

UCLA

UCLA Previously Published Works

Title

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

Permalink

<https://escholarship.org/uc/item/5cw931qw>

Journal

Genome Biology and Evolution, 14(2)

ISSN

1759-6653

Authors

Zhang, Xinhui

Jacobs, David

Publication Date

2022-02-04

DOI



10.1093/gbe/evab284

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

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_{over}, 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.

© The Author(s) 2022. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

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 DNMT1 and 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 ($CpG_{obs/e}$) 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

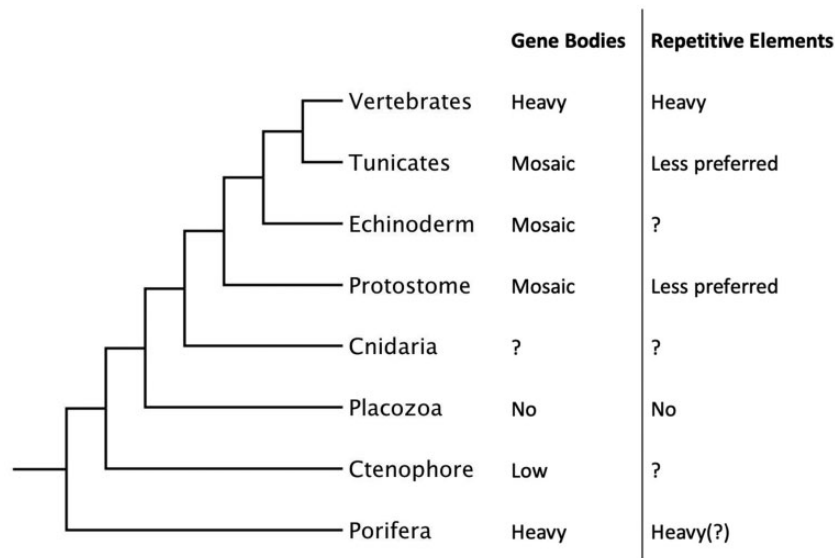


FIG. 1.—Methylation on the gene bodies and repetitive elements vary greatly across the tree of animals (Gavery and Roberts 2010; Hunt et al. 2010; Zemach et al. 2010; Glastad et al. 2011; Nanty et al. 2011; Sarda et al. 2012; Falckenhayn et al. 2013; Dabe et al. 2015).

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 1Average 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	<i>Actinia tenebrosa</i>	0.69	0.90	
		<i>Aiptasia pallida</i>	0.69	0.98	
		<i>Anthopleura elegantissima</i>	0.68		
	Alcyonacea ①	<i>Edwardsiella lineata</i>	0.758		
		<i>Nematostella vectensis</i>	0.77		
		<i>Acanthogorgia aspera</i>	0.83		
		<i>Briareum asbestinum</i>	0.79		
		<i>Clavularia</i> sp.	0.82		
		<i>Corallium rubrum</i>	0.84		
		<i>Dendronephthya gigantea</i>	0.84	1.03	
		<i>Eleutherobia rubra</i>	0.88		
		<i>Eunicella cavolinii</i>	0.85		
		<i>Eunicella verrucosa</i> ^a	0.84		
		<i>Gorgonia ventalina</i>	0.75		
		<i>Leptogorgia sarmentosa</i>	0.85		
		<i>Xenia</i> sp.	0.79		
		Corallimorpharia ④	<i>Corynactis australis</i>	0.81	
			<i>Rhodactis indosinensis</i>	0.71	
	<i>Ricordea yuma</i>		0.72		
	Helioporacea ②	<i>Heliopora coerulea</i>	0.66		
	Scleractinia ⑤	<i>Acropora digitefera</i>	0.70	0.90	
		<i>Acropora millepora</i>	0.72	0.91	
		<i>Ctenactis echinata</i>	0.64		
		<i>Favia lizardensis</i>	0.66		
		<i>Lobactis scutaria</i>	0.67		
		<i>Madracis auretenra</i>	0.75		
		<i>Montastraea cavernosa</i>	0.68		
		<i>Orbicella faveolata</i>	0.73	0.93	
		<i>Platygyra carnosus</i>	0.72		
		<i>Pocillopora damicornis</i>	0.70	0.87	
		<i>Seriatopora hystrix</i>	0.68		
		<i>Stylophora pistillata</i>	0.70	0.92	
	Hydrozoa	Aplanulata ⑩	<i>Ectopleura larynx</i>	0.97	
<i>Hydra oligactis</i>			0.75		
<i>Hydra viridissima</i>			0.64		
<i>Hydra vulgaris</i>			0.63	1.03	
Capitata ⑳		<i>Porpita porpita</i>	0.67		
		<i>Veleva veleva</i>	0.80		
Filifera III ⑰		<i>Hydractinia polyclina</i>	0.82		
		<i>Hydractinia symbiolongicarpus</i>	0.81		
Filifera IV ㉑		<i>Podocoryna carnea</i>	0.63		
		<i>Turritopsis</i> sp.	0.83		
Leptothecata ㉒		<i>Clytia hemisphaerica</i>	0.81		
		<i>Dynamena pumila</i>	0.83		
Limnomedusae ⑭		<i>Craspedacusta sowerbyi</i>	0.90		
Narcomedusae ⑮		<i>Aegina citrea</i> ^b	0.85		
Siphonophorae ⑧		<i>Abylopsis tetragona</i>	0.72		
		<i>Agalma elegans</i>	0.76		
	<i>Craseoa lathetica</i>	0.74			
	<i>Nanomia bijuga</i>	0.83			
	<i>Physalia physalis</i>	0.76			

(continued)

Table 1 Continued

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

NOTE.—Circled numbers correspond to figures 2 and 4.

^aSpecies in which DNMTs are absent.

^bSpecies in which MBDs are absent.

^cSpecies 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_{ov/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_{ov/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_{ov/e} (supplementary table S2, Supplementary Material online) with *Calvadosia* and *Clytia* being the two exceptions, both of which have overall average gene body CpG_{ov/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_{ov/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_{ov/e} than nondifferentially expressed genes ($P < 2.2 \times 10^{-16}$; Welch two sample *t*-test).

CpG_{ov/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_{ov/e} negatively correlate with repeat content (fig. 5A 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. 5C) ($t = 5.0099$, $P = 0.0003963$, 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.

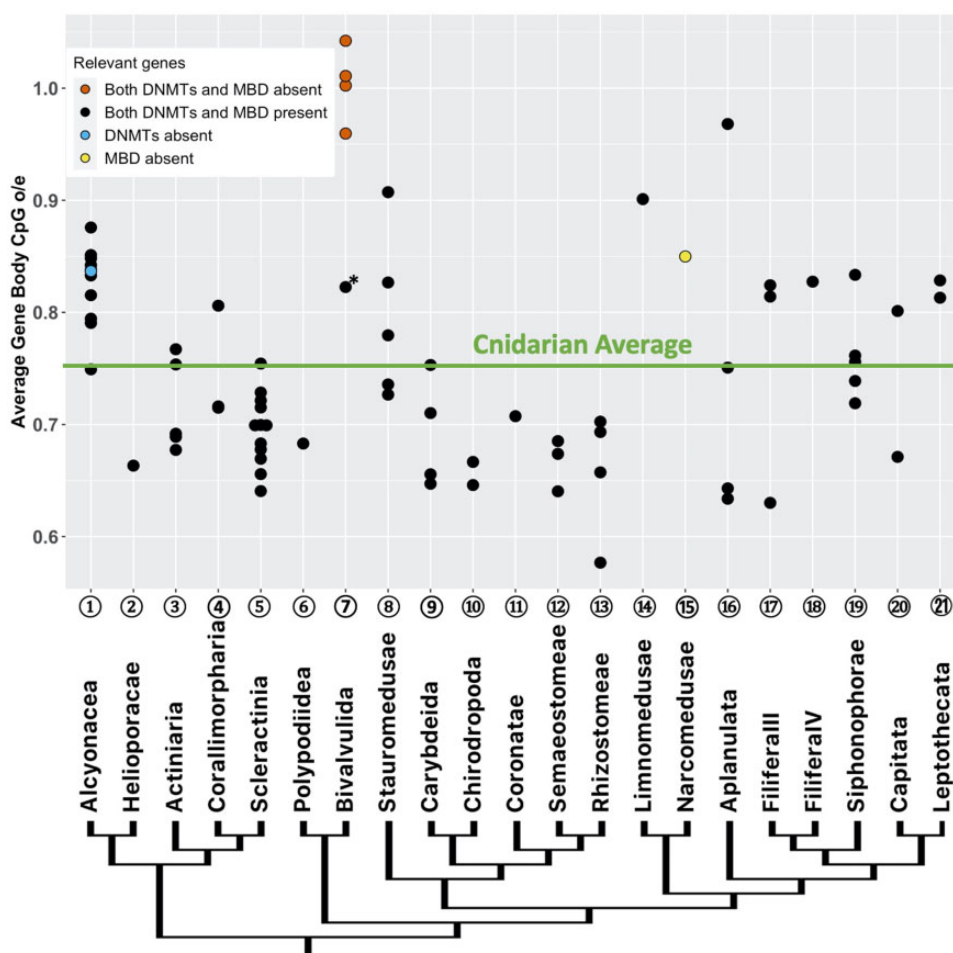


FIG. 2.—Average gene body CpG_{o/e} of each cnidarian species organized in ordinal rank groups. Green horizontal line indicates the average among all species. Phylogenetic relationship between the major cnidarian groups rendered from Kayal et al. (2018) and Bentlage and Collins (2021). Circled numbers 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.

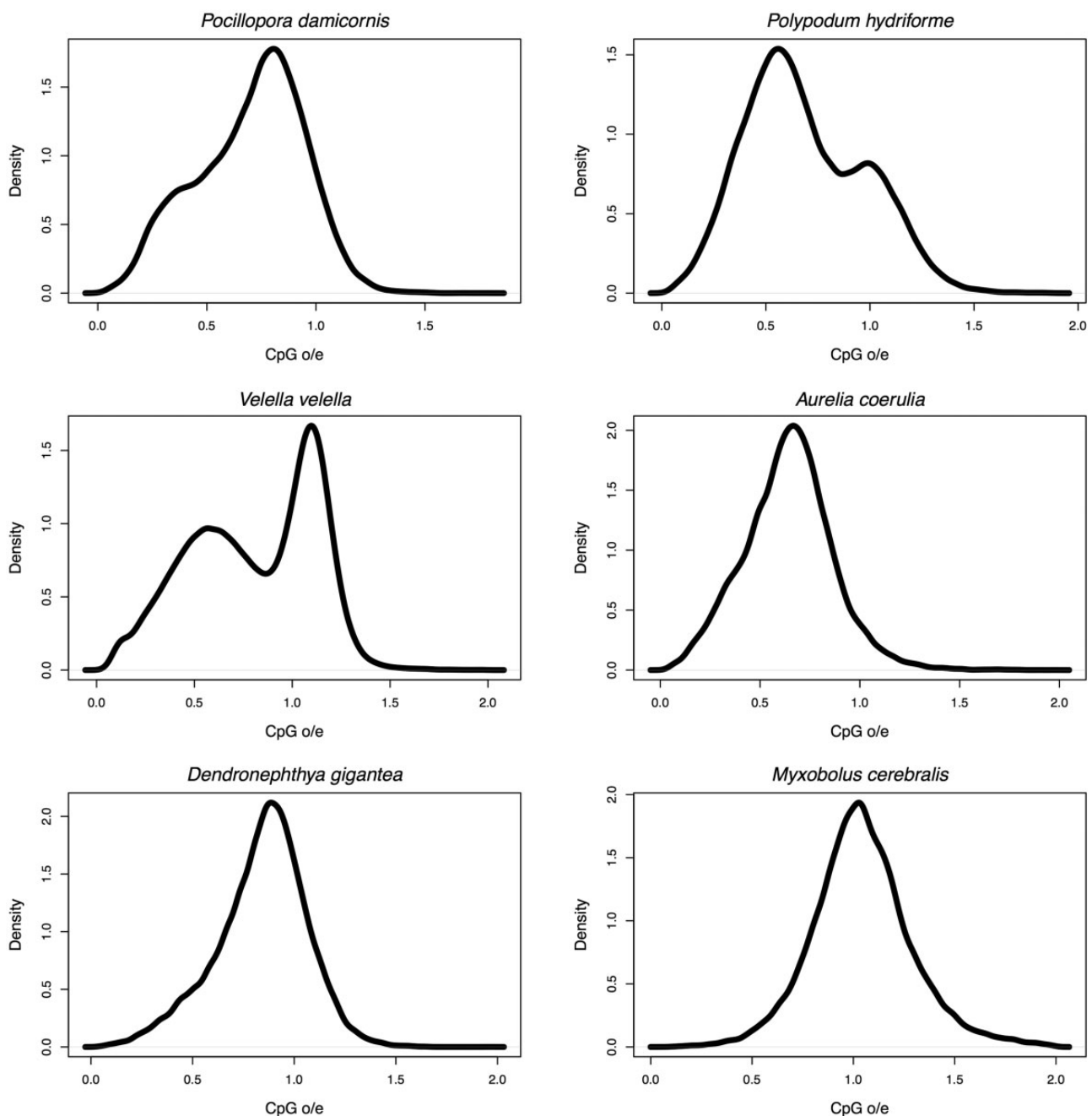


Fig. 3.—Density plots of gene body CpG_{o/e} of selected species as examples. The discrete bimodality typically found in bilaterian invertebrates occurs only in 11 of the 41 cnidarian species where CpG_{o/e} levels indicate substantial methylation. Among the six examples shown, *Polypodium* and *Velella* 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”; *Polypodium* 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.

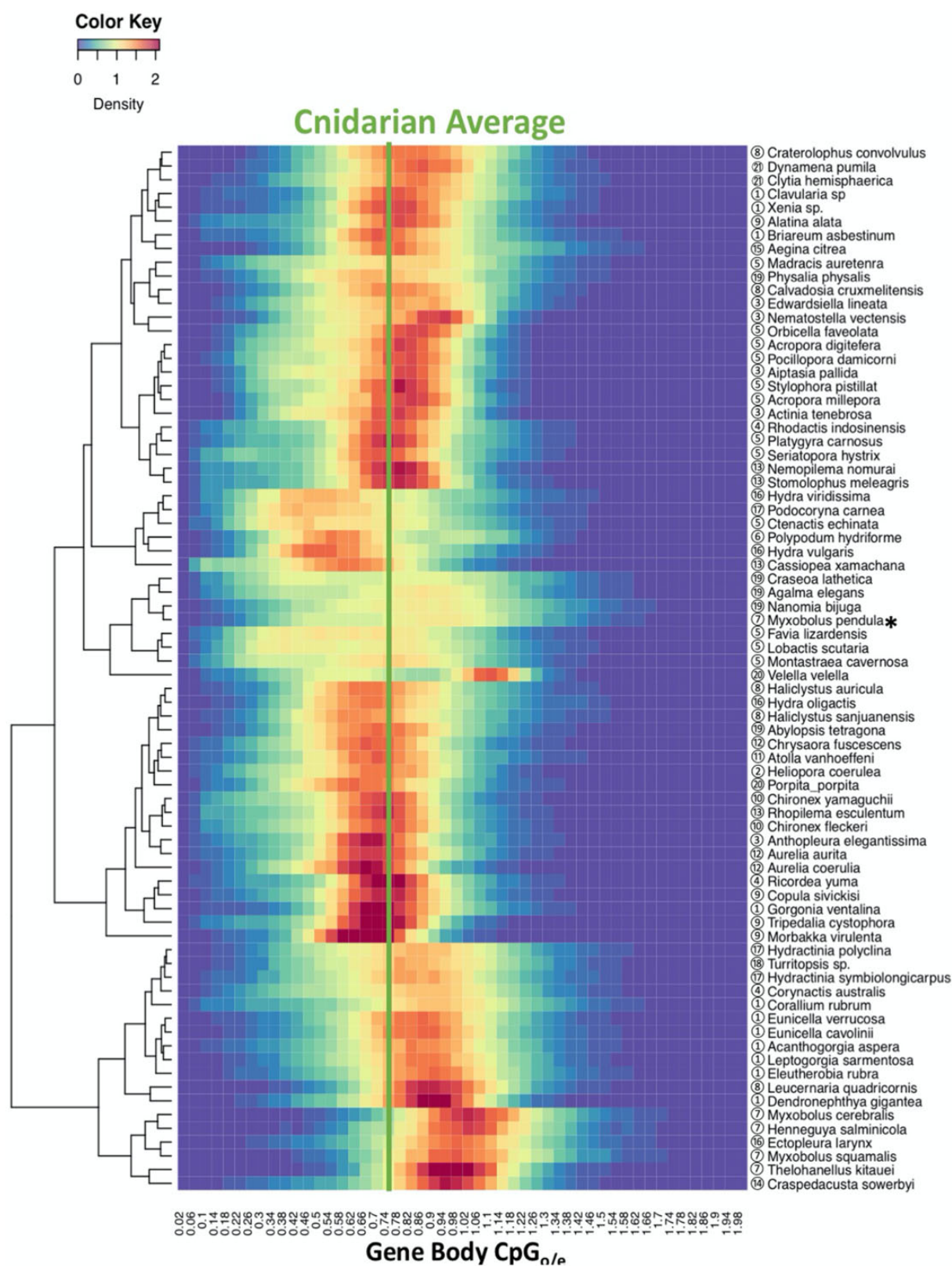


Fig. 4.—Heatmap of gene body CpG_{0.1e} distribution patterns in Cnidaria. Circled numbers next to the species names correspond to figure 2 and table 1. The species marked by the asterisk is *Myxobolus pendula*, where the distribution pattern might have been affected by possible contamination from the 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
<i>Acropora digitifera</i>	0.7093909	0.6666281	15.794	2.2e-16
<i>Aiptasia pallida</i>	0.7152435	0.6313672	27.393	2.2e-16
<i>Alatina alata</i>	0.7597698	0.7381270	11.037	2.2e-16
<i>Aurelia coerulea</i>	0.6475442	0.6267479	6.8606	7.135e-12
<i>Calvadosia cruxmelitensis</i>	0.7769339	0.8065653	-15.144	2.2e-16
<i>Clytia hemisphaerica</i>	0.8164449	0.8091184	3.2878	0.001012
<i>Hydra vulgaris</i>	0.6437276	0.6156787	8.1367	4.315e-16
<i>Morbakka virulenta</i>	0.6508604	0.6392557	5.6923	1.271e-08
<i>Nematostella vectensis</i>	0.7860412	0.7203910	25.841	2.2e-16
<i>Physalia physalis</i>	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, *Polypodium 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

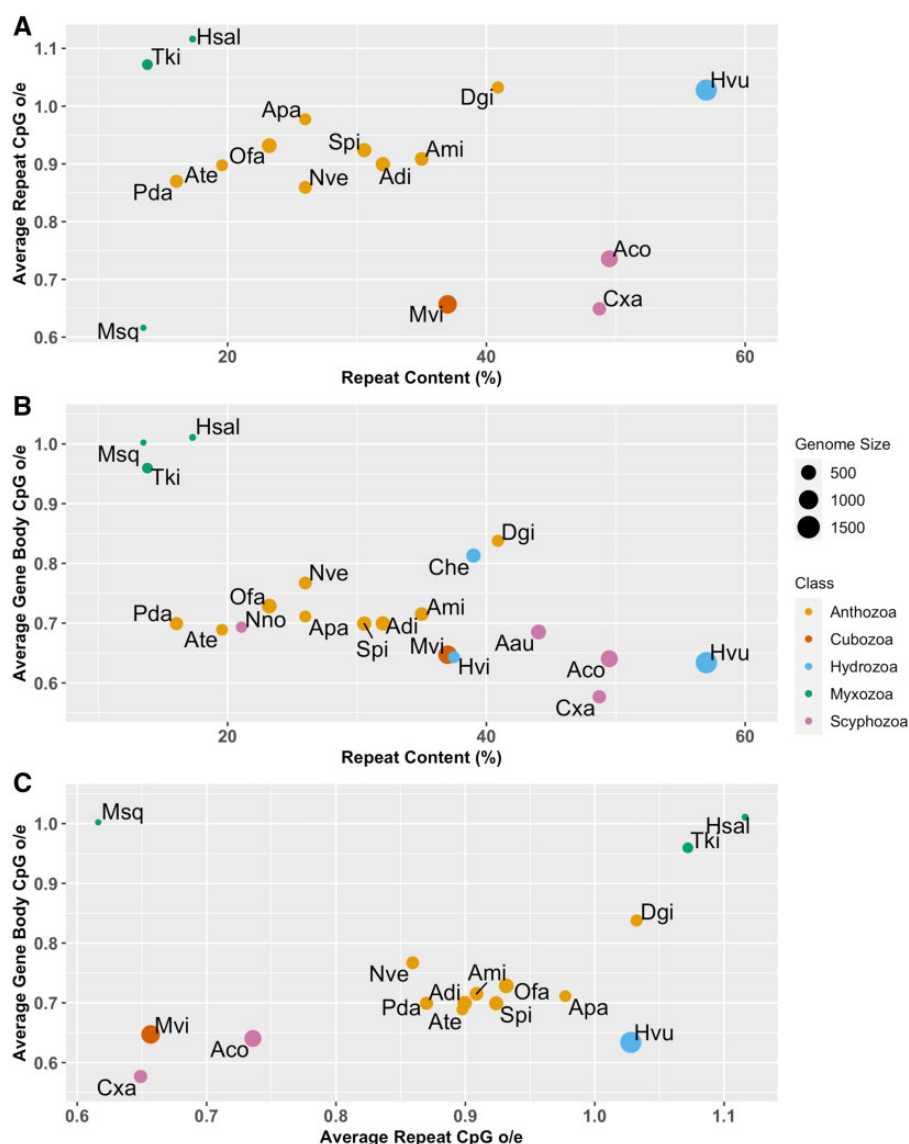


Fig. 5.—Both repeat and gene body CpG_{o/e} correlate with repeat content. (A) The average CpG_{o/e} of repeats is negatively correlated with the repeat content of the genome, with two notable exceptions (*Hydra vulgaris* and *Myxobolus squamalis*). (B) The average gene body CpG_{o/e} is negatively correlated with repeat content. (C) Gene body CpG_{o/e} and repeat CpG_{o/e} are positively correlated 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 coerulea*; Adi, *Acropora digitefera*; Ami, *Acropora millepora*; Apa, *Aiptasia pallida*; Ate, *Actinia tenebrosa*; Che, *Clytia hemisphaera*; 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 telost 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 coerulea*

(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_{0/e} of ten-way orthologs was compared with genes that are not shared by all ten species. For each pair of species, CpG_{0/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.
- Muscante 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