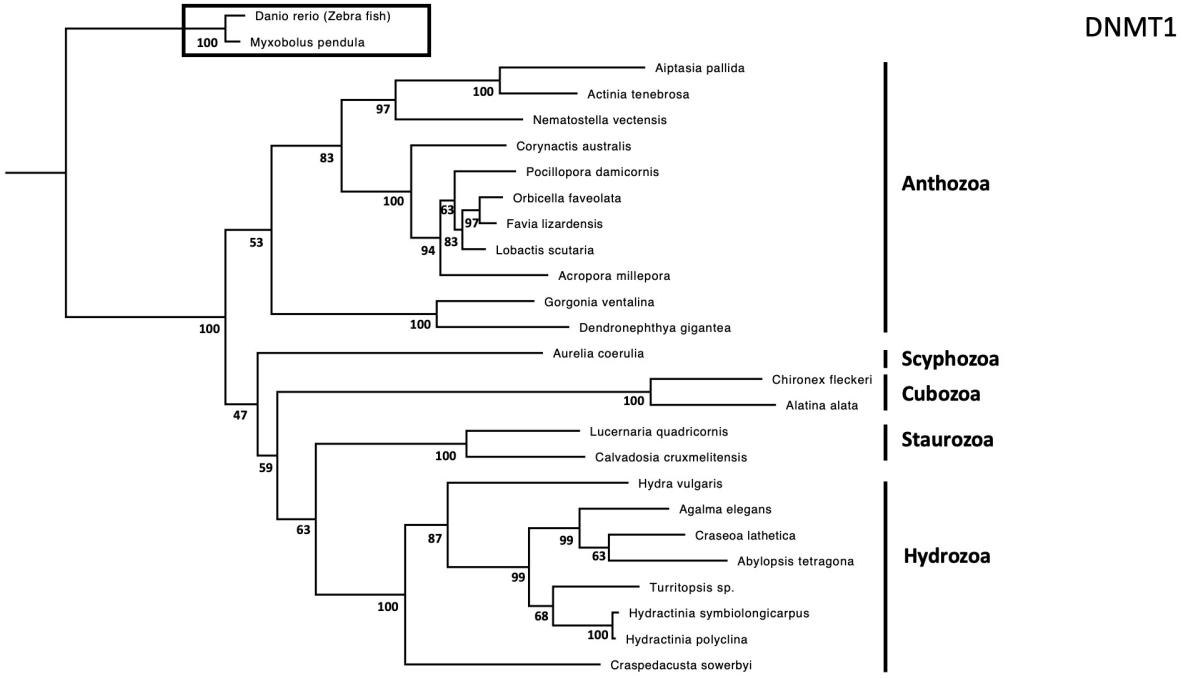
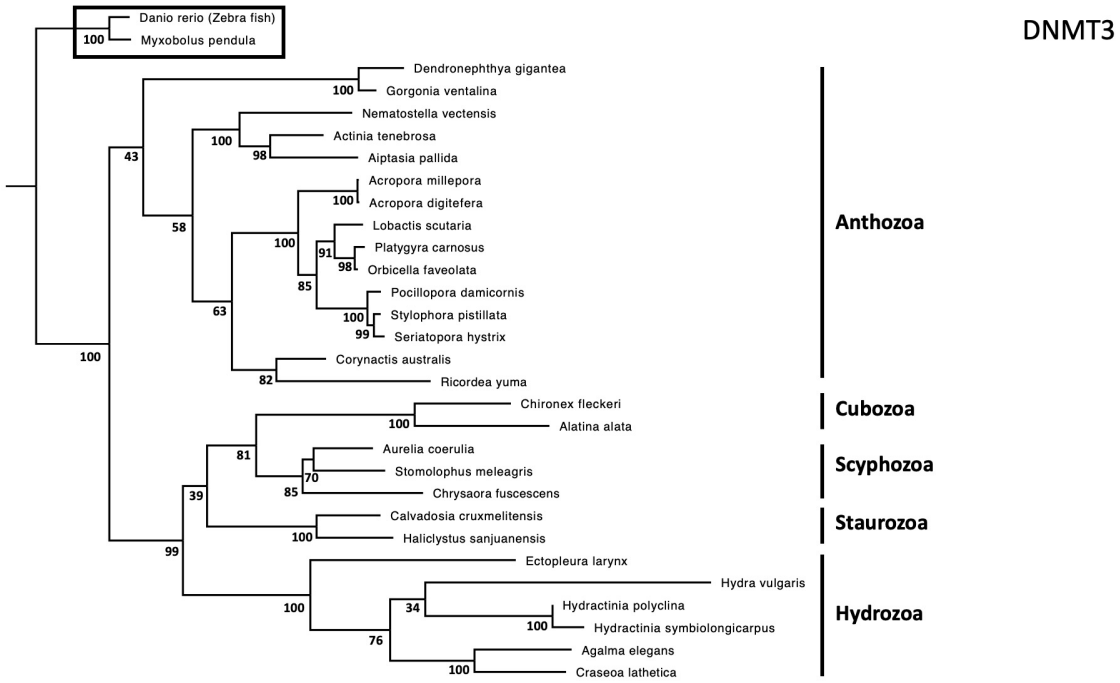

TpG o/e
0.0
0.5
1.0
1.5
2.0
2.5
3.0
CpG o/e

Figure S2. Phylogenetic analyses of DNMT1 and DNMT3. Sequences found in *Myxobolus pendula* are more closely related to corresponding sequences found in the zebrafish than the other Cnidarians.





Supplemental References:

Brekhman V, Malik A, Haas B, Sher N, Lotan T. 2015. Transcriptome profiling of the dynamic life cycle of the scypohozoan jellyfish Aurelia aurita. *BMC genomics* 16:74.

Hasegawa Y, Watanabe T, Takazawa M, Ohara O, Kubota S. 2016. De novo assembly of the transcriptome of Turritopsis, a jellyfish that repeatedly rejuvenates. *Zoological science* 33:366–371.

Okubo N, Hayward DC, Forêt S, Ball EE. 2016. A comparative view of early development in the corals Favia lizardensis, Ctenactis echinata, and Acropora millepora-morphology, transcriptome, and developmental gene expression. *BMC evolutionary biology* 16:48.

Stefanik DJ, Lubinski TJ, Granger BR, Byrd AL, Reitzel AM, DeFilippo L, Lorenc A, Finnerty JR. 2014. Production of a reference transcriptome and transcriptomic database (EdwardsiellaBase) for the lined sea anemone, Edwardsiella lineata, a parasitic cnidarian. *BMC genomics* 15:71.

CHAPTER 2

Stage-Specific Methylomes of a Medusa-Bearing Jellyfish

Abstract

As a Scyphozoan jellyfish, *Aurelia coerulea* has a complex life cycle underlined by dynamic expression profiles. To better understand the potential role of DNA methylation in regulating gene expressions, we conducted whole-genome bisulfite sequencing studies of different stages of the *Aurelia* life cycle. *Aurelia* DNA methylation shares similarities with bilaterian invertebrates, including gene bodies being the primary target of CpG methylation, a sharp peak of CpG methylation immediately following the transcription start site, and that genes with low methylation being more dynamically expressed. Interestingly, we identify differentially methylated genes across stages, yet changes in methylation do not correlate with changes in expression, a phenomenon previously noted in bilaterian invertebrates as well. Overall, our results suggest that the role of DNA methylation on gene bodies relates more to stabilizing expression via repressing spurious transcription start sites.

Introduction

To better understand the role of DNA methylation in regulating life history transitions, and the evolution of DNA methylation patterns in Metazoa, we report the first whole-genome methylome study on *Aurelia coerulea* ('species 1' sensu, Dawson and Jacobs). Previous studies on *Aurelia coerulea* have shown that the complex life history is underlain by dynamic gene expression changes across stages (Gold et al 2018). This study surveys DNA methylation at different stages of *Aurelia* life history with bisulfite sequencing to assess the potential role of methylation in regulating gene expression across life history transitions.

The *Aurelia* life cycle incorporates a larval stage ('planula'), a bottom-dwelling stage ('polyp'), and a free-swimming stage ('ephyra') that subsequently develops into the adult medusa stage (Figure 1). The *Aurelia* medusa can reproduce sexually and produce planula (Gold et al 2018; Brekhman et al 2015). Additionally, medusae possesses complex sensory structures called 'rhopalia', which are located are the bell margins of the medusa (Nakanishi et al 2009). *Aurelia* undergoes dramatic metamorphic changes between stages in its life history, and previously Fuchs et al (2014) showed that the transition from the polyp stage to the ephyra stage in *Aurelia*, also known as 'strobilation', was blocked when polyps were treated with 5-azacytidine, a DNA methyltransferase inhibitor. This suggests a role for DNA methylation in regulating life history transitions (Fuchs et al 2014). Studies in several bilaterian invertebrates also indicate that DNA methylation is important in regulating life history, such as in the honeybee (Kucharski et al 2008) and the Pacific oyster (Riviere et al 2013).

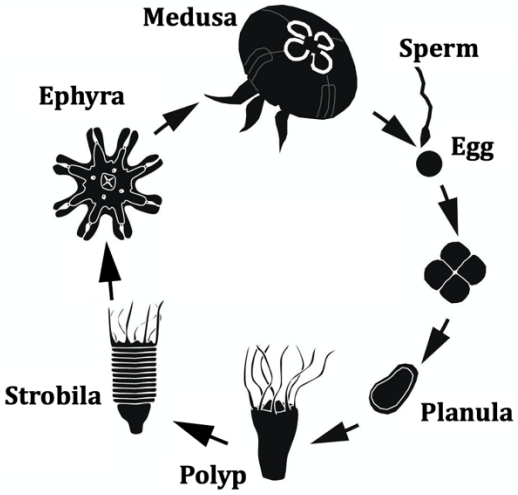


Figure 1. The life cycle of *Aurelia* comprises multiple stages of drastically different morphology and ecology, including a free-swimming medusa stage. Figure adapted from Fuchs et al 2014.

Besides the complex life history, *Aurelia* is also of interest for DNA methylation studies because of its phylogenetic placement. As a member of Cnidaria, the sister group to Bilateria, *Aurelia* could provide insights into the evolution of DNA methylation in early animals. In particular, previous work on DNA methylation in Cnidaria is limited to a few species in Anthozoa, Myxozoa, and Hydrozoa, and none has a complex life history with a medusa stage. Thus, it is unknown how DNA methylation changes throughout the life history in any medusa-bearing species which includes the Scyphozoa, the true jellyfish such as *Aurelia*. Interestingly, in a recent broad survey of DNA methylation in Cnidaria, species with complex life history showed genomic signature of higher methylation, suggesting a correlation between DNA methylation and life history complexity (Zhang and Jacobs 2022). The present study on stage-specific DNA methylation in *Aurelia* therefore has implications for the evolution of DNA methylation and its role in life history evolution in animals. For the rest of the introduction, we review the patterns of DNA methylation in Metazoa, and the hypothesized functions of DNA methylation in regulating gene expression in invertebrates.

General Description

The addition of a methyl group to cytosines is a common epigenetic modification found in eukaryotes (Feng et al 2010). In Metazoa, cytosines in cytosine-phospho-guanine dinucleotides ('CpG sites') are the predominant target of methylation, instead of CHH or CHG sites (H=A, C, or T). Interestingly, the overall level and pattern of CpG methylation varies greatly across Metazoa. Vertebrate genomes are heavily methylated except for promoters with high CpG content (Zemach et al 2010). In bilaterian invertebrates studied thus far, methylation at gene bodies ('transcription units') tends to be higher than methylation at transposable elements (TEs), and exhibits a 'mosaic' or

'bimodal' distribution pattern (Zemach et al 2010; Feng et al 2010; Suzuki et al 2007; Zhang & Jacobs 2022). Loss of cytosine methylation have been observed in several species, including the fruit fly *Drosophila melanogaster,* the nematode *Caenorhabditis elegans*, and several parasitic cnidarian species (Wenzel et al 2011; Takayama et al 2014; Kyger et al 2020; Zhang & Jacobs, 2022). Sarda et al. among others (2012) hypothesized that gene body methylation (GbM) was the ancestral DNA methylation pattern in animals. Using a proxy of DNA methylation, Zhang & Jacobs, 2022 reported prevalent GbM in Cnidaria, the sister group of Bilateria, suggesting that GbM was present in the last common ancestor of Eumetazoa.

Complex relationships between methylation and gene expression

There are multiple patterns relating the methylation and gene expression in animals. Some taxa show a direct relationship where expression increases with GbM, such as *Aiptasia,* an anemone with photosynthetic endosymbionts (Li et al 2018). Other taxa show a nonlinear relationship where both high and low methylation are correlated with low expression. This pattern is evident in several coral species, the honeybee *Apis mellifera*, the tunicate *Ciona intestinalis*, as well as human (Dimond and Roberts 2016; Zemach et al 2010; Jjingo et al 2012). Given this variation it is hard to ascribe universal function of gene body methylation across animals. It has been hypothesized that GbM represses spurious transcription initiation by repressing intragenic transcription start sites (Neri et al 2017; Huh et al 2013; Li et al 2018) and/or regulates alternative splicing events (Flores et al 2012; Shayevitch et al 2018; Lister et al 2009), but direct evidence for the exact mechanism is still needed.

Previous bisulfite studies in Cnidaria have been largely limited to the class Anthozoa, the parasitic clade Myxozoa, and *Hydra* in Hydrozoa, all of which lack a

medusa stage in their life cycle (Zemach et al 2010; Dixon et al 2014; Dixon et al 2016; Dimond and Roberts 2016; Li et al 2018; Kyger et al 2020; Liew et al 2018; Ying et al 2022). Medusa bearing forms including the true jellies, Scyphozoa, that have complex life histories have not been studied. Zhang and Jacobs (2022) reported a broad survey of DNA methylation in Cnidaria using a proxy, and described elevated levels of GbM in Scyphozoa suggesting GbM associated with life history complexity. Here we present the first whole-genome methylome study in a scyphozoan jellyfish *Aurelia*.

Results

The global methylation rate of cytosines in all contexts (CpG, CHG, CHH) is 1.20%, with significant enrichment in CpG sites (the methylation rates of CpG, CHG, and CHH are 7.08%, 0.32%, and 0.29%, respectively). CpG methylation ranges from 6.40% to 7.11% between stages. Consistent with other invertebrate taxa, CpG methylation is enriched in gene bodies; and within gene bodies, intron methylation tends to be heavier than exons, which is consistent with bisulfite studies on other cnidarians (Figure 2) (Ying et al 2022).

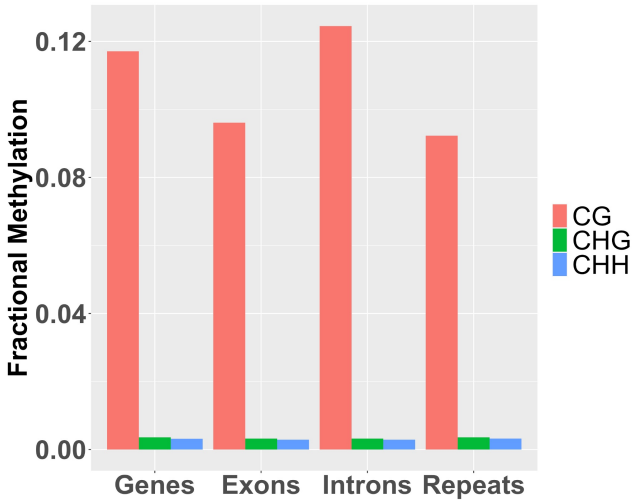


Figure 2. Methylation is predominantly focused on cytosines in CpG context, and enriched in gene bodies, especially introns, relative to repetitive elements.

Methylation on gene bodies tends to be heavier towards the 5' end with a peak immediately following the transcription start site ('TSS') and decreases gradually towards the transcription end site ('TES') (Figure 3a). Coincidently, CpG sites are more frequent near the transcription start site, and the increase in CpG density in the region 5' to TSS coincides with the decrease in methylation (Figure 3b). Interestingly, despite the decrease in GbM from TSS to TES, the density of CpG remains constant across gene bodies. Like other invertebrates with substantial DNA methylation, GbM exhibits a bimodal distribution, where genes generally fall into two classes with either high or low methylation (Figure 3c). Differentially expressed genes identified in Gold et al 2018 tend to have lower methylation. Zhang and Jacobs (2022) noted a lack of discrete bimodality in the distribution of gene body CpG o/e. Here we note that GbM measured by bisulfite sequencing demonstrates more discrete bimodality with an obvious local minimum. Nevertheless, GbM is negatively correlated with gene body CpG o/e as expected given the hypermutability of methylated cytosines (Figure 3d).

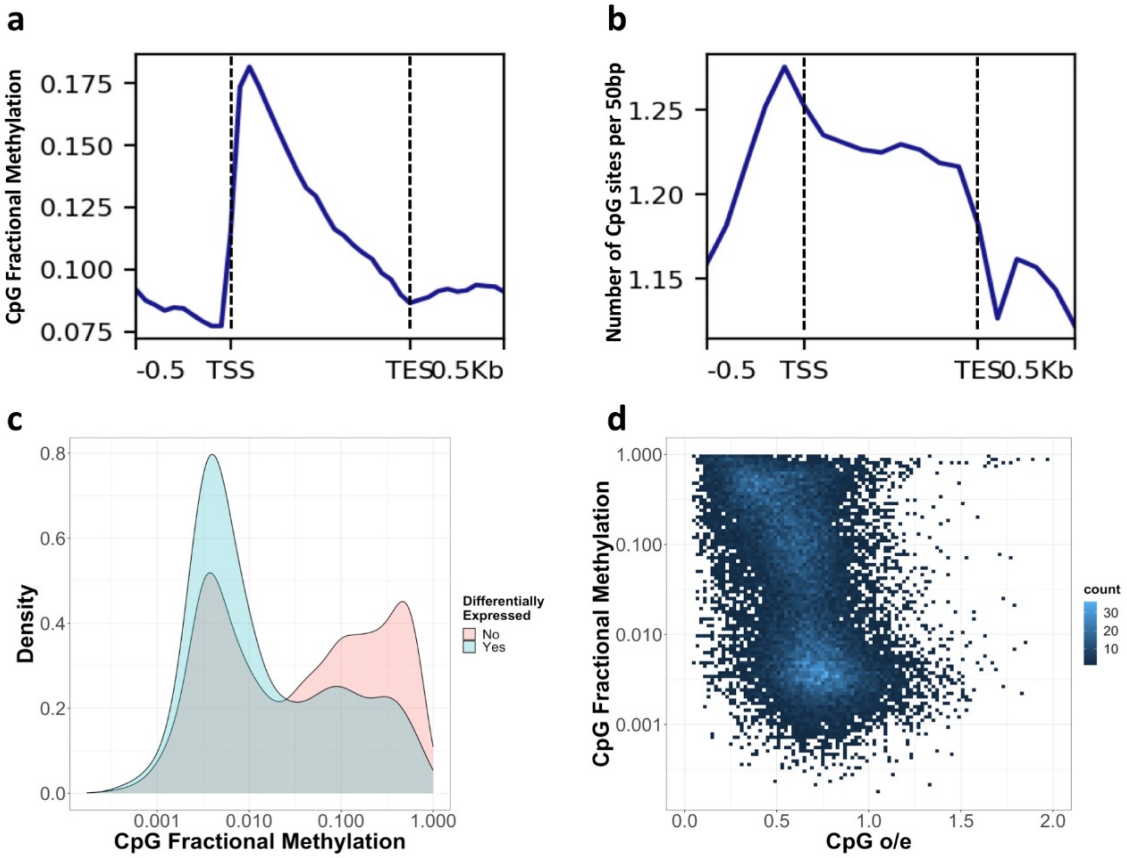


Figure 3. a. CpG methylation is enriched in gene bodies, with a sharp peak at the transcription start site and gradually decreases towards the 3' end of genes. b.CpG content in *Aurelia* genes peaks preceding the transcription start site. c. *Aurelia* GbM shows a bimodal distribution that is typically seen in invertebrates; differentially expressed genes are enriched in the low methylation section. d. GbM and CpG o/e are negatively correlated due to the hypermutability of methylated cytosines, which leads to the depletion of cytosines over evolutionary time.

At each stage, GbM has a bell-shaped correlation with expression (Figure 4). Expression increases with methylation initially, then decreases as methylation continues to increase, and both high and low methylation are correlated with low expression whereas moderately methylated genes show the highest expression. We identified 17218, 5028, 19580 differentially methylated regions (DMRs) at the planula-polyp, polyp-ephyra, and ephyra-medusa transitions respectively; of these, 3483, 1394, and 3734 overlap with genes, and many of these differentially methylated genes are shared

between the transitions (Figure S2); however, the differences in GbM across stages do not significantly correlate with changes in transcription (Figure 5).

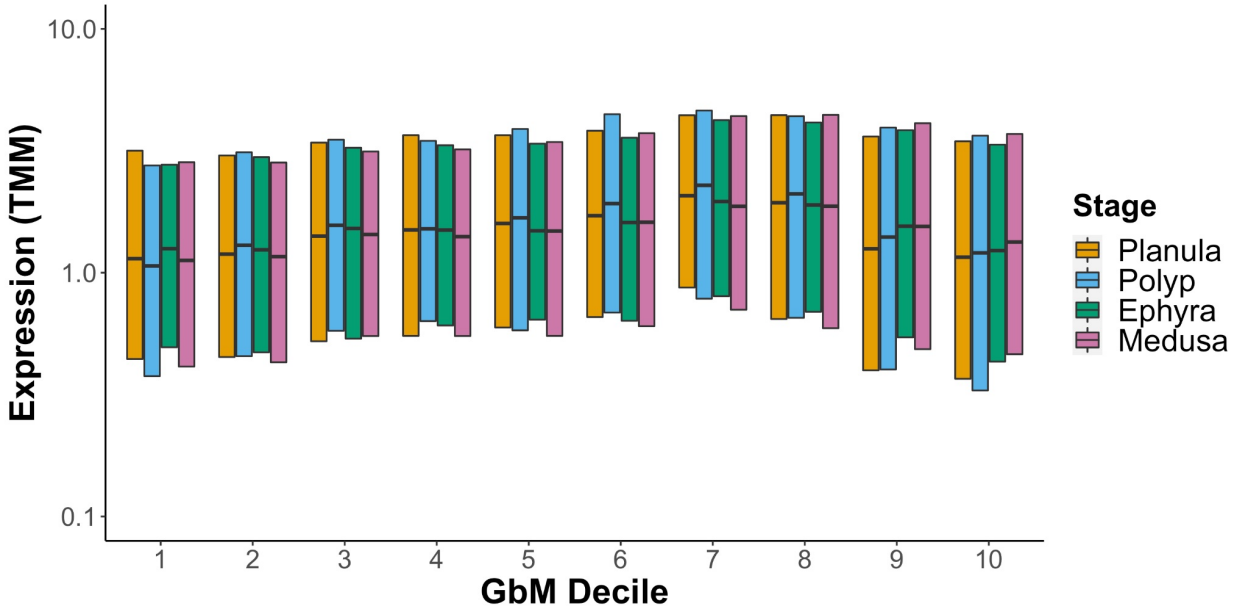


Figure 4. Expression and GbM have a bell-shaped correlation.

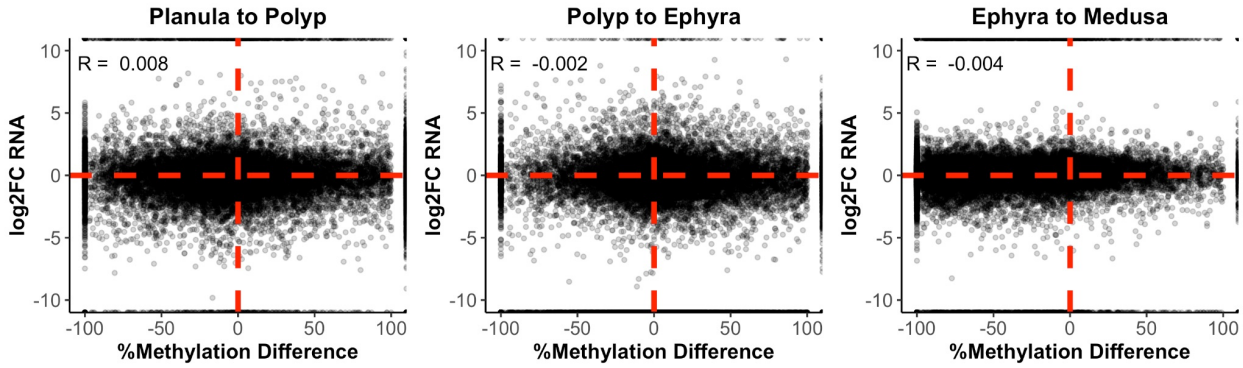


Figure 5. Changes in GbM are not correlated with changes in transcription at each life history transition.

Discussion

In this study, we find that the general patterns of DNA methylation in *Aurelia* are consistent with previous studies in bilaterian invertebrates and cnidarians. Cytosine methylation targets primarily CpG sites within gene bodies, and there exists a sharp

increase at the transcription start site. Similar sharp peaks immediately following the TSS have also been reported in *Aiptasia* and *Stylophora* (Li et al 2018; Liew et al 2018). Like other cnidarians, *Aurelia* GbM is heavier on introns than exons, which reduces the mutation burden on coding sequences (Li et al 2018; Ying et al 2022). The negative correlation between empirically measured DNA methylation and CpG o/e adds to the body of work that supports the use of CpG o/e as a reliable proxy for DNA methylation. In addition, GbM follows a bimodal distribution, a pattern that has been commonly observed in invertebrates. Consistent with Zhang and Jacobs (2022), methylation of gene bodies is heavier than that of repetitive elements. Combined with previous methylation studies in Cnidaria including *Nematostella*, *Acropora, Hydra,* etc., these results support that GbM is likely to be present in the last common ancestor of Cnidaria, and possibly that of Eumetazoa.

Although DNA methylation (and particularly GbM) is indicated in life history transitions in animals such as the honeybee and the Pacific oyster, changes in GbM have minimal correlation with changes in gene expression (Kucharski et al 2008; Riviere et al 2013). In *Aurelia*, we discover a large number of genes that are differentially methylated between the stages studied; however, similar to the honeybee and the pacific oyster, changes in GbM do not predict changes in expression. To our knowledge, this is the first documentation of such a phenomenon in a cnidarian with complex life histories. Combined with previous study by Fuchs et al (2014), where *Aurelia* strobilation (the transition from polyp to ephyra) halted upon DNA methyltransferases being inhibited, our results suggest that the role of DNA methylation in regulating life history transitions is underlied by functions other than regulating transcription levels, and one of these functions could be modulating expression plasticity.

While GbM is indicated in biological processes such as regulating development and responding to environmental stresses in various invertebrate taxa (e.g., Wang et al 2020; Dixon et al 2018), its exact relationship with transcription remains puzzling. In some invertebrate animals, it has been shown that genes with functions such as cell-cell signaling and development tend to be sparsely methylated, while genes with basic housekeeping functions tend to be strongly methylated (Elango et al 2009; Park et al 2011; Sarda et al 2012; Gavery et al 2010; Dixon et al 2014; Harris et al 2019). Our results show that hypomethylation is strongly associated with more dynamic gene expression and high methylation is associated with more stable expression profiles across *Aurelia* life history, and thus provides the first evidence of GbM involved in modulating gene expression plasticity in a cnidarian with a complex life history.

In addition, GbM might also regulate life history transitions by repressing spurious transcription initiation, regulating alternative splicing, and/or interacting with other epigenetic modifications. Our results show that the relationship between GbM and transcription could be rather complex. However, it is important to note that the relationship could be hard to interpret given that the signals could be mixed across different cell types. GbM has been shown to interact with other epigenetic modifications such as H3K36me3 in vertebrates (Teissandier et al 2017), but evidence in invertebrates is lacking. Aside from more detailed studies using specific cell types, surveys of histone modifications in Cnidaria could also provide further insights into the evolution of GbM in animals.

The present work is the first description of genome-wide methylome dynamics in a cnidarian model with a complex life history. The overall distribution of DNA methylation on the *Aurelia* genome shares many similarities with previously studied

bilaterian invertebrates and cnidarians (Anthozoa), such as the enrichment of methylation on CpG dinucleotides and particularly those on gene bodies. Interestingly, our results suggest that GbM has a non-linear relationship with transcription, and that changes in GbM during life history transitions do not correlate with changes in expression even though it has been shown to be critical during the polyp-to-ephyra transition. We highlight that low GbM is associated with more plastic expression and high GbM associated with more stable expression across life history stages.

Materials and Methods

Planula and medusa bell margin samples were obtained from Cabrillo Aquarium and were confirmed to be *Aurelia coerulea* by sequencing the CO1 gene. Total DNA was extracted from pooled whole animals of planula, polyps, and ephyra stages, and the bell margins of medusae using the protocol described in Gold et al (2018). Libraries were constructed using Illumina Truseq DNA Prep with a bisulfite treatment using Qiagen Epitech Kit. The bisulfite treated DNA was then amplified by PCR for 12 cycles. WGBS libraries were sequenced using Illumina HiSeq single-end 100bp at the UCLA Broad Stem Cell Research Center.

The sequencing reads were adapter and quality trimmed using TrimGalore (Cutadapt) (version 0.6.7) with default settings, and subsequently mapped to the *Aurelia coerulea* reference genome using BSseeker2 (Guo et al 2013), based on the Bowtie2 aligner in end-to-end mode. Cytosines with at least 4X coverage were selected for downstream analyses.

Methylation data were visualized using the IGV genome browser (Thorvaldsdóttir et al 2013). The metagene plots were produced by deepTools (Ramírez et al 2013)

plotHeatmap and plotProfile functions, using a bin size of 50 bp while gene bodies scaled to 500 bp.

Differentially methylated regions between planula and polyp, polyp and ephyra, ephyra and medusa were identified using Metilene (version 0.2-8) (Jühling et al 2018), and annotated by intersecting with gene models derived from Gold et al (2018). Percent methylation change for each gene was calculated as (methylation at stage 2 - methylation at stage 1)/methylation at stage 1.

References

Brekhman, V., Malik, A., Haas, B., Sher, N., & Lotan, T. (2015). Transcriptome profiling of the dynamic life cycle of the scypohozoan jellyfish Aurelia aurita. *BMC genomics*, 16(1), 1-15.

Dawson, M. N., & Jacobs, D. K. (2001). Molecular evidence for cryptic species of Aurelia aurita (Cnidaria, Scyphozoa). *The Biological Bulletin*, 200(1), 92-96.

Dimond, J. L., & Roberts, S. B. (2016). Germline DNA methylation in reef corals: patterns and potential roles in response to environmental change. *Molecular Ecology*, 25(8), 1895-1904.

Dixon, G. B., Bay, L. K., & Matz, M. V. (2014). Bimodal signatures of germline methylation are linked with gene expression plasticity in the coral Acropora millepora. *BMC genomics*, 15(1), 1-11.

Dixon, G. B., Bay, L. K., & Matz, M. V. (2016). Evolutionary consequences of DNA methylation in a basal metazoan. *Molecular biology and evolution*, 33(9), 2285-2293.

Dixon, G., Liao, Y., Bay, L. K., & Matz, M. V. (2018). Role of gene body methylation in acclimatization and adaptation in a basal metazoan. *Proceedings of the National Academy of Sciences*, *115*(52), 13342-13346.

Elango, N., Hunt, B. G., Goodisman, M. A., & Soojin, V. Y. (2009). DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, Apis mellifera. *Proceedings of the National Academy of Sciences*, 106(27), 11206-11211.

Feng, S., Cokus, S. J., Zhang, X., Chen, P. Y., Bostick, M., Goll, M. G., ... & Jacobsen, S. E. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences*, 107(19), 8689-8694.

Flores, K., Wolschin, F., Corneveaux, J. J., Allen, A. N., Huentelman, M. J., & Amdam, G. V. (2012). Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics*, *13*(1), 1-9.

Fuchs, B., Wang, W., Graspeuntner, S., Li, Y., Insua, S., Herbst, E. M., ... & Khalturin, K. (2014). Regulation of polyp-to-jellyfish transition in Aurelia aurita. *Current Biology*, 24(3), 263-273.

Gavery, M. R., & Roberts, S. B. (2010). DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (Crassostrea gigas). *BMC Genomics*, 11(1), 1-9.

Gold, D. A., Katsuki, T., Li, Y., Yan, X., Regulski, M., Ibberson, D., ... & Greenspan, R. J. (2019). The genome of the jellyfish Aurelia and the evolution of animal complexity. *Nature Ecology & Evolution*, 3(1), 96-104.

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M. Q., ... & Pellegrini, M. (2013). BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC genomics*, *14*(1), 1-8.

Harris, K. D., Lloyd, J. P., Domb, K., Zilberman, D., & Zemach, A. (2019). DNA methylation is maintained with high fidelity in the honey bee germline and exhibits global non-functional fluctuations during somatic development. *Epigenetics & chromatin*, 12(1), 1-18.

Huh, I., Zeng, J., Park, T., & Soojin, V. Y. (2013). DNA methylation and transcriptional noise. *Epigenetics & chromatin*, *6*(1), 1-10.

Jjingo, D., Conley, A. B., Soojin, V. Y., Lunyak, V. V., & Jordan, I. K. (2012). On the presence and role of human gene-body DNA methylation. *Oncotarget*, 3(4), 462.

Jühling, F., Kretzmer, H., Bernhart, S. H., Otto, C., Stadler, P. F., & Hoffmann, S. (2016). Metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome research*, 26(2), 256-262.

Kucharski, R., Maleszka, J., Foret, S., & Maleszka, R. (2008). Nutritional control of reproductive status in honeybees via DNA methylation. *Science*, *319*(5871), 1827-1830.

Kyger, R., Luzuriaga-Neira, A., Layman, T., Milkewitz Sandberg, T. O., Singh, D., Huchon, D., ... & Alvarez-Ponce, D. (2021). Myxosporea (Myxozoa, Cnidaria) lack DNA cytosine methylation. *Molecular biology and evolution*, 38(2), 393-404.

Li, Y., Liew, Y. J., Cui, G., Cziesielski, M. J., Zahran, N., Michell, C. T., ... & Aranda, M. (2018). DNA methylation regulates transcriptional homeostasis of algal endosymbiosis in the coral model Aiptasia. *Science advances*, 4(8), eaat2142.

Liew, Y. J., Zoccola, D., Li, Y., Tambutté, E., Venn, A. A., Michell, C. T., ... & Aranda, M. (2018). Epigenome-associated phenotypic acclimatization to ocean acidification in a reef-building coral. *Science advances*, 4(6), eaar8028.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., ... & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), 315-322.

Nakanishi, N., Hartenstein, V., & Jacobs, D. K. (2009). Development of the rhopalial nervous system in Aurelia sp. 1 (Cnidaria, Scyphozoa). *Development genes and evolution*, 219(6), 301-317.

Neri, F., Rapelli, S., Krepelova, A., Incarnato, D., Parlato, C., Basile, G., ... & Oliviero, S. (2017). Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, 543(7643), 72-77.

Park, J., Peng, Z., Zeng, J., Elango, N., Park, T., Wheeler, D., ... & Yi, S. V. (2011). Comparative analyses of DNA methylation and sequence evolution using Nasonia genomes. *Molecular biology and evolution*, 28(12), 3345-3354.

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research*, *42*(W1), W187-W191.

Riviere, G., Wu, G. C., Fellous, A., Goux, D., Sourdaine, P., & Favrel, P. (2013). DNA methylation is crucial for the early development in the oyster C. gigas. *Marine Biotechnology*, 15(6), 739-753.

Sarda, S., Zeng, J., Hunt, B. G., & Yi, S. V. (2012). The evolution of invertebrate gene body methylation. *Molecular biology and evolution*, 29(8), 1907-1916.

Shayevitch, R., Askayo, D., Keydar, I., & Ast, G. (2018). The importance of DNA methylation of exons on alternative splicing. *RNA*, *24*(10), 1351-1362.

Suzuki, M. M., Kerr, A. R., De Sousa, D., & Bird, A. (2007). CpG methylation is targeted to transcription units in an invertebrate genome. *Genome research*, 17(5), 625-631.

Takayama, S., Dhahbi, J., Roberts, A., Mao, G., Heo, S. J., Pachter, L., ... & Boffelli, D. (2014). Genome methylation in D. melanogaster is found at specific short motifs and is independent of DNMT2 activity. *Genome research*, 24(5), 821-830.

Teissandier, A., & Bourc'his, D. (2017). Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. *The EMBO journal*, *36*(11), 1471-1473.

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics*, *14*(2), 178-192.

Wang, H., Liu, Z., Wang, Y., Ma, L., Zhang, W., & Xu, B. (2020). Genome-wide differential DNA methylation in reproductive, morphological, and visual system differences between queen bee and worker bee (Apis mellifera). *Frontiers in genetics*, *11*, 770.

Wenzel, D., Palladino, F., & Jedrusik-Bode, M. (2011). Epigenetics in C. elegans: facts and challenges. *genesis*, 49(8), 647-661.

Ying, H., Hayward, D. C., Klimovich, A., Bosch, T. C., Baldassarre, L., Neeman, T., ... & Miller, D. J. (2022). The role of DNA methylation in genome defense in Cnidaria and other invertebrates. *Molecular Biology and Evolution*, 39(2), msac018.

Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, 328(5980), 916-919.

Zhang, X., & Jacobs, D. (2022). A Broad Survey of Gene Body and Repeat Methylation in Cnidaria Reveals a Complex Evolutionary History. *Genome biology and evolution*, 14(2), evab284.

Supplemental Materials

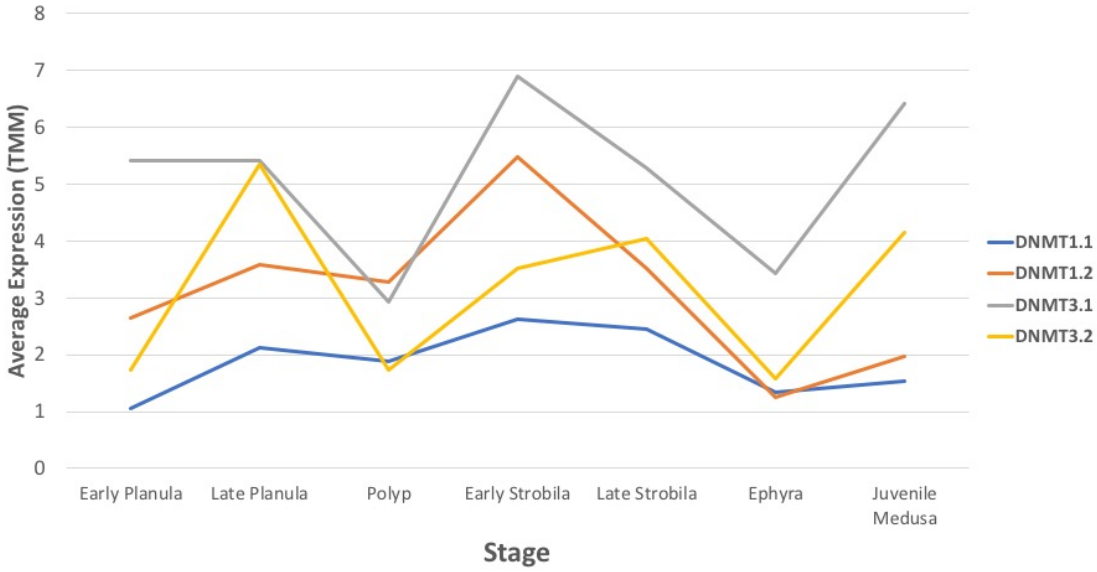


Figure S1. Expression of DNMT1 and DNMT3 across *Aurelia* life history stages.

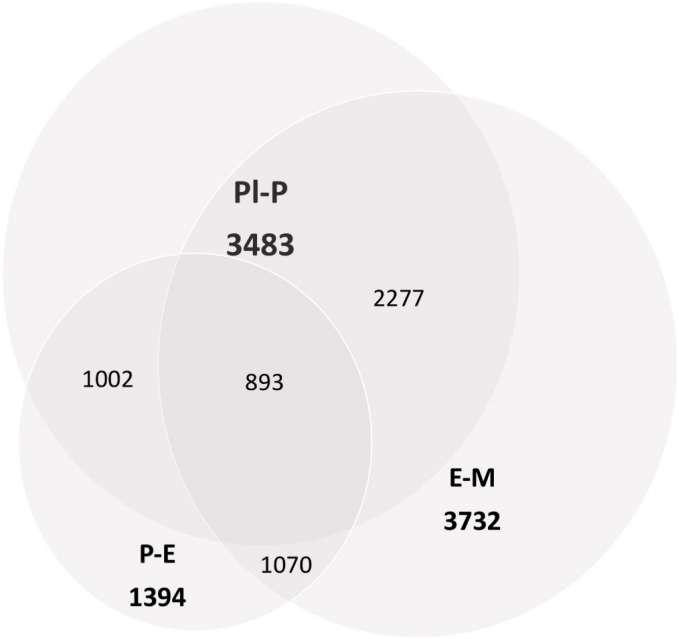


Figure S2. Number of DMRs that overlap with genes at each transition. Pl - planula; P - polyp; E - ephyra; M - medusa.

CHAPTER 3

The Evolution of Transposable Elements in the *Aurelia* Genus

Abstract

Aurelia, a group of 'true jellyfish' in Cnidaria, provides an interesting system to study the evolution of transposable elements (TEs) because of their antiquity, known genomic divergence between some of the species, and their life history with alternating sexual and asexual reproduction. We use a homology-free method to annotate and quantify 12 species in this genus. We note a large proportion of TEs in all species surveyed seem to be unique to *Aurelia*. We did not find signatures of particular TE expansion events contributing to the genome size difference between two species (*A. aurita* and *A. coerulea*). Our results also show that the two main clades in *Aurelia* have undergone different evolutionary histories regarding TE expansion as suggested by sequence divergence analyses.

Introduction

Transposable elements (TEs) are mobile elements ubiquitous in eukaryotic genomes and make up a significant proportion of the nuclear DNA in most species (Wells and Feschotte 2020). Once thought to be genetic oddities, TEs have profound impact on genome evolution as demonstrated by an increasing number of studies (Ågren and Wright 2011; Batzer and Deininge 2002; Bourque et al 2018; Chuong et al 2017; Cosby et al 2019; Fedoroff 2012; Feschotte and Pritham 2007; Wong et al 2019). Due to their repetitive nature, annotation and quantification of TEs are notoriously difficult. In this study, we use a homology-free method to examine the TEs present in 12 species in *Aurelia*, a genus of Scyphozoa (the 'true jellyfish') in Cnidaria.

Aurelia contains numerous cryptic species (Dawson and Jacobs 2001; Dawson 2003). Interestingly, the two species of *Aurelia* with genome assemblies, *A. aurita* and

A. coerulea, are significantly different in their genome sizes (492Mbp and 713Mbp, respectively) (Khalturin et al 2019; Gold et al 2019). Khalturin et al (2019) noted that a large proportion of TEs in the *A. aurita* genome did not have similar sequences in RepBase, hence are likely to be 'novel' or specific to the species. In this study, we additionally generated next-generation sequencing data (coverage 5-10X) for 10 more *Aurelia* species and used a homology-free method to annotate and quantify TEs to test whether TEs contribute to the huge difference in genome sizes in *Aurelia*.

TEs are vastly diverse, and can be divided into Class I retrotransposons, and Class II DNA transposons (Finnegan 1989; Wicker et al 2007). The main difference between the two is how they transpose in the host genome. Retrotransposons proliferate through an mRNA intermediate which then reverse transcribes and inserts into the host genome ('copy and paste'). In contrast, DNA transposons mobilize in the genome by moving themselves to a new location ('cut and paste'). Within each class, there are numerous superfamilies and families of elements. Both TE content and TE composition have been found to vary in closely related species (Wong et al 2019; Naville et al 2019).

TEs are thought to be a major contributor to genome evolution and a source for genetic variation. TE content has been known to strongly correlate with genome size (Kidwell 2002; Naville et al 2019; Tenaillon et al 2010; Zhang and Jacobs 2022). There have been a number of studies showing genome expansion as a result of TE amplification (Sun and Mueller 2014; Vitte and Panaud 2005; Blommaert et al 2019; Wong et al 2019). In addition to genome evolution, TEs have also been shown to contribute to speciation in several vertebrate lineages (de Boer et al 2007; Pace et al 2008; Serrato-Capuchina and Matute, 2018; Pritham et al 2007). However, studies outside of vertebrates and plants are few.

In addition to the perspective of genome evolution, *Aurelia* offers an interesting case to study TEs because of their life history. As Scyphozoan jellyfish, *Aurelia* alternates sexual and asexual reproduction throughout its life cycle (for an illustration of the life cycle, see Gold et al 2019). It has long been thought that sexual reproduction allows for a stable number of TEs, whereas asexually reproduced populations would have to purge their genomes of TEs (or at least the most deleterious ones) because they lack recombination that could otherwise mitigate the detrimental effects from TE activities (Arkhipova 2001; Dolgin and Charleswort 2006). This hypothesis has been tested in ancient asexual taxa such as a class of rotifers (Arkhipova and Meselson 2000). *Aurelia*, and Scyphozoa in general, are some of the few animals that have both sexually and asexually reproducing stages in the life cycle that are not parasites. They thus provide a unique opportunity to study sex and TEs without the factor of extremely compacted genomes typical of parasites.

Results

Among the 12 species surveyed, the repeat content ranges from 15.22% (*A. sp7*) to 52.75% (*A. aurita*, Roscoff strain), and shows a positive correlation with genome sizes (Figure 1). Consistent with previous studies in *A. aurita* and *A. coerulea*, our results also show that the majority of TEs in *Aurelia* were 'unknown' elements. Less than 2% of TEs were annotated when using RepBase by finding homologous sequences. Using a custom library instead that is independent of RepBase substantially increased the amount of reads that are classified, yet there are still a large number of TEs that are 'unknown' (Figure 2; Table S1). The rest of the results presented are all from analyses using custom libraries. Amongst the annotated TEs, long interspersed nuclear elements (LINEs) are

consistently the most frequent in all species except for *A. sp7* (Figure 2; Table S1; Figure S1).

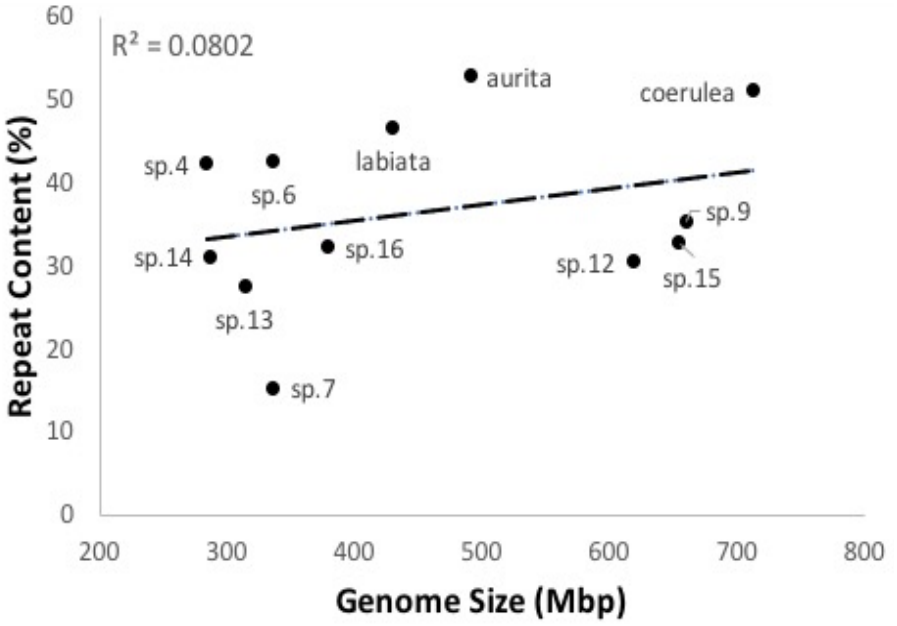


Figure 1. Repeat contents and genome sizes of 12 *Aurelia* species.

We then investigated the ages of TE superfamilies by computing the divergence between sampled reads and the annotated consensus contigs. Interestingly, we highlight that the (*sp.13, sp.12, sp.15, sp9, sp.16*) clade (the top clade on Figure 2) and the (*aurita, sp.6, sp.4, sp.14, sp.7, labiata,* and *coerulea*) clade (the bottom clade on Figure 2) show different distributions of TE superfamilies, suggesting the two groups underwent different evolutionary histories (Figure 3). Specifically, the top clade has more TEs that show higher divergence to the consensus sequences (10-20%), suggesting that the expansion of these TEs is ancient. In contrast, TEs in the bottom clade are more concentrated towards the younger end of the spectrum, suggesting that the expansion activities in these species are more recent.

Aside from the different distributions of TE ages, as suggested by divergence, certain TEs also show distinctive patterns of evolution. First, long terminal repeats (LTRs) consistently show signature of recent expansion activities regardless of the overall distribution of TE ages, with *A.aurita* and *A. coerulea* being the only two exceptions where a significant proportion of LTRs are ancient (Figure 3). Second, short interspersed nuclear elements (SINEs) are significantly more frequent in sp. 13 (0.91%; Figure 2; Table S1), and are a result of relatively ancient expansion activities (Figure 3). Third, sp. 7 is distinctly different from all the other species examined in that DNA transposons are the most abundant class instead of retrotransposons, and we show that the expansion of these DNA transposons is extremely recent (0-2% divergent), suggesting that the invasion of these TEs happened in recent evolutionary history of this species.

Directly comparing *A. aurita* and *A. coerulea*, the two emerging *Aurelia* model organisms with genome assemblies, we find that a large number of TEs are shared by the two species, and their frequencies in both species are highly correlated (Figure 4). Although there are some TEs that show varied frequencies in the two genomes, they tend to be quite rare (<0.01%), and therefore caution should be taken when interpreting these results. For TEs that have frequencies higher than 0.01% in both genomes, there is no indication that they underwent species-specific expansions.

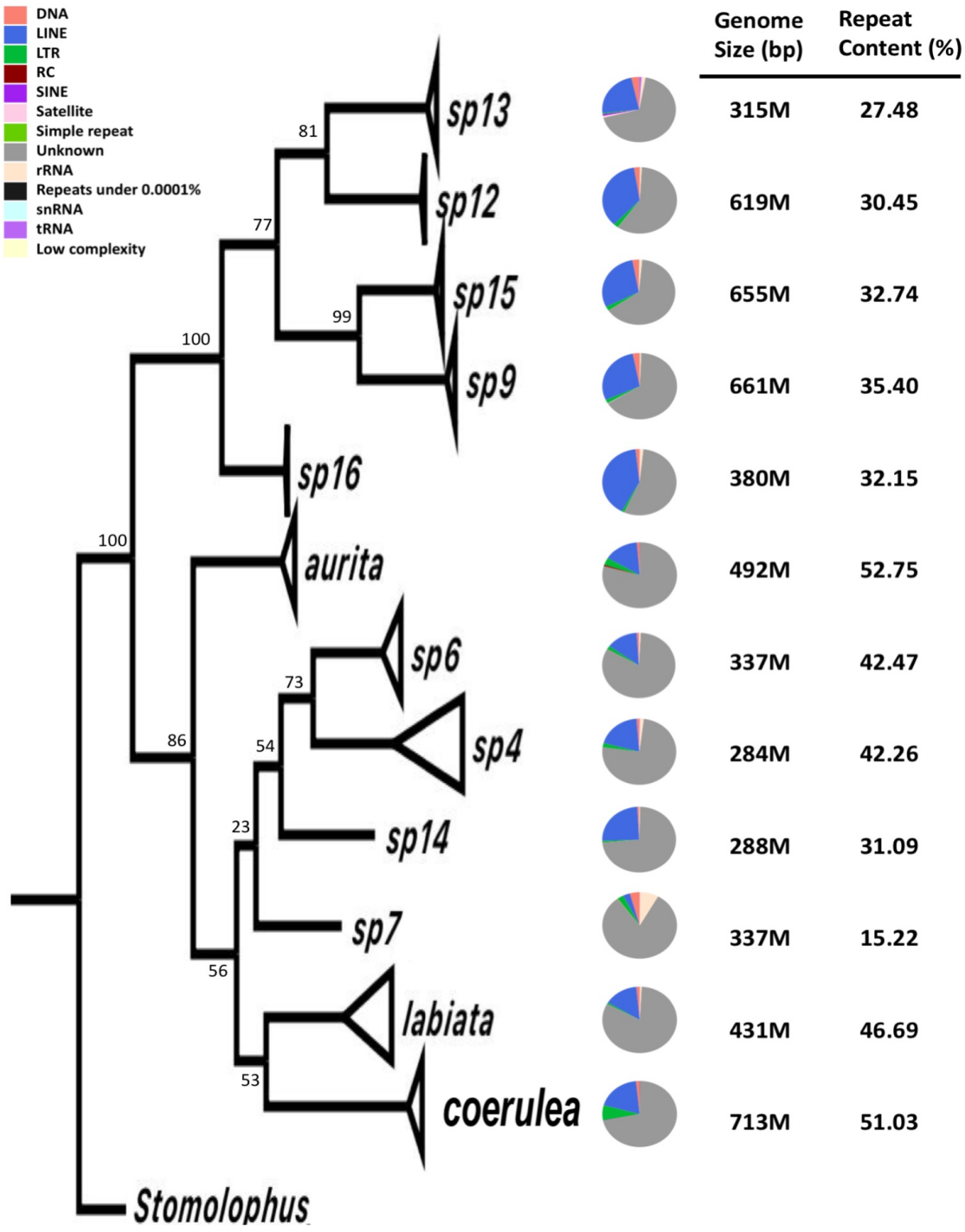
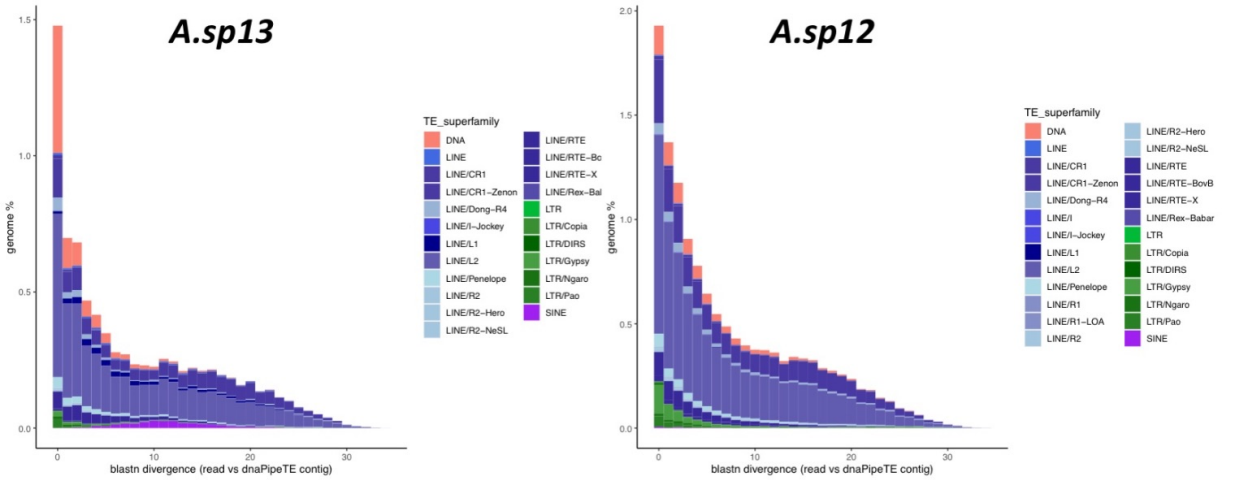
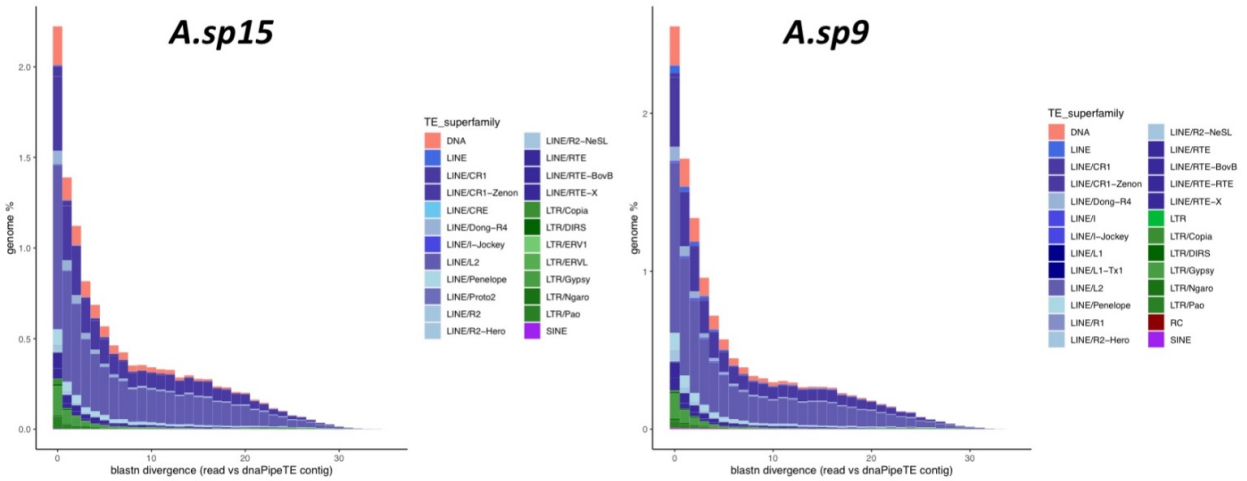
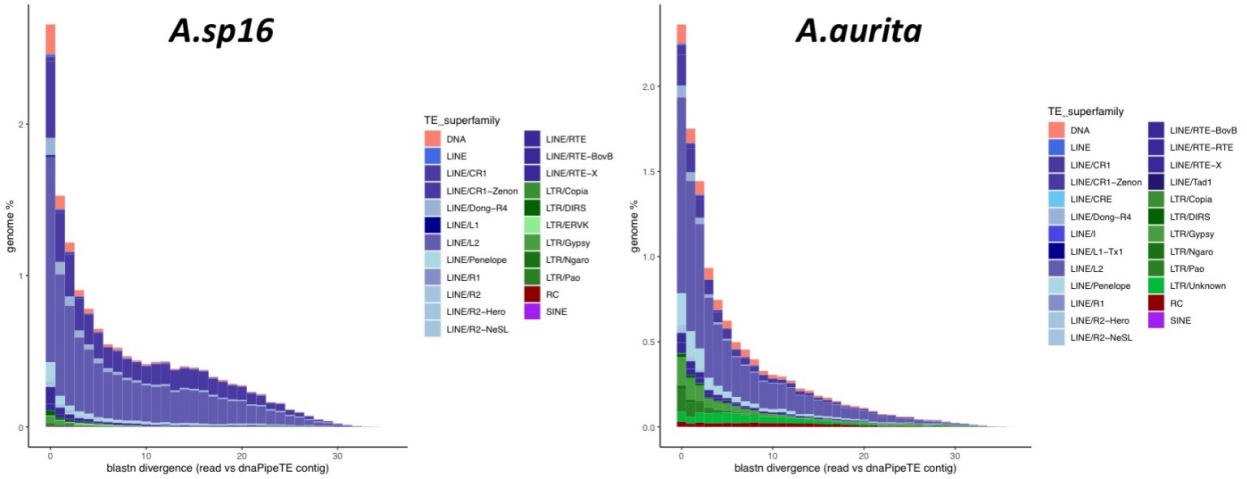


Figure 2. Composition of TEs in *Aurelia* species. The pie charts show the relative abundance of TE families. 'Unknown' sequences are repetitive sequences without annotations. Genome sizes and repeat contents are noted on the right. *Stomolophus* is the outgroup on this CO1 phylogenetic tree. Bootstrap values are noted at each node on the tree.


A.sp13
genome %
blastn divergence (read vs dnaPipeTE contig)
TE_superfamily
DNA
LINE
LINE/CR1
LINE/CR1-Zenon
LINE/Dong-R4
LINE/I-Jockey
LINE/L1
LINE/L2
LINE/Penelope
LINE/R2
LINE/R2-Hero
LINE/R2-NeSL
LINE/RTE
LINE/RTE-Bc
LINE/RTE-X
LINE/Rex-Bal
LTR
LTR/Copia
LTR/DIRS
LTR/Gypsy
LTR/Ngaro
LTR/Pao
SINE
A.sp12
genome %
blastn divergence (read vs dnaPipeTE contig)
TE_superfamily
DNA
LINE
LINE/CR1
LINE/CR1-Zenon
LINE/Dong-R4
LINE/I
LINE/I-Jockey
LINE/L1
LINE/L2
LINE/Penelope
LINE/R1
LINE/R1-LOA
LINE/R2
LINE/R2-Hero
LINE/R2-NeSL
LINE/RTE
LINE/RTE-BovB
LINE/RTE-X
LINE/Rex-Babar
LTR
LTR/Copia
LTR/DIRS
LTR/Gypsy
LTR/Ngaro
LTR/Pao
SINE


A.sp15
genome %
blastn divergence (read vs dnaPipeTE contig)
TE_superfamily
DNA
LINE
LINE/CR1
LINE/CR1-Zenon
LINE/CRE
LINE/Dong-R4
LINE/I-Jockey
LINE/L2
LINE/Penelope
LINE/Proto2
LINE/R2
LINE/R2-Hero
LINE/R2-NeSL
LINE/RTE
LINE/RTE-BovB
LINE/RTE-X
LTR/Copia
LTR/DIRS
LTR/ERV1
LTR/ERVL
LTR/Gypsy
LTR/Ngaro
LTR/Pao
SINE
A.sp9
genome %
blastn divergence (read vs dnaPipeTE contig)
TE_superfamily
DNA
LINE
LINE/CR1
LINE/CR1-Zenon
LINE/Dong-R4
LINE/I
LINE/I-Jockey
LINE/L1
LINE/L1-Tx1
LINE/L2
LINE/Penelope
LINE/R1
LINE/R2-Hero
LINE/R2-NeSL
LINE/RTE
LINE/RTE-BovB
LINE/RTE-RTE
LINE/RTE-X
LTR
LTR/Copia
LTR/DIRS
LTR/Gypsy
LTR/Ngaro
LTR/Pao
RC
SINE


A.sp16
genome %
blastn divergence (read vs dnaPipeTE contig)
TE_superfamily
DNA
LINE
LINE/CR1
LINE/CR1-Zenon
LINE/Dong-R4
LINE/L1
LINE/L2
LINE/Penelope
LINE/R1
LINE/R2
LINE/R2-Hero
LINE/R2-NeSL
LINE/RTE
LINE/RTE-BovB
LINE/RTE-X
LTR/Copia
LTR/DIRS
LTR/ERVK
LTR/Gypsy
LTR/Ngaro
LTR/Pao
RC
SINE
A.aurita
genome %
blastn divergence (read vs dnaPipeTE contig)
TE_superfamily
DNA
LINE
LINE/CR1
LINE/CR1-Zenon
LINE/CRE
LINE/Dong-R4
LINE/I
LINE/L1-Tx1
LINE/L2
LINE/Penelope
LINE/R1
LINE/R2-Hero
LINE/R2-NeSL
LINE/RTE-BovB
LINE/RTE-RTE
LINE/RTE-X
LINE/Tad1
LTR/Copia
LTR/DIRS
LTR/Gypsy
LTR/Ngaro
LTR/Pao
LTR/Unknown
RC
SINE

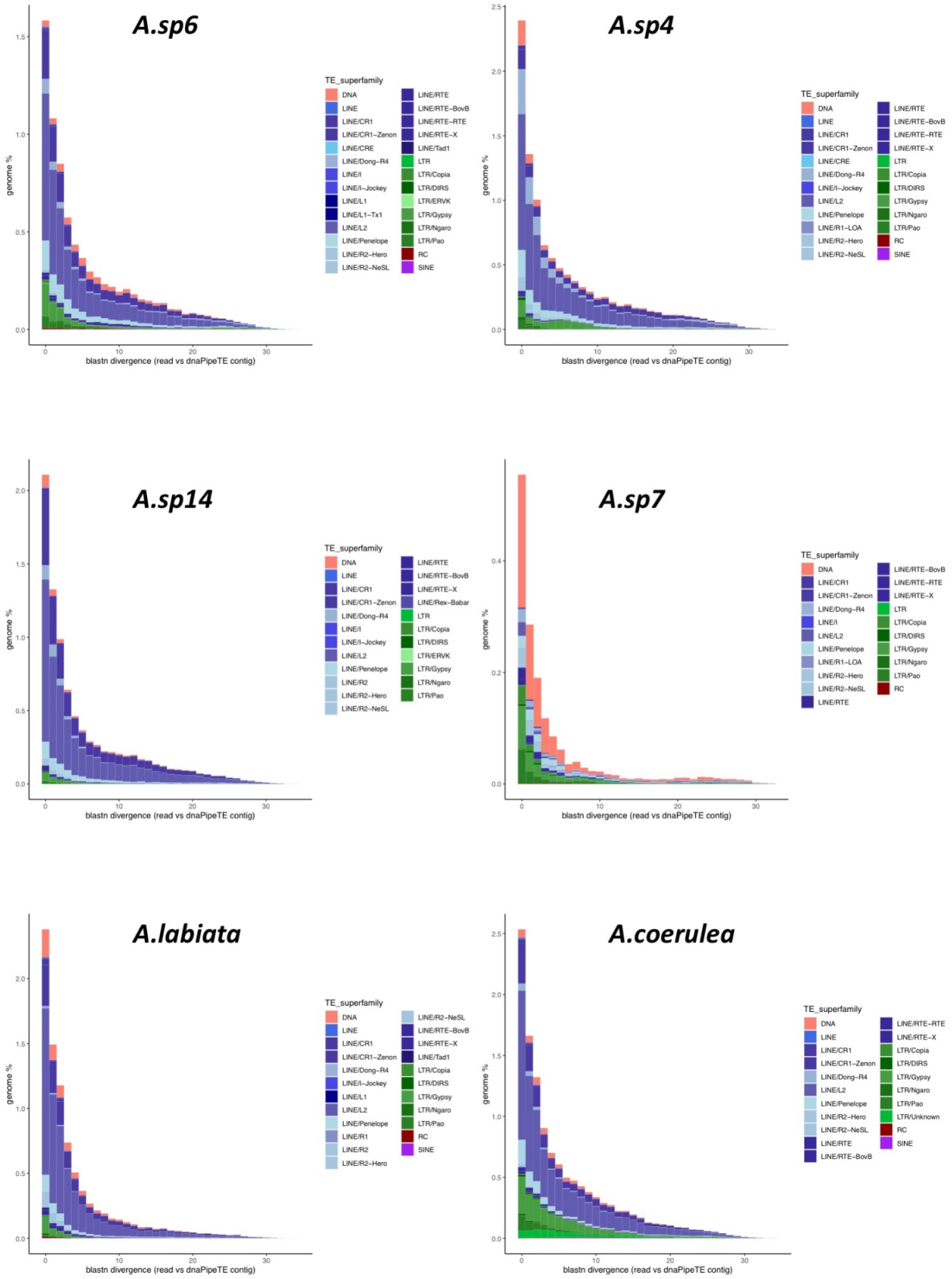


Figure 3. Distribution of TE ages in *Aurelia*.

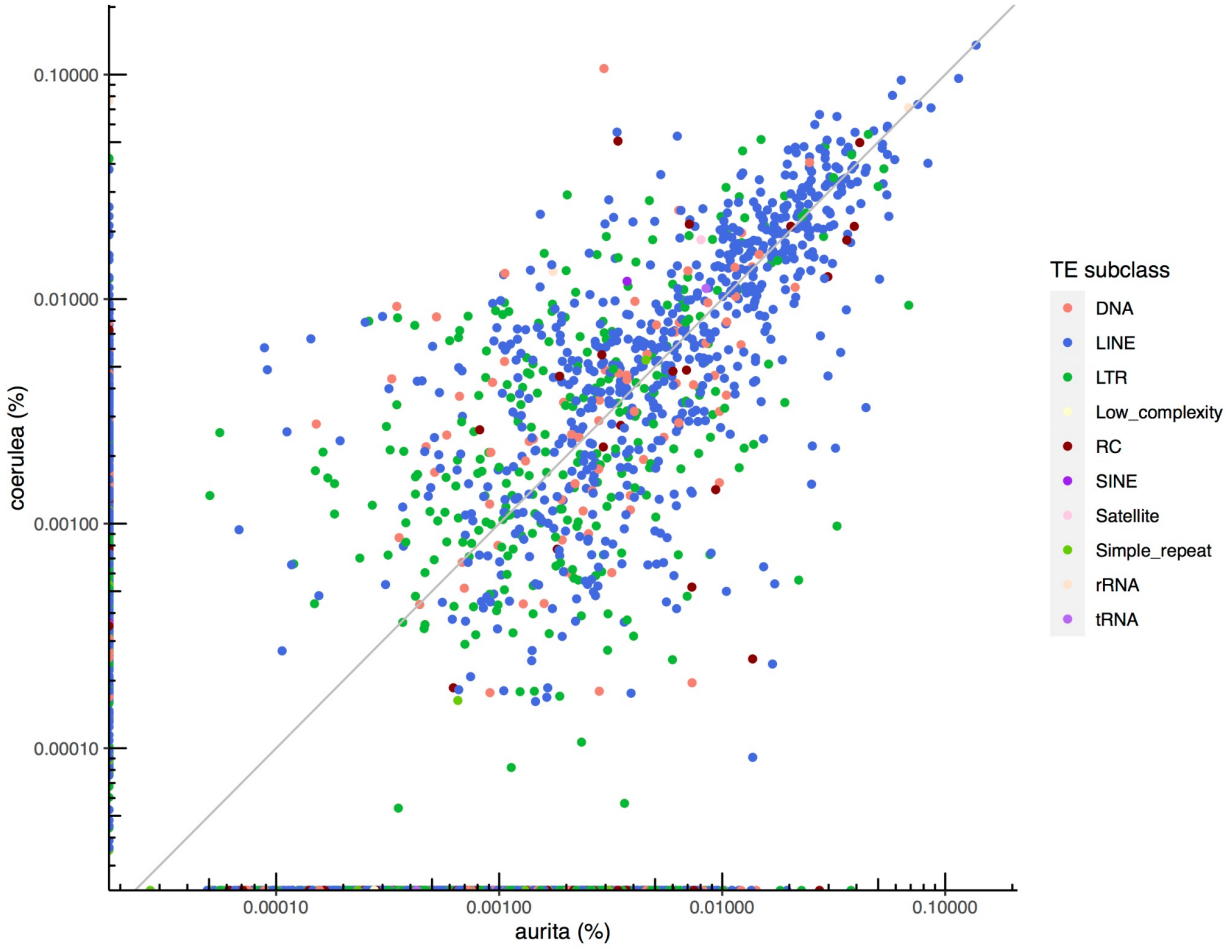


Figure 4. TEs shared by *A. aurita* and *A. coerulea*, and their corresponding frequencies in either genome.

Discussion

Given the inherent difficulties to accurately annotate and quantify TEs in genome assemblies, we employed a *de novo* technique and our estimation of TE content in *A. aurita* and *A. coerulea* are slightly higher than previously estimated, but not drastically different, which is expected (Khalturin et al 2019; Gold et al 2019). A large proportion of TEs seem to be unique in *Aurelia* in all species surveyed, a phenomenon previously noted by Khalturin et al (2019).

The two clades of *Aurelia* have distinct patterns of TE evolution. The (*sp.12, sp.13, sp.15, sp.9, sp.16*) clade has more TEs resulting from ancient expansion events, while TE expansion appears more recent in the (*aurita, sp.6, sp.4, sp.14, sp.7, labiata, coerulea*) clade. Since the two clades separated 15-20MYA, those ancient expansion events in the first clade likely happened in that time frame.

Genomic differences between *Aurelia* species have been noted previously. Most notably, *A. aurita* and *A. coerulea* not only show significant difference between their genome sizes, but also single copy orthologous genes show sequence divergence to a larger degree than that between mice and rats (Gold et al 2019). This study adds to this body of literature in two ways. First, we report that the studied species have likely undergone two different histories in their repeatome evolution, where some species were subjected to more ancient TE expansion activities while others did not experience those ancient events. Second, looking more closely at *A. aurita* and *A. coerulea,* our results suggest that the shared TE families between the two species are highly correlated in frequencies in the two genomes. We did not find any evidence for expansion events that are specific to either species, indicating that the recent TE activities likely happened before the two diverged, and the activities of these shared TEs are unlikely to have contributed to the genome size difference. Meanwhile, our results agree with Gold et al (2019) and Khalturin et al (2019) in that there are a large number of TEs in both species that seem to be unique and therefore not annotated. Our analyses do not exclude the possibility that some of these unique TEs have undergone lineage-specific expansions and contributed to the bigger genome size in *A. coerulea*.

Aurelia provides an interesting system to study sex and TE evolution as they alternate between asexual reproduction (the polyp stage) and sexual reproduction (the

medusa stage). Interestingly, in *Aurelia coerulea,* the Piwi proteins and Piwi interacting proteins are highly expressed in planula and polyp stages, and decrease sharply towards the medusa stage, suggesting that the asexually reproducing stages are repressing TE activities more vigilantly (Figure S2). On the other hand, *sp.7* with its low TE content, might have undergone long periods of asexual reproduction, and thus TE activities, especially those of the Class II retrotransposons, have been under extreme controls by the host genome.

Conclusion

We report the first comparative study of TEs in Scyphozoa. Compared to other studies of similar scope, *Aurelia* stands out in a number of ways. First, *Aurelia* seems to possess many repetitive sequences that are not homologous to known TEs, resulting in a large proportion of the repeatome not being annotated (Figure 2; Figure S3). To our knowledge, this is the first systemic report of a genus that processes large amounts of unique TEs. Future work is needed to better understand these *Aurelia*-specific TEs. Second, the vast majority of TEs in *Aurelia* are extremely young, whereas animals such as certain insects have more TEs from ancient expansion activities (Castro et al 2020; Blommaert et al 2019). Third, unlike its hydrozoan relative, *Hydra*, where the expansion of a single family of TEs contributed to the genome expansion of the brown hydras (Wong et al 2019), we do not find any particular TEs as the cause of the larger genome size of *A. coerulea*.

Materials and Methods

Sequencing

Total DNA was extracted from *Aurelia* species using the method described in Gold et al (2019) (Table 1). Libraries were constructed using Illumina TruSeq DNA Prep and DNA was sheared to 300-600bp. Sequencing was done using paired end 100bp. Raw sequencing reads were quality and adapter trimmed using TrimGalore (Cutadapt) (version 0.6.7) with default settings. The sequencing coverage ranges between 5.0X and 10.6X.

Species	Location	Coverage
A. labiata	Canada	7.5X
A. coerulea	Japan	10.6X
sp. 4	Kakaban	8.9X
sp. 6	Indonesia	9.8X
sp. 7	New Zealand	10.0X
sp. 9	Gulf of Mexico	7.6X
sp. 12	Gulf of California-La Paz	5.5X
sp. 13	El Salvador -El espino	8.1X
sp. 14	Gulf of Panama, Panama (Pacific)	10.6X
sp. 15	Bocas del Toro, Panama (Caribbean)	5.0X
sp. 16	Argentina	8.9X

Table 1. Species sequenced in this study and the locale from which the sample originated.

TE Annotation and Quantification

The TEs of newly sequenced *Aurelia* species in Table 1 and *Aurelia aurita* (SRR8040391) were assembled, annotated, and quantified from raw reads using the pipeline dnaPipeTE v1.3 (Goubert et al 2015). To avoid false positives due to mitochondrial DNA, possible mitochondrial reads were first removed from the raw reads by blasting to published cnidarian mitochondrial genes (supplemental file 1). All

species were sampled at 0.15X coverage. In absence of genome sizes for all species other than *A. aurita* and *A. coerulea*, estimation of genome sizes was done by kmer using Jellyfish and subsequently CovEst in the repeats mode to account for the repetitive elements in the reads (Marcais and Kingsford 201; Hozza et al 2015). The estimated genome sizes were then given to the dnaPipeTE command using the -genome_size flag, and coverage using the -genome_coverage flag.

The initial run of dnaPipeTe returned majority of repeats as 'Unknown' as no homologous sequences could be found in the default RepBase library. To account for this, we further annotated the contigs assembled by dnaPipeTe using RepeatClassifier, as well as the *Aurelia aurita* (GCA_004194395.1, GCA_004194415.1) and *Aurelia coerulea* (https://davidadlergold.faculty.ucdavis.edu/jellyfish/, last accessed Mar.6, 2022) genome assemblies using RepeatModular (Flynn et al 2020). The annotated contigs were then used as a custom library for the final round of dnaPipeTE by using the -RM_lib flag.

Aurelia Phylogeny

CO1 sequences used for the phylogeny were obtained from NCBI. The sequences were aligned using MUSCLE (Edgar 2004), and the alignment was manually trimmed in PAUP (supplemental file 2). The maximum likelihood phylogeny was computed by RAxML with 100 bootstraps (Stamatakis 2014).

References

Ågren JA, Wright SI. 2011. Co-evolution between transposable elements and their hosts: A major factor in genome size evolution? *Chromosom. Res.* 19:777–786.

Arkhipova I, Meselson M. 2000. Transposable elements in sexual and ancient asexual taxa. *Proc. Natl. Acad. Sci. U. S. A.*97:14473–14477.

Arkhipova IR. 2001. Transposable elements in the animal kingdom. *Mol. Biol.*35:157–167.

Batzer MA, Deininger PL. 2002 . Alu repeats and human genomic diversity. *Nat. Rev. Genet.* May;3(5):370-9.

Blommaert J, Riss S, Hecox-Lea B, Mark Welch DB, Stelzer CP. 2019. Small, but surprisingly repetitive genomes: transposon expansion and not polyploidy has driven a doubling in genome size in a metazoan species complex. *BMC Genomics* 20:466.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, et al. 2018. Ten things you should know about transposable elements 06 Biological Sciences 0604 Genetics. *Genome Biol.*19.

Castro MRJ, Goubert C, Carareto CMA, Monteiro FA, Vieira C. 2020. Homology-free detection of transposable elements unveils their dynamics in three ecologically distinct rhodnius species. *Genes (Basel).*11:1–14.

Charlesworth B, Sniegowski P, Stephan W. 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature.* 371:215–220.

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* 18:71–86.

Cosby RL, Chang NC, Feschotte C. 2019. Host–transposon interactions: conflict, cooperation, and cooption. *Genes Dev*. 33:1098–1116.

Dawson MN, Jacobs DK. 2001. Molecular Evidence for Cryptic Species of Aurelia aurita(Cnidaria, Scyphozoa). *Biol. Bull.* 200:92–96.

Dawson MN. 2003. Macro-morphological variation among cryptic species of the moon jellyfish, Aurelia (Cnidaria: Scyphozoa). *Mar. Biol.* 143:369–379.

de Boer JG, Yazawa R, Davidson WS, Koop BF. 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* 8:1–10.

Dolgin ES, Charlesworth B. 2006. The fate of transposable elements in asexual populations. *Genetics*. 174:817–827.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 32:1792–1797.

Fedoroff N V. 2012. Transposable elements, epigenetics, and genome evolution. *Science (80-.).*338:758–767.

Feschotte C, Pritham EJ. 2007. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu. Rev. Genet.* 41:331–368.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.*9:397–405.

Finnegan DJ. 1989. Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5:103–107.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U. S. A.*117:9451–9457.

Gold DA, Katsuki T, Li Y, Yan X, Regulski M, Ibberson D, Holstein T, Steele RE, Jacobs DK, Greenspan RJ. 2019. The genome of the jellyfish Aurelia and the evolution of animal complexity. *Nat. Ecol. Evol.* 3:96–104

Goubert C, Modolo L, Vieira C, Moro CV, Mavingui P, Boulesteix M. 2015. De Novo Assembly and Annotation of the Asian Tiger Mosquito (Aedes albopictus) Repeatome with dnaPipeTE from Raw Genomic Reads and Comparative Analysis with the Yellow Fever Mosquito (Aedes aegypti). *Genome Biol. Evol.* 7:1192–1205.

Hozza M, Vinař T, Brejová B. 2015. How Big is that Genome? Estimating Genome Size and Coverage from k-mer Abundance Spectra. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 9309:199–209.

Khalturin K, Shinzato C, Khalturina M, Hamada M, Fujie M, Koyanagi R, Kanda M, Goto H, Anton-Erxleben F, Toyokawa M, et al. 2019. Medusozoan genomes inform the evolution of the jellyfish body plan. *Nat. Ecol. Evol.* 3:811–822.

Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica*. 2002 May;115(1):49-63.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27:764–770.

Naville M. Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr. Biol.*29:1161-1168.e6.

Pace JK, Gilbert C, Clark MS, Feschotte C. 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc. Natl. Acad. Sci. U. S. A.*105:17023–17028.

Pritham EJ, Putliwala T, Feschotte C. 2007. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene*. 390:3–17.

Serrato-Capuchina A, Matute DR. 2018. The Role of Transposable Elements in Speciation. *Genes (Basel)*. 9.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 30:1312–1313.

Sun C, Mueller RL. 2014. Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. *Genome Biol. Evol.*6:1818–1829.

Tenaillon MI, Hollister JD, Gaut BS. 2010. A triptych of the evolution of plant transposable elements. *Trends Plant Sci.*15:471–478.

Vitte C, Panaud O. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenet. Genome Res.* 110:91–107.

Wells JN, Feschotte C. 2020. A Field Guide to Eukaryotic Transposable Elements. *Annu. Rev. Genet.*54:539–561.

Wong WY, Simakov O, Bridge DM, Cartwright P, Bellantuono AJ, Kuhn A, Holstein TW, David CN, Steele RE, Martínez DE. 2019. Expansion of a single transposable element family is associated with genome-size increase and radiation in the genus Hydra. *Proc. Natl. Acad. Sci. U. S. A.*116:22915–22917.

Zhang X, Jacobs D. 2022. A Broad Survey of Gene Body and Repeat Methylation in Cnidaria Reveals a Complex Evolutionary History. *Genome Biol. Evol.*14:1–14.

Supplemental Materials

Supplemental File 1: mitochondrial sequences from NCBI used to clean reads of mitochondrial DNA.

```
>YP_783951.1 ATP synthase F0 subunit 6 (mitochondrion) [Aurelia aurita]
MTASYFDQFGIVSIIPGLTNSSLMLILSIIFIIGFFNVNKLIPGRWQSIIEIVYDHWIVLVKDSLGVKGS
KYFPFVFSLFLLIAWLNVLGLFPYVFTVGTHIIVTFGLSLSILIIVTILGIKNFKADFFSMLMPQGAPMG
LAPLLVLIETASYLSRAISLGVRLAANLSAGHLLFAILASFAYQMLVGDLVLLSIFPIAIMVFITILEIA
VAMIQAYVFCLLTTIYLGDTINLH
>YP_783951.1 ATP synthase F0 subunit 6 (mitochondrion) [Aurelia aurita]
MTASYFDQFGIVSIIPGLTNSSLMLILSIIFIIGFFNVNKLIPGRWQSIIEIVYDHWIVLVKDSLGVKGS
KYFPFVFSLFLLIAWLNVLGLFPYVFTVGTHIIVTFGLSLSILIIVTILGIKNFKADFFSMLMPQGAPMG
LAPLLVLIETASYLSRAISLGVRLAANLSAGHLLFAILASFAYQMLVGDLVLLSIFPIAIMVFITILEIA
VAMIQAYVFCLLTTIYLGDTINLH
>YP_783952.1 ATP synthase F0 subunit 8 (mitochondrion) [Aurelia aurita]
MPQLDIVTFINQYIWVIGSVSIVVIVMLSVILPSIKKLTQIRGVVEEEVMESKKEKDFNSFKKLISF
>YP_009747479.1 ATP synthase F0 subunit 8 (mitochondrion) [Aurelia coerulea]
MPQLDVVTFVNQYVWIIGSVSVVVIIMLSVILPSIKKLTQIRGVTEEEIIESKEERDFTSFKKLISL
>YP_009867126.1 cob (mitochondrion) [Entacmaea quadricolor]
MQLRKQNPILSIVNGLIIDLPAPANLSYMWNFGSLLGVCLVIQIATGIFLAMHYCADVSLAFASVDHIMR
DVNYGFLLRYFHANGASMFFLCLYIHIGRGLYYGSYSKVEVWSVGVVLLILTMATAFIGYVLPWGQMSFW
GATVITNLLSAIPYVGTDIVQWVWGGFSVSNATLNRFFSLHYLFPFLLAALGVIHLICLHVDGSNNPIGV
RSDPDKVSFHVYYTSKDWYGIVVFAVFFCSLVYLAPNLLGDPENFIQANPLVTPVHIQPEWYFLFAYAIL
RSIPNKLGGVVAMFASLLVLFLLPWLHSSRFRGLTFRPLGRIAFWFLIADFLLLTWIGSQPVEEPFILVG
QLASVFYFSYFLILTPLLGYIENRLLFSRGGR
>YP_009867113.1 cob (mitochondrion) [Anemonia sulcata]
MQLRKQNPILSIVNGLIIDLPAPANLSYMWNFGSLLGVCLVIQIATGIFLAMHYCADVSLAFASVDHIMR
DVNYGFLLRYFHANGASMFFLCLYIHIGRGLYYGSYARVEVWNVGVVLLILTMATAFIGYVLPWGQMSFW
GATVITNLLSAIPYVGTDIVQWVWGGFSVSNATLNRFFSLHYLFPFLLVALAAVHLICLHVDGSNNPIGV
RSDLDKVSFHVYYTSKDWYGIVAFAVFFCSLVYLAPNLLGDPENFIQANPLVTPVHIQPEWYFLFAYAIL
RSIPNKLGGVVAMFASLLVLFLLPWLHSSRFRGLTFRPLGRMAFWFLIADFLLLTWIGSQPVEEPFILVG
QLASVFYFSYFLILTPLLGYLENRLLFPKGEW
>ATV80887.1 cytochrome c oxidase subunit I, partial (mitochondrion) [Aurelia sp. 12
LGD-2017]
LYIIFGAFSAMVGTAFSMIIRLELSGPGSMLGDDQLYNVVVTAHAFVMIFFFVMPVLIGGFGNWFVPLYI
GAPDMAFPRLNNISFWLLPPALLLLLGSSLIEQGAGTGWTVYPPLSSIQAHSGGSVDMAIFSLHLAGASS
IMGAINFITTILNMRAPGMTMDKIPLFVWSVLVTAVLLLLSLPVLAGAITMLLTDRNFNTSFFDPAGGGD
PILFQHLF
>AAX94464.1 cytochrome c oxidase subunit I, partial (mitochondrion) [Aurelia sp. 6 sensu
Dawson et al. (2005)]
TLYLIFGAFSAMVGTAFSMIIRLELSGPGSMLGDDQLYNVVVTAHALIMIFFFVMPVLIGGFGNWLVPLY
IGAPDMAFPRLNNISFWLLPPALLLLLGSSLIEQGAGTGWTVYPPLSSIQAHSGGSVDMAIFSLHLAGAS
SIMGAINFITTILNMRAPGMTMDRIPLFVWSVLVTAILLLLSLPVLAGAITMLLTDRNFNTSFFDPAGGG
DPILFQHLF
>ATV80869.1 cytochrome c oxidase I, partial (mitochondrion) [Aurelia sp. 16 LGD-2017]
LYIIFGAFSAMVGTAFSMIIRLELSGPGSMLGDDQLYNVVVTAHAFVMIFFFVMPVLIGGFGNWFVPLYI
GAPDMAFPRLNNISFWLLPPALLLLLGSSLIEQGAGTGWTVYPPLSSIQAHSGGSVDMAIFSLHLAGASS
IMGAINFITTILNMRAPGMSMDKIPLFVWSVLVTAVLLLLSLPVLAGAITMLLTDRNFNTSFFDPAGGGD
PILFQHLF
```

```
>AAX94406.1 cytochrome c oxidase subunit I, partial (mitochondrion) [Aurelia labiata]
TLYLIFGAFSAMVGTAFSMIIRLELSGPGSMLGDDQLYNVVVTAHAFIMIFFFVMPVLIGGFGNWLVPLY
IGAPDMAFPRLNNISFWLLPPALLLLLGSSLIEQGAGTGWTVYPPLSSIQAHSGGSVDMAIFSLHLAGAS
SIMGAINFITTILNMRAPGMTMDKIPLFVWSVLVTAVLLLLSLPVLAGAITMLLTDRNFNTSFFDPAGGG
DPILFQHLF
>AAX94474.1 cytochrome c oxidase subunit I, partial (mitochondrion) [Aurelia sp. 7 sensu
Dawson et al. (2005)]
TLYLIFGAFSAMVGTAFSMIIRLELSGPGSMLGDDQLYNVVVTAHALIMIFFFVMPVLIGGFGNWLVPLY
IGAPDMAFPRLNNISFWLLPPALLLLLGSSLIEQGAGTGWTIYPPLSSIQAHSGGSVDMAIFSLHLAGAS
SIMGAINFITTILNMRAPGMTMDKIPLFVWSVLVTAILLLLSLPVLAGAITMLLTDRNFNTSFFDPAGGG
DPILFQHLF
>YP_783953.1 cytochrome c oxidase subunit II (mitochondrion) [Aurelia aurita]
MLINLLNNDIPLNWQLSFQDSVSPIMEEIVFLHDQIMFVVPVISVLVVWLMVRSILSKYKYKFLVEGTVI
EVVWTVIPAIILIFIAFPSLKLLYLMDETMEPAITIKAVGHQWYWSYEYSDYDVNDLEFDSYMVPTSDLE
LGDNRLLEVDNCLVLPIQAHIRVLVTGADVIHSFAIPSLGIKMDAVPGRLNQSNMFMKRSGLFYGQCSEI
CGTNHSFMPIAVRGVSLDTFINWVESKISEE
>YP_783950.1 cytochrome c oxidase subunit III (mitochondrion) [Aurelia aurita]
MSNTFHPYHLVDPSPWPYVSACGAFFLTLGSVIYFHYSFIWIMLLGLLILGTTFVVWCRDIIREATFQGH
HTEIVKKGLKIGMLLFIVSEICFFVSFFWAFFHSSLAPAIELGALWPPTGIQILNPFSVPLLNTAILLSS
GATVTWAHHAIINGDRKQAIVGLVSTVLLGMIFTALQIVEYIEAPFAISDSVYGSTFFMATGFHGFHVLV
GTTFLLVCLLRLVFHQYTRHHHFGFEAASWYWHFVDVVWLFLYVCIYWWGS
>YP_009747481.1 cytochrome c oxidase subunit III (mitochondrion) [Aurelia coerulea]
MSNTFHPYHLVDPSPWPYVSACGAFFLTLGSVIYFHYSFIWIMILGFLIFGTTFTVWCRDIIREATFQGH
HTEIVKKGLKIGMLLFIVSEICFFVSFFWAFFHSSLAPAIELGALWPPIGIQILNPFSVPLLNTAILLSS
GATVTWAHHAIINGQRKQAIAGLITTVLLGVIFTSLQGIEYIEAPFAISDSVYGSTFFMATGFHGFHVLV
GTTFLLVCLLRLVFHQYTRHHHFGFEAASWYWHFVDVVWLFLYVCIYWWGS
>ABG56241.1 NADH dehydrogenase subunit 1 (mitochondrion) [Aurelia aurita]
MIYLLVNLIKLLMIIIPVLIAVAYLTLAERKVLGYMQSRKGPNVVGIYGLLQPLADGVKLFSKEMTVPNH
SSLILYFISPILFLTLALVLWGVLPYTLNNVYSELSLGVLFILAISSVSVYAILMSGWASTSKYAYFGSL
RAAAQMISYEVSMGLILISVVICIGSLNVSTISWFQTNSSLFIFTLIPAGIMFMVSLIAETNRAPFDLTE
GESELVSGYNVEYSSMSFALLFLAEYANIIFMSALFSLIFLGGWQGLFSSSVIWLGIKTSMIAMFFVWVR
ASYPRIRYDQLMMLLWKSYLPLSLALLALCSAILWSFGGLPVL
>ABG56246.1 NADH dehydrogenase subunit 2 (mitochondrion) [Aurelia aurita]
MLYNYYEVLLLTVLIASLVLKVTKKLFNILMFLTITFLLLSVKSAGGDIISWEPLVKIFCVLIGVIVINS
NRARIFDSSNVLILCVILASILMVSTTNMLALYLCIELQSLSIFILIAQERELTSRIEASLKYFVLSSIS
SGLLLLGSSLLFIHSGSCEITTLITQEYTPEKVLILVALLFKLAASPFHFWLPDVYQGSNNNTLLLLGTL
PKISVLGIFMVLFPGNKLVLVATIFSLIIGCIGAINQSKTKRLLAYSSIVGMGFILLGINISSYQGIEAG
TIYLTIYIITFVSILIVINSSIDEKPTLGELSSLLGNNLALLSTFGILILSVAGIPPLGGFFSKWLVLTS
AITTEFIFSSILSIICAIIAGVYYVRLVKIGYFQEEKTFLIWRKVLIKEKEMEPIQATILGMLLYSVTFI
LVCPQILWKPIHYGVLSLF
>ABG56243.1 NADH dehydrogenase subunit 3 (mitochondrion) [Aurelia aurita]
MISITFQKVTIILVLSLLLSTVIGVASYVLGDTSPDKEKVSVYECGFAPYDNPGNPISIRFFLIGILFLV
FDLEISLLLPWSVSSHLISLPGIWLIFIFVFILTWGLLYEWIKGGLEWE
>ABG56240.1 NADH dehydrogenase subunit 4 (mitochondrion) [Aurelia aurita]
MILLISTVLITSIFLVMLVGNSNPLLSKKIALGSSFLVLWLSIVLWTSHNTESLFYSNIITGWLASVPLS
TDWGSLYFGVDHLSLPFILLTAILTPICIIISWNSITYLVKEFLVFLLAIHLLLIGVFTSLNILLFYIMF
EGILIPMFLMIGIWGSRKEKERAAFYFFFFTLAGSLFMLLGIFTIYKHTGTLDYTILSLQTLPSELQFWV
FIGFFLSLAVKIPKIPFHIWLPQAHVEAPVAGSVLLAGVLLKLGGYGFLRFSWTLFPVASQYFSPVIILL
GSIAVIYASLSTCRQTDAKKLVAYSSVAHMGLVTIAIFSRTSEGLVAAIILMLAHGFVSSGLFITVTNLY
DRFHTRIIRYYKGVVYTMPIYSTLFFLLILANIAFPISMNFVGEFMSILSALQFSLLAILCPLAGMVLSA
AYSLLFYNKVSFGYPSSFLQFARDVNRRELWPPLLLIVLTITLGVMPTCLDLTLCLPYIY
```

```
>ABG56242.1 NADH dehydrogenase subunit 4L (mitochondrion) [Aurelia aurita]
MGIKSLTTYSFLLFILGILGISINRTNLILLLMSVELMLLSTSLLFMAGASFYDLILGQVFTIFIFTVAA
AESAIGLAIIIAYFRLRGKISIKLLNLLKG
>ABG56245.1 NADH dehydrogenase subunit 5 (mitochondrion) [Aurelia aurita]
MFTSIVLLPLISSITLLLNGRKLGGKGSNIVSCTCMLICLCVSLLLYHNVVLLETGISINYWSWVSYNLF
NTNLSFFVDSLSASMILAVNIVSFLVHMYSVSYMKGDPHTSRFMGYLSLFTFFMLILVSSQNYLQLFIGW
EGVGLCSYLLINFWYSRIQANKAAIKAMLVNRLGDAALIAAVIIMWVNYGSINFLSVLPISQFKFDWVIP
ILLLVGAIGKSAQIGLHTWLPDAMEGPTPVSALIHAATMVTAGVYLIIRSSPLFELSSLALIVTALVGSI
TALFAATIGLAQNDIKKIIAYSTCSQLGFMVLSCGISYYSVALFHLVNHAFFKALLFLGAGSIIHSMLDE
QDIRKMGGINITQPISYVFIVIGSLSLAGFPFMAGYYSKDLLLELSSQKYFVIFGFWLGIFATLLTALYS
FKLILRSFLQFQNSPKIYWKAAHESDWVLLAPLLVLAMGSITSGFLLFNPTLNNMVHPIIPVSIKLLPLL
VTIAGGTTGLLLFWITKRLWNIFLGSSSKKLYQFLVSAWDFDKIISNIIVVPIFKFGYKVTFKLFDTGII
ETFGPRGASIKIVKFSSFISSIQSGLLFNYATIIILFCSLFILFG
>ABG56244.1 NADH dehydrogenase subunit 6 (mitochondrion) [Aurelia aurita]
MISPIFMILTIISGIMVVTAIHPVHSIIWLVLSFIGAAIVFIQLQADFIALATLIIYVGAIAILFIFVLM
MLNLSSFEFDGNISSIIPLGITSILGLITLIIGLEYSNINDPEVEEILIKELLDIQTIGFQLYNSYAPHL
IVASLILLVAMIGAIILTLQPHLAAKRQDNFIQITRDFHNISKSHNNSSS
>KXJ27654.1 DNA polymerase beta [Exaiptasia diaphana]
MSKRKASDQNPNKEFCDFLNELAEYEKNVTRQIHKYNAYRKAASVLAKHPTKVKNGKEAKELPGIGDKIA
KKIDEFSKTGKLEKLEKIRADEGSSIIKDLTRVSGIGPAVAKKLIDDGIKSIADLKKAQGRLNHHQKIGL
KYVEEFEKRIPREEMVKLKDIVLEHVEKYDKKLVATVCGSFRRGAATSGDIDIMLTHPSFTSTSKTFEKK
VLLKNLVSKMEKSQFITDTLSLGETKFNGVCILPSRPKEKAPLHRRIDIRLIPHDQYYCGILYFTGSDEF
NKRMRHEALEKGFTINEYSIRCVLFPGTPGEPLPVSCEKDVFDIIGMKYKDPEERNL
```

Supplemental File 2: CO1 sequence alignment by MUSCLE

```
79 666
sp1_AY903204.1          ----------------ATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAACTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAACTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TTCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTAGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAACATGAGGGCCCCCGGAATGACTATGGATAAAATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGTTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAATTTTATTTCAGCATCTTTTC---------
sp1_AY903091.1          AACATTATATCTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAACTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAACTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TTCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTAGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAAATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGTTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAATTTTATTTCAGCATCTTTTC---------
sp1_AY903169.1          AACATTATATCTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAACTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAATTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TTCCGTTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTAGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAAATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGTTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAATTTTATTTCAGCATCTTTTC---------
sp1_AY903191.1          AACATTATATATAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAATTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAATTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TCCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTM
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTGGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAGATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGCTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGGGGAGATCCAATT----------------------------
sp1_AY903192.1          AACATTATATMTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAATTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAATTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TCCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTGGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAGATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGCTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCYTGCTGGAGGAGGAGATCC--------------------------------
sp1_AY903197.1          AACATTATATMTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAATTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAATTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TCCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTGGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAGATACCTC
                        TATTCGTATGATCCGTASTGGTAACCGCAATATTATTGYTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAA-----------------------------
sp1_AY903130.1          AACATTATATCTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAATTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAATTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TCCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTATTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTGGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAGATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGTTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAATTTTATTTCAGCATCTTTTC---------

sp1_KF962060.1          --------ATCTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAATTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAATTGTACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TCCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTACTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTGGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAGATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGTTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAATTTTATTTCAGCATCTTTTCTGATTTTT
sp1_AY903143.1          AACATTATATCTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAATTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAATTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TCCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTGGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAGATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGTTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAATTTTATTTCAGCATCTTTTC---------
sp1_LK022703.1          -ACATTATATCTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAATTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAATTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TCCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTGGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAGATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGTTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAATTTTATTTCAGCATCTTTTC---------
```

```
coerulea_KX691554.1     ---ATTATATCTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAACTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAACTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TTCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTAGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAAATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGTTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAATTTTATTTCAGCATCTTTTC--------
coerulea_MZ052188.1     AACATTATATCTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAACTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAATTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TTCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTAGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAAATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGTTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAATTTTATTTCAGCATCTTTTCTGATTTTT
coerulea_MZ052187.1     --------ATCTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAATTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAATTATACAATGTTGTAGTAACTGCACATGCTTTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TCCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTGGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAGATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGTTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAATTTTATTTCAGCACCTTTTCTGATTTTT
coerulea_MZ052193.1     AACATTATATCTAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAATTGTCAGGACCTGGATCTATGCTAGGA
                        GATGACCAATTATACAATGTTGTAGTAACTGCACATGCTCTAATAATGATTTTCTTTTTTGTTATGCCCGTCTTAATAGGGGGGTTTGGAAATTGACTGG
                        TCCCATTATATATTGGAGCCCCAGATATGGCTTTCCCAAGGCTTAACAATATTAGTTTTTGATTACTACCCCCAGCTCTGCTTTTACTATTAGGGTCTTC
                        TCTTATAGAACAGGGAGCAGGTACTGGATGAACAATTTATCCCCCCCTAAGCGCAATTCAGGCCCATTCCGGTGGTTCAGTGGATATGGCTATATTTAGT
                        TTGCATCTGGCAGGGGCCTCCTCAATAATGGGAGCTATAAATTTTATCACTACCATATTAAATATGAGGGCCCCCGGAATGACTATGGATAAGATACCTC
                        TATTCGTATGATCCGTACTGGTAACCGCAATATTATTGTTATTATCTTTACCTGTCTTGGCTGGGGCTATTACAATGTTATTAACCGACAGAAACTTTAA
                        TACATCCTTCTTTGACCCTGCTGGAGGAGGAGATCCAATTTTATTTCAGCATCTTTTCTGATTTTT
aurita_AY903095.1       AACACTATACTTAATATTTGGTGCTTTTTCCGCCATGGTGGGAACTGCCTTCAGTATGATTATAAGACTGGAACTATCAGGCCCAGGATCCATGTTGGGG
                        GACGATCAGCTATATAACGTTGTAGTAACCGCTCATGCTCTTATAATGATTTTCTTTTTCGTAATGCCCGTTTTGATAGGGGGATTTGGAAACTGGCTAG
                        TTCCCCTATATATAGGAGCTCCAGATATGGCCTTTCCAAGGCTTAACAATATTAGTTTCTGATTATTACCTCCAGCTTTATTACTATTATTAGGGTCTTC
                        TCTTATAGAACAAGGAGCAGGTACTGGTTGAACCATTTACCCTCCTTTAAGTTCAATACAAGCTCATTCCGGAGGTTCAGTAGATATGGCCATATTTAGT
                        CTTCATTTAGCAGGAGCTTCCTCTATTATGGGTGCTATTAACTTTATTACCACTATTTTAAATATGAGAGCCCCTGGTATGACCATGGATAGAATACCTT
                        TATTCGTATGATCTGTATTAGTTACTGCAATCTTATTATTGTTGTCCTTACCCGTATTAGCTGGGGCAATTACCATGTTGTTGACTGATAGAAATTTCAA
                        CACATCCTTCTTTGACCCTGCTGGGGGAGGAGATCCAATACTATTCCAACATTTATTT---------
aurita_AY903093.1       AACACTATACTTAATATTTGGTGCTTTTTCCGCCATGGTGGGAACTGCCTTCAGTATGATTATAAGACTGGAACTATCAGGCCCAGGATCCATGTTGGGG
                        CACGATCAACTATATAACGTTGTAGTAACCGCTCATGCTCTTATAATGATTTTCTTTTTCGTAATGCCCGTTTTGATAGGGGGATTTGGAAACTGGCTAG
                        TTCCCCTATATATAGGAGCTCCAGATATGGCCTTTCCAAGGCTTAACAATATTAGTTTCTGATTATTACCTCCAGCTTTATTACTATTATTAGGGTCTTC
                        TCTTATAGAACAAGGAGCAGGTACTGGTTGAACCATTTACCCTCCTTTAAGTTCAATACAAGCTCATTCCGGAGGTTCAGTAGATATGGCCATATTTAGT
                        CTTCATTTAGCAGGAGCTTCCTCTATTATGGGTGCTATTAACTTTATTACCACTATTTTAAATATGAGAGCCCCTGGTATGACCATGGATAGAATACCTT
                        TATTCGTATGATCTGTATTAGTTACTGCAATCTTATTATTGTTGTCCTTACCCGTATTAGCTGGGGCAATTACCATGTTGTTGACTGATAGAAATTTCAA
                        CACATCCTTCTTTGACCCTGCTGGGGGAGGAGATCCAATACTATTCCAACATTTATTT---------
aurita_AY903094.1       AACACTATACTTAATATTTGGTGCTTTTTCCGCCATGGTGGGAACTGCCTTCAGTATGATTATAAGACTGGAACTATCAGGCCCAGGATCCATGTTGGGG
                        GACGATCAACTATATAACGTTGTAGTGACCGCTCATGCCCTTATAATGATTTTCTTTTTCGTAATGCCCGTTTTGATAGGGGGATTTGGAAACTGACTAG
                        TTCCCCTATATATAGGAGCTCCAGATATGGCCTTTCCAAGGCTTAACAATATCAGTTTCTGATTATTACCTCCAGCTTTATTACTATTATTAGGGTCTTC
                        CCTTATAGAACAAGGAGCAGGTACTGGTTGAACCATTTACCCTCCTTTAAGTTCAATACAAGCTCATTCTGGAGGTTCAGTAGATATGGCCATATTTAGT
                        CTTCATTTAGCAGGAGCTTCCTCTATTATGGGTGCTATTAACTTTATTACCACTATTTTAAATATGAGAGCCCCTGGTATGACCATGGATAGAATACCTT
                        TATTCGTATGATCTGTATTAGTCACTGCAATCTTATTATTGTTGTCCTTACCCGTATTAGCTGGGGCAATTACCATGTTGTTGACTGATAGAAATTTCAA
                        CACATCCTTCTTTGACCCTGCTGGGGGAGGAGATCCAATACTATTCCAACATTTATTT---------

labiata_AY903071.1      GACACTATATCTAATATTCGGTGCTTTTTCTGCTATGGTAGGAACCGCTTTTAGTATGATTATAAGATTGGAACTATCAGGACCTGGGTCCATGTTAGGA
                        GACGATCAATTATATAATGTTGTAGTAACTGCCCACGCTTTCATAATGATATTTTTTTTCGTGATGCCCGTTCTAATAGGAGGGTTTGGAAACTGACTGG
                        TTCCTTTATATATAGGGGCTCCAGATATGGCTTTTCCGAGACTTAACAATATTAGTTTCTGACTCTTGCCCCCGGCTTTATTACTTCTATTAGGGTCCTC
                        CCTTATAGAACAAGGAGCAGGTACTGGATGAACTGTATATCCCCCACTAAGCTCGATTCAGGCTCATTCCGGAGGTTCAGTGGATATGGCTATATTTAGT
                        CTACATTTGGCAGGAGCTTCCTCTATTATGGGTGCTATTAATTTTATTACCACAATTCTAAACATGAGAGCCCCCGGAATGACAATGGATAAGATACCCC
                        TATTTGTTTGGTCTGTATTAGTCACTGCAGTATTATTATTGCTATCCTTGCCTGTCTTGGCTGGGGCAATAACCATGTTGTTAACCGATAGGAACTTCAA
                        TACATCTTTTTTCGATCCTGCTGGAGGAGGAGACCCTATTTTGTTTCAACATCTTTTT--------
labiata_AY903069.1      GACACTATATCTAATATTCGGTGCTTTTTCTGCTATGGTAGGAACCGCTTTTAGTATGATTATAAGATTGGAACTATCAGGACCTGGGTCCATGTTAGGA
                        GACGATCAATTATATAATGTTGTAGTAACTGCCCACGCTTTCATAATGATATTTTTTTTCGTGATGCCCGTTCTAATAGGAGGGTTTGGAAACTGACTGG
                        TTCCTTTATATATAGGGGCTCCAGATATGGCTTTTCCGAGACTTAACAATATTAGTTTCTGACTCTTGCCCCCGGCTTTATTACTTCTATTAGGGTCCTC
                        CCTTATAGAACAAGGAGCAGGTACTGGATGAACTGTACATCCCCCACTAAGCTCGATTCAGGCTCATTCCGGAGGTTCAGTGGATATGGCTATATTTAGT
                        CTACATTTGGCAGGAGCTTCCTCTATTATGGGTGCTATTAATTTTATTACCACAATTCTAAACATGAGAGCCCCCGGAATGACAATGGATAAGATACCCC
                        TATTTGTTTGGTCTGTATTAGTCACTGCAGTATTATTATTGCTATCCTTGCCTGTCTTGGCTGGGGCAATAACCATGTTGTTAACCGATAGGAACTTCAA
                        TACATCTTTTTTCGATCCTGCTGGAGGAGGGGACCCTATTTTGTTTCAACATCTTTTT--------
labiata_AY903070.1      GACACTATATCTAATATTCGGTGCTTTTTCTGCTATGGTAGGAACCGCTTTTAGTATGATTATAAGATTGGAACTATCAGGACCTGGGTCCATGTTAGGA
                        GACGATCAATTATATAATGTTGTAGTAACTGCCCACGCTTTCATAATGATATTTTTTTTCGTGATGCCCGTTCTAATAGGAGGGTTTGGAAACTGACTGG
                        TTCCTTTATATATAGGGGCTCCAGATATGGCTTTTCCGAGACTTAACAATATTAGTTTCTGACTCTTGCCCCCGGCTTTATTACTTCTATTAGGGTCCTC
                        CCTTATAGAACAAGGAGCAGGTACTGGATGAACTGTATATCCCCCACTAAGCTCGATTCAGGCTCATTCCGGAGGTTCAGTGGATATGGCTATATTTAGT
                        CTACATTTGGCAGGAGCTTCCTCTATTATGGGTGCTATTAATTTTATTACCACAATTCTAAACATGAGAGCCCCTGGAATGACAATGGATAAGATACCCC
                        TATTTGTTTGGTCTGTATTAGTCACTGCAGTATTATTATTGCTATCCTTGCCTGTNTTGGCTGGGGCGATAACCATGTTGTTAACCGATAGGAACTTCAA
                        TACATCCTTTTTCGATCCTGCTGGAGGAGGGGACCCTATTTTGTTTCAACATCTTTTT--------
labiata_AY903072.1      GACACTATATCTAATATTCGGTGCTTTTTCTGCTATGGTAGGAACCGCTTTTAGTATGATTATAAGATTGGAACTATCAGGACCTGGGTCCATGTTAGGA
                        GACGATCAATTATATAATGTTGTAGTAACTGCCCACGCTTTCATAATGATATTTTTTTTCGTGATGCCCGTTCTAATAGGAGGGTTTGGAAACTGACTGG
                        TTCCTTTATATATAGGGGCTCCAGATATGGCTTTTCCGAGACTTAACAATATTAGTTTCTGACTCTTGCCCCCGGCTTTATTACTTCTATTAGGGTCCTC
                        CCTTATAGAACAAGGAGCAGGTACTGGATGAACTGTATATCCCCCACTAAGCTCGATTCAGGCTCATTCCGGAGGTTCAGTGGATATGGCTATATTTAGT
                        CTACATTTGGCAGGAGCTTCCTCTATTATGGGTGCTATTAATTTTATTACCACAATTCTAAACATGAGAGCCCCTGGAATGACAATGGATAAGATACCCC
                        TATTTGTTTGGTCTGTATTAGTCACTGCAGTATTATTATTGCTATCCTTGCCTGTCTTGGCTGGGGCGATAACCATGTTGTTAACCGATAGGAACTTCAA
                        TACATCTTTTTTCGATCCTGCTGGAGGAGGGGACCCTATTTTGTTTCAACATCTTTTT--------
```

```
labiata_AY903068.1      GACACTATATCTAATATTCGGTGCTTTTTCTGCTATGGTAGGAACCGCTTTTAGTATGATTATAAGATTGGAACTATCAGGACCTGGGTCCATGTTAGGA
                        GACGATCAATTATATAATGTTGTAGTAACTGCCCACGCTTTCATAATGATATTTTTTTTCGTGATGCCCGTTCTAATAGGAGGGTTTGGAAACTGACTGG
                        TTCCTTTATATATAGGGGCTCCAGATATGGCTTTTCCGAGACTTAACAATATTAGTTTCTGACTCTTGCCCCCGGCTTTATTACTTCTATTAGGGTCCTC
                        CCTTATAGAACAAGGAGCAGGTACTGGATGAACTGTATATCCCCCACTAAGCTCGATTCAGGCTCATTCCGGAGGTTCAGTGAATATGGCTATATTTAGT
                        CTACATTTGGCAGGAGCTTCCTCTATTATGGGTGCTATTAATTTTATTACCACAATTCTAAACATGAGAGCCCCTGGAATGACAATGGATAAGATACCCC
                        TATTTGTTTGGTCTGTATTAGTCACTGCAGTATTATTATTGCTATCCTTGCCTGTCTTGGCTGGGGCGATAACCATGTTGTTAACCGATAGGAACTTCAA
                        TACATCTTTTTTCGATCCTGCTGGAGGAGGGGACCCTATTTTGTTTCAACATCTTTTT--------
labiata_AY903074.1      AACACTATATCTAATATTCGGTGCTTTTTCTGCTATGGTAGGAACCGCTTTTAGTATGATTATAAGATTGGAACTATCAGGACCCGGGTCCATGTTAGGA
                        GACGATCAATTATATAATGTTGTAGTAACTGCCCACGCTTTCATAATGATATTTTTTTTCGTGATGCCCGTCCTAATAGGAGGGTTCGGAAACTGGTTGG
                        TTCCCCTATATATAGGAGCCCCAGATATGGCTTTCCCGAGACTTAACAATATTAGTTTTTGACTCTTACCCCCAGCTTTATTACTTCTATTAGGATCTTC
                        TCTTATAGAACAAGGAGCAGGTACCGGATGAACTGTATACCCTCCACTAAGCTCGATTCAGGCTCATTCCGGAGGTTCGGTGGATATGGCTATATTTAGT
                        TTACATTTGGCAGGGGCTTCTTCTATCATGGGTGCTATTAATTTTATAACTACCATTCTGAATATGAGAGCCCCTGGAATGACAATGGATAAGATACCCC
                        TATTTGTTTGGTCTGTATTAGTTACTGCAGTATTATTATTGTTATCCTTGCCTGTTTTGGCTGGAGCAATAACCATGTTACTAACCGACAGGAACTTTAA
                        TACATCTTTTTTCGATCCTGCTGGGGGAGGAGACCCTATTTTGTTTCAACATCTTTTT--------
labiata_AY903077.1      AACACTATATCTAATATTCGGTGCTTTTTCTGCTATGGTAGGAACCGCTTTTAGTATGATTATAAGATTGGAACTATCAGGACCCGGGTCCATGTTAGGA
                        GACGATCAATTATATAATGTTGTAGTAACTGCCCACGCTTTCATAATGATATTTTTTTTCGTAATGCCCGTCCTAATAGGAGGGTTCGGAAACTGGTTGG
                        TTCCCCTATATATAGGAGCCCCAGATATGGCTTTCCCGAGACCTAACAATATTAGTTTTTGACTCTTACCCCCAGCTTTATTACTTCTATTAGGATCTTC
                        TCTTATAGAACAAGGAGCAGGTACCGGATGAACTGTATACCCTCCACTAAGCTCGATTCAGGCTCATTCCGGAGGTTCGGTGGATATGGCTATATTTAGT
                        TTACATTTGGCAGGGGCTTCTTCTATCATGGGTGCTATTAATTTTATAACTACCATTCTGAACACGAGAGCCCCTGGAATGACAATGGATAAGATACCCC
                        TATTTGTTTGGTCTGTATTAGTTACTGCAGTATTATTATTGTTATCCTTGCCTGTTTTGGCTGGAGCAATAACCATGTTACTAACCGACAGGAACTTTAA
                        TACATCTTTTTTCGATCCTGCTGGGGGAGGAGACCCTATTTTGTTTCAACATCTTTTT--------
sp4_AY903111.1          ----------ATTTTCGGGGCCTTTTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTAGAGCTATCAGGACCTGGATCCATGTTAGGG
                        GACGATCAACTATACAATGTTGTGGTCACTGCTCATGCCCTTATAATGATCTTCTTTTTTGTAATGCCAGTTTTAATAGGGGGATTTGGAAACTGACTGG
                        TGCCTTTATACATAGGAGCTCCAGATATGGCCTTCCCTAGGCTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATTGAACAAGGGGCAGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCTGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATGCCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAATTCTATTTCAACATCTTTTC---------
sp4_AY903108.1          AACATTATATTTAATATTCGGAGCTTTCTCCGCTATGGTAGGAACTGCATTCAGTATGATTATCAGATTGGAACTATCAGGTCCTGGATCCATGCTAGGA
                        GATGATCAACTATACAACGTTGTAGTTACCGCTCATGCTTCTATAATGATCTTTTTCTTCGTAATGCCCGTTCTAATAGGAGGTTTTGGAAATTGATTGG
                        TACCCTTGTATATAGGAGCTCCAGATATGGCCTTCCCTAGGCTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATTGAACAAGGGGCAGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCTGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATACCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAATTCTATTTCAACATCTTTTC---------
sp4_AY903109.1          AACATTATATTTAATATTCGGAGCTTTCTCCGCTATGGTAGGAACTGCATTCAGTATGATTATCAGATTGGAGCTATCAGGTCCTGGATCCATGCTAGGA
                        GATGATCAACTATACAACGTTGTAGTTACCGCTCATGCTCTTATAATGATCTTTTTCTTCGTAATGCCCGTTCTAATAGGAGGTTTTGGAAATTGATTGG
                        TACCCTTGTATATAGGAGCTCCAGATATGGCCTTCCCTAGGCTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATTGAACAAGGGGCGGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCTGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATACCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAATTCTATTTCAACATCTTTTC---------
sp4_AY903102.1          AACATTATATTTAATATTCGGAGCTTTCTCCGCTATGGTAGGAACTGCATTCAGTATGATTATCAGATTGGAACTATCAGGTCCTGGATCCATGCTAGGA
                        GATGATCAACTATACAACGTTGTAGTTACCGCTCATGCTCTTATAATGATCTTTTTCTTCGTAATGCCCGTTCTAATAGAAGGTTTTGGAAATTGATTGG
                        TACCCTTGTATATAGGAGCTCCAGATATGGCCTTCCCTAGGCTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATNGAACAAGGGGCAGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCCGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATACCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAATTCTATTTCAACATCTTTTC---------

sp4_AY903145.1          AACATTATATTTAATATTCGGAGCTTTCTCCGCTATGGTAGGAACTGCATTCAGTATGATTATCAGATTGGAACTATCAGGTCCTGGATCTATGCTAGGA
                        GATGATCAACTATACAACGTTGTAGTTACCGCTCATGCTCTTATAATGATCTTTTTCTTCGTAATGCCCGTTCTAATAGGAGGTTTTGGAAATTGATTGG
                        TACCCTTGTATATAGGAGCTCCAGATATGGCCTTCCCTAGGTTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATTGAACAAGGGGCAGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCCGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATACCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAATTCTATTTCAACATCTTTTC---------
sp4_AY903110.1          AACATTATATTTAATATTCGGAGCTTTCTCCGCTATGGTAGGAACTGCATTCAGTATGATTATCAGATTGGAACTATCAGGTCCTGGATCCATGCTAGGA
                        GATGATCAACTATACAACGTTGTAGTTACCGCTCATGCTCTTATAATGATCTTTTTCTTCGTAATGCCCGTTCTAATAGGAGGTTTTGGAAATTGATTGG
                        TACCCTTGTATATAGGAGCTCCAGATATGGCCTTCCCTAGGCTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATTGAACAAGGGGCAGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCTGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATACCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAAT----------
sp4_AY903098.1          AACATTATATTTAATATTCGGAGCTTTCTCCGCTATGGTAGGAACTGCATTCAGTATGATTATCAGATTGGAACTATCAGGTCCTGGATCCATGCTAGGA
                        GATGATCAACTATACAACGTTGTAGTTACCGCTCATGCTCTTATAATGATCTTTTTCTTCGTAATGCCCGTTCTAATAGGAGGTTTTGGAAATTGATTGG
                        TACCCTTGTATATAGGAGCTCCAGATATGGCCTTCCCTAGGCTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATTGAACAAGGGGCAGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCTGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATACCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAATTCTATTTCAACATCTTTTC---------
sp4_AY903097.1          AACGTTATATTTAATATTCGGAGCTTTCTCCGCTATGGTAGGAACTGCATTCAGTATGATTATCAGATTGGAACTATCAGGTCCTGGATCCATGCTAGGA
                        GATGATCAACTATACAACGTTGTAGTTACCGCTCATGCTCTTATAATGATCTTTTTCTTCGTAATGCCCGTTCTAATAGGAGGTTTTGGAAATTGATTGG
                        TACCCTTGTATATAGGAGCTCCAGATATGGCCTTCCCTAGGCTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATTGAACAAGGGGCAGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCTGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATACCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAATTCTATTTCAACATCTTTTC---------
```

```
sp4_AY903107.1          --------------------GGAGCTTTCTCCGCTATGGTAGGAACTGCATTCAGTATGATTATCAGATTGGAACTATCAGGTCCTGGATCCATGCTAGGA
                        GATGATCAACTATACAACGTTGTAGTTACCGCTCATGCTCTTATAATGATCTTTTTCTTCGTAATGCCCGTTCTAATAGGAGGTTTTGGAAATTGATTGG
                        TACCCTTGTATATAGGAGCTCCAGATATGGCCTTCCCTAGGCTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATTGAACAAGGGGCAGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCTGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATACCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAATTCTATTTCAACATCTTTTC---------
sp4_AY903144.1          AACATTATATTTAATATTCGGAGCTTTCTCCGCTATGGTAGGAACTGCATTCAGTATGATTATCAGATTGGAACTATCAGGTCCTGGATCCATGCTAGGA
                        GATGATCAACTATACAACGTTGTAGTTACCGCTCATGCTCTTATAATGATCTTTTTCTTCGTAATGCCCGTTCTAATAGGAGGTTTCGGAAATTGATTGG
                        TACCCTTGTATATAGGAGCTCCAGATATGGCCTTCCCTAGGCTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATTGAACAAGGGGCAGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCCGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATACCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAATTCTATTTCAACATCTTTTC---------
sp4_AY903136.1          AACATTATATTTAATATTCGGAGCTTTCTCCGCTATGGTAGGAACTGCATTCAGTATGATTATCAGATTGGAACTATCAGGTCCTGGATCCATGCTAGGA
                        GATGATCAACTATACAACGTTGTGGTTACCGCTCATGCTCTTATAATGATCTTTTTCTTCGTAATGCCCGTTCTAATAGGAGGTTTTGGAAATTGATTGG
                        TACCCTTGTATATAGGAGCTCCAGATATGGCCTTCCCTAGGCTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATTGAACAAGGGGCAGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCCGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATACCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAATTCTATTTCAACATCTTTTC---------
sp4_AY903101.1          AACATTATATTTAATATTCGGAGCTTTCTCCGCTATGGTAGGAACTGCATTCAGTATGATTATCAGATTGGAACTATCAGGTCCTGGATCCATGCTAGGA
                        GATGATCAACTATACAACGTTGTAGTTACCGCTCATGCTCTTATAATGATCTTTTTCTTCGTAATGCCCGTTCTAATAGGAGGTTTTGGAAATTGATTGG
                        TACCCTTGTATATAGGAGCTCCAGATATGGCCTTCCCTAGGCTGAACAATATAAGTTTCTGACTTCTACCTCCAGCTTTATTGCTTTTATTAGGATCTTC
                        ATTGATTGAACAAGGGGCAGGTACCGGATGAACCATATATCCTCCTCTAAGTTCTATTCAGGCTCATTCCGGAGGTTCTGTAGATATGGCTATATTCAGC
                        TTACATTTGGCAGGAGCCTCATCCATAATGGGTGCTATTAATTTTATTACAACTATCCTAAATATGAGAGCCCCTGGAATGACGATGGACAGAATACCTT
                        TATTTGTTTGGTCTGTTTTGGTCACTGCAATACTACTGCTATTATCTCTACCCGTCCTCGCCGGAGCAATTACTATGCTATTAACTGACAGGAATTTTAA
                        CACATCTTTCTTTGATCCTGCTGGAGGAGGGGATCCAATTCTATTTCAACATCTTTTC---------
sp6_AY903129.1          AACACTATATTTAATTTTCGGGGCCTTTTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTAGAGCTATCAGGACCTGGATCCATGTTAGGG
                        GACGATCAACTATACAATGTTGTGGTCACTGCTCATGCCCTTATAATGATCTTCTTTTTTGTAATGCCAGTTTTAATAGGGGGATTTGGAAACTGACTGG
                        TGCCTTTATACATAGGAGCTCCGGATATGGCCTTTCCAAGACTTAACAATATTAGTTTTTGGCTTTTACCTCCAGCTCTACTACTTCTATTGGGATCTTC
                        ACTAATAGAACAAGGAGCAGGTACTGGATGAACTGTTTACCCTCCTTTAAGTTCAATTCAAGCTCATTCAGGAGGTTCTGTAGATATGGCAATATTTAGC
                        CTACATCTGGCAGGGGCTTCATCTATAATGGGTGCTATTAATTTTATAACTACTATTTTAAATATGAGAGCCCCTGGAATGACAATGGATAGAATACCTC
                        TATTTGTATGATCTGTACTAGTTACGGCAATCCTGTTATTACTTTCTCTTCCAGTCCTTGCTGGAGCAATTACAATGCTTCTAACTGATAGAAATTTCAA
                        TACATCTTTCTTTGATCCCGCTGGAGGAGGAGACCCAATACTATTCCAACATCTCTTT---------
sp6_AY903099.1          AACACTATATTTAATTTTCGGGGCCTTTTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTAGAGTTATCAGGACCTGGATCCATGTTAGGG
                        GACGATCAACTATACAATGTTGTGGTCACTGCCCATGCCCTTATAATGATCTTCTTTTTTGTAATGCCAGTTTTAATAGGGGGATTTGGAAACTGACTGG
                        TGCCTTTATATATAGGAGCTCCGGATATGGCCTTTCCAAGACTTAACAATATTAGTTTTTGACTTTTACCTCCAGCTCTGCTACTTCTATTGGGATCTTC
                        ACTAATAGAACAAGGAGCAGGTACTGGGTGAACTGTTTACCCTCCTTTGAGTTCAATTCAAGCTCATTCAGGAGGTTCTGTAGATATGGCAATATTTAGC
                        CTACACCTGGCAGGGGCTTCATCCATAATGGGTGCTATTAATTTTATAACTACTATTTTAAATATGAGAGCCCCTGGAATGACAATGGATAGAATACCTC
                        TATTTGTATGATCTGTATTAGTTACAGCAATCCTGTTATTACTTTCCCTTCCAGTCCTTGCTGGAGCAATTACAATGCTCCTAACTGATAGAAATTTCAA
                        TACATCTTTCTTTGATCCCGCTGGAGGAGGAGACCCAATACTATTCCAACATCTCTTT---------
sp6_AY903100.1          AACACTATATTTAATTTTCGGGGCCTTTTCTGCTATGGTAGGAACTGCCTTTAGTATGATTATAAGATTAGAGTTATCAGGACCTGGATCCATGTTAGGG
                        GACGATCAACTATACAATGTTGTGGTCACTGCTCATGCCCTTATAATGATCTTCTTTTTTGTAATGCCAGTTTTAATAGGGGGATTTGGAAACTGACTGG
                        TGCCTTTATATATAGGGGCCCCGGATATGGCCTTTCCAAGACTTAACAATATTAGTTTTTGACTTTTGCCTCCAGCTCTGCTACTTCTATTGGGATCTTC
                        ACTAATAGAACAAGGAGCAGGTACTGGGTGAACTGTTTACCCTCCTTTAAGTTCAATTCAAGCTCATTCAGGAGGTTCTGTAGATATGGCAATATTTAGC
                        CTACATCTGGCAGGGGCTTCATCCATAATGGGTGCTATTAATTTTATAACTACTATTTTAAATATGAGAGCCCCTGGAATGACAATGGATAGAATACCTC
                        TATTTGTATGATCTGTATTAGTTACGGCAATCCTGTTATTACTTTCCCTTCCAGTCCTTGCTGGAGCAATTACAATGCTCCTAACTGATAGAAATTTCAA
                        TACATCTTTCTTTGATCCCGCTGGAGGAGGAGACCCAATACTATTCCAACATCTCTTT---------

sp7_AY903139.1          AACACTATATTTGATATTCGGTGCTTTTTCTGCCATGGTAGGAACTGCCTTTAGTATGATTATAAGATTGGAACTATCAGGACCTGGATCCATGTTGGGA
                        GACGATCAATTATATAATGTCGTAGTTACTGCCCACGCTCTTATAATGATCTTCTTCTTTGTAATGCCCGTTCTAATAGGAGGATTTGGAAACTGGCTGG
                        TTCCTCTATACATAGGAGCTCCAGACATGGCCTTCCCAAGACTTAATAACATCAGTTTCTGGCTACTACCTCCAGCTTTATTACTTTTATTGGGATCTTC
                        CTTAATAGAACAAGGAGCAGGAACCGGATGAACTATATATCCGCCTCTCAGCTCAATTCAAGCTCATTCTGGAGGTTCTGTAGATATGGCAATATTTAGT
                        CTGCATTTGGCGGGGGCTTCATCCATTATGGGTGCTATTAACTTTATAACCACCATTTTGAATATGAGAGCTCCTGGGATGACAATGGATAAAATACCTT
                        TATTTGTTTGATCTGTATTAGTTACTGCAATATTATTGTTATTGTCTTTACCTGTACTTGCCGGAGCAATTACTATGTTATTAACTGATAGGAATTTTAA
                        TACATCTTTCTTCGATCCTGCTGGAGGAGGAGATCCAATCTTGTTTCAGCATCTCTTT---------
sp9_KU984356.1          ---ACTATATATTTTATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTAGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTTATTCCAACACCTCTTC---------
sp9_KU984345.1          ---ACTATATATTTTATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTGGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTGGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGAGGAGGGGATCCTATCTTATTCCAACACCTCTTC---------
sp9_AY903174.1          ---------------------------------GCTATGGNAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTAGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTNTCTT-------------------------
```

```
sp9_MN531716.1          AACACTATATATTTTATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTAGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCTTTCTTTGAT----------------------------------------------------------
sp9_KU984344.1          ---ACTATATATTTTATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTTCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTAGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTTATTCCAACACCTCTTC---------
sp9_KU984347.1          ---ACTATATATTTTATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TGCCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTAGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTTATTCCAACACCTCTTC---------
sp9_KU984363.1          ---ACTATATATTTTATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTAGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTTATTCCAACACCTCTTC---------
sp9_KY564357.1          ---ACTATATATTTTATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTAGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTTATTCCAACACCTCTTC---------
sp9_AY903175.1          ----------------------GCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTAGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTTAT------------------------
sp9_AY903173.1          ------------------------------TGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTAGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGG------------------------------------------
```

```
sp9_MN531717.1          ------------------------------TCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCTTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTGGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        C-----------------------------------------------------------------------
sp9_AY903176.1          ---------------------------------GCTATGGNAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTGGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTT-------------------------
sp9_KY564356.1          ---ACTATATATTTTATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTAATCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTTATTCCAACACCTCTTC---------
sp9_KY564354.1          ---ACTATATATTTTATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTAGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTTATTCCAACACCTCTTC---------
```

```
sp9_KU984357.1          ---ACTATATATTTTATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAACTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTGGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTTATTCCAACACCTCTTC---------
sp9_KU984343.1          ---ACTATATATTTTATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTATTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTGGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTTATTCCAACACCTCTTC---------
sp9_AY903172.1          ------------------TCGGTGCTTTCTCTGCTATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTGGAGCTGTCTGGACCTGGCTCCATGCTCGGT
                        GATGACCAGCTTTATAACGTCGTGGTCACCGCTCACGCCTTCATAATGATTTTCTTTTTCGTAATGCCTGTGCTAATAGGGGGTTTTGGAAATTGATTTG
                        TACCCTTATATATAGGGGCTCCTGATATGGCCTTCCCAAGACTTAATAACATTAGTTTCTGGTTACTACCACCCGCTCTGCTTCTGTTGTTAGGATCTTC
                        TTTAATAGAGCAAGGCGCAGGAACGGGTTGAACCGTTTATCCTCCCCTAAGTTCTATACAAGCTCATTCTGGAGGTTCCGTCGATATGGCAATATTTAGT
                        CTACATCTAGCTGGGGCTTCCTCTATAATGGGGGCCATAAATTTCATTACCACCATTTTGAACATGAGAGCCCCTGGTATGAGTATGGATAAAATACCCT
                        TGTTTGTTTGATCTGTTCTGGTCACCGCAATTTTACTACTACTCTCCTTACCTGTATTAGCTGGTGCTATAACTATGCTCTTGACAGATAGGAATTTTAA
                        CACATCCTTCTTTGATCCTGCCGGGGGAGGGGATCCTATCTTATTCCAACACCTCTTC---------
sp12_KY564344.1         ---ACTATATATTATATTCGGTGCCTTCTCTGCTATGGTGGGAACTGCTTTTAGTATGATAATAAGATTAGAGTTATCTGGGCCTGGCTCTATGCTTGGT
                        GACGATCAATTATACAACGTTGTAGTAACTGCCCACGCTTTTGTTATGATCTTCTTTTTTGTAATGCCTGTTTTAATAGGTGGTTTCGGGAATTGGTTTG
                        TACCCCTATACATAGGTGCTCCTGATATGGCCTTCCCAAGGTTAAACAATATCAGTTTCTGGTTATTGCCACCTGCTCTTCTACTGCTGCTAGGATCTTC
                        GCTCATAGAGCAGGGGGCAGGAACTGGCTGAACTGTCTATCCCCCCTTAAGTTCGATACAAGCTCACTCTGGAGGTTCTGTTGACATGGCCATTTTTAGC
                        CTACATTTAGCAGGAGCTTCTTCCATAATGGGAGCAATAAACTTCATTACCACCATATTGAATATGAGGGCCCCCGGGATGACCATGGACAAAATACCGC
                        TATTCGTTTGGTCTGTTCTGGTGACCGCAGTTCTGCTACTGCTTTCTTTACCCGTCCTTGCTGGTGCCATAACTATGCTTTTGACGAGACAGGAACTTTAA
                        CACGTCCTTTTTTGATCCTGCCGGAGGAGGAGATCCTATTTTGTTCCAGCACCTATTT---------
sp12_KY564346.1         ---ACTATATATTATATTCGGTGCCTTCTCTGCTATGGTGGGAACTGCTTTTAGTATGATAATAAGATTAGAGTTATCTGGGCCTGGCTCTATGCTTGGT
                        GACGATCAATTATACAACGTTGTAGTAACTGCCCACGCTTTTGTTATGATCTTCTTTTTTGTAATGCCTGTTTTAATAGGTGGTTTCGGGAATTGGTTTG
                        TACCCCTATACATAGGTGCTCCTGATATGGCCTTCCCAAGGTTAAACAATATCAGTTTCTGGTTATTGCCACCTGCTCTTCTGCTGCTGCTAGGATCTTC
                        GCTCATAGAGCAGGGGGCAGGAACTGGCTGAACTGTCTATCCCCCCTTAAGTTCGATACAAGCTCACTCTGGAGGTTCTGTTGACATGGCCATTTTTAGC
                        CTACATTTAGCAGGAGCTTCTTCCATAATGGGAGCAATAAACTTCATTACCACCATATTGAATATGAGGGCCCCCGGGATGACCATGGACAAAATACCGC
                        TATTCGTTTGGTCTGTTCTGGTGACCGCAGTTCTGCTACTGCTTTCTTTACCCGTCCTTGCTGGTGCCATAACTATGCTTTTGACGGACAGGAACTTTAA
                        CACGTCCTTTTTTGATCCTGCCGGAGGAGGAGATCCTATTTTGTTCCAGCACCTATTT---------
sp13_KY564338.1         ---ACTATATATTATATTCGGCGCTTTCTCTGCCATGGTGGGAACTGCTTTTAGCATGATAATTAGATTAGAATTATCTGGACCTGGCTCTATGCTTGGT
                        GATGATCAATTATACAATGTTGTAGTAACTGCCCACGCTTTCGTAATGATTTTCTTTTTCGTAATGCCTGTTCTAATAGGTGGTTTTGGGAATTGGTTTG
                        TACCCCTATACATAGGTGCCCCCGACATGGCCTTTCCGAGACTGAATAATATCAGTTTCTGGCTATTGCCCCCCGCTCTTCTATTACTACTAGGATCCTC
                        ATTAATAGAACAAGGAGCAGGAACTGGCTGAACTGTTTATCCTCCCTTAAGTTCCATTCAAGCTCATTCTGGGGGTTCCGTTGATATGGCCATTTTCAGC
                        CTGCACTTAGCAGGAGCTTCTTCTATAATGGGAGCAATAAACTTTATTACAACCATATTGAACATGAGGGCCCCTGGGATGAGTATGGACAAAATACCTT
                        TATTTGTCTGGTCTGTTTTAGTAACTGCAATTCTGTTATTACTTTCTCTACCTGTCTTGGCGGGCGCCATTACTATGCTTTAACAGATAGAAATTTTAA
                        CACATCCTTTTTCGATCCTGCTGGGGGAGGAGACCCCATCCTATTCCAACATTTATTT---------
sp13_KY564335.1         ---ACTATATATTATATTCGGCGCTTTCTCTGCCATGGTGGGAACTGCTTTTAGCATGATAATTAGATTAGAATTATCTGGACCTGGCTCTATGCTTGGT
                        GATGATCAATTATACAATGTTGTAGTAACTGCCCACGCTTTCGTAATGATTTTCTTTTTCGTGATGCCTGTTCTAATAGGTGGTTTTGGGAATTGGTTTG
                        TACCCCTATACATAGGTGCCCCCGACATGGCCTTTCCGAGACTGAATAATATCAGTTTCTGGCTATTGCCCCCCGCTCTTCTATTACTATTAGGATCCTC
                        ATTAATAGAACAAGGAGCAGGAACTGGCTGAACTGTTTATCCTCCCTTAAGTTCCATTCAAGCTCATTCTGGAGGTTCCGTTGATATGGCCATTTTCAGC
                        CTGCACTTAGCAGGAGCTTCTTCTATAATGGGAGCAATAAACTTTATTACAACCATATTGAACATGAGGGCCCCTGGGATGAGTATGGACAAAATACCTT
                        TATTTGTCTGGTCTGTTTTAGTAACTGCGATTCTGTTATTACTTTCTCTACCTGTCTTGGCGGGCGCCATTACTATGCTTCTAACAGATAGAAATTTTAA
                        CACATCCTTTTTCGATCCTGCTGGGGGAGGAGACCCCATCCTATTCCAACATTTATTT---------

sp13_KY564331.1         ---ACTATATATTATATTCGGCGCTTTCTCTGCCATGGTGGGAACTGCTTTTAGCATGATAATTAGATTAGAATTATCTGGACCTGGCTCTATGCTTGGT
                        GATGATCAATTATACAATGTTGTAGTAACTGCCCACGCTTTCGTAATGATTTTCTTTTTCGTGATGCCTGTTCTAATAGGTGGTTTTGGGAATTGGTTTG
                        TACCCCTATACATAGGTGCCCCCGACATGGCCTTTCCGAGACTGAATAATATCAGTTTCTGGCTATTGCCCCCCGCTCTTCTATTACTACTAGGATCCTC
                        ATTAATAGAACAAGGAGCAGGAACTGGCTGAACTGTTTATCCTCCCTTAAGTTCCATTCAGGCTCATTCTGGAGGTTCCGTTGATATGGCCATTTTCAGC
                        CTGCACTTAGCAGGAGCTTCTTCTATAATGGGAGCAATAAACTTTATTACAACCATATTGAACATGAGGGCCCCTGGGATGAGTATGGACAAAATACCTT
                        TATTTGTCTGGTCTGTTTTAGTAACTGCGATTCTGTTATTACTTTCTCTACCTGTCTTGGCGGGCGCCATTACTATGCTTCTAACAGATAGAAATTTTAA
                        CACATCCTTTTTCGATCCTGCTGGGGGAGGAGACCCCATCCTATTCCAACATTTATTT---------
sp13_KY564337.1         ---ACTATATATTATATTCGGCGCTTTCTCTGCCATGGTGGGAACTGCTTTTAGCATGATAATTAGATTAGAATTATCTGGACCTGGCTCTATGCTTGGT
                        GATGATCAATTATACAATGTTGTAGTAACTGCCCACGCTTTCGTAATGATTTTCTTTTTCGTGATGCCTGTTCTAATAGGTGGTTTTGGGAATTGGTTTG
                        TACCCCTATACATAGGTGCCCCCGACATGGCCTTCCCGAGACTGAATAATATCAGTTTCTGGCTATTGCCCCCCGCTCTTCTATTACTACTAGGATCCTC
                        ATTAATAGAACAAGGAGCAGGAACTGGCTGAACTGTTTATCCTCCCTTAAGTTCCATTCAAGCTCATTCTGGAGGTTCCGTTGATATGGCCATTTTCAGC
                        CTGCACTTAGCAGGAGCTTCTTCTATAATGGGAGCAATAAACTTTATTACAACCATATTGAACATGAGGGCCCCTGGGATGAGTATGGACAAAATACCTT
                        TATTTGTCTGGTCTGTTTTAGTAACTGCGATTCTGTTATTACTTTCTCTACCTGTCTTGGCGGGCGCCATTACTATGCTTCTAACAGATAGAAATTTTAA
                        CACATCCTTTTTCGATCCTGCTGGGGGAGGAGACCCCATCCTATTCCAACATTTATTT---------
sp13_KY564334.1         ---ACTATATATTATATTCGGCGCTTTCTCTGCCATGGTGGGAACTGCTTTTAGCATGATAATTAGATTAGAATTATCTGGACCTGGTTCTATGCTTGGT
                        GATGATCAATTATACAATGTTGTAGTAACTGCCCACGCTTTCGTAATGATTTTCTTTTTCGTGATGCCTGTTCTAATAGGTGGTTTTGGGAATTGGTTTG
                        TACCCCTATACATAGGTGCCCCCGACATGGCCTTTCCGAGACTGAATAATATCAGTTTCTGGCTATTGCCCCCCGCTCTTCTATTACTACTAGGATCCTC
                        ATTAATAGAACAAGGAGCAGGAACTGGCTGAACTGTTTATCCTCCCTTAAGTTCCATTCAAGCTCATTCTGGAGGTTCCGTTGATATGGCCATTTTCAGC
                        CTGCACTTAGCAGGAGCTTCTTCTATAATGGGAGCAATAAACTTTATTACAACCATATTGAACATGAGGGCCCCTGGGATGAGTATGGACAAAATACCTT
                        TATTTGTCTGGTCTGTTTTAGTAACTGCGATTCTGTTATTACTTTCTCTACCTGTCTTGGCGGGCGCCATTACTATGCTTCTAACAGATAGAAATTTTAA
                        CACATCCTTTTTCGATCCTGCTGGGGGAGGAGACCCCATCCTATTCCAACATTTATTT---------
sp13_KY564332.1         ---ACTATATATTATATTCGGCGCTTTCTCTGCCATGGTGGGAACTGCTTTTAGCATGATAATTAGATTAGAATTATCTGGACCTGGCTCTATGCTTGGT
                        GATGATCAATTATACAATGTTGTAGTAACTGCCCACGCTTTCGTAATGATTTTCTTTTTCGTGATGCCTGTTCTAATAGGTGGTTTTGGGAATTGGTTTG
                        TACCCCTATACATAGGTGCCCCCGACATGGCCTTTCCGAGACTGAATAATATCAGTTTCTGGCTATTGCCCCCCGCTCTTCTATTACTACTAGGATCCTC
                        ATTAATAGAACAAGGAGCAGGAACTGGCTGAACTGTTTATCCTCCCTTAAGTTCCATTCAAGCTCATTCTGGAGGTTCCGTTGATATGGCCATTTTCAGC
                        CTGCACTTAGCAGGAGCTTCTTCTATAATGGGAGCAATAAACTTTATTACAACCATATTGAACATGAGGGCCCCTGGGATGAGTATGGACAAAATACCTT
                        TATTTGTCTGGTCTGTTTTAGTAACTGCGATTCTGTTATTACTTTCTCTACCTGTCTTGGCGGGCGCCATTACTATGCTTCTAACAGATAGAAATTTTAA
                        CACATCCTTTTTCGATCCTGCTGGGGGAGGAGACCCCATCCTATTCCAACATTTATTT---------
```

```
sp14                      AACACTATATATAATATTCGGTGCTTTTTCTGCCATGGTGGGAACTGCCTTTAGTATGATTATAAGATTAGAGTTGTCAGGACCTGGATCTATGCTAGGG
                          GACGATCAACTATATAATGTTGTAGTTACTGCTCATGCCCTTATAATGATTTTCTTTTTTGTAATGCCCGTTTTGATAGGGGGATTTGGAAACTGATTAG
                          TACCCTTATACATAGGAGCTCCAGATATGGCATTCCCAAGACTCAATAATATTAGTTTCTGGCTATTACCCCCGGCCTTATTACTTTTATTAGGTTCTTC
                          ATTAATAGAGCAAGGAGCAGGCACAGGATGAACCGTATATCCTCCTCTGAGCTCAATTCAAGCCCATTCCGGAGGTTCTGTAGATATGGCTATATTTAGT
                          CTTCATCTAGCTGGAGCTTCATCCATAATGGGTGCTATCAATTTTATAACTACCATTTTCAACATGAGAGCCCCCGGAATGACAATGGATAAAATACCTT
                          TATTTGTTTGATCTGTCTTAGTCACCGCAATACTATTATTATTATCCCTACCCGTTCTTGCCGGAGCAATTACCATGCTACTGACCGACAGAAACTTCAA
                          TACATCTTTTTTCGATCCTGCTGGAGGAGGAGATCCAATTTTATTTCAACATCTTTTTTGATTCTT
sp15_KY564352.1           ---ACTATATATTATATTTGGTGCTTTCTCCGCCATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTAGAGTTGTCTGGACCTGGCTCCATGCTTGGC
                          GATGACCAACTTTATAACGTTGTGGTCACTGCCCACGCTTTCGTAATGATTTTCTTCTTTGTAATGCCTGTACTGATAGGCGGCTTTGGGAACTGATTTG
                          TCCCCTTGTATATAGGTGCTCCCGATATGGCTTTTCCAAGACTTAACAATATTAGTTTCTGGTTACTACCACCTGCTCTACTTCTACTACTAGGATCCTC
                          CCTGATAGAACAAGGGGCAGGAACGGGTTGAACCGTTTACCCTCCTCTGAGCTCTATACAAGCTCATTCTGGAGGTTCTGTCGATATGGCTATATTTAGT
                          CTACACTTGGCAGGGGCTTCCTCTATAATGGGAGCTATAAATTTCATTACTACCATTTTGAATATGAGAGCTCCTGGCATGAGCATGGATAAAATACCAC
                          TGTTCGTTTGATCTGTTTTGGTCACCGCAGTTCTGCTATTACTCTCCTTGCCTGTGCTAGCTGGTGCTATAACTATGCTTTTAACAGACAGGAATTTTAA
                          CACGTCCTTCTTTGATCCTGCCGGGGGAGGAGATCCTGTCTTGTTCCAGCACCTGTTC---------
sp15_KY564351.1           ---ACTATATATTATATTTGGTGCTTTCTCCGCCATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTAGAGTTGTCTGGGCCTGGCTCCATGCTTGGC
                          GATGACCAACTTTATAACGTTGTGGTCACTGCCCACGCTTTCGTAATGATTTTCTTCTTTGTAATGCCTGTACTGATAGGTGGCTTTGGGAACTGATTTG
                          TCCCCTTGTATATAGGTGCTCCCGATATGGCTTTTCCAAGACTTAACAATATTAGTTTCTGGTTACTACCACCTGCTCTACTTCTACTACTAGGATCCTC
                          CCTGATAGAACAAGGGGCAGGAACGGGTTGAACCGTTTACCCTCCTCTGAGCTCTATACAAGCTCATTCTGGAGGTTCTGTCGATATGGCTATATTTAGT
                          CTACACTTGGCAGGGGCTTCCTCTATAATGGGAGCTATAAATTTCATTACTACCATTTTGAATATGAGAGCTCCTGGCATGAGCATGGATAAAATACCAC
                          TGTTCGTTTGATCTGTTTTGGTCACCGCAGTTCTGCTATTACTCTCCTTGCCTGTGCTAGCTGGTGCTATAACTATGCTTTTAACAGACAGGAATTTTAA
                          CACGTCCTTCTTTGATCCTGCCGGGGGAGGAGATCCTGTCTTGTTCCAGCACCTGTTC---------
sp15_KY564350.1           ---ACTATATATTATATTTGGTGCTTTCTCCGCCATGGTAGGAACTGCTTTTAGTATGATAATTAGGTTAGAGTTGTCTGGGCCTGGCTCCATGCTTGGC
                          GATGACCAACTTTATAACGTTGTGGTCACTGCCCACGCTTTCGTAATGATTTTCTTCTTTGTAATGCCTGTACTGATAGGTGGCTTTGGGAACTGATTTG
                          TCCCCTTGTATATAGGTGCTCCCGATATGGCTTTTCCAAGACTTAACAATATTAGTTTCTGGTTACTACCACCTGCTCTACTTCTACTACTAGGATCCTC
                          CCTGATAGAACAAGGGGCAGGAACGGGTTGAACCGTTTACCCTCCTCTGAGCTCTATACAAGCTCATTCTGGAGGTTCTGTCGATATGGCTATATTTAGT
                          CTACACTTGGCAGGGGCTTCCTCTATAATGGGAGCTATAAATTTCATTACTACCATTTTGAATATGAGAGCTCCTGGCATGAGCATGGATAAAATACCAC
                          TGTTCGTTTGATCTGTTTTGGTCACCGCAGTTCTGCTATTACTCTCCTTGCCTGTGCTAGCTGGTGCTATAACTATGCTTTTAACAGACAGGAATTTTAA
                          CACATCCTTCTTTGATCCTGCCGGGGGAGGAGATCCTGTCTTGTTCCAGCACCTGTTC---------
sp16_KY564348.1           ---ACTATACATAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATAATCAGATTAGAATTATCTGGACCAGGCTCAATGCTTGGT
                          GACGATCAACTCTACAATGTTGTAGTAACTGCTCACGCTTTTGTAATGATTTTCTTTTTTGTAATGCCTGTTTTAATAGGTGGCTTTGGGAACTGGTTTG
                          TACCTCTGTACATAGGCGCTCCTGATATGGCTTTCCCAAGGTTGAATAATATTAGTTTCTGACTACTACCGCCTGCTCTCTTGCTATTATTAGGATCTTC
                          CCTAATAGAACAGGGGGCAGGAACTGGTTGAACCGTCTACCCCCCTTTAAGTTCGATACAAGCTCATTCCGGAGGTTCTGTTGACATGGCCATATTTAGT
                          CTACATTTAGCAGGGGCTTCTTCCATAATGGGAGCCATAAACTTTATTACTACTATATTAAACATGAGAGCTCCTGGGATGAGCATGGATAAAATACCAT
                          TATTTGTATGATCTGTTCTAGTAACTGCGGTTTTATTATTACTTTCTTTACCCGTATTAGCCGGTGCTATAACCATGCTTTTAACGGATAGAAATTTCAA
                          TACATCCTTTTTCGATCCCGCCGGAGGAGGAGACCCTATTTTGTTTCAACATTTGTTC---------
sp16_KY564349.1           ---ACTATACATAATATTCGGTGCTTTCTCTGCTATGGTAGGAACTGCCTTTAGTATGATAATCAGATTAGAATTATCTGGACCAGGCTCAATGCTTGGT
                          GACGATCAACTCTACAATGTTGTAGTAACTGCTCACGCTTTTGTAATGATTTTCTTTTTTGTAATGCCTGTTTTAATAGGTGGCTTTGGGAACTGGTTTG
                          TACCTCTGTACATAGGCGCTCCTGATATGGCTTTCCCAAGGTTGAATAATATTAGTTTCTGACTACTACCGCCTGCTCTCTTGCTATTATTAGGATCTTC
                          CCTAATAGAACAGGGGGCAGGAACTGGTTGAACCGTCTACCCCCCTTTAAGTTCGATACAAGCTCATTCCGGAGGTTCTGTTGACATGGCCATATTTAGT
                          CTACATTTAGCAGGGGCTTCTTCCATAATGGGAGCCATAAACTTTATTACTACTATATTAAACATGAGAGCTCCTGGGATGAGCATGGATAAAATACCAT
                          TATTTGTATGATCTGTTCTAGTAACTGCGGTTTTATTATTACTTTCTTTACCCGTATTAGCCGGTGCTATAACCATGCTTTTAACGGATAGAAATTTCAA
                          TACGTCCTTTTTCGATCCCGCCGGAGGAGGAGACCCTATTTTGTTTCAACATTTGTTC---------
Stomolophus_KY611123.1    ------------------------------------------AGGGACAGCTTTTAGTATGATTATTAGACTAGAATTGTCTGGTCCAGGATCAATGCTAGGA
                          GACGACCAACTATACAATGTTGTAGTAACAGCTCATGCTTTAATAATGATATTCTTTTTTGTAATGCCTGTTTTAATAGGAGGATTTGGAAATTGGTTTG
                          TCCCTTTATATATAGGGGCTCCCGATATGGCTTTCCCTAGATTAAACAATATAAGTTTTTGATTACTACCTCCTGCTTTATTATTATTATTGGGATCATC
                          TTTAATAGAACAAGGAGCAGGAACCGGTTGAACCATTTATCCTCCCTTAAGTTCAATACAGGCCCATTCAGGAGGATCTGTAGATATGGCTATATTCAGC
                          TTACATTTAGCTGGAGCTTCTTCCATTATGGGAGCAATTAATTTTATTACCACAATTTTAAATATGAGAGCTCCTGGAATGACAATGGATAGAATTCCTC
                          TATTTGTATGATCTGTATTAGTTACTGCAATATTGTTATTATTGTCATTACCAGTTTTAGCAGGGGCTATAACTATGTTGTTAACAGATAGAAATTTCAA
                          TACTTCATTTTTCGATCCTGCTGGTGGAGGAGATCCTATTTGTTTCAACATTTATTC---------
```

```
Stomolophus_KY611117.1    ------------------------------------------AGGGACAGCTTTTAGTATGATTATTAGACTAGAATTGTCTGGTCCAGGATCAATGCTAGGA
                          GACGACCAACTATACAATGTTGTAGTAACAGCTCATGCTTTGATAATGATATTCTTTTTTGTAATGCCTGTTTTAATAGGAGGATTCGGAAATTGGTTTG
                          TCCCCTTATATATAGGAGCTCCCGATATGGCTTTCCCTAGATTAAACAATATAAGTTTTTGATTACTACCTCCTGCTTTATTATTATTATTGGGATCATC
                          TTTAATAGAACAAGGAGCAGGAACCGGTTGGACCATTTATCCTCCCTTAAGTTCAATACAAGCCCATTCAGGAGGATCTGTAGATATGGCTATATTCAGC
                          TTACATTTAGCTGGAGCTTCTTCCATTATGGGAGCAATTAATTTTATTACCACAATTTTAAATATGAGAGCTCCTGGAATGACAATGGATAGAATTCCTC
                          TATTTGTATGATCTGTATTAGTTACTGCAATATTGTTATTATTGTCATTACCAGTTTTAGCGGGGGCTATAACTATGTTGTTAACAGATAGGAATTTCAA
                          TACTTCATTCTTCGACCCTGCTGGTGGAGGAGATCCTATTTTGTTTCAACATTTATTC---------
Stomolophus_KY611115.1    ------------------------------------------AGGGACAGCTTTTAGTATGATTATTAGACTAGAATTGTCTGGTCCAGGATCAATGCTAGGA
                          GACGACCAACTATACAATGTTGTAGTAACAGCTCATGCTTTAATAATGATATTCTTTTTTGTAATGCCTGTTTTAATAGGGGGATTCGGAAATTGGTTTG
                          TCCCCTTATACATAGGAGCTCCCGATATGGCTTTCCCTAGATTAAACAATATAAGTTTTTGATTACTACCTCCTGCTTTATTATTATTATTGGGATCATC
                          TTTAATAGAACAAGGAGCAGGAACCGGTTGGACCATTTATCCTCCCTTAAGTTCAATACAAGCCCATTCAGGAGGATCTGTAGATATGGCTATATTCAGC
                          TTACATTTAGCTGGAGCTTCTTCCATTATGGGAGCAATTAATTTTATTACCACAATTTTAAATATGAGAGCTCCTGGAATGACAATGGATAGAATTCCTC
                          TATTTGTATGATCTGTATTAGTTACTGCAATATTGTTATTATTGTCACTACCAGTTTTAGCGGGGGCTATAACTATGTTGTTAACAGATAGGAATTTCAA
                          TACTTCATTCTTCGACCCTGCTGGTGGAGGAGATCCTATTTTGTTTCAACATTTATTC---------
Stomolophus_KY611119.1    ------------------------------------------AGGGACAGCTTTTAGTATGATTATTAGACTAGAATTGTCTGGTCCAGGATCAATGCTAGGA
                          GACGACCAACTATACAATGTTGTAGTAACAGCTCATGCTTTAATAATGATATTCTTTTTTGTAATGCCTGTTTTAATAGGAGGATTCGGAAATTGGTTTG
                          TCCCCTTATACATAGGAGCTCCCGATATGGCTTTCCCTAGATTAAACAATATAAGTTTTTGATTACTACCTCCTGCTTTATTATTATTATTGGGATCATC
                          TTTAATAGAACAAGGAGCAGGAACCGGTTGGACCATTTATCCTCCCTTAAGTTCAATACAAGCCCATTCAGGAGGATCTGTAGATATGGCTATATTCAGC
                          TTACATTTAGCTGGAGCTTCTTCCATTATGGGAGCAATTAATTTTATTACCACAATTTTAAATATGAGAGCTCCTGGAATGACAATGGATAGAATTCCTC
                          TATTTGTATGATCTGTATTAGTTACTGCAATATTGTTATTATTATCATTACCAGTTTTAGCGGGGGCTATAACTATGTTGTTAACAGATAGGAATTTCAA
                          TACTTCATTCTTCGACCCTGCTGGTGGAGGAGATCCTATTTTGTTTCAACATTTATTC---------
Stomolophus_KY611114.1    ------------------------------------------AGGGACAGCTTTTAGTATGATTATTAGACTAGAATTGTCTGGTCCAGGATCAATGCTAGGA
                          GACGACCAACTATACAATGTTGTAGTAACAGCTCATGCTTTAATAATGATATTCTTTTTTGTAATGCCTGTTTTAATAGGAGGATTCGGAAATTGGTTTG
                          TCCCCTTATACATAGGAGCTCCCGATATGGCTTTCCCTAGATTAAACAATATAAGTTTTTGATTACTACCTCCTGCTTTATTATTATTATTGGGATCATC
                          TTTAATAGAACAAGGAGCAGGAACCGGTTGGACCATTTATCCTCCTTTAAGTTCAATACAAGCCCATTCAGGAGGATCTGTAGATATGGCTATATTCAGC
                          TTACATTTAGCTGGAGCTTCTTCCATTATGGGAGCAATTAATTTTATTACCACAATTTTAAATATGAGAGCTCCTGGAATGACAATGGATAGAATTCCTC
                          TATTTGTATGATCTGTATTAGTTACTGCAATATTGTTATTATTGTCATTACCAGTTTTAGCGGGGGCTATAACTATGTTGTTAACAGATAGGAATTTCAA
                          TACTTCATTCTTCGACCCTGCTGGTGGAGGAGATCCTATTTTGTTTCAACATTTATTC---------
```

```
Stomolophus_KY611113.1  ----------------------------------------AGGGACAGCTTTTAGTATGATTATTAGACTAGAATTGTCTGGTCCAGGATCAATGCTAGGA
                        GACGACCAACTATACAATGTTGTAGTAACAGCTCATGCTTTAATAATGATATTCTTTTTTGTAATGCCTGTTTTAATAGGAGGATTCGGAAATTGGTTTG
                        TCCCCTTATACATAGGAGCTCCCGATATGGCTTTCCCTAGATTAAACAATATAAGTTTTTGATTACTACCTCCTGCTTTATTATTATTATTGGGATCATC
                        TTTAATAGAACAAGGAGCAGGAACCGGTTGGACCATTTATCCTCCCTTAAGTTCAATACAAGCCCATTCAGGAGGATCTGTAGATATGGCTATATTCAGC
                        TTACATTTAGCTGGAGCTTCTTCCATTATGGGAGCAATTAATTTTATTACCACAATTTTAAATATGAGAGCTCCTGGAATGACAATGGATAGAATTCCTC
                        TATTTGTATGATCTGTATTAGTTACTGCAATATTGTTATTATTGTCATTACCAGTTTTAGCGGGGGCTATAACTATGTTGTTAACAGATAGGAATTTCAA
                        TACTTCATTCTTCGACCCTGCTGGTGGAGGAGATCCTATTTTGTTTCAACATTTATTC--------
Stomolophus_KY611121.1  ----------------------------------------AGGGACAGCTTTTAGTATGATTATTAGACTAGAATTGTCTGGTCCAGGATCAATGCTAGGA
                        GACGACCAACTATACAATGTTGTAGTAACAGCTCATGCTTTAATAATGATATTCTTTTTTGTAATGCCTGTTTTAATAGGAGGGTTCGGAAATTGGTTTG
                        TCCCCTTATACATAGGAGCTCCCGATATGGCTTTCCCTAGATTAAACAATATAAGTTTTTGATTACTACCTCCTGCTTTATTATTATTATTGGGATCATC
                        TTTAATAGAACAAGGAGCAGGAACCGGTTGGACCATCTATCCTCCCTTAAGTTCAATACAGGCCCATTCAGGAGGATCTGTAGATATGGCTATATTCAGC
                        TTACATTTAGCTGGAGCCTCTTCCATTATGGGAGCAATTAATTTTATTACCACAATTTTAAATATGAGAGCTCCTGGAATGACAATGGATAGAATTCCTT
                        TATTTGTATGATCTGTATTAGTTACTGCAATATTGTTATTATTGTCATTACCAGTTTTAGCGGGGGCTATAACTATGTTGTTAACAGATAGAAATTTCAA
                        TACTTCATTCTTCGACCCTGCTGGTGGAGGAGATCCTATTTTGTTTCAACATTTATTC--------
Stomolophus_KY611122.1  ----------------------------------------AGGGACAGCTTTTAGTATGATTATTAGACTAGAATTGTCTGGTCCAGGATCAATGCTAGGA
                        GACGACCAACTATACAATGTTGTAGTAACAGCTCATGCTTTAATAATGATATTCTTTTTTGTAATGCCTGTTTTAATAGGGGGGTTCGGAAATTGGTTTG
                        TCCCCTTATACATAGGAGCTCCCGATATGGCTTTCCCTAGATTAAACAATATAAGTTTTTGATTACTACCTCCTGCTTTATTATTATTATTGGGATCATC
                        TTTAATAGAACAAGGAGCAGGAACCGGTTGGACCATTTATCCTCCCTTAAGTTCAATACAGGCCCATTCAGGAGGATCTGTAGATATGGCTATATTCAGC
                        TTACATTTAGCTGGAGCCTCTTCCATTATGGGAGCAATTAATTTTATTACCACAATTTTAAACATGAGAGCTCCTGGAATGACAATGGATAGAATTCCTT
                        TATTTGTATGATCTGTATTAGTTACTGCAATATTGTTATTATTGTCATTACCAGTTTTAGCGGGGGCTATAACTATGTTGTTAACAGATAGAAATTTCAA
                        TACTTCATTCTTCGACCCTGCTGGTGGAGGAGATCCTATTTTGTTTCAACATTTATTC--------
```

Figure S1. Principal component analysis of frequencies of TEs in *Aurelia*.

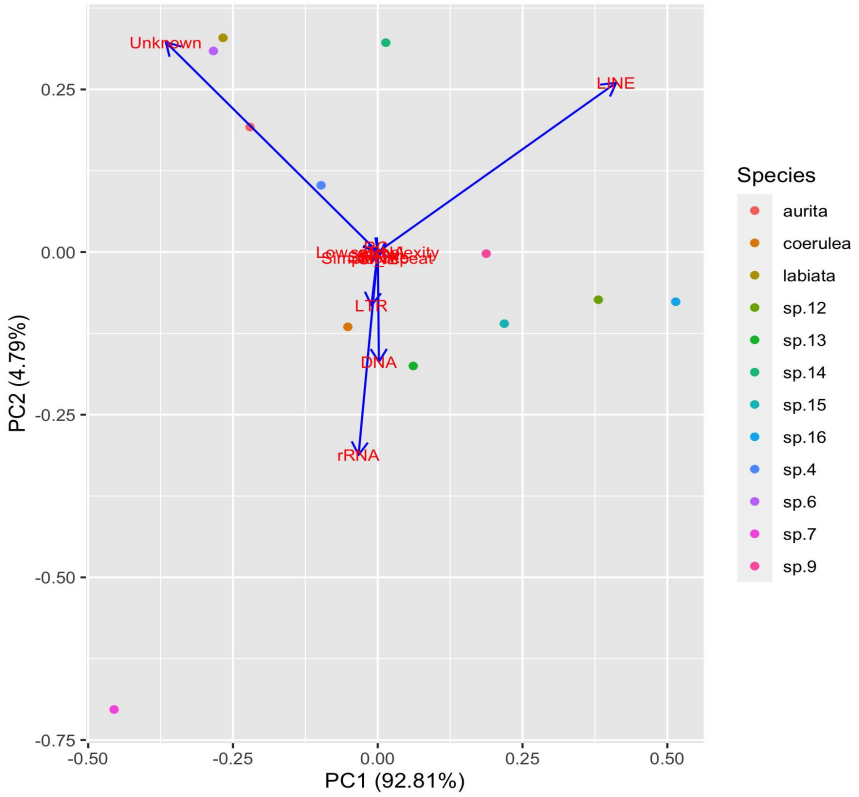


Figure S2. Piwi protein expressions are high in the early stages of *Aurelia coerulea* life history, and decreases towards the sexually reproducing stage, medusa.

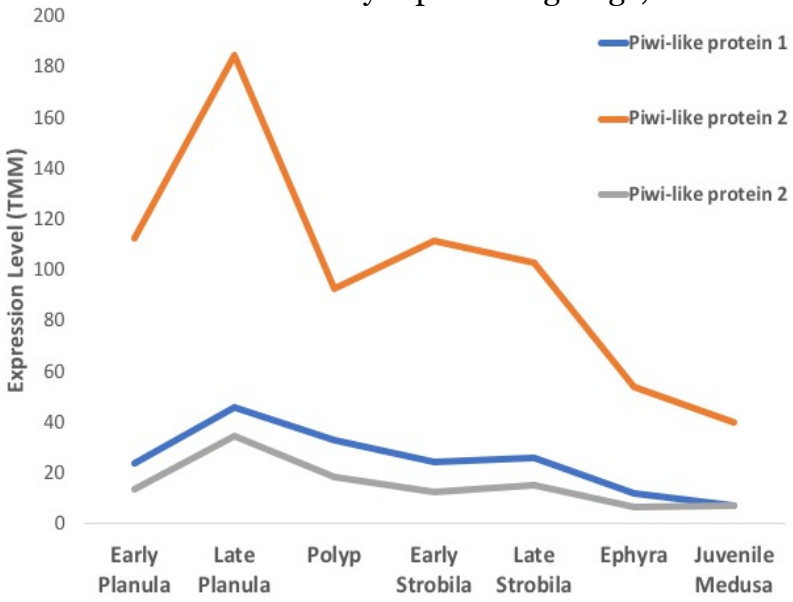
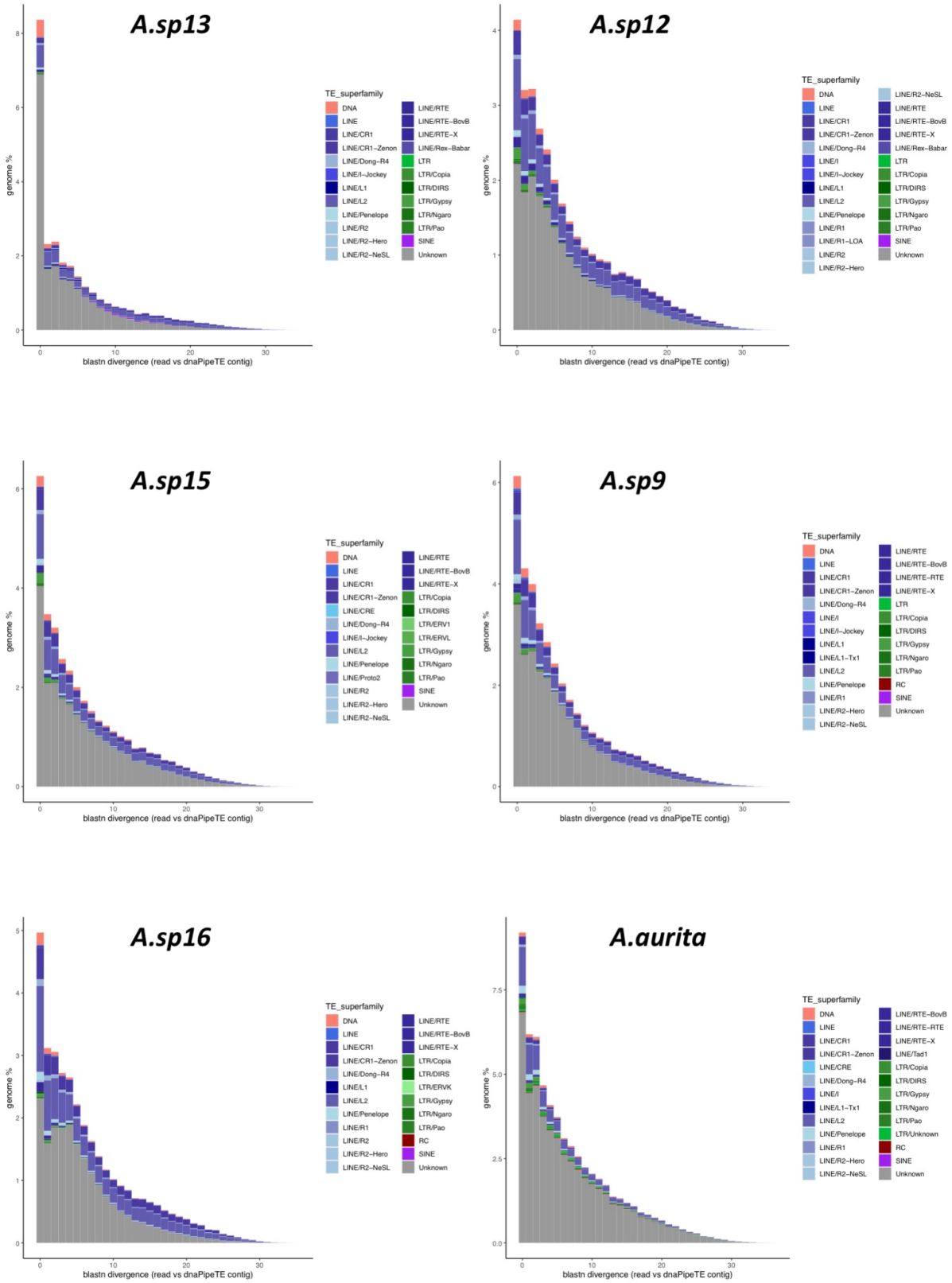


Figure S3. Distribution of TE ages with 'unknown' sequences.



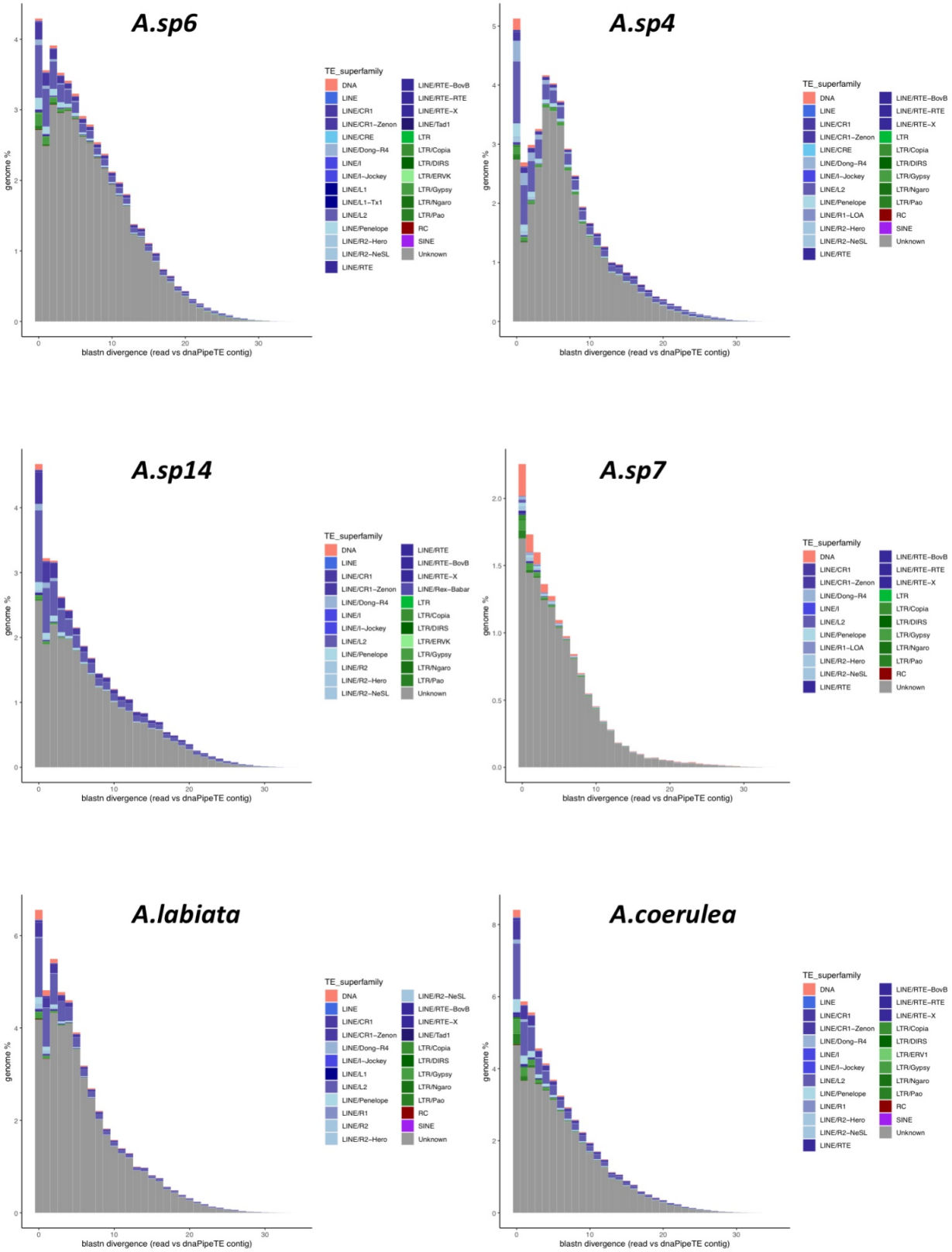

A.sp6
A.sp4
A.sp14
A.sp7
A.labiata
A.coerulea
genome %
blastn divergence (read vs dnaPipeTE contig)
TE_superfamily
DNA
LINE
LINE/CR1
LINE/CR1-Zenon
LINE/CRE
LINE/Dong-R4
LINE/I
LINE/I-Jockey
LINE/L1
LINE/L1-Tx1
LINE/L2
LINE/Penelope
LINE/R1
LINE/R1-LOA
LINE/R2
LINE/R2-Hero
LINE/R2-NeSL
LINE/RTE
LINE/RTE-BovB
LINE/RTE-RTE
LINE/RTE-X
LINE/Rex-Babar
LINE/Tad1
LTR
LTR/Copia
LTR/DIRS
LTR/ERV1
LTR/ERVK
LTR/Gypsy
LTR/Ngaro
LTR/Pao
RC
SINE
Unknown

Table S1. Types of TEs and relative abundance in each species in percentages.

	DNA	LINE	LTR	Low complexity	SINE	Satellite	Simple repeat	rRNA	tRNA	RC	snRNA	Unknown
aurita	1.27	15.1	3.55	0	0.01	0	0.07	0	0.07	0.69	0	79.14
labiata	1.48	14.73	0.76	0	0.03	0.03	0.02	0.73	0.06	0.06	0	82.08
coerulea	1.63	19.07	7.32	0	0.18	0	0.04	0	0.05	0.12	0	71.45
sp.4	1.26	19.5	2.07	0	0.01	0	0	1.71	0.2	0.01	0	75.23
sp.6	0.96	13.93	1.74	0	0.02	0.07	0	0.73	0.08	0.12	0.01	82.33
sp.14	0.87	24.68	0.81	0	0	0.01	0	0.35	0.05	0	0	73.24
sp.15	2.85	29.55	3.05	0.01	0.03	0.01	0.07	1.38	0.01	0	0	63.97
sp.16	1.75	40.35	1.11	0.01	0.03	0.03	0.03	1.6	0.06	0.03	0	55.02
sp.12	2.5	35.37	2.12	0.01	0.06	0.02	0.05	0.83	0	0	0	59
sp.13	3.44	23.17	0.61	0	0.91	0.92	0.1	1.77	1.03	0	0	68.06
sp.7	4.3	3.01	2.67	0	0	0	0.21	8.07	0.05	0.06	0.01	81.61
sp.9	3.04	28.78	1.73	0	0.06	0.17	0.08	0.46	0.01	0.04	0.12	65.45