Published in final edited form as:
Nat Ecol Evol. 2019 October 01; 3(10): 1464–1473. doi:10.1038/s41559-019-0983-2.

Convergent evolution of a vertebrate-like methylome in a marine sponge

Alex de Mendoza[1,2,*], **William L Hatleberg**[3], **Kevin Pang**[4], **Sven Leininger**[4], **Ozren Bogdanovic**[5,6], **Jahnvi Pflueger**[1,2], **Sam Buckberry**[1,2], **Ulrich Technau**[7], **Andreas Hejnol**[4], **Maja Adamska**[4,8], **Bernard M Degnan**[3], **Sandie M Degnan**[3], **Ryan Lister**[1,2,*]

[1]Australian Research Council Centre of Excellence in Plant Energy Biology, School of Molecular Sciences, The University of Western Australia, Perth, Western Australia, Australia

[2]Harry Perkins Institute of Medical Research, Perth, Western Australia, Australia

[3]School of Biological Sciences, The University of Queensland, Brisbane, Queensland, Australia

[4]Sars International Centre for Marine Molecular Biology, University of Bergen, Bergen, Norway

[5]Genomics and Epigenetics Division, Garvan Institute of Medical Research, Sydney, New South Wales, Australia

[6]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales, Australia

[7]Department for Molecular Evolution and Development, Centre of Organismal Systems Biology, University of Vienna, Vienna, Austria

[8]ARC Centre for Excellence for Coral Reef Studies, Research School of Biology, Australian National University, Canberra, Australian Capital Territory, Australia

Abstract

Vertebrates have highly methylated genomes at CpG positions whereas invertebrates have sparsely methylated genomes. This increase in methylation content is considered a major regulatory innovation of vertebrate genomes. However, here we report that a marine sponge, proposed as the sister group to the rest of animals, has a highly methylated genome. Despite major differences in genome size and architecture, we find similarities between the independent acquisitions of the hypermethylated state. Both lineages show genome wide CpG depletion, conserved strong transcription factor methyl-sensitivity, and developmental methylation dynamics at 5-hydroxymethylcytosine enriched regions. Together, our findings trace back patterns associated

*Correspondence to Alex de Mendoza (alex.demendoza@uwa.edu.au) or Ryan Lister (ryan.lister@uwa.edu.au).

Author contributions

A.d.M. and R.L. designed the study. A.d.M. prepared MethylC-seq, TAB-seq and DAP-seq libraries, with the help of O.B. and J.P. The data were analysed by A.d.M., with help from S.B. *Amphimedon* materials were provided by S.M.D., B.M.D. and W.L.H. *Mnemiopsis* materials were provided by K.P. and A.H. *Sycon* material was provided by S.L. and M.A. *Nematostella* material was provided by U.T. The manuscript was written by A.d.M. and R.L. All authors commented on the final manuscript.

Competing interests

The authors declare no competing interests.

Europe PMC Funders Author Manuscripts

with DNA methylation in vertebrates to the early steps of animal evolution. Thus, the sponge methylome challenges prior hypotheses concerning the uniqueness of vertebrate genome hypermethylation and its implications for regulatory complexity.

5-methylcytosine (mC) is a stable and heritable DNA base modification mostly deposited at CpG dinucleotides in eukaryotes. Despite being the most common base modification in animal genomes, the patterns of mC deposition are very distinct between animal lineages. Most invertebrate genomes are sparsely methylated, while others have completely lost DNA methylation[1,2]. Methylation in invertebrates is regarded as “mosaic”, as its location is mostly restricted to active gene bodies and silenced repetitive elements[3,4]. In contrast, vertebrates show widespread high methylation (hypermethylation), which is largely only absent at CpG island containing promoters and active distal regulatory elements[2,5]. Furthermore, demethylation in vertebrate genomes frequently occurs at distal regulatory elements upon transcription factor binding and enhancer activation[6,7]. It is not well understood how DNA methylation changed from a locally restricted mark to become the predominant default state of CpGs in vertebrate genomes[8]. Given that this transition apparently occurred only once during animal evolution, it has been difficult to determine to what extent hypermethylation acted as the driver of defining features of vertebrate genomes. Here, we sought to understand the evolution of DNA methylation systems, and their role in animal genomes, by profiling the methylomes of early animal lineages.

Results

The DNA methylation toolkit is conserved across the animal kingdom

We first surveyed the published genomes of animals and their unicellular relatives for genes that are essential for DNA methylation. The deposition of mC is dependent on cytosine DNA methyltransferases (DNMTs). DNMT1 is responsible for methylation maintenance, partnering with UHRF1 to target nascent hemimethylated CpGs at replication forks, while DNMT3 is generally associated with *de novo* methylation[9]. Conversely, active demethylation is regulated by Ten-eleven translocation (TET) enzymes, which oxidize mC into 5-hydroxymethylcytosine (hmC)[10]. Our survey shows that the genomes of non-bilaterian metazoans (except placozoans) encode the full repertoire of mC associated genes (Figure 1a, Supplementary Figure 1). Despite DNMT1, DNMT3, UHRF1 and TET predating animal origins[11], the specific protein domain configurations characteristic of each of these families originated at the root of animals (Supplementary Figure 2). These conserved domains include PWWP or ADD domains in DNMT3 and the Tandem Tudor Domain in UHRF1, capable of recognising histone tail modifications. Furthermore, the zinc finger CXXC domain, known to bind to unmethylated CpG-rich regions, is conserved in DNMT1 and TET orthologues, except in the *Mnemiopsis leidyi* TET orthologue. Therefore, the basic set of genes involved in mC deposition and removal is conserved from the origin of metazoans to vertebrates.

The sponge *Amphimedon* has a hypermethylated genome

To assess the diversity of methylation profiles across non-bilaterian lineages (excluding DNMT-lacking placozoans), we performed whole genome bisulfite sequencing (WGBS) on

whole-organism samples of four species: the demosponge *Amphimedon queenslandica* [12], the calcareous sponge *Sycon ciliatum* [13], the ctenophore *Mnemiopsis leidyi* [14] and the cnidarian *Nematostella vectensis* [15]. These data revealed that the global levels of mC vary dramatically, despite these species sharing the same set of genes involved in methylation deposition. Remarkably, *Amphimedon* shows 80% global methylation (fraction of mC basecalls at all covered CpGs) and most CpGs are highly methylated (mCG/CG ≥ 0.8); such a hypermethylated state has previously been found only in vertebrate genomes (Figure 1b, Supplementary Table 1). Furthermore, unlike most other invertebrates, in *Amphimedon* all gene bodies are methylated at levels that are independent from the abundance of their transcripts (Figure 1c, Supplementary Figure 3). Finally, *Amphimedon* exhibits sharp mC depletion at promoters, a feature that was previously exclusively attributed to vertebrates[2].

Most vertebrate promoters contain CpG islands, unmethylated regions that exhibit high CpG densities[5]. Despite having a compact genome and short intergenic regions[16] (Table 1), *Amphimedon* also shows CpG density enrichment around the Transcriptional Start Sites (TSS) of unmethylated promoters (Figure 1d). In fact, other invertebrates with genome methylation show a similar CpG enrichment at unmethylated promoters[17], while species lacking mC have lost CpG enrichment at TSS (Supplementary Figure 3). Thus, the tendency to accumulate CpGs at unmethylated promoters is widespread throughout animals, but only demosponges and vertebrates exhibit a low density of CpGs everywhere else in the genome (<2 CpG/100 bp, Figure 1d, Supplementary Figure 4), as already noted in the genome description of *Amphimedon* [12]. The widespread CpG dinucleotide depletion in vertebrate genomes has been explained by the propensity of methylated CpGs to undergo spontaneous deamination[2,8]. CpG depletion is more pronounced in *Amphimedon* than in zebrafish, whereas other invertebrates show much less pronounced biases, or none in the case of species lacking mC (Figure 1e). Interestingly, *Sycon* has considerably high methylation levels compared to other invertebrates, but does not show global CpG depletion, at odds with the mutability expectations. In contrast, the genome of the demosponge *Tethya wilhelma* shows a similar depletion to *Amphimedon*, suggesting that hypermethylation could be conserved across demosponges (Figure 1e, Supplementary Figure 4)[18]. In sum, genome hypermethylation has led to similar global depletion of CpG dinucleotides in vertebrates and *Amphimedon*, while unmethylated promoters are protected from mC mutability and accumulate higher CpG densities.

Unmethylated promoters are enriched for methyl-sensitive transcription factor binding sites in *Amphimedon*

However, accumulation of CpGs at promoters might not be associated only with protection from mutational biases[19], as it also influences binding affinities of regulatory proteins. Proteins encoding a zinc finger CXXC domain such as KDM2A/B/FBXL19 are known to protect CpG regions from mC deposition[2,20]. It is noteworthy that some of these zinc finger CXXC-containing chromatin modifiers are conserved throughout animal evolution (Supplementary Figure 1). Furthermore, some transcription factors have distinct binding affinities dependent on the methylation status of their binding sites[21,22]. Thus, unmethylated CpG-rich regions could encode regulatory information sensitive to methylation. To test this hypothesis, we identified the unmethylated regions (UMRs) of the genome in several species

(Figure 2a). *Amphimedon* UMRs are shorter than in other animals, and mostly overlap promoter regions (Figure 2a, Supplementary Figure 5). We then performed *de novo* motif identification for *Amphimedon* UMRs, revealing highly enriched motifs that match those of vertebrate transcription factors (5 out of 6 of the top motifs in *Amphimedon* promoters). These transcription factors include members of the Specificity protein (Sp) family, which are known to prevent methylation at CpG islands[23,24], and Nuclear Respiratory Factor (NRF), one of the best characterized examples of methyl-sensitive transcription factors in vertebrates[21].

As most motifs in *Amphimedon* UMRs contain at least one CpG (8 out of 9 in Figure 2b), its methylation status could affect the binding affinity of the transcription factor. To test this, we *in vitro* translated *Amphimedon* transcription factors and conducted affinity purification with natively methylated *Amphimedon* genomic DNA coupled to sequencing (DAP-seq)[25]. Motifs enriched in transcription factor DAP-seq peaks accurately matched those of their vertebrate orthologues (match score > 0.85, Figure 2c, Supplementary Figure 6). Furthermore, when performing affinity purification with PCR-amplified DNA (unmethylated, ampDAP-seq[25]), we found many peaks unique to the ampDAP-seq libraries, indicating distinct binding affinities that are dependent on methylation status. For transcription factors Ying Yang 1 (YY1), Early growth response protein (EGR) and GLI, the enrichment of mCpG at ampDAP-seq specific peaks is significant but modest (p value < 0.05 Fisher exact test, fold change < 2, Figure 2d), which indicates a weak negative effect of methylation on binding affinity. This weak methyl-sensitivity is consistent with those of the human orthologues *in vitro* [22]. In contrast, the *Amphimedon* Sp orthologue showed a weak preference for mC in the DAP-seq peaks, which is also consistent with human SP1 and SP2 binding preferences[22]. However, *Amphimedon* NRF shows almost no methylation in its DAP-seq peaks, whereas the hundreds of NRF peaks specific to ampDAP-seq (comprising 54% of all ampDAP-seq peaks) are originally methylated regions that become accessible to NRF binding due to methylation depletion by PCR amplification (Figure 2e). Given that transcription factors tend to conserve binding affinities for millions of years[26], we propose that the strong methyl-sensitivity of NRF has been conserved from sponges to humans. Interestingly, NRF is a transcription factor restricted to animal genomes, thus it could have been a methylation reader from early steps of animal evolution. Therefore, given the conserved methylation-dependent modulation of transcription factor binding affinities, species with hypermethylated genomes have the potential for methylation to restrict transcription factor-dependent regulatory networks.

Conservation of transcription factor methyl-sensitivity, together with enrichment of CpGs at unmethylated promoters, could have a role in shaping promoter sequence architecture across metazoans. To test this idea, we searched for enrichment of the *Amphimedon* CpG-bearing UMR motifs in a subset of animal promoters. *Amphimedon* and vertebrates show higher enrichments for these motifs at unmethylated promoters, when compared to species with lower levels of genomic mC or those that have lost mC (Supplementary Figure 7). This suggests that sequence composition of promoters is shaped to accommodate restricted methyl-sensitive transcription factor binding in hypermethylated genomes[27].

Europe PMC Funders Author Manuscripts

Local but not global methylation developmental dynamics in *Amphimedon*

In vertebrates, developmental changes in gene expression are mostly anticorrelated to methylation levels at regulatory regions[6,7]. To determine whether such anticorrelation can be observed in *Amphimedon*, we sequenced the DNA methylomes of early development (cleavage), mid development (larval and juvenile) and adult stages. As reported for zebrafish and frog[7], *Amphimedon* also lacks global mC level changes during these developmental transitions (Figure 1b), although it remains possible that unsampled developmental stages do present such global mC changes. However, hundreds of differentially methylated regions (DMRs) follow 5 distinct developmental trajectories (hypomethylation, hypermethylation and transient hypomethylation, Figure 3a). *Amphimedon* DMRs were mostly found in promoter regions. However, when computing the correlation between DMR methylation level with the nearest gene expression level in all five developmental stages, we observed a modest enrichment for anticorrelation values (Pearson $r < -0.9$, Figure 3c,d). When restricting the analysis to DMRs found in promoter regions, anti-correlation was only found in 23% of cases (Pearson $r < -0.5$), and presence of methyl-sensitive transcription factor binding motifs poorly predicted transcriptional response (Supplementary Figure 8). These observations indicate that mC dynamics mildly correlate with transcriptional repression in *Amphimedon,* which indicates that it is not a widespread mechanism to regulate gene transcription in this species.

Activation of transcription along development correlates with the deposition of the histone modification H3K4me3, which has also been shown to be anticorrelated with mC [28,29]. However, most vertebrate promoters are permanently unmethylated, and H3K4me3 is predominantly deposited during later stages of development, such as the onset of zygotic gene activation[30]. Similarly, deposition of H3K4me3 in *Amphimedon* adult stages occurs at stably unmethylated promoters (Figure 3e). Therefore, H3K4me3 is not responsible for protecting promoters from mC. Thus, despite mC being the default state in vertebrate and *Amphimedon* genomes, which could potentially spread to methylate transcriptionally silent regions of the genome, most promoters are constantly protected from methylation in a transcription-independent manner.

Genomic 5-hydroxymethylcytosine correlates with transcription factor binding sites in *Amphimedon*

Because we detect localized regions that lose DNA methylation during development, and because TET is expressed throughout development in *Amphimedon* (Supplementary Figure 9), we investigated whether hydroxymethylation (hmC) could have a role in DNA demethylation. To do this, we performed Tet Assisted Bisulfite sequencing (TAB-seq)[31] on two mid-developmental samples to capture the dynamics of hydroxymethylation. Interestingly, the global hmC levels in *Amphimedon* were equivalent to that of vertebrate embryos (2-3%), and an order of magnitude higher than in the cnidarian *Nematostella* (0.2%, Figure 4a, Supplementary Table 2). We found 932 regions that were enriched in hmC (> 0.1 hmC/C) in both stages, indicating that some regions are stably marked by hmC (Supplementary Figure 9). Total methylation levels measured by WGBS (which include mC and hmC) on those stably hydroxymethylated regions showed a decrease in all mid-developmental samples, indicating that hmC shows some correlation with transient

demethylation (Figure 4b). Surprisingly, the hmC-rich regions were highly enriched in motifs very similar to that of T-box and Homeobox transcription factors. Interestingly, these motifs lack CpG sites and are not enriched in the UMR dataset, indicating that they do not belong to the basal regulatory transcription factor lexicon. Overall, this could suggest that hmC deposition is linked to transcription factor binding[32] in sponges, as has been previously proposed for vertebrates and amphioxus [32,33]. Consistently, hmC-rich regions are enriched in promoters (odds ratio 1.25 versus expectation) and introns (odds ratio 2.5) of developmental genes, reinforcing their potential as regulatory sites (Figure 4c,d,e). Furthermore, it shows that TET enzymes are not restricted to RNA hydroxymethylation in invertebrates as previously suggested[34]. However, only hypermethylated genomes would require hmC as a general and frequent mechanism to modulate their methylomes at regulatory regions.

Discussion

The highly methylated genome of *Amphimedon* reveals striking similarities with vertebrates. However, it is very unlikely that the last common ancestor of animals had a hypermethylated genome state that was subsequently lost in all the extant invertebrate lineages profiled to date (Figure 4f). The invertebrate “mosaic” methylation state resembles patterns found in plant genomes, and is therefore likely to be the ancestral state in eukaryotes, and thus also in animals[1,3]. Furthermore, the methylomes of the calcareous sponge *Sycon* and the sea lamprey[35], the sister groups to *Amphimedon* and jawed vertebrates respectively, show intermediate methylation levels, suggesting gradual steps towards genomic hypermethylation. Similarly, on the other end of the spectrum, total loss of DNA methylation in *Drosophila melanogaster* or *Caenorhabditis elegans* was preceded by a gradual decrease of methylation levels in insects and nematodes[36,37]. Given that methylation loss does not seem to occur rapidly, transition to hypermethylation should also be a gradual process, since a rapid transition could be detrimental for the regulatory outcomes of a genome. Furthermore, species that are known to display very slow evolutionary rates and be representative of ancestral genomic characteristics, such as the cephalochordate amphioxus, also show typical invertebrate “mosaic” methylation patterns [33]. Also, none of the vertebrate species profiled to date shows loss of the hypermethylated state, just reduction of global levels [38]. Thus, an independent transition to hypermethylation in both *Amphimedon* and vertebrates is the most parsimonious explanation supported by current data. More methylation data from additional sponges and metazoan lineages should offer key insights to test the conflicting hypotheses. Further examples of the hypermethylated state across the animal tree of life, or secondary losses in hypermethylated lineages, would provide support for an alternative scenario in which hypermethylation is an ancestral but unstable state, having been lost repeatedly in most lineages. Based on current data, we propose that the likely independent transition to hypermethylation may have shaped the genomes and the regulatory complexity of sponges and vertebrates in a similar manner, causing global loss of CpGs and accumulation of CpGs at unmethylated promoters, and imposing burdens to methyl-sensitive transcription factor binding.

Given that the machinery responsible for methylation, demethylation, or methylation recognition is widely conserved throughout animals, some regulatory similarities between

Amphimedon and vertebrates are not likely to be convergent, but rather are potentially traces of conserved mechanisms. But only in hypermethylated genomes might mechanisms such as TET-mediated demethylation or transcription factor methyl-sensitivity become more prevalent and readily detectable. This highlights how DNA methylation is a pleiotropic genome modification that evolves modularly; some species have exclusively lost methylation either on gene bodies or transposable elements[3,17,37,39,40], whereas *Amphimedon* and vertebrates have gained hypermethylation. This modularity of methylation profiles, however, is not linked to the conservation of DNMT genes and their domain architectures. Only the loss of both DNMT1 and DNMT3 orthologues correlates with methylation loss. This pattern of DNMT conservation not being a good predictor of the methylation pattern in animals is consistent with data from land plants and fungi [41–43]. This is not surprising given that DNMTs do not bind DNA in a sequence-specific manner. Thus, a complex crosstalk between DNMTs and chromatin components is likely to determine the methylation landscapes in a lineage-specific manner. This specificity is likely to evolve through affinity changes in the chromatin interaction domains found in DNMTs, such as PWWP in DNMT3B[44]. Furthemore, the central role that some of the DNA methylation enzymes play in chromatin might be uncoupled to DNA methylation, as recently shown for DNMT enzymes in insects[45,46] or corroborated by the presence of TET and MBD proteins in species that have lost DNA methylation such as *Drosophila* (Supplementary Figure 1). A better understanding of the mechanisms that determine DNMT deposition is required to formulate predictive hypothesis on methylation patterns based on gene content alone.

Intriguingly, most of the hypotheses proposed to explain the evolution of genome hypermethylation in vertebrates cannot explain *Amphimedon*'s case. The ancestral vertebrate whole genome duplication could have required hypermethylation to compensate for chromosome imbalance, but *Amphimedon* did not undergo a whole genome duplication. *Amphimedon* also reveals how the retention of TET and DNMT3 duplicates in vertebrates is not required for hypermethylation. Most importantly, genome hypermethylation cannot be associated with the higher levels of regulatory complexity, gene length or genome size of vertebrates[2,8,47,48], because a sponge that possess only a small number of cell types[49] and a very compact genome has evolved a similar mC pattern. Given that *Amphimedon* has a relatively high repeat content (Table 1), hypermethylation could have evolved as a protective measure against transposon expansion or external DNA recognition, a mechanism previously proposed for vertebrates [8]. However, the genome of species such as the pufferfish *Tetraodon nigroviridis* have less than 5% of repeats and still have a typical vertebrate hypermethylated genome [3]. In contrast, the oyster *Crassostrea gigas* has a typical invertebrate "mosaic" methylome [50] and a similar repeat content to *Amphimedon* (36%)[51]. Gene body methylation is believed to have a role in avoiding aberrant transcription initiation[2]. This role could avoid transcriptional read-through into nearby genes in the gene-dense *Amphimedon* genome, since any transcription initiation event is likely to result in transcription of a protein coding gene or its anti-sense, unlike in genomes with longer intergenic distances. However, other compact animal genomes like those of nematodes did not undergo hypermethylation, but rather methylation loss. Furthermore, the calcareous sponge *Sycon* is not as gene-dense as *Amphimedon* and yet shows high methylation levels (Table 1).

Given that we do not fully understand the function of DNA methylation in either vertebrates or invertebrates, it is currently difficult to speculate about the possible causes of the origin and evolution of hypermethylation. It is likely that the evolution of hypermethylation is a complex process, in response to multiple evolutionary pressures, perhaps in a lineage-specific manner. The evolutionary costs that methylation imposes on hypermethylated genomes, such as increased CpG mutability or off-target alkylation damage[37], suggests that retention of the hypermethylated state likely has yet unknown evolutionary causes and functions. For example, hypermethylation could have evolved in an ancestral species in response to a specific threat (e.g. viral or transposon element invasion), but has been subsequently co-opted for new functions and thus retained in the descendant species. Also, the chromatin changes that had to co-evolve with the hypermethylated state could be highly difficult to revert, which would explain why there is no evidence for loss of the hypermethylated state in any jawed vertebrate genome. Regardless of the mechanisms of origin, the independent acquisition of a highly methylated genome in the sponge *Amphimedon* challenges and expands our knowledge of the evolution and function of DNA methylation, and reveals the hidden complexity of animal lineages that were thought to be simple[52].

Methods

Animal collection and nucleic acid extraction

Adult *Amphimedon* queenslandica were collected on Heron Island Reef, Great Barrier Reef, Australia, in February 2015. Embryos, larvae (both pre-competent and competent) and juveniles from multiple adults were collected as previously described [53] and frozen in liquid nitrogen. Pools of embryos, larvae and juveniles, and a 1 cm^3 tissue biopsy from one adult were used to extract gDNA via proteinase K digestion in an SDS lysis buffer, phenol:chloroform:isoamyl phase separation, and RNase treatment. Together, these developmental stages span the full life cycle of *Amphimedon* and comprise multiple morphologically distinct cell types [49,54].

Mnemiopsis leydi were originally collected from Kristineberg, Sweden, but kept in culture at the Sars Institute in Bergen. *Mnemiopsis* embryos from a self-fertilized single adult were collected at 3, 4.5, 8.5, 14 and 30 hours post-fertilization[55], while the juvenile sample was an independent individual (1 cm adult). The DNA was extracted using DNAzol and cleaned up with phenol-chloroform and RNase treatment.

Sycon ciliatum adults were collected from the Bergen fjords, 3 adults were used to extract genomic DNA using the Qiagen AllPrep kit.

Nematostella vectensis embryos were obtained from crossing of adult polyps kept at 18 °C in 16‰ artificial sea water in the dark. The animals were placed at 25 °C under light for 10 hours to induce spawning. Developing embryos were kept at 21 °C, and collected at major developmental timepoints: blastula (9 hours post fertilization), gastrula (24 hours post fertilization) and planula larva (72 hours post fertilization). Genomic DNA was isolated from embryos and larvae by standard phenol-chloroform extraction.

Whole genome bisulfite sequencing

For each species and developmental time point, 500 ng to 1 μg of genomic DNA spiked with 0.1%(w/w) of unmethylated lambda genomic DNA was sonicated to mean 200 bp fragments. Fragmented DNA was then end-repaired, A-tailed and ligated to NEXTflex methylated sequencing adaptors (BIOO Scientific). Bisulfite conversion of DNA was performed using EZ DNA Methylation-Gold Kit (Zymo Research), followed by library amplification using KAPA HiFi HotStart Uracil+ DNA polymerase (Kapa Biosystems). Depending on the amount of material after bisulfite conversion, libraries were amplified by 5 to 8 cycles of PCR. Libraries were then sequenced on an Illumina HiSeq 1500. The resulting fastq files were adaptor trimmed using BBduk (BBmap tools), and mapped to the reference genomes using BSseeker2[56], based on the Bowtie2 aligner (bs_seeker2-align.py --aligner=bowtie2 --bt2--end-to-end). Non-conversion rates were obtained using the lambda genomic DNA reads, where all mC calls are counted as false positives and divided by the total coverage on C positions. One replicate was profiled for each developmental stage. Publicly available datasets for zebrafish and mouse were obtained from a previous publication[7].

Methylation data was visualised using IGV genome browser. Positional heatmaps and average plots were computed using deepTools2[57] *plotHeatmap* and *plotProfile* functions, using a bin size of 100 bp (-bs 100) and allowing scaling gene bodies to 3000 bp. Methylation fractions on CpGs was computed using R, classifying each CpG in the 4 categories specified in Figure 1b.

To obtain CpG density plots, we first generated a track of all CpG positions in each genome using the *coverage2cytosine* function in Bismark[58]. These were converted to bigwigs using the UCSC *bedGraphToBigWig* function, and processed with deepTools2 requiring --*missingDataAsZero* option and 100 bp bins.

TAB-seq

The time points profiled with TAB-seq were selected among mid-developmental stages of *Amphimedon* and *Nematostella*, to allow comparison of hmC to mC levels of samples from previous and later developmental stages. For all samples, 500 ng of matched genomic DNA for each species and developmental timepoint was spiked with 0.5% hydroxymethylated pUC19 plasmid and 0.25% lambda genomic DNA methylated at CpGs. The mix was then sheared to mean 200 bp, followed by the ß-glucosyltransferase reaction and Tet1-based oxidation as per the manufacturer's instructions of the 5hmC TAB-seq Kit (WiseGene, K001). The resulting DNA was bisulfite converted and amplified as for MethylC-seq, and bioinformatically processed using the same software, but including pUC19 in the reference genome. The protection rate was calculated as the total number of C basecalls in reads aligned to pUC19, divided by the total amount of hmC in the control pUC19. The hmC levels of the pUC19 control were established through a separate MethylC-seq library preparation, as only 87% of the Cs in the pUC19 are actually hmC. The non-conversion rate was obtained as the sum of C calls in CpH positions of the lambda genome (where H = A, C or T), and the non-oxidation rate was obtained as the sum of C calls in the CpG positions of the lambda genome. Global hmCG levels for each sample were corrected by subtracting the

non-conversion rate from the hmC level and dividing the value by (1 - non-conversion rate - non-protection rate) to correct for false negatives.

UMR and DMR analysis

To test whether mCG was symmetrical on both strands of the DNA in these species, CpG positions with MethylC-seq coverage > 10 on each strand were assessed for correlation of mCG values. Given that all species retrieved high correlations (Pearson p > 0.9 as calculated by R "cor" function), we merged information for both strands into single CpG values, which were later used for downstream analysis. Identification of UMRs was performed using MethylSeekR[59] *segmentUMRsLMRs* function with default parameters (meth.cutoff = 0.5, nCpG.cutoff = 4) for each whole genome bisulfite sequencing sample, and later filtered for UMR mCG levels < 0.1.

For comparing promoters across species, we selected UMRs intersecting promoters in species with DNA methylation. These UMRs were required to be less than 3 kilobases, to avoid long unmethylated intergenic regions in species such as *Nematostella* and *Sycon*, or "CpG canyons" in vertebrates, reasoning that shorter regions protected from DNA methylation are more likely to contain constrained regulatory information. For species without DNA methylation, we used H3K4me3 peaks from modENCODE and previous publications (GSE71131)[60], given that proximal H3K4me3 demarcates unmethylated regions in species with DNA methylation. For *Mnemiopsis* we used -1 Kb upstream from the TSS, given that UMRs are rare in promoters. Motif searches were performed using findMotifsGenome function and the known.motif dataset in HOMER (-nomotif -mknown).

To obtain DMRs, we used the DSS[61] package from Bioconductor to retrieve DMRs with a minimum difference in mCG/CG level (delta) of 0.2, p-value < 0.05, and ≥ 5 CpGs in an all-versus-all pairwise comparison between samples. DSS was called allowing a span of 100 bs for smoothing in the *DMLtest* function (smoothing=TRUE, smoothing.span=100) and *callDMR* function allowing a merge distance of 100 bp (delta=0.2, p.threshold=0.05, minCG=5, dis.merge=100). This set of DMRs was then further filtered by requiring a minimal WGBS coverage of 4 in all samples, a minimal length of 3 CpGs and a minimal difference of 0.4 between the maximum mCG level and the minimum. Developmental trajectories were obtained by clustering DMR methylation values using k-means (*kmeans* function in R) seeded with a number equal to the number of samples being analyzed (n = 5). Imposing a higher number of k-means clusters on the DMR dataset resulted in redundant trajectories among clusters.

UMRs and DMRs were classified by intersecting the coordinates with genome annotations of each species based on RNA-seq data[14,29,62,63] using BEDTools[64]. UMR/DMRs were associated to genes by taking the nearest TSS, or in the case of intersecting a bidirectional promoter, were assigned to both genes sharing the promoter. Publicly available RNA-seq data was mapped to the longest-isoform for each gene using Kallisto[65] *quant* function with default parameters and CEL-seq raw count data was obtained from GEO dataset GSE70185 and transformed into TPMs with R, dividing the total number of reads per gene by the total number of mapped reads on the sample. For samples with RNA-seq/CEL-seq replicates, values were summarised for each stage using the mean of all replicates.

Transposon and repetitive region annotation

For each species, de novo gene prediction through Augustus[66] was obtained using the "intronless" mode. The resulting genes were then scanned for known transposon protein domains as defined by PFAM. Repetitive regions were obtained for each species running RepeatModeler[67] as default, and filtering out the transposable element coding regions obtained with the Augustus approach.

ChIP-seq and motif enrichment analysis

Sequencing data was downloaded from GSE79645[28] and GSE32483[68] trimmed using BBduk and mapped to the reference genome using bowtie with parameters "-m 1 -v 1 -q -S". Peaks were called using MACS2[69] callpeak with $q < 0.01$ using input as background. Narrow peaks were then merged allowing for a maximal distance of 1000 bp between peaks using BEDTools *merge* function (-d 1000). Coverage for each ChIP-seq was normalised by total number of mapped reads into reads per million (RPM) using the deepTools2 *bamCoverage* function (--normalizeUsing CPM -bs 10). Differential H3K4me3 peak calling was performed using MACS2 *bdgdiff* function.

Motif enrichment analysis for different sets of genomic intervals was performed using the HOMER[70] *findMotifsGenome* function, allowing lengths of 6, 8, 10 and 12 bp and setting the size as "given" (-size given -len 6,8,10,12).

Gene annotation and evolutionary analysis

Every gene family was annotated using HMMER3[71] (hmmsearch --cut-ga) to search for core PFAM domains (PF00145 for DNMTs, PF12851 for TET, PF02182 for UHRF1 and PF01429 for MBDs) in the proteomes of 34 metazoan and 4 unicellular holozoan species with genome sequence available. The resulting hits were then aligned using MAFFT (L-INS-I mode)[72] and trimmed using TrimAL(-automated mode)[73], and maximum likelihood phylogenies were constructed using IQ-TREE[74]. IQ-TREE was ran using the option "-m TEST", which calculates the best fitting amino acid substitution model for each gene family. The resulting phylogenetic trees were then used to assign orthology to the distinct gene families. For zinc finger CXXC containing proteins, we searched the human orthologues of KDM2A, KMT2A, CXXC1 and CXXC4 using blastp against the 38 holozoan proteomes. We gathered all the hits with an e-value < 0.0001, limiting the sequences to 150. Then we constructed phylogenetic trees as for the other gene families. The resulting sequences were then searched using HMMER3 (--cut-ga threshold) and the PFAM model PF02008 to determine the presence of zinc fingers CXXC.

All the transcription factors that were identified in the motif enrichment analysis of Figure 2d were searched using the human orthologue as a query in a blastp search against the 34 metazoan proteomes. The best 100 hits were then aligned, trimmed and phylogenies were constructed as described above. The closest human paralog was used as an out-group, and every *Amphimedon* ortholog was confirmed using the phylogenetic tree as reference.

DAP-seq and ampDAP-seq

Amphimedon orthologues of transcription factors enriched at UMRs were *in vitro* synthesized and cloned into pIX-HALO plasmids. For Aqu_NRF, Aqu_YY1 and Aqu_SP, both the full length transcription factor, as well as the DNA binding domains (DBDs) plus 50 padding amino acids were used, as reported in previous approaches[22]. Full length transcription factors and DBDs were *in vitro* translated using the TNT SP6 Coupled Wheat Germ Extract System (Promega). The subsequent steps were performed following the standard DAP-seq protocol[25]. AmpDAP-seq libraries were generated using 15 ng of unamplified adaptor ligated DNA and amplified by PCR for 11 cycles to deplete the DNA methylation. For each pIX-HALO plasmid, 40 ng of *Amphimedon* DAP-seq and ampDAP-seq libraries with unique Illumina multiplexing indexes were used in the affinity pull down. The pooled libraries were sequenced with an Illumina NextSeq 500 instrument. The resulting reads were trimmed using fastp[75] (default parameters), mapped to the *Amphimedon* genome using bowtie[76] (-q -m 1 -v 1 -q -S -1), and uniquely mapped reads were used to call peaks using MACS2[69] (-B -f BAMPE -q 0.05 --down-sample). For peak calling, DAP-seq and ampDAP-seq libraries were obtained by incubating the DNA with an empty pIX-HALO plasmid, which was subsequently used as background in the *callpeaks* function ("-c DAPseq_emptyHaloPlasmid.bam" or "-c ampDAPseq_emptyHaloPlasmid.bam" for DAP-seq and ampDAP-seq samples respectively). Motif enrichment on peaks and motif scans were obtained using HOMER[70] as described for UMRs. Methylation levels on motifs were obtained from the *Amphimedon* precompetent larva whole genome bisulfite sequencing data, which was generated from the same pool of genomic DNA used for making the DAP-seq libraries. Enrichment levels of DAP-seq and ampDAP-seq libraries were obtained using deepTools2 *bamCompare* function (-bs=10 --operation log2). Visualization of DAP-seq and ampDAP-seq was obtained using deepTools2 *computeMatrix* and *plotHeatmap* functions, previously filtering for peaks that had the NRF motif (--referencePoint center -bs 100),

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Jose M. Polo for critical reading of this manuscript. This work was supported by the Australian Research Council (ARC) Centre of Excellence program in Plant Energy Biology (CE140100008). RL was supported by a Sylvia and Charles Viertel Senior Medical Research Fellowship, ARC Future Fellowship (FT120100862), and Howard Hughes Medical Institute International Research Scholarship. SMD and BMD were supported by grants from the Australian Research Council. Research in AH's group was supported by the European Research Council Community's Framework Program Horizon 2020 (2014–2020) ERC grant agreement 648861 and an NSF IRFP Postdoctoral Fellowship (1158629) to KP. AdM was funded by an EMBO long term fellowship (ALTF 144-2014). UT was funded by a grant from the Austrian Science Fund FWF (P27353).

Data Availability

Sequencing data have been deposited in Gene Expression Omnibus (GEO) under the following accession number GSE124016.

Code Availability

The analysis code is available on https://github.com/AlexdeMendoza/SpongeMethylation.

References

1. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet. 2010; 11:204–220. [PubMed: 20142834]
2. Schübeler D. Function and information content of DNA methylation. Nature. 2015; 517:321–326. [PubMed: 25592537]
3. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science. 2010; 328:916–919. [PubMed: 20395474]
4. Feng S, et al. Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A. 2010; 107:8689–8694. [PubMed: 20395551]
5. Jones, Pa. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012; 13:484–492. [PubMed: 22641018]
6. Stadler MB, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011; 480:490–495. [PubMed: 22170606]
7. Bogdanović O, et al. Active DNA demethylation at enhancers during the vertebrate phylotypic period. Nat Genet. 2016; 48:417–426. [PubMed: 26928226]
8. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet. 2008; 9:465–476. [PubMed: 18463664]
9. Lyko F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. Nat Rev Genet. 2018; 19:81–92. [PubMed: 29033456]
10. Wu X, Zhang Y. TET-mediated active DNA demethylation: mechanism, function and beyond. Nat Rev Genet. 2017; 18:517–534. [PubMed: 28555658]
11. Iyer, LM, Abhiman, S, Aravind, L. Natural history of eukaryotic DNA methylation systems. Vol. 101. Elsevier Inc; 2011.
12. Srivastava M, et al. The Amphimedon queenslandica genome and the evolution of animal complexity. Nature. 2010; 466:720–726. [PubMed: 20686567]
13. Fortunato SAV, et al. Calcisponges have a ParaHox gene and dynamic expression of dispersed NK homeobox genes. Nature. 2014; 514:620–623. [PubMed: 25355364]
14. Ryan JF, et al. The Genome of the Ctenophore Mnemiopsis leidyi and Its Implications for Cell Type Evolution. Science. 2013; 342:1242592–1242592. [PubMed: 24337300]
15. Putnam NH, et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science. 2007; 317:86–94. [PubMed: 17615350]
16. Fernandez-Valverde SL, Degnan BM. Bilaterian-like promoters in the highly compact Amphimedon queenslandica genome. Sci Rep. 2016; 6:22496–22496. [PubMed: 26931148]
17. Suzuki MM, Kerr ARW, De Sousa D, Bird A. CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res. 2007; 17:625–631. [PubMed: 17420183]
18. Francis WR, et al. The genome of the contractile demosponge Tethya wilhelma and the evolution of metazoan neural signalling pathways. bioRxiv. 2017; doi: 10.1101/120998
19. Cohen NM, Kenigsberg E, Tanay A. Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. Cell. 2011; 145:773–786. [PubMed: 21620139]
20. Boulard M, Edwards JR, Bestor TH. FBXL10 protects Polycomb-bound genes from hypermethylation. Nat Genet. 2015; 47:479–485. [PubMed: 25848754]
21. Domcke S, et al. Competition between DNA methylation and transcription factors determines binding of NRF1. Nature. 2015; 528:575–579. [PubMed: 26675734]
22. Yin Y, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 2017; 356
23. Brandeis M, et al. Sp1 elements protect a CpG island from de novo methylation. Nature. 1994; 371:435–438. [PubMed: 8090226]

24. Macleod D, Charlton J, Mullins J, Bird AP. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. 1994; 8:2282–2292. [PubMed: 7958895]
25. Bartlett A, et al. Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat Protoc. 2017; 12:1659–1672. [PubMed: 28726847]
26. Nitta KR, et al. Conservation of transcription factor binding specificities across 600 million years of bilateria evolution. Elife. 2015; 4:1–20.
27. Krebs AR, Dessus-Babus S, Burger L, Schübeler D. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. Elife. 2014; 3:e04094. [PubMed: 25259795]
28. Gaiti F, et al. Landscape of histone modifications in a sponge reveals the origin of animal cis-regulatory complexity. Elife. 2017; 6
29. Schwaiger M, et al. Evolutionary conservation of the eumetazoan gene regulatory landscape. Genome Res. 2014; 24:639–650. [PubMed: 24642862]
30. Hontelez S, et al. Embryonic transcription is controlled by maternally defined chromatin state. Nat Commun. 2015; 6:10148. [PubMed: 26679111]
31. Yu M, et al. Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. Nat Protoc. 2012; 7:2159–2170. [PubMed: 23196972]
32. Sardina JL, et al. Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. Cell Stem Cell. 2018; 23:727–741.e9. [PubMed: 30220521]
33. Marlétaz F, et al. Amphioxus functional genomics and the origins of vertebrate gene regulation. Nature. 2018; 564:64–70. [PubMed: 30464347]
34. Delatte B, et al. RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. Science. 2016; 351:282–285. [PubMed: 26816380]
35. Zhang Z, et al. Genome-wide and single-base resolution DNA methylomes of the Sea Lamprey (Petromyzon marinus) Reveal Gradual Transition of the Genomic Methylation Pattern in Early Vertebrates. bioRxiv. 2015; doi: 10.1101/033233
36. Bewick AJ, Vogel KJ, Moore AJ, Schmitz RJ. Evolution of DNA Methylation across Insects. Mol Biol Evol. 2017; 34:654–665. [PubMed: 28025279]
37. Rošić S, et al. Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. Nat Genet. 2018; 50:452–459. [PubMed: 29459678]
38. Mugal CF, Arndt PF, Holm L, Ellegren H. Evolutionary consequences of DNA methylation on the GC content in vertebrate genomes. G3. 2015; 5:441–447. [PubMed: 25591920]
39. Bewick AJ, et al. On the origin and evolutionary consequences of gene body DNA methylation. Proc Natl Acad Sci U S A. 2016; 113:9111–9116. [PubMed: 27457936]
40. Wang X, et al. Function and Evolution of DNA Methylation in Nasonia vitripennis. PLoS Genet. 2013; 9:e1003872–e1003872. [PubMed: 24130511]
41. Bewick AJ, et al. Diversity of cytosine methylation across the fungal tree of life. Nature Ecology & Evolution. 2019; doi: 10.1038/s41559-019-0810-9
42. Takuno S, Ran J-H, Gaut BS. Evolutionary patterns of genic DNA methylation vary across land plants. Nature Plants. 2016; 2:15222–15222. [PubMed: 27249194]
43. Niederhuth CE, et al. Widespread natural variation of DNA methylation within angiosperms. Genome Biol. 2016; 17:194. [PubMed: 27671052]
44. Baubec T, et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature. 2015; 520:243–247. [PubMed: 25607372]
45. Bewick AJ, et al. Dnmt1 is essential for egg production and embryo viability in the large milkweed bug, Oncopeltus fasciatus. Epigenetics Chromatin. 2019; 12:6. [PubMed: 30616649]
46. Schulz NKE, et al. Dnmt1 has an essential function despite the absence of CpG DNA methylation in the red flour beetle Tribolium castaneum. Sci Rep. 2018; 8
47. Lechner M, et al. The correlation of genome size and DNA methylation rate in metazoans. Theory Biosci. 2013; 132:47–60. [PubMed: 23132463]
48. Regev A, Lamb MJ, Jablonka E. The Role of DNA Methylation in Invertebrates: Developmental Regulation\ror Genome Defense? Mol Biol Evol. 1998; 15:880–891.

49. Sebé-Pedrós A, et al. Early metazoan cell type diversity and the evolution of multicellular gene regulation. Nature Ecology & Evolution. 2018
50. Wang X, et al. Genome-wide and single-base resolution DNA methylomes of the Pacific oyster Crassostrea gigas provide insight into the evolution of invertebrate CpG methylation. BMC Genomics. 2014; 15:1119. [PubMed: 25514978]
51. Zhang GG, et al. The oyster genome reveals stress adaptation and complexity of shell formation. Nature. 2012; doi: 10.1038/nature11413
52. Dunn CW, Leys SP, Haddock SHD. The hidden biology of sponges and ctenophores. Trends Ecol Evol. 2015; 30:282–291. [PubMed: 25840473]
53. Leys SP, et al. Isolation of amphimedon developmental material. CSH Protoc. 2008; 2008
54. Leys SP, Degnan BM. Embryogenesis and metamorphosis in a haplosclerid demosponge: gastrulation and transdifferentiation of larval ciliated cells to choanocytes. Invertebr Biol. 2005; 121:171–189.
55. Pang K, Martindale MQ. Comb jellies (ctenophora): a model for Basal metazoan evolution and development. CSH Protoc. 2008; 2008
56. Guo W, et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics. 2013; 14:774–774. [PubMed: 24206606]
57. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014; 42:W187–91. [PubMed: 24799436]
58. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011; 27:1571–1572. [PubMed: 21493656]
59. Burger L, Gaidatzis D, Schübeler D, Stadler MB. Identification of active regulatory regions from DNA methylation data. Nucleic Acids Res. 2013; 41:e155–e155. [PubMed: 23828043]
60. Sebé-Pedrós A, et al. The Dynamic Regulatory Genome of Capsaspora and the Origin of Animal Multicellularity. Cell. 2016; 165:1224–1237. [PubMed: 27114036]
61. Wu H, et al. Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. Nucleic Acids Res. 2015; 43:e141–e141. [PubMed: 26184873]
62. Fernandez-Valverde SL, Calcino AD, Degnan BM. Deep developmental transcriptome sequencing uncovers numerous new genes and enhances gene annotation in the sponge Amphimedon queenslandica. BMC Genomics. 2015; 16:1–11. [PubMed: 25553907]
63. Leininger S, et al. Developmental gene expression provides clues to relationships between sponge and eumetazoan body plans. Nat Commun. 2014; 5:3905–3905. [PubMed: 24844197]
64. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
65. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016; 34:525–527. [PubMed: 27043002]
66. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008; 24:637–644. [PubMed: 18218656]
67. Smit, AFA; Hubley, R. RepeatModeler Open-1.0. 2008. *Available fom*http://www.repeatmasker.org
68. Bogdanovic O, et al. Dynamics of enhancer chromatin signatures mark the transition from pluripotency to cell specification during embryogenesis. Genome Res. 2012; 22:2043–2053. [PubMed: 22593555]
69. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012; 7:1728–1740. [PubMed: 22936215]
70. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010; 38:576–589. [PubMed: 20513432]
71. Eddy SR. Accelerated Profile HMM Searches. PLoS Comput Biol. 2011; 7:e1002195–e1002195. [PubMed: 22039361]
72. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol. 2013; 30:772–780. [PubMed: 23329690]

73. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009; 25:1972–1973. [PubMed: 19505945]

74. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015; 32:268–274. [PubMed: 25371430]

75. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018; 34:i884–i890. [PubMed: 30423086]

76. Langmead B, Trapnell C, Pop M, Salzberg S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25–R25. [PubMed: 19261174]

77. Simion P, et al. A Large and Consistent Phylogenomic Dataset Supports Sponges as the Sister Group to All Other Animals. Curr Biol. 2017; 27:958–967. [PubMed: 28318975]

78. Whelan NV, et al. Ctenophore relationships and their placement as the sister group to all other animals. Nat Ecol Evol. 2017; 1:1737–1746. [PubMed: 28993654]

Europe PMC Funders Author Manuscripts

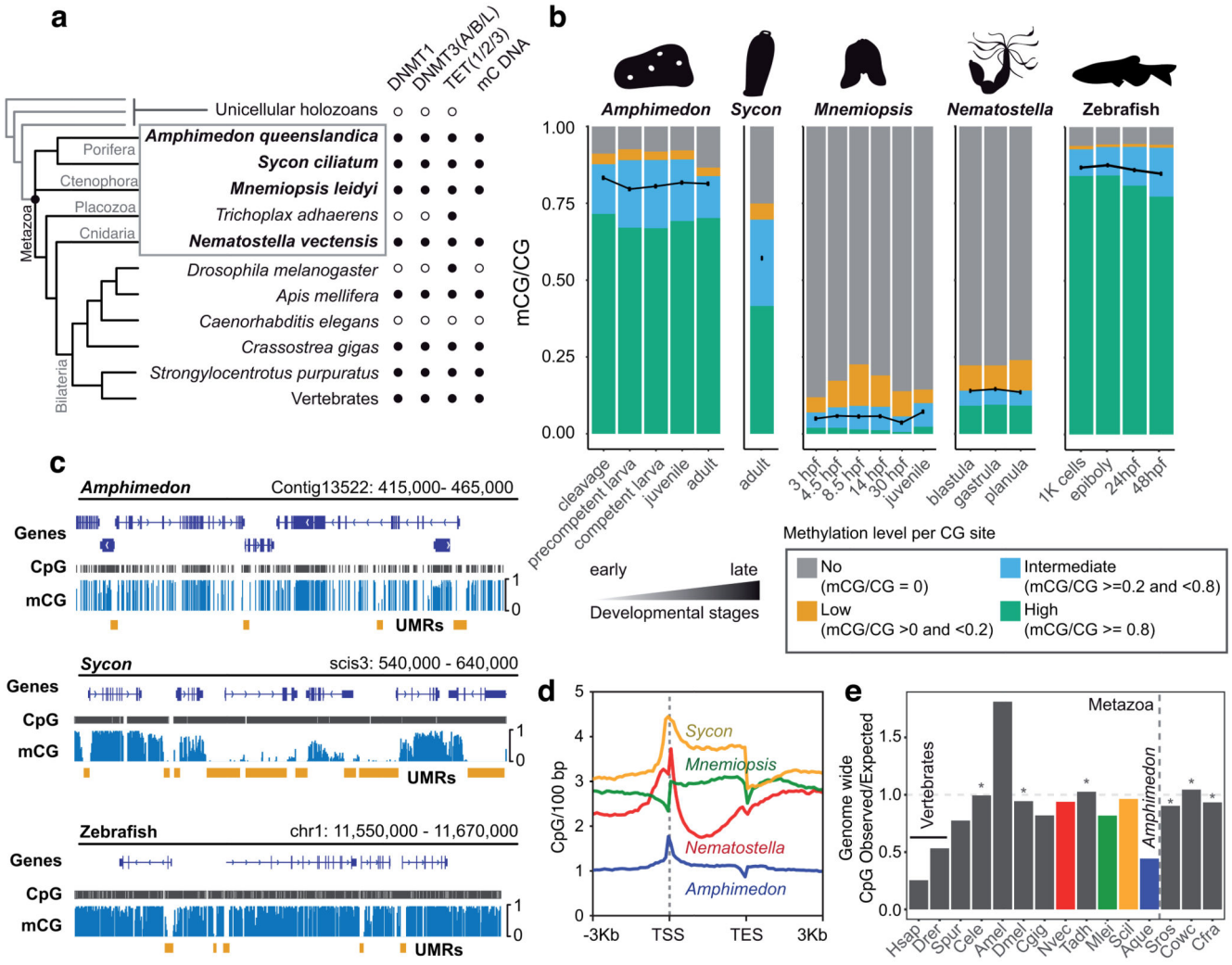

Figure 1. *Amphimedon* has a vertebrate-like methylome.

(**a**) Distribution of cytosine methylation related genes in metazoan genomes. Presence/absence of methylated DNA is based on the species analysed in this study (highlighted in bold) and previous reports. The cladogram is based on current consensus (or lack thereof) on animal phylogeny [77,78]. (**b**) Global mCG levels (represented as black dots) and fraction of total CpGs displaying high (≥0.8), intermediate (≥0.2 and <0.8), low (>0 and <0.2) and no (0) 5mC at developmental stages of four non-bilaterian species and zebrafish. (**c**) Genome browser showing globally high methylation in *Amphimedon* and zebrafish, and "mosaic" methylation in Sycon. Orange bars indicate unmethylated regions (UMRs). (**d**) CpG density mean profile on methylated gene bodies of the four non-bilaterian species. (**e**) Genomic CpG Observed versus Expected ratios of diverse metazoans and unicellular holozoans. Vertical dashed line indicates the boundary between animals and unicellular genomes, horizontal line indicates no CpG bias (1.0) and an asterisk indicates species without genomic mC. Abbreviations: Hsap (*Homo sapiens*), Drer (*Danio rerio*), Spur (*Strongylocentrotus purpuratus*), Cele (*Caenorhabditis elegans*), Amel (*Apis mellifera*), Dmel (*Drosophila melanogaster*), Cgig (*Crassostrea gigas*), Nvec (*Nematostella vectensis*), Tadh (*Trichoplax*

Europe PMC Funders Author Manuscripts

adhaerens), Mlei (*Mnemiopsis leidyi*), Scil (*Sycon ciliatum*), Aque (*Amphimedon queenslandica*), Sros (*Salpingoeca rosetta*), Cowc (*Capsaspora owczarzaki*), Cfra (*Creolimax fragrantissima*), TSS (Transcriptional Start Site), TES (Transcriptional End Site), hpf (hours post fertilization).

Europe PMC Funders Author Manuscripts

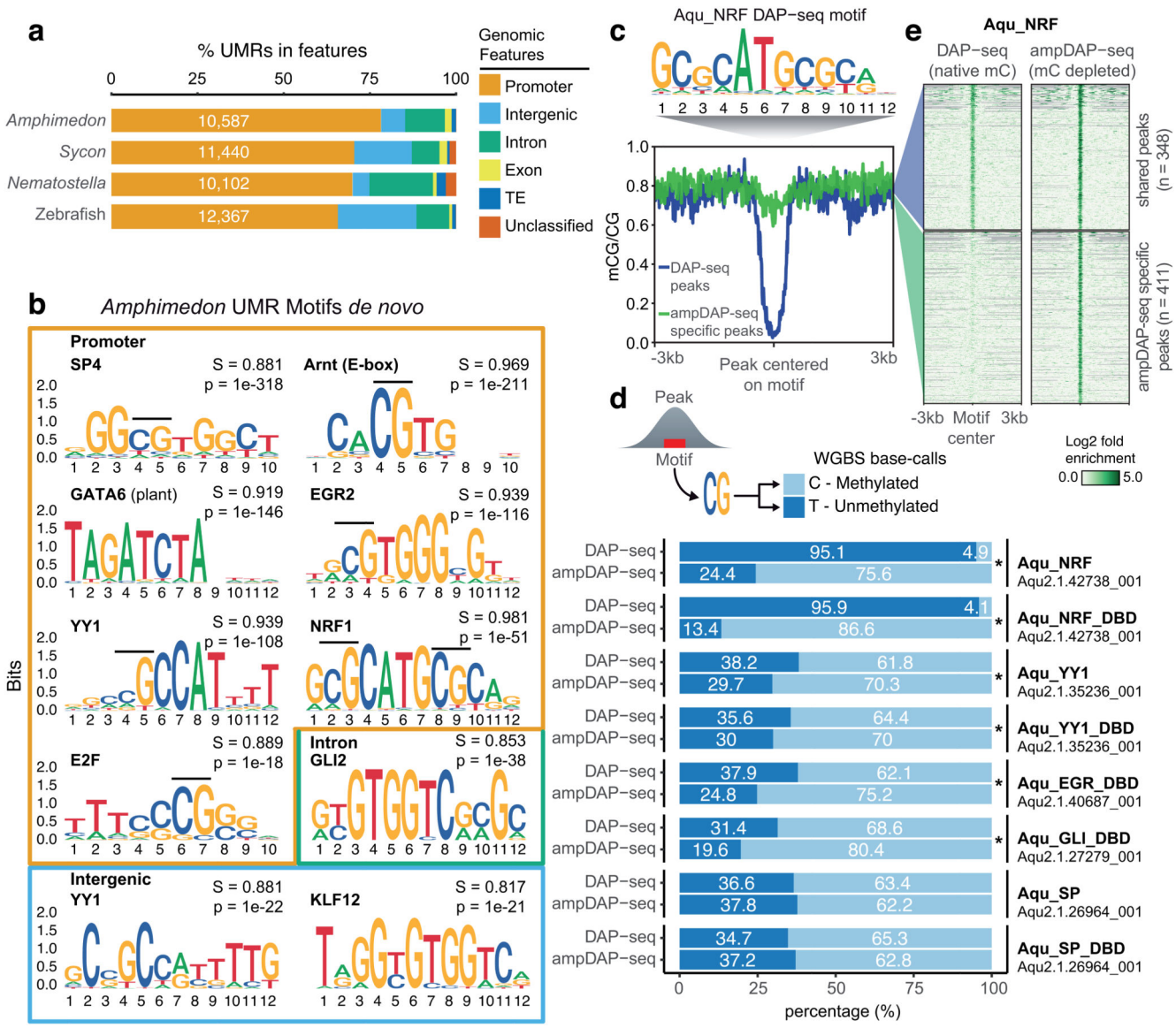

Figure 2. Methyl-sensitive transcription factors are enriched at unmethylated *Amphimedon* promoters.

(**a**) Unmethylated region (UMR) intersections with genomic features. Unclassified are regions found in scaffolds/contigs without any genes/transposable elements. (**b**) Sequence logos showing *de novo* motif enrichments for UMRs in promoters, introns and intergenic regions of *Amphimedon*. In bold is the best motif match, S is the match score to the known motif, p is the p-value enrichment (one-sided binomial test). (**c**) Motif enriched in *Amphimedon* NRF (Aqu_NRF) DAP-seq peaks. (**d**) Proportions of methylated calls versus unmethylated calls from *Amphimedon* whole genome bisulfite sequencing data at all DAP-seq peaks, and ampDAP-seq specific peaks. CpGs were obtained from the best motif within each peak. An asterisk shows significant enrichment for methylated calls in the ampDAP-seq specific peaks (one-sided Fisher exact test < 0.05). DBD (DNA Binding Domain). (**e**) Mean methylation profile at *Amphimedon* NRF DAP-seq peaks (blue) and ampDAP-seq

specific peaks (green), and heatmap showing enrichment levels of NRF DAP-seq and ampDAP-seq at those peaks. DAP-seq and ampDAP-seq data shown as log2 fold enrichment of NRF against empty pIX-HALO plasmid.

Europe PMC Funders Author Manuscripts

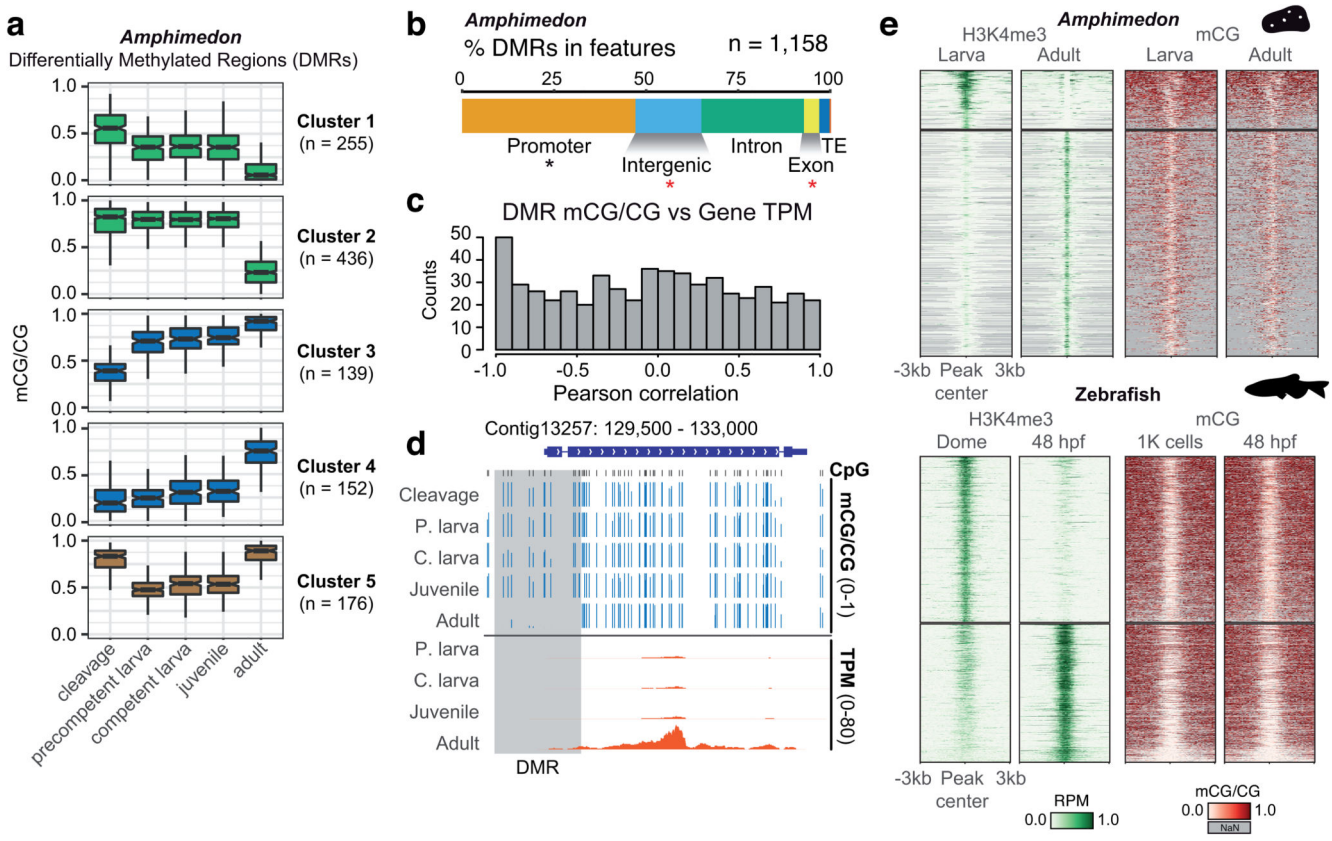

Figure 3. Methylation dynamics during *Amphimedon* development.

(**a**) K-means clustering of Differentially Methylated Regions (DMRs) (ΔmCG ≥ 0.4) throughout *Amphimedon* development, coloured by directionality of mCG change over time (green = decrease; blue = increase, brown = transient decrease in mid-developmental stages). Boxplot centre lines are medians, box limits are quartiles 1 (Q1) and 3 (Q3), whiskers are 1.5 × interquartile range (IQR) and points are outliers. (**b**) Intersection of DMRs with genomic features in *Amphimedon*. An asterisk indicates $p < 0.01$ in a two-sided Fisher's exact test between the DMRs and expected, in black enrichment against genomic background, in red depletion. (**c**) Distribution of Pearson correlation values between DMR mCG level and the transcript abundance (TPM) of the associated gene(s) for each developmental stage. (**d**) Genome browser representation of a DMR situated in a promoter displaying strong anti-correlation between DMR methylation and gene transcript abundance. (**e**) Heatmap of H3K4me3 ChIP-seq signal (Reads Per Million, RPM) and methylation levels in differential H3K4me3 peaks in *Amphimedon* and zebrafish.

Europe PMC Funders Author Manuscripts

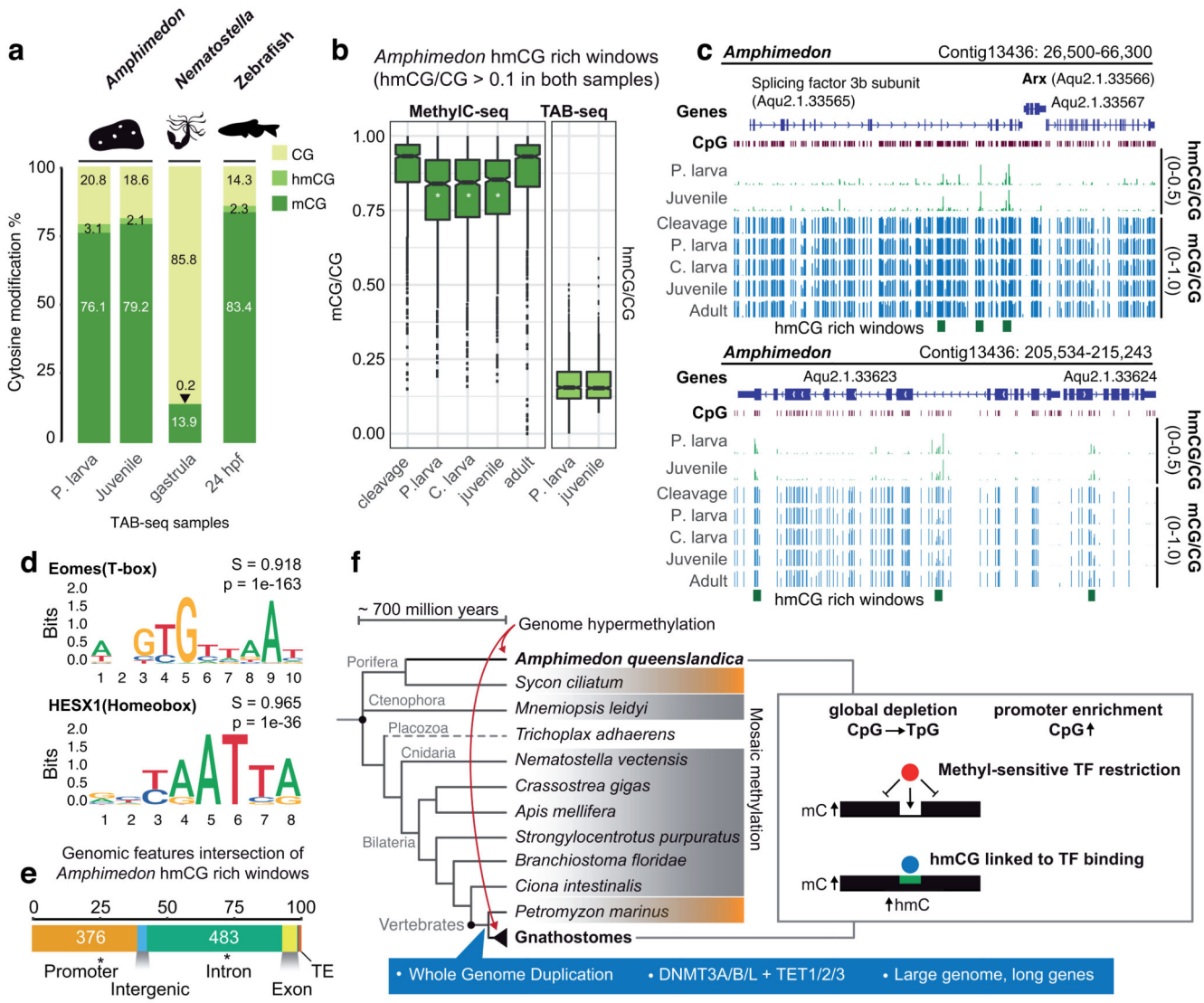

Figure 4. Genomic DNA hydroxymethylation is enriched at transcription factor binding sites in *Amphimedon*.

(**a**) Global mCG, hmCG and unmethylated CpG levels in *Amphimedon*, *Nematostella* and zebrafish assessed by MethylC-seq and TAB-seq corrected by non-conversion rate and protection rate. (**b**) Methylation and hydroxymethylation levels on hmCG rich windows (hmCG > 0.1 in both Precompetent larva and Juvenile stages). A white asterisk on the boxplot indicates FDR < 0.01 in a multiple one-sided wilcoxon enrichment test (Benjamini-Hochberg corrected) between the methylation level at hmCG rich windows against a set of random 150bp windows (n = 7,580) with equivalent coverage and CpG densities. Boxplot centre lines are medians, box limits are quartiles 1 (Q1) and 3 (Q3), whiskers are 1.5 × interquartile range (IQR) and points are outliers. (**c**) Genome browser representations of the *Amphimedon Arx* locus displaying distal hmC rich windows in the adjacent gene and transient demethylation in hmCG rich windows at mid-developmental samples (**d**) Sequence logos of de novo enriched motifs in *Amphimedon* hmCG rich windows. In bold is the best

Europe PMC Funders Author Manuscripts

match to known motifs, S is the match score and p is the p-value enrichment (one-sided binomial test). (**e**) Intersection of genomic features with *Amphimedon* hmCG rich windows. Asterisk indicates two-sided Fisher's exact test $p < 0.05$ enrichment against expected. (**f**) Cladogram representing phylogenetic relationships of animals. Species that have a sparse or "mosaic" DNA methylome are shaded in grey, those that have intermediate levels (~50%) are shaded in orange, and a dashed line indicates lack of methylation. The right-hand box summarizes the similarities observed between *Amphimedon* and vertebrate methylomes. In blue are listed some of the vertebrate characteristics that have been correlated to genome hypermethylation but are absent in *Amphimedon*.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Genome characteristics of sampled animal genomes.

Variation of genome features in the four species under study and zebrafish as a vertebrate with a moderately sized genome for comparison. Total interspersed repeats as reported by RepeatModeler and RepeatMasker. mCG/CG levels shown as an interval between max and minimum levels as in Supplementary Figure 1.

species	*Amphimedon queenslandica*	*Sycon ciliatum*	*Mnemiopsis leidyi*	*Nematostella vectensis*	*Danio rerio*
genome size (Mb)	166	357	155	356	1,674
GC content (%)	30.97	36.6	37.5	33.9	36.6
Total Interspersed Repeats (%)	35.56	25.39	21.88	25.42	46.40
mCG/CG (%)	79-83	57	3.7-7.2	13-14	78
Number of Transcription Factors	447	792	378	635	2,429
Gene number	40,122	26,206	16,548	25,729	25,592
Median intergenic length (bp)	901	3,179	1,756	3,485.5	7,376
Median gene length (bp)	1,084	2,928.5	3,024	4,135	12,725
Exonic sequence (%)	24.45	9.95	17.37	11.68	4.67