DNA methylation regulates transcriptional homeostasis of algal endosymbiosis in the coral model *Aiptasia*

Yong Li[1], Yi Jin Liew[1], Guoxin Cui[1], Maha J. Cziesielski[1], Noura Zahran[1], Craig T. Michell[1], Christian R. Voolstra[1], Manuel Aranda[1]

[1]Red Sea Research Center, Division of Biological and Environmental Science and Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal, KSA

*Correspondence to: manuel.aranda@kaust.edu.sa

Abstract: The symbiotic relationship between cnidarians and dinoflagellates is the cornerstone of coral reef ecosystems. Although research is focusing on the molecular mechanisms underlying this symbiosis, the role of epigenetic mechanisms, which have been implicated in transcriptional regulation and acclimation to environmental change, is unknown. To assess the role of DNA methylation in the cnidarian-dinoflagellate symbiosis, we analyzed genome-wide CpG methylation, histone associations, and transcriptomic states of symbiotic and aposymbiotic anemones in the model system *Aiptasia*. We find methylated genes are marked by histone H3K36me3 and show significant reduction of spurious transcription and transcriptional noise, revealing a role of DNA methylation in the maintenance of transcriptional homeostasis. Changes in DNA methylation and expression show enrichment for symbiosis-related processes such as immunity, apoptosis, phagocytosis recognition and phagosome formation, and unveil intricate interactions between the underlying pathways. Our results demonstrate that DNA methylation provides an epigenetic mechanism of transcriptional homeostasis during symbiosis.

Introduction

Coral reefs are ecologically important marine ecosystems, which cover less than 0.2% of our oceans but sustain an estimated ~25% of the world's marine species and 32 of 33 animal phyla [1, 2, 3]. Coral reefs are also economically important by providing food and livelihood opportunities to at least 500 million people; worldwide, they have a net present value of almost USD 800 billion, and they generate USD 30 billion in net economic benefits annually [3]. Unfortunately, these ecosystems are under severe threat from anthropogenic stressors including global warming and water pollution, among others, which can cause coral bleaching (loss of intracellular endosymbionts from coral) and overall coral reef decline. Despite increasing efforts on studying the mechanisms underlying the regulation and environmental stress related breakdown of this symbiotic association [4, 5], we still lack knowledge on basic molecular processes, for instance whether epigenetic mechanisms are involved in symbiosis regulation and could potentially contribute to increased resilience in response to environmental stress as reported in other organisms [6, 7].

DNA methylation plays an important role in many biological processes of plants and animals [8, 9, 10, 11]. It has been proposed as a mechanism for organisms to adjust their phenotype in response to their environment in order to optimize organismal response to changing environmental conditions [7, 12]. For instance, recent findings in mice show an important function for DNA methylation in inhibiting spurious transcription along the gene body, allowing for reduction of nonsense transcripts from highly expressed loci [13]. However, its role and function in cnidarians is, at present, unknown [14]. The sea anemone *Aiptasia* is an emerging model to study the cnidarian-dinoflagellate symbiosis. Like corals, it establishes a stable but temperature sensitive symbiosis with dinoflagellates of the genus Symbiodinium but, unlike corals, can also be naturally maintained in an aposymbiotic state. Its ease of culture and facultative symbiosis provides a tractable system to study the molecular mechanism underlying symbiosis without the impeding stress responses associated with coral bleaching stress [15, 16].

Using the model system *Aiptasia* (strain CC7, sensu *ExAiptasia pallida*), we obtained whole-genome CpG DNA methylation, ChIP-Seq and RNA-Seq data from aposymbiotic (Apo) and symbiotic (Sym) individuals to study the function of DNA methylation in transcriptional regulation and its role in the cnidarian-dinoflagellate symbiosis.

55

56 **Results**

57 ***Aiptasia* DNA Methylation patterns change with symbiotic states**

58 To assess changes in DNA methylation in response to symbiosis, we performed whole-
59 genome bisulfite sequencing with an average coverage of 53× per individual on 12 anemones,
60 providing 6 biological replicates per treatment (symbiotic vs. aposymbiotic). Methylation calling
61 using the combined dataset identified 710,768 CpGs (6.37% of all CpGs in *Aiptasia* genome),
62 i.e. methylated sites in the *Aiptasia* genome. Notably, the percentage of CpGs is much lower than
63 in mammals (60–90%) [17], but comparable to the coral *Stylophora pistillata* (7%) [18]. We
64 identified 10,822 genes (37% of all 29,269 gene models identified in the *Aiptasia* genome) with
65 at least 5 methylated positions that were subsequently defined as methylated genes. On average,
66 these genes had 18.4% CpGs methylated, 3-fold higher than the average methylation density
67 across the entire genome (Chi-squared test p value $< 2.2 \times 10^{-16}$) and 167-fold higher than the
68 methylation levels in non-coding regions. These findings indicate that the distribution of CpG
69 methylation is non-random and mainly located in gene bodies, similar to corals [18, 19] and other
70 invertebrate species [20, 21].

71 To analyze the relationship between methylation density (percentage of CpGs) and gene
72 density (the number of genes per 10,000 bp), we ran a sliding window (window size: 40 kb, step:
73 30 kb) and visualized the results in a Circos plot (Fig. S1) [22]. The correlation of CpG content
74 and distribution of methylation showed a negative correlation (Pearson correlation coefficient: r
75 $= -0.31$, p value $< 2.2 \times 10^{-16}$) suggesting that methylation tends to preferentially occur in CpG-
76 poor regions (Fig. S2). Gene density had a positive correlation with methylation density ($r =$
77 0.21, p value $< 2.2 \times 10^{-16}$) consistent with the finding that methylation is predominantly located
78 in gene bodies (Fig. 1). We also observed that within gene bodies, introns showed significantly
79 higher methylation densities than exons (Fig. 1B).

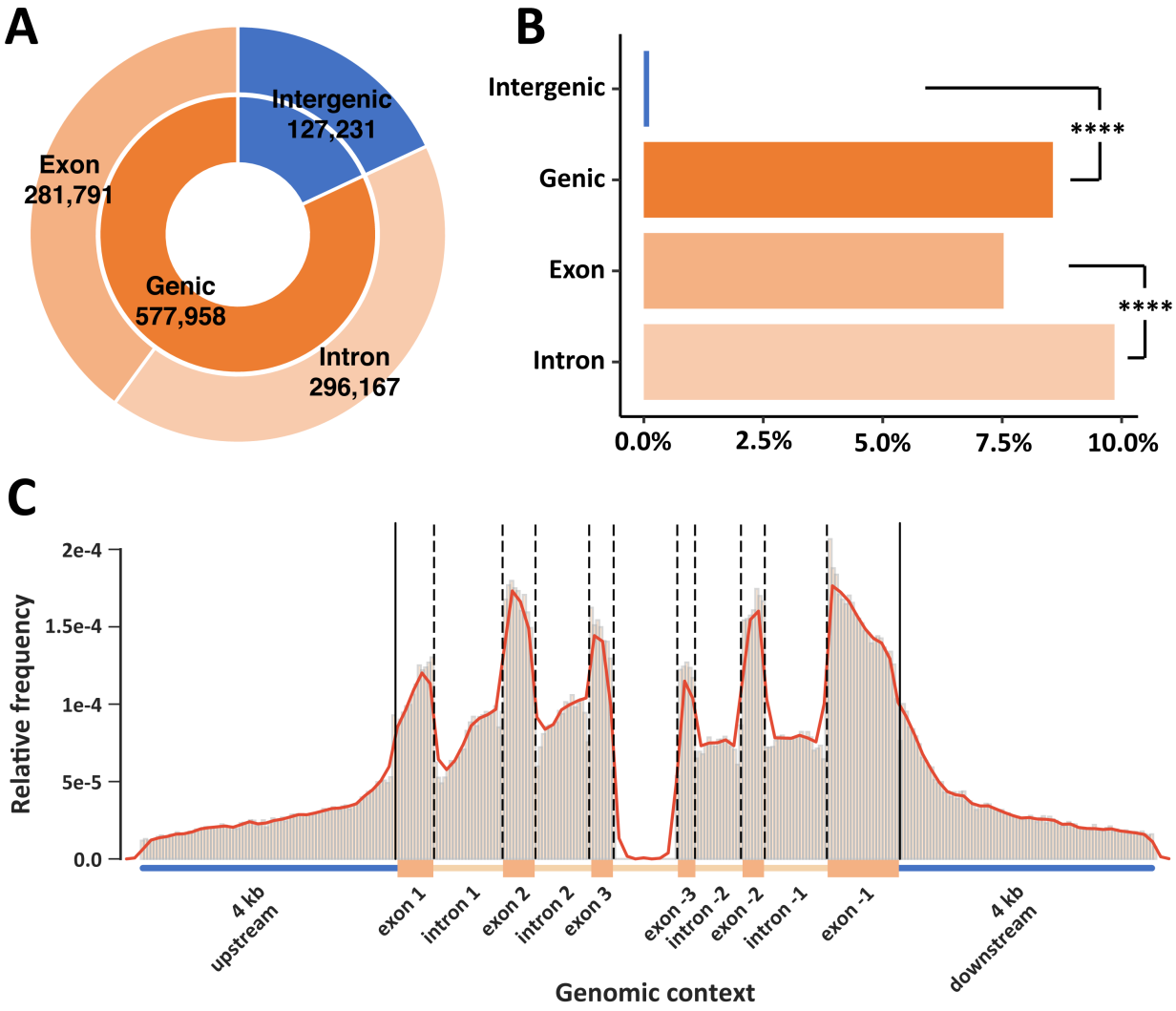


Fig. 1. DNA methylation landscape

(**A**) Distribution of CpG across intergenic (18%), genic (82%), intronic (42%) and exonic (40%) regions in the *Aiptasia* sp. genome. (**B**) Normalized percentage of methylated CpGs in different regions. Chi-squared test shows significant differences between intergenic and genic regions, and between exons and intron (****p<0.0001). (**C**) Relative frequencies of methylated positions across a normalized gene model.

Methylated genes are marked by H3K36me3

Analysis of methylation patterns (see above) within gene bodies showed rapidly increasing methylation levels after the transcription start site (TSS) that are maintained before slowly decreasing towards the transcription termination site (TTS) (Fig. S3A). Interestingly, we found that gene body methylation in *Aiptasia* is positively correlated with expression (Fig. 2A),

93 suggesting that DNA methylation either increases the expression of genes or that DNA
94 methylation is established as a consequence of transcription whereby increased expression
95 results in increasing methylation levels. The latter interpretation would be in line with recent
96 findings in mouse embryonic stem cells [13], which demonstrated that gene body methylation is
97 established and maintained as a result of active transcription by RNA polymerase II (Pol II) and
98 recruitment of the histone modifying protein SetD2 that trimethylates histone H3 at lysine 36
99 (H3K36me3). This histone mark is specifically bound via the PWWP domain present in the
100 DNA methyltransferase Dnmt3b, which in turn methylates the surrounding DNA accordingly,
101 resulting in the inhibition of transcription initiation from cryptic promoters within the gene body
102 and thus a significant reduction of spurious transcription.

103 Analysis of the *Aiptasia* gene set identified a DNMT3 gene (AIPGENE24404) that also
104 encodes a PWWP domain as reported for the mouse homolog. In order to test if the previously
105 described mechanism is conserved in *Aiptasia*, we performed a ChIP-Seq experiment using a
106 validated antibody against H3K36me3 (Fig. S4). As predicted, our analysis confirmed a
107 significantly higher association of H3K36me3 with methylated genes ($p = 2.48 \times 10^{-20}$ for highly
108 methylated genes and all methylated genes, Fig. 2B and C). We then analyzed if methylated
109 genes also exhibited significantly lower levels of spurious transcription in *Aiptasia*. Analysis of
110 transcriptional profiles of methylated and unmethylated genes indeed showed significantly lower
111 levels of spurious transcription along the gene body of methylated genes ($p < 2 \times 10^{-6}$, Fig. 3A).

112 A dampening effect of DNA methylation on transcription was also observed with regard
113 to transcriptional noise similar to findings in the coral *Stylophora pistillata* [18]. Regression
114 analysis of median methylation levels and the coefficient of transcriptional variation of genes
115 showed that, given the same expression level, methylated genes always exhibited lower levels of
116 transcriptional variation (Fig. 3B).

117
118

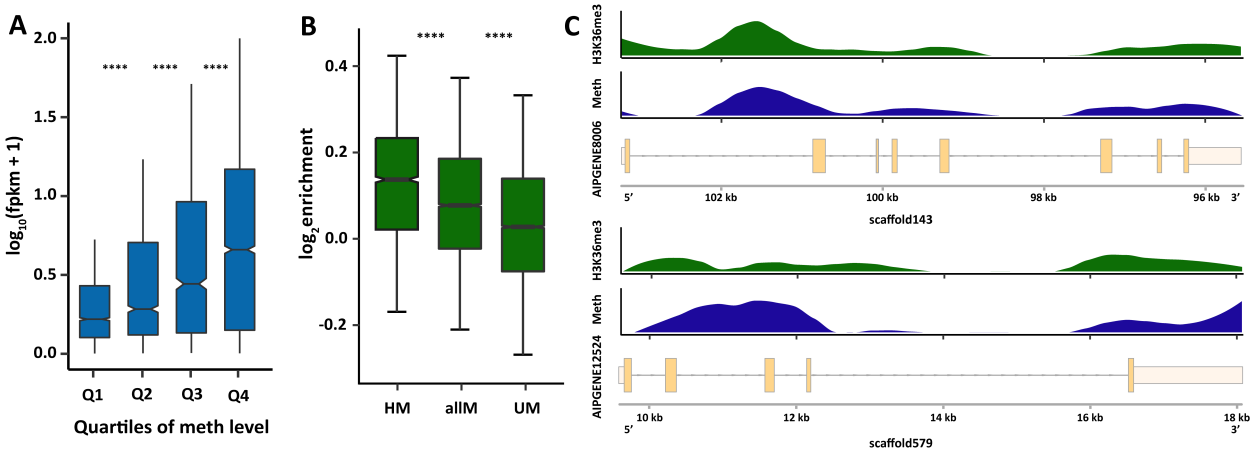


Fig. 2. DNA methylation is associated with higher expression

(**A**) Gene expression is positively correlated with median methylation levels, *t*-test *p* values are 7.65×10^{-21}, 3.75×10^{-14} and 1.75×10^{-13} for the first quartile (Q1) and the second quartile (Q2) of methylation levels, Q2 and Q3, and Q3 and Q4, respectively. (**B**) ChIP-Seq analysis of H3K36me3 signals show significant enrichment in methylated genes (*t*-test *p* values: 2.48×10^{-20} for highly methylated genes (HM) and all methylated genes (allM), and 1.06×10^{-72} for unmethylated genes (UM) and allM). Highly methylated genes show the strongest enrichment with H3K36me3 followed by all methylated genes. In contrast unmethylated genes show only weak enrichment of H3K36me3 over input controls. (**C**) Distribution of H3K36me3 enrichment and DNA methylation levels across two exemplary gene models. H3K36me3 and DNA methylation show coinciding distribution patterns over genes.

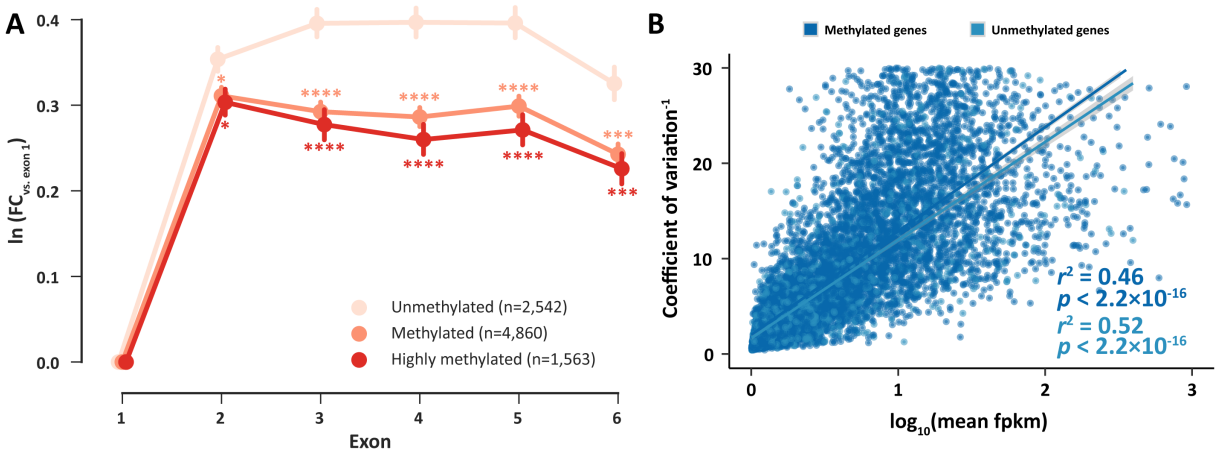


Fig. 3. DNA methylation regulates transcriptional homeostasis

(A) Spurious transcription in gene bodies is significantly lower in methylated and highly methylated genes. The *y*-axis shows the natural logarithm of the coverage fold change of exons 1–6 vs. exon 1. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$; ****: $p < 0.0001$. (B) There is a linear relationship between the inverse of transcriptional noise (CV^{-1}) and log expression level ($\log_{10}$fpkm). Given same expression level, methylated genes always show lower levels of

141 transcriptional noise. For methylated genes, n = 8,561, $r^2 = 0.46$, $p < 2.2 \times 10^{-16}$, for
142 unmethylated genes, n = 2,491, $r^2 = 0.52$, $p < 2.2 \times 10^{-16}$.
143
144
145 **DNA methylation regulates transcriptional homeostasis during symbiosis**

146 Based on our previous findings, we investigated if DNA methylation is also involved in
147 the regulation of symbiosis by identifying differentially methylated genes (DMGs) between
148 symbiotic and aposymbiotic *Aiptasia*. Comparison of DNA methylation patterns using Principal
149 Component Analysis (PCA) clearly separated symbiotic and aposymbiotic individuals by the
150 first principal component, which accounted for ~18% of the variance (Fig. 4A). This analysis
151 echoed the findings from a PCA analysis on gene expression where symbiosis state was
152 separated by the second principal component accounting for ~21% of the variance (Fig. 4B) [23]
153 and highlighted that specific changes in DNA methylation patterns occurred in response to
154 symbiosis.

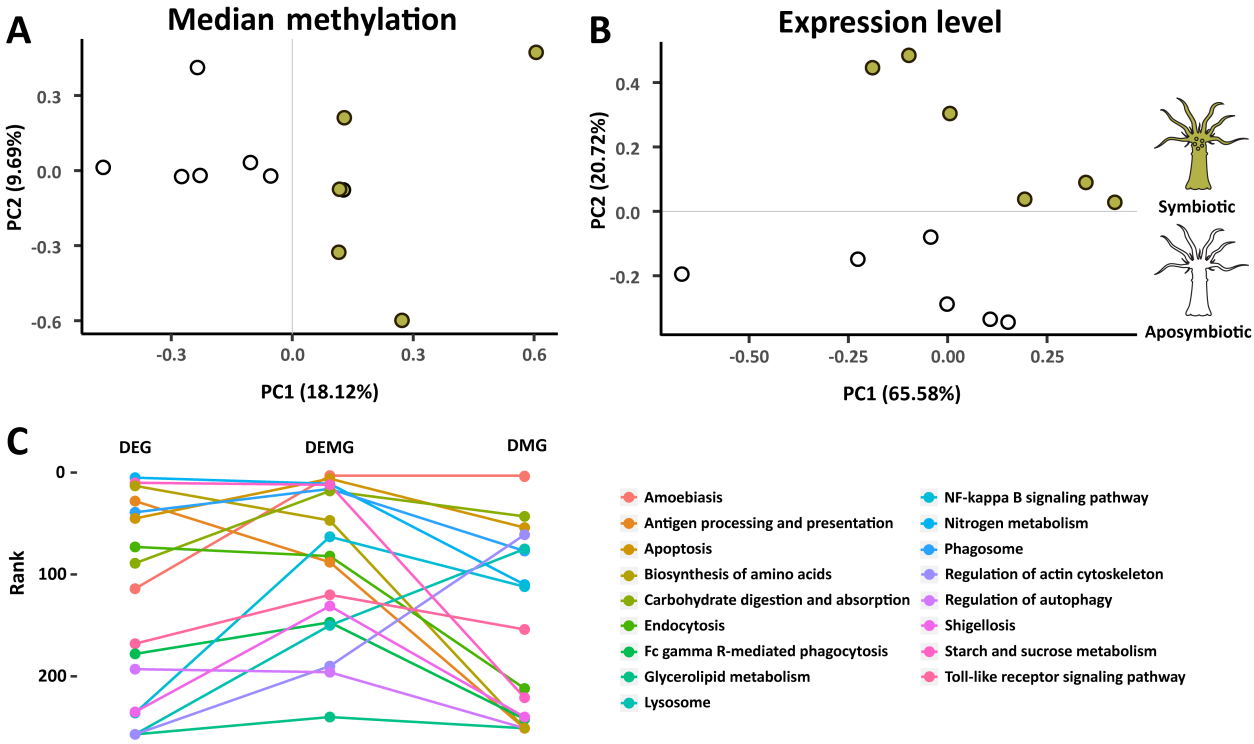
155 Subsequently we analyzed changes in DNA methylation and gene expression between
156 symbiotic and aposymbiotic *Aiptasia* to assess their correlation on potential biological functions
157 in symbiosis. We determined differentially methylated genes using a generalized linear model
158 from Foret *et al*. [24] that was modified to allow for replicate-aware analysis. This approach
159 identified 2,133 DMGs (FDR ≤ 0.05, Supplement Table S1) that specifically changed their
160 methylation status in response to symbiosis. To verify these results, we sequenced a subset of 14
161 DMGs using bisulfite PCRs. The results show a strong correlation ($r^2 = 0.815$ and $p = 1\times10^{-5}$ for
162 Apo, $r^2 = 0.922$ and $p = 5.2 \times 10^{-8}$ for Sym) to our WGBS and confirm the observed methylation
163 changes within these loci (Fig. S5).

164 Analysis of gene expression changes in the same 12 samples (i.e., 6 symbiotic and 6
165 aposymbiotic anemones) identified 1,278 differentially expressed genes (DEGs, FDR ≤ 0.05,
166 Supplement Table S2), of which 14 genes were subsequently confirmed via qPCR (Fig. S6).
167 However, analysis of the overlap between DMGs and DEGs showed only 103 genes that were
168 shared, suggesting that differentially expressed genes are not necessarily the same cohort of
169 genes that are differentially methylated. Functional enrichment analyses based on Gene Ontology
170 (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways of all DMGs and
171 DEGs identified several symbiosis relevant functions and pathways in both groups (Supplement

172 Table S3-S10).

173 Based on the finding that gene body DNA methylation is likely a consequence of active
174 transcription, we hypothesized that changes in DNA methylation patterns might also provide a
175 record of transcriptional activity over longer periods of time. We therefore tested if differential
176 methylation and acute transcriptional changes, obtained from our RNA-Seq analysis, provide a
177 complementary view of the processes underlying symbiosis. For this we compared enrichment of
178 symbiosis-specific pathways across the sets of 2,133 DMGs, 1,278 DEGs, and the combined set
179 of both DMGs and DEGs (3,308 DEMGs). Interestingly, we observed that the combined data set
180 (DEMGs) provided significantly lower p-values for previously identified symbiosis-related
181 pathways, including apoptosis, phagosome, nitrogen metabolism, and arginine biosynthesis,
182 among others (paired *t*-test: DEMG vs. DEG $p = 0.015$; DEMG vs. DMG $p = 0.009$) (Fig. 4C
183 and Supplement Table S11). This suggested that changes in methylation and transcription indeed
184 provide complementary information with regard to transcriptional adjustments in response to
185 symbiosis.

186



187

188 **Fig.4. PCA and KEGG pathway enrichment analysis**

189 (**A**, **B**) PCA (Principal Component Analysis) of gene expression and median methylation levels
190 of *Aiptasia* genes. Both gene expression and DNA methylation separate samples by symbiosis
191 state. (**C**) KEGG pathway enrichment analysis. The combined sets of differentially expressed
192 and differentially methylated genes (DEMG) provides significant lower *p* values (front ranks) for
193 symbiosis related pathways.

194

195 **DMGs and DEGs are involved in all stages of symbiosis**

196 Analysis of the combined DMG and DEG gene set showed significant enrichment of
197 genes involved in the distinct phases of symbiosis, that is symbiosis establishment, maintenance,
198 and breakdown [4]. Using an integrated pathway analysis based on known molecular interactions
199 between proteins we found that these processes are linked through several DMGs and/or DEGs
200 (Fig. S7 and Fig. S8, and see Supplement Table S11-S12 and Supplementary discussion).

201 For instance, we found numerous symbiosis-related receptors to respond to symbiosis on
202 a transcriptional and/or methylation level (Fig. S7), including C-type lectins (Fig. S7.3), Toll-like
203 receptors (Fig. S7.5), and the scavenger receptor SRB1 (Fig. S7.2) that has previously been
204 implicated in symbiont recognition in the sea anemone *Anthopleura elegantissima* [25]. Following
205 symbiont recognition, we also found several known engulfment and sorting-related genes to
206 change in methylation and/or expression such as Rab5 (Fig. S7.10), sorting nexin (Fig. S7.17),
207 Rac1 (Fig. S7.6), the lysosomal-associated membrane protein 1/2 (Fig. S7.22), and many genes
208 related to the cytoskeleton and movement (Fig. S7.33-39).

209 As expected in a metabolic symbiosis [4, 26] we also identified a large number of genes involved in
210 nutrient exchange. These included genes involved in the provision of inorganic carbon in the
211 form of CO_2 or bicarbonate (HCO_3^-) to fuel symbiont driven photosynthesis [27] (Fig. S8.1) as well
212 as genes involved in the exchange of fixed carbon in the form of lipids (Fig. S8.11), sugars and
213 amino acids (Fig. S8.10, S8.4) [28]. Concordantly, we also found that genes involved in nitrogen
214 acquisition, such as ammonium transporter (Fig. S8.2) and genes involved in glutamate
215 metabolism (Fig. S8.5-7), respond to symbiosis.

216 Finally, our analysis also highlighted genes putatively involved in the expulsion or
217 degradation of symbionts in response to environmental stress or as a means to control symbiont
218 densities. Autophagy is of interest in this regard because it links to other membrane trafficking
219 pathways and to apoptosis, and evidence suggests that autophagy also plays a role in removal of
220 symbionts during bleaching [29, 30]. Intracellular degradation of the symbiont is a result of
221 reengagement of the phagosomal maturation process or autophagic digestion of the symbiont by
222 the host cell [4], and we find both apoptosis- and autophagy-related genes to significantly change
223 in their methylation and/or expression level. These include the apoptosis genes RAC
224 serine/threonine-protein kinase (Fig. S7.25), Caspase 7 (Fig. S7.31), Caspase 8 (CASP8) (Fig.
225 S7.30), Nitric oxide synthase (Fig. S7.21) and Bcl2 (Fig. S7.27), as well as the Autophagy
226 proteins 5 and 10 (Fig. S7.14-15), among others.

227

228 **Discussion**

229 To assess the role of CpG methylation in the cnidarian-dinoflagellate symbiosis, we
230 undertook a global analysis of changes in the DNA methylomes and transcriptomes of
231 aposymbiotic and symbiotic *Aiptasia*. In contrast to their vertebrate counterparts, only 6.37% of
232 the CpGs in the *Aiptasia* genome are methylated, but their distribution is highly non-random ($p <$
233 3×10^{-300}) and that methylated CpGs are most highly localized in gene bodies (18.4% of CpGs).
234 Analysis of the distribution of the histone modification H3K36me3 further showed significant
235 enrichment of this epigenetic mark in methylated genes, echoing findings in mammals and
236 invertebrates [31]. More importantly, we find that methylated genes show significant reduction of
237 spurious transcription and transcriptional noise (Fig. 2B), suggesting that both the underlying
238 mechanism of epigenetic crosstalk as well as the biological function of DNA methylation is
239 evolutionary conserved throughout metazoans. These results highlight a tight interaction of
240 transcription and epigenetic mechanisms in optimizing gene expression in response to changing
241 transcriptional needs [13]. Further support for such a role is provided by the analysis of
242 differentially methylated and differentially expressed genes, which, when combined, showed
243 significant increase in enrichment of symbiosis relevant processes. This suggests that DNA
244 methylation and transcriptome analyses provide complementary views of cellular responses to

245 symbiosis whereby methylation changes provide a transcriptional record of longer-term
246 transcriptional adjustments.

247 While our analysis identified several genes, processes, and pathways previously reported
248 to be involved in symbiosis, it further highlights their intricate molecular interactions. Symbiosis
249 recognition, sorting and breakdown are interconnected processes, which is reflected in the
250 observed changes in methylation and expression. The molecular machinery involved in
251 phagosome maturation is tightly linked to autophagy and apoptosis enabling the host to respond
252 to potential pathogen invasion but also to degrade and remove dead or unsuitable symbionts.
253 This is strongly supported by immunofluorescence examinations of *Aiptasia pulchella*
254 gastrodermal cell macerates, showing that Rab5 appears around healthy, newly ingested and
255 already established *Symbiodinium*, but is replaced by Rab7 in heat-killed or DCMU-treated
256 newly ingested *Symbiodinium*. Conversely, Rab7 is absent from untreated newly infected or
257 already-established *Symbiodinium* [32, 33].

258 Rab5 is also required for the exosomal release of CD63 [34], which mediates the
259 endocytotic sorting process and transport to lysosomes [35]. This process is further regulated by
260 Rac1 [36] in conjunction with sorting nexin and the GTPase Rho, all of which were also identified
261 in our analyses. The sorting of phagocytosed *Symbiodinium* is critical to symbiosis establishment
262 as *Symbiodinium* is phagocytosed at the apical end and transported to the base of the cell, where
263 they are protected from digestion. In contrast, *Symbiodinium* staying at the apical end of the cell
264 are degraded [37].

265 Similar to the processes of symbiosis initiation and breakdown, we also found significant
266 enrichment of genes involved in nutrient exchange and many of these transporters have
267 previously been implicated in symbiosis maintenance [4, 38]. Notably, this also included genes
268 involved in the transport and assimilation of ammonium. Nitrogen is a main limiting nutrient in
269 coral reefs [39, 40, 41], and the coral-dinoflagellates symbiosis has been proposed to increase the
270 efficiency of nitrogen utilization by both partners [42] whereby the underlying nature of this
271 mechanism is currently debated [43, 44].

272

Conclusions

This study provides the first analysis of the function and role of DNA methylation in a symbiotic anthozoan. Our results show that the epigenetic crosstalk between the histone mark H3K36me3 and gene body methylation is conserved in cnidarians and reveal a role of gene body methylation in reducing of spurious transcription and transcriptional noise. Furthermore, we show that changes in DNA methylation patterns are specific to symbiosis and imply a functional in the establishment, maintenance, and breakdown of this important symbiotic association. Our findings therefore provide evidence for a role of DNA methylation as epigenetic mechanism involved in the maintenance of transcriptional homeostasis during the cnidarian-dinoflagellate symbiosis. The premise that epigenetic mechanisms play a role in organismal acclimation warrants future experiments targeted to investigate if DNA methylation could also contribute to resilience through the epigenetic adjustment of transcription in response to environmental stress in *Aiptasia* and corals.

Author contributions

M.A. conceived and coordinated the project. Y.L., G.C., M.J.C. and N.Z. performed experiments. M.A., C.R.V. and Y.J.L. provided tools and/or data. C.T.M. constructed libraries for whole genome bisulfite sequencing, ChIP-Seq and RNA-Seq. Y.L., Y.J.L. and M.A. analyzed expression, methylation and ChIP-Seq data. M.A. and Y.L. wrote the manuscript with input from Y.J.L. and C.R.V. All authors read and approved the final manuscript.

Acknowledgements

Research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST). We thank Prof. John Pringle for the provision of the initial *Aiptasia* CC7 and *Symbiodinium* SSB01 cultures. We thank Sebastian Baumgarten for his help with establishing the Aiptasia cultures and critical reading of the manuscript.

References and Notes:

1. Spalding MD, Grenfell AM. New estimates of global and regional coral reef areas. *Coral Reefs* **16**, 225-230 (1997).

2. Davidson MG. Protecting coral reefs: The principal national and international legal instruments. *The Harvard Environmental Law Review* **26**, 48 (2002).

3. Sylvain JC. How to Protect A Coral Reef: The Public Trust Doctrine and the Law of the Sea. *Sustainable Development Law & Policy* **7**, 4 (2006).

4. Davy SK, Allemand D, Weis VM. Cell biology of cnidarian-dinoflagellate symbiosis. *Microbiology and molecular biology reviews : MMBR* **76**, 229-261 (2012).

5. Meyer E, Weis VM. Study of cnidarian-algal symbiosis in the "omics" age. *The Biological bulletin* **223**, 44-65 (2012).

6. Lämke J, Bäurle I. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. *Genome Biology* **18**, 124 (2017).

7. Rando OJ, Verstrepen KJ. Timescales of Genetic and Epigenetic Inheritance. *Cell* **128**, 655-668 (2007).

8. He X-J, Chen T, Zhu J-K. Regulation and function of DNA methylation in plants and animals. *Cell Res* **21**, 442-465 (2011).

9. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484-492 (2012).

10. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet* **9**, 465-476 (2008).

11. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21 (2002).

12. Richards EJ. Inherited epigenetic variation - revisiting soft inheritance. *Nat Rev Genet* **7**, 395-401 (2006).

13. Neri F*, et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72-77 (2017).

14. Torda G*, et al.* Rapid adaptive responses to climate change in corals. *Nature Clim Change* **7**, 627-636 (2017).

15. Baumgarten S*, et al.* The genome of Aiptasia, a sea anemone model for coral symbiosis. *Proc Natl Acad Sci U S A* **112**, 11893-11898 (2015).

16. Voolstra CR. A journey into the wild of the cnidarian model system Aiptasia and its symbionts. *Molecular Ecology* **22**, 4366-4368 (2013).

17. Tucker KL. Methylated cytosine and the brain: a new base for neuroscience. *Neuron* **30**, 649-652 (2001).

18. Liew YJ*, et al.* Epigenome-associated phenotypic acclimatization to ocean acidification in a reef-building coral. *bioRxiv*, (2017).

19. Dixon GB, Bay LK, Matz MV. Patterns of gene body methylation predict coral fitness in new environments. *bioRxiv*, (2017).

20. Wang X, *et al.* Function and Evolution of DNA Methylation in *Nasonia vitripennis*. *PLoS Genet* **9**, e1003872 (2013).

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

22. Krzywinski M, *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645 (2009).

23. Venables WN, Ripley BD. Exploratory Multivariate Analysis. In: *Modern Applied Statistics with S* (ed^(eds). 4 edn. Springer-Verlag New York (2002).

24. Foret S, *et al.* DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proc Natl Acad Sci U S A* **109**, 4968-4973 (2012).

25. Rodriguez-Lanetty M, Phillips WS, Weis VM. Transcriptome analysis of a cnidarian-dinoflagellate mutualism reveals complex modulation of host gene expression. *BMC Genomics* **7**, 23 (2006).

26. Muscatine L. *The role of symbiotic algae in carbon and energy flux in reef corals*. Elsevier Science Publishing Company (1990).

27. Rädecker N, Pogoreutz C, Wild C, Voolstra CR. Stimulated Respiration and Net Photosynthesis in Cassiopeia sp. during Glucose Enrichment Suggests in hospite CO2 Limitation of Algal Endosymbionts. *Frontiers in Marine Science* **4**, (2017).

28. Oakley CA, Ameismeier MF, Peng L, Weis VM, Grossman AR, Davy SK. Symbiosis induces widespread changes in the proteome of the model cnidarian Aiptasia. *Cellular Microbiology* **18**, 1009-1023 (2016).

29. Dunn SR, Schnitzler CE, Weis VM. Apoptosis and autophagy as mechanisms of dinoflagellate symbiont release during cnidarian bleaching: every which way you lose. *Proceedings of the Royal Society B: Biological Sciences* **274**, 3079-3085 (2007).

30. Downs CA, *et al.* Symbiophagy as a cellular mechanism for coral bleaching. *Autophagy* **5 2**, 211-216 (2009).

31. Nanty L, *et al.* Comparative methylomics reveals gene-body H3K36me3 in Drosophila predicts DNA methylation and CpG landscapes in other invertebrates. *Genome Research* **21**, 1841-1850 (2011).

32. Chen MC, Cheng YM, Sung PJ, Kuo CE, Fang LS. Molecular identification of Rab7 (ApRab7) in Aiptasia pulchella and its exclusion from phagosomes harboring zooxanthellae. *Biochem Biophys Res Commun* **308**, 586-595 (2003).

33. Chen MC, Cheng YM, Hong MC, Fang LS. Molecular cloning of Rab5 (ApRab5) in Aiptasia pulchella and its retention in phagosomes harboring live zooxanthellae. *Biochem Biophys Res Commun* **324**, 1024-1033 (2004).

34. Baietti MF, *et al.* Syndecan-syntenin-ALIX regulates the biogenesis of exosomes. *Nat Cell Biol* **14**, 677-685 (2012).

35. Latysheva N, *et al.* Syntenin-1 is a new component of tetraspanin-enriched microdomains: mechanisms and consequences of the interaction of syntenin-1 with CD63. *Mol Cell Biol* **26**, 7707-7718 (2006).

36. Anitei M, *et al.* Protein complexes containing CYFIP/Sra/PIR121 coordinate Arf1 and Rac1 signalling during clathrin-AP-1-coated carrier biogenesis at the TGN. *Nat Cell Biol* **12**, 330-340 (2010).

37. McAuley PJ, Smith DC. The Green Hydra Symbiosis. V. Stages in the Intracellular Recognition of Algal Symbionts by Digestive Cells. *Proceedings of the Royal Society B: Biological Sciences* **216**, 7-23 (1982).

38. Lin S*, et al.* The Symbiodinium kawagutii genome illuminates dinoflagellate gene expression and coral symbiosis. *Science* **350**, 691-694 (2015).

39. Cook CB, Mullerparker G, Delia CF. Ammonium Enhancement of Dark Carbon Fixation and Nitrogen Limitation in Symbiotic Zooxanthellae - Effects of Feeding and Starvation of the Sea-Anemone Aiptasia-Pallida. *Limnology and Oceanography* **37**, 131-139 (1992).

40. Grover R, Maguer J-F, Allemand D, Ferrier-Pagès C. Uptake of dissolved free amino acids by the scleractinian coral *Stylophora pistillata*. *Journal of Experimental Biology* **211**, 860-865 (2008).

41. Radecker N, Pogoreutz C, Voolstra CR, Wiedenmann J, Wild C. Nitrogen cycling in corals: the key to understanding holobiont functioning? *Trends Microbiol* **23**, 490-497 (2015).

42. Wang J, Douglas AE. Nitrogen recycling or nitrogen conservation in an alga-invertebrate symbiosis? *J Exp Biol* **201 (Pt 16)**, 2445-2453 (1998).

43. Wang J, Douglas A. Nitrogen recycling or nitrogen conservation in an alga-invertebrate symbiosis? *The Journal of Experimental Biology* **201**, 2445-2453 (1998).

44. Aranda M*, et al.* Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports* **6**, 39734 (2016).

460 Supplementary Materials for

461

462 # DNA methylation regulates transcriptional homeostasis of algal endosymbiosis
463 # in the coral model *Aiptasia*

464 Yong Li[1], Yi Jin Liew[1], Guoxin Cui[1], Maha J. Cziesielski[1], Noura Zahran[1], Sebastian
465 Baumgarten[1], Craig T. Michell[1], Christian R. Voolstra[1], Manuel Aranda[1]

466 [1]Red Sea Research Center, King Abdullah University of Science and Technology (KAUST),
467 Thuwal, KSA

468

469 correspondence to: manuel.aranda@kaust.edu.sa

470

471

472 **Supplementary information**

473 ***Aiptasia* Culture and DNA/RNA Extraction**

474 *Aiptasia* sp. of the clonal strain CC7 [1] was used for this study. Anemones were maintained
475 in polycarbonate tubs with autoclaved seawater at ~25 °C on a 12 h: 12 h light: dark cycle at 20-
476 40 µmol m^{-2} s^{-1} light intensity and fed freshly hatched *Artemia nauplii* (brine-shrimp)
477 approximately twice per week. To generate aposymbiotic anemones, animals were subjected to
478 multiple cycles of cold-shock treatment and photosynthesis inhibitor diuron (Sigma-Aldrich, St.
479 Louis, MO) as described in Baumgarten (2015) [2]. Aposymbiotic *A. pallida* were kept
480 individually in 15 ml autoclaved seawater in 6-well plates and inspected by fluorescence
481 stereomicroscopy to confirm complete absence of dinoflagellates. Four separate batches of
482 aposymbiotic anemones were generated and maintained for a period of 1 year before beginning
483 of the experiment described below.

484 To generate symbiotic anemones, we separately infected aposymbiotic CC7 individuals
485 from each of the four aposymbiotic cultures described above using the compatible Clade B
486 *Symbiodinium* strain SSB01, originally isolated from *Aiptasia* strain H2 [2, 3]. The four batches of
487 symbiotic anemones were maintained for further 12 months under regular culture condition as

488 described above. The corresponding four aposymbiotic cultures were maintained in dark until 3
489 months before collection. For the last 3 months individuals from these aposymbiotic cultures
490 were subjected to the same culture conditions as the symbiotic cultures in order to monitor for
491 unwanted spontaneous re-establishment of symbiosis under light.
492 After the 12-month experimental period, we collected six biological replicates from each of
493 the four aposymbiotic and symbiotic cultures (one additional replicate was taken from batches 1
494 and 2 of each treatment) for subsequent DNA and RNA extraction as described below.
495 For each treatment, 6 biological replicates, weighing 20-28 mg (wet weight), were extracted
496 using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, Hilden, Germany). The
497 manufacturers protocol was followed with the omission of the optional step 4 (temporal storage
498 at 4°C if not performing DNA purification immediately). DNA concentrations were determined
499 using a Qbit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA). RNA
500 concentrations and integrity were determined using a Bioanalyzer Nano RNA Kit (Agilent
501 Technologies, Santa Clara, CA).

502

503 **RNA-Seq and Bisulfite Sequencing**

504 Directional mRNA libraries were produced using the NEBNext® Ultra™ Directional RNA
505 Library Prep Kit for Illumina® (NEB) following the manufacturer's protocol.
506 Bisulfite DNA libraries were prepared following a modified version of the NEBNext®
507 Ultra™ II DNA Library Prep Kit for Illumina® (NEB). Methylated TruSeq Illumina® adapters
508 (Illumina) were used during the adapter ligation step followed by bisulfite conversion with the
509 EpiTect Bisulfite kit (QIAGEN), with the following cycling conditions (95°C – 5 min, 60°C – 25
510 min, 95°C – 5 min, 60°C – 85 min, 95°C – 5 min, 60°C – 175 min, and 3 cycles of 95°C – 5 min,
511 60°C – 180 min, hold at 20°C ≤ 5 hours) (reference is Illumina Bisulfite).
512 The final libraries were enriched with the KAPA HiFi HotStart Uracil+ ReadyMix (2X)
513 (KAPA Biosystems) following the standard protocol for bisulfite-converted NGS library
514 amplification. Final libraries were quality checked using the Bioanalyzer DNA 1K chip
515 (Agilent), and quantified using Qubit 2.0 (Thermo Fisher Scientific), then pooled in equimolar
516 ratios and sequenced on the HiSeq2000.
517

518 **Identification of methylated CpGs**

519 Sequencing of the 12 libraries (2 conditions, 6 biological replicates each) resulted in 819
520 million read pairs from 8 lanes of the Illumina HiSeq2000 platform. Adapters were trimmed
521 from the raw sequences using cutadapt v1.8 [4]. Subsequently, trimmed reads were mapped to the
522 *Aiptasia* genome [2] using Bowtie2 v2.2.3 [5], and methylation calls was performed using Bismark
523 v0.13 [6].
524 Three filters were used to reduce false positives. Firstly, for each position with k methylated
525 reads mapping to it, the probability of it occurring through sequencing error (i.e. unmethylated
526 position appearing as methylated) was modelled using a binomial distribution $B(n, p)$, where n is
527 the coverage (methylated + unmethylated reads) and p the probability of sequencing error (set to
528 0.01). We kept positions with k methylated reads if $P(X \geq k) < 0.05$ (post-FDR correction).
529 Secondly, retained methylated positions had to have ≥ 1 methylated read in all six biological
530 replicates of at least one growth condition. Finally, median coverage of retained positions across
531 all 12 samples had to be ≥ 10.

532

533 **Assignment of genomic context to methylated cytosines**

534 Based on the gene annotation of the *Aiptasia* genome (GFF3 file) [2] and the positional
535 coordinates of the methylated cytosines produced by Bismark, we annotated every methylated
536 cytosine based on the genomic context, including whether the methylated position resides in a
537 genic or intergenic region, and the distances to the 5' and 3' end of each genomic feature
538 (gene/intergenic region/exon/intron).

539

540 **CpG bias**

541 Methylated cytosines are frequently spontaneously deaminated to uracil which can be
542 subsequently converted to thymine after DNA repair. As a result of this process, methylated
543 CpGs are expected to decrease in abundance over evolutionary time, and the ratio of observed to
544 expected CpGs (CpG O/E) has previously been used to predict putatively methylated and
545 unmethylated genes [7, 8]. CpG O/E of *Aiptasia* protein coding genes were calculated according to

546 J. Zeng *et al* [9].

547

548 **Identification of differentially methylated genes**

549 Using the methylation level of aposymbiotic genes as a control, generalized linear models
550 (GLMs) [10] were implemented in R [11] to identify genes that were differentially methylated in the
551 symbiotic treatment. The general formula used was:

552 glm(methylated, non_methylated ~ treatment * position, family="binomial")

553 where "methylated, non_methylated" was a two-column response variable denoting the
554 number of methylated and non-methylated reads at a particular position. For predictor variables,
555 "position" denoted relative position of the methylated site in the gene, while "treatment" denoted
556 symbiotic or aposymbiotic conditions. Data from individual replicates were entered separately to
557 assign equal weightage to each replicate, as pooling results in a disproportionate skew towards
558 the replicate with the highest coverage. Genes with < 5 methylated positions were filtered out to
559 reduce type I errors; and genes with FDR ≤ 0.05 were considered as differentially methylated
560 genes (DMGs).

561

562 **Identification of differentially expressed genes**

563 RNA-Seq generated 889 million raw read pairs from six lanes on the Illumina Hiseq2000
564 platform. Adaptors, primers and low quality bases were removed from the ends of raw reads
565 using Trimmomatic v0.33 (ILLUMINACLIP:TruSeq2-PE.fa:4:25:9 LEADING:28
566 TRAILING:28 SLIDINGWINDOW:4:30 MINLEN:50). The resulting trimmed reads were
567 mapped to the *Aiptasia* genome using HISAT v2.0.1 [12] and transcripts were assembled based on
568 the *Aiptasia* gene models (GFF3 file) using StringTie v1.2.2 [13]. Trinity
569 (align_and_estimate_abundance.pl – Bowtie2 v2.2.7/RSEM v1.2.22/edgeR v3.10.5) [5, 14, 15, 16, 17]
570 was run against the transcripts using trimmed reads for expression abundance estimation, then
571 differentially expressed genes (DEGs) were identified with FDR ≤ 0.05.

572

573 **Spurious transcription analysis**

574 Trimmed reads were mapped to the *Aiptasia* genome using HISAT2 v2.1.0 and mapping
575 coverage per position was extracted using BEDtools v2.17.0. Coverage per exon was calculated
576 and normalized across all 6 replicates (assuming every replicate had 1 million coverage in total),
577 then average coverage ratios of exon 2 to 6 versus exon 1 per gene were calculated to determine
578 spurious transcription levels.
579
580 **GO enrichment of DMGs and DEGs**

581 GO (Gene Ontology) [18] annotation was based on the previously published *Aiptasia* genome
582 [2]. Functional enrichment of DMGs and DEGs were carried out with topGO respectively [19] using
583 default settings. GO terms with $p \leq 0.05$ were considered significant, and the occurrence of at
584 least ≥ 5 times in the background set was additionally required for DMGs. Multiple testing
585 correction was not applied on the resulting *p*-values as the tests are considered to be non-
586 independent [19].

587
588 **KEGG enrichment of DMGs and DEGs**

589 KEGG (Kyoto Encyclopedia of Genes and Genomes)[20, 21] orthology (KO) annotation was
590 carried out by combining the KEGG annotations provided in the original *Aiptasia* genome
591 publications and a separate set of annotations based on the KAAS (KEGG Automatic Annotation
592 Server, http://www.genome.jp/tools/kaas/) (parameters: GHOSTZ, Eukaryotes, Bi-directional
593 Best Hit) [22]. A KEGG pathway enrichment analysis of both DMGs and DEGs was carried out
594 using Fisher exact test and pathways with $p \leq 0.05$ were considered significant.

595
596 **Validation of gene expression changes from RNA-Seq by qPCR**

597 Three randomly picked RNA libraries per treatment were used for qPCR validation of
598 RNA-Seq results. cDNA was synthesized using Invitrogen SuperScript III First-Strand Synthesis
599 SuperMix kit. A total of 14 genes were validated for differential expression using qPCR
600 (Supplement Table S13-S15). RPS7, RPL11 and NDH5 were used as internal reference
601 standards [23]. qPCR was carried out using Invitrogen Platinum SYBR Green qPCR SuperMix-

602 UDG kit on Applied Biosystems 7900HT Fast Real-Time PCR System. All protocols were
603 strictly followed.

604

605 **Validation of methylation changes using bisulfite PCR**

606 Three randomly picked DNA libraries per treatment were used for methylation validation.
607 Bisulfite conversion was done using the EZ-96 DNA Methylation-Gold Kit (Zymo Research). 18
608 genes were used to design primers, 14 of 18 obtained effective amplifications (Supplement Table
609 S16), then the fragments were enriched by PCR amplification using Promega PCR Master Mix.
610 Sequencing indexes were added to enriched fragments using Illumina Nextera XT Index Kit.
611 Enriched fragments were sequenced on the Illumina MiSeq platform. All protocols were strictly
612 followed. 1,870x data per replicate were obtained, methylated CpGs were identified using
613 Bismark as described above. The correlations between whole genome bisulfite conversion and
614 bisulfite PCR were calculated using generalized linear model.

615

616 **Chromatin Immunoprecipitation – ChIP**

617 We used the Zymo-Spin ChIP Kit to conduct histone bound chromatin extraction, with
618 minor adjustments to manufacturers protocol. Briefly, three biological replicates, each consisting
619 of two symbiotic anemones, were used for this experiment. Each anemone was first washed with
620 PBST (phosphate-buffered saline with 0.1% triton). Anemones were then fixed in 1X PBS with
621 1% formaldehyde for 15 minutes. To stop cross-linking reactions glycine was added to the
622 solution and left to rest for 10 more minutes. Following manufacturers protocol, we centrifuged
623 and washed whole anemones. We prepared the Nuclei Prep Buffer according to protocol and
624 crushed the two anemones of each replicate together using a douncer for homogenization.
625 Samples were then sonicated for 15 cycles on ice (15 sec ON, 30 sec cooling) to ensure
626 fragmentation to 200-500 bp. Thereafter the protocol was followed without further
627 modifications.

628 Immunoprecipitation was achieved using a target specific antibody to histone 3 lysine 36
629 tri-methylation (H3K36me3) (ab9050, Abcam). Corresponding input controls for each of the 3
630 replicates were generated as suggested by the manufacturer. DNA fragment quality and quantity

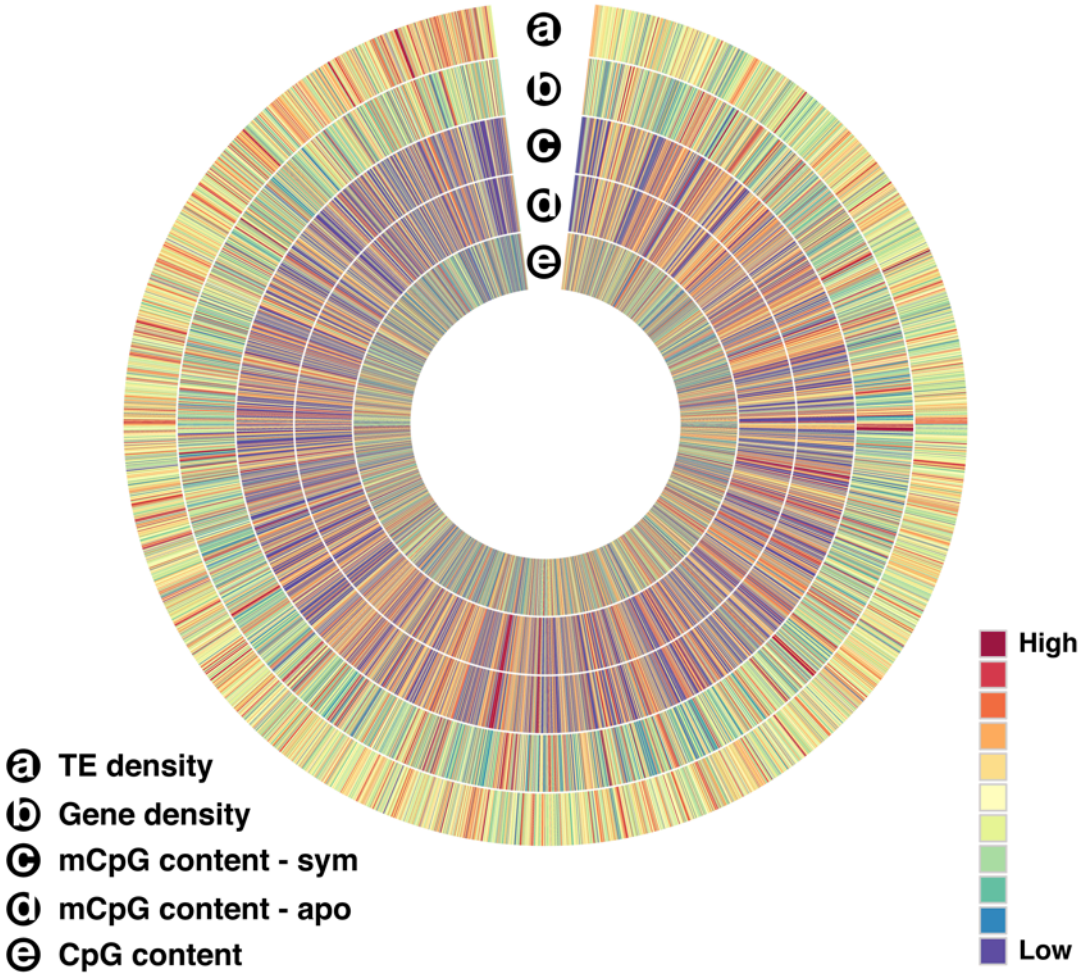
631 was confirmed using High Sensitivity DNA Reagents (Agilent Technologies, California, United
632 States) on a bioanalyzer, after which ChIP libraries were constructed using NEBNext® ChIP-Seq
633 Library Prep Master Mix Set (#E7645, New England Biolabs, Massachusetts, United States).
634 Sequencing resulted in 10M read pairs per replicate. These read pairs were trimmed using
635 Trimmomatic and mapped to the *Aiptasia* genome using HISAT2 as described above.
636 H3K36me3 enrichments were calculated as $\log_2$(average signal/average input control) for all
637 genes, unmethylated genes and highly methylated genes (methylation level > 70 and methylation
638 density > 40). P-values were calculated using t-test.
639

640 **Antibody affinity validation through Western blotting**

641 Total protein was extracted from a snap-frozen anemone crushed in 10% TCA
642 (Trichloroacetic acid). The homogenized sample was left to incubate overnight at -20 °C to allow
643 proteins to precipitate. The solution was centrifuged at 20,000g at 4 °C for 20 minutes to collect
644 suspended proteins. The pellet was then washed three times in 80% acetone and then spun down
645 again as previously. The final pellet was then air-dried for 10-15 minutes to remove residual
646 acetone. The final protein was suspended in Urea lysis buffer (7 M urea, 2 M thiourea, pH 7.5)
647 by vortexing for 2 hours.
648 Samples were then prepared for western blot by adding 4x sampling buffer (0.38 M Tris
649 base, 8% SDS, 4mM EDTA, 40% glycerol, 4mg/ml bromphenol blue) to a final concentration of
650 1X. After a 2 minutes incubation at 90°C samples were ready to be run gel at 10-12 mA. The gel
651 was transferred to a PVDF nitrocellulose membrane, rinsed with TBS buffer (150mM NaCl,
652 25mM Tris pH7.4, 0.1% Triton X-100) and blocked for 30 min at RT in TBS containing 5% fat-
653 free powder milk. The primary antibody was diluted in TBS/milk and incubated on an undulating
654 orbital shaker overnight at 4 °C. After three washes in TBS for 10 minutes each, the membrane
655 was again blocked in TBS/milk for 20 minutes at RT before proceeding with secondary antibody
656 staining. The horseradish peroxidase-linked-antibody (Anti-Rabbit IgG HRP conjugate W4011
657 and Anti-Mouse IgG HRP conjugate W4031, Promega. Wisconsin, United States) was diluted in
658 TBS/milk (1:10000) and incubated for 2 hours at RT. After final triplicate 10 minute washes in
659 TBS, membranes were developed.
660
661

662 **Supplementary Figures**

663



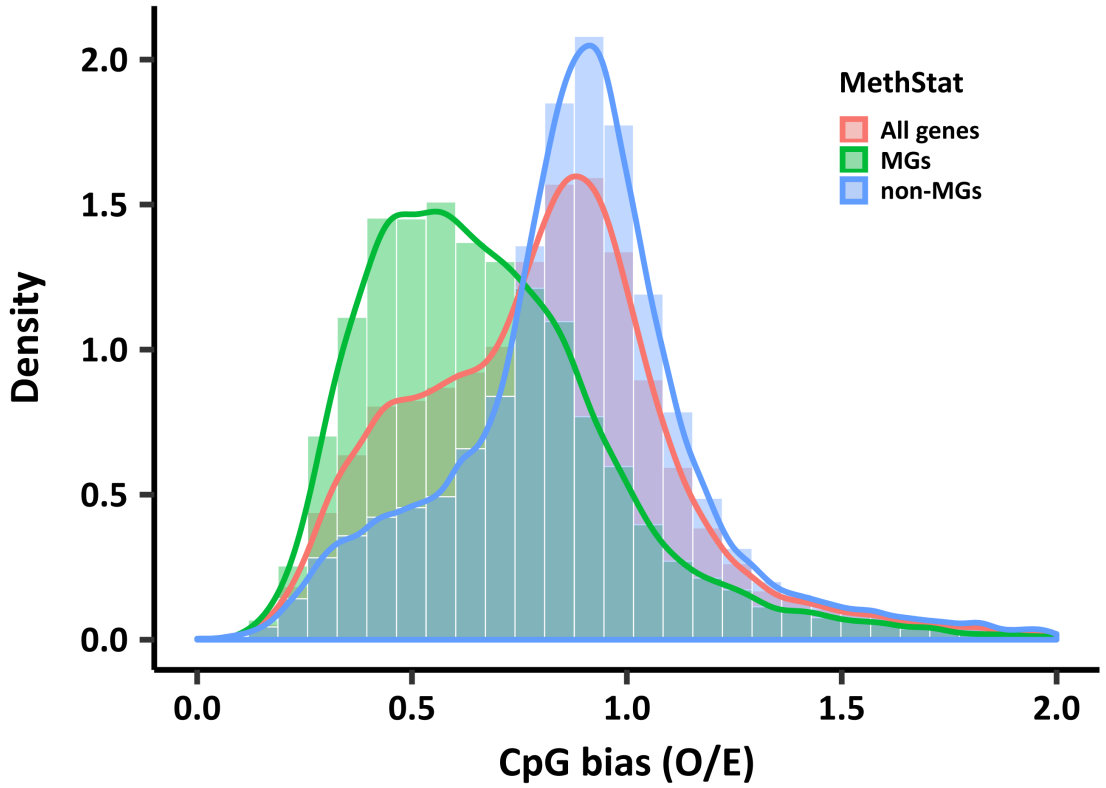
664

665 **Fig. S1. Circos visualization of different data at the genome-wide level**

666 (**a**) TE density. (**b**) Gene density. (**c**) Fraction of methylated CpGs in symbiotic treatment. (**d**)
667 Fraction of methylated CpGs in aposymbiotic treatment. (**e**) CpG content.

668

669



670

671 **Fig. S2. Methylated genes in *Aiptasia* have lower CpG O/E**

672 CpG distribution of methylated genes (represented by red curve) peaks at around 0.5, which is
673 lower than in unmethylated genes (represented by green curve) peaking at around 0.9. mC to T
674 conversion skews the CpG O/E distribution of all genes as expected (represented by blue curve),
675 but methylated and unmethylated genes still show a large overlap of their CpG O/E distributions.
676 These results indicate that gene body methylation cannot be accurately inferred from CpG O/E in
677 *Aiptasia*.

678

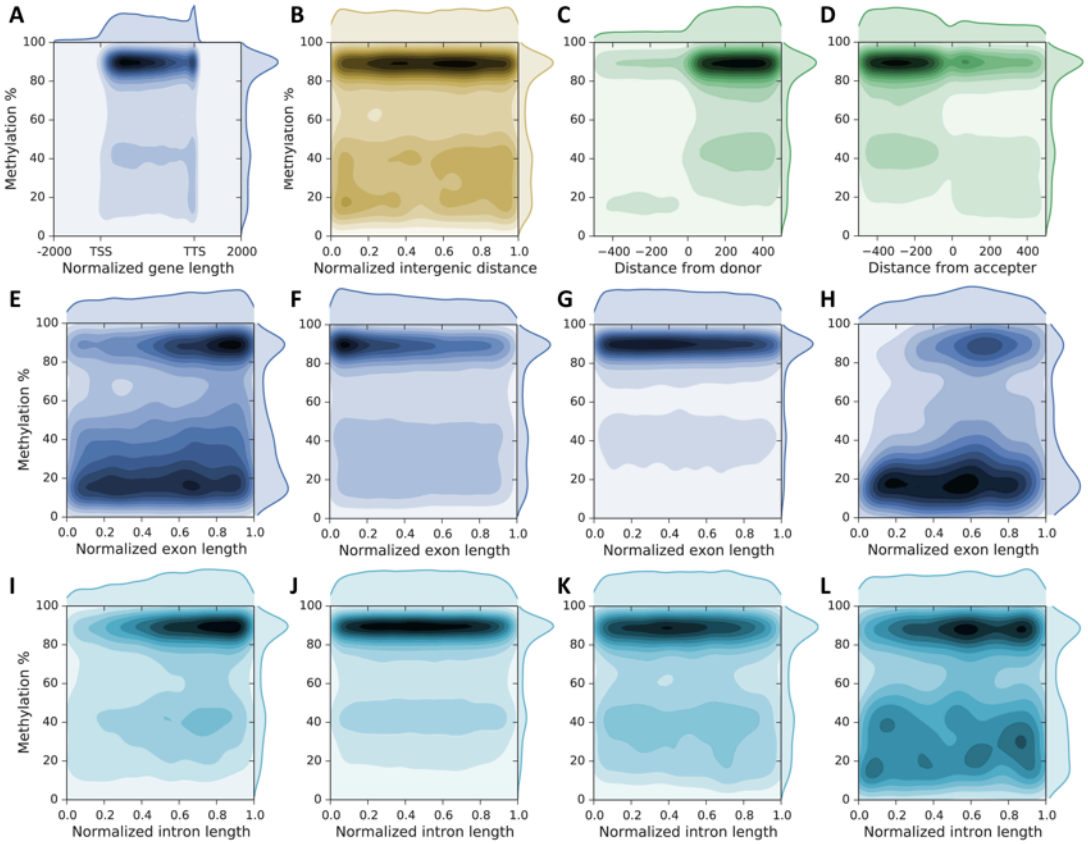
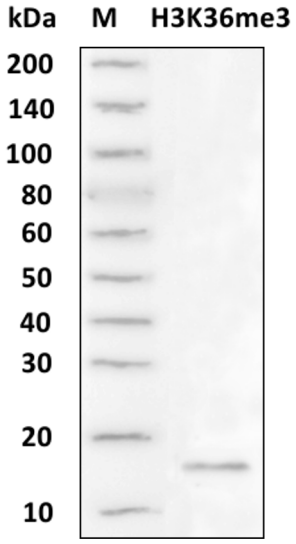


Fig. S3. Methylation patterns

(**A**) DNA methylation is mainly located in the proximal part of gene bodies with slightly decreasing levels towards the end. (**B**) Methylation pattern over intergenic regions. (**C**) Methylation pattern around splice donor sites show increasing levels immediately after donor sites. (**D**) Methylation pattern around acceptor sites show decreasing levels immediately after splice acceptor sites. (**E**) Methylation pattern over initial exons show increasing methylation levels (3,147 exons with 35,885 methylation sites were used). (**F**) Methylation pattern over internal exons show decreasing methylation levels (7,977 exons with 139,009 methylation sites were used). (**G**) Methylation pattern over terminal exons show decreasing methylation levels (7,905 exons with 102,162 methylation sites were used). (**H**) Methylation pattern over introns from single-exon genes follow a similar trend as observed for multi exon genes with increasing methylation levels in the proximal and decreasing levels in the posterior part of the exon (298 exons with 4,735 methylation sites were used). (**I**) Methylation pattern over initial introns show increasing methylation levels (3,381 introns with 39,262 methylation sites were used). (**J**) Methylation pattern over internal introns maintain stable methylation levels (7,371 introns with

695 211,950 methylation sites were used). (**K**) Methylation levels over terminal introns decrease
696 slightly (3,959 introns with 34,246 methylation sites were used). (**L**) Methylation levels over
697 introns from one-intron genes change gently with initial increase followed by a decrease (1,055
698 introns with 10,709 methylation sites).
699

700



701

702 **Fig. S4. Western blot**

703 Western blot result for antibody affinity validation, target band is 15kDa in size as expected from
704 molecular weight analysis.

705

706

707

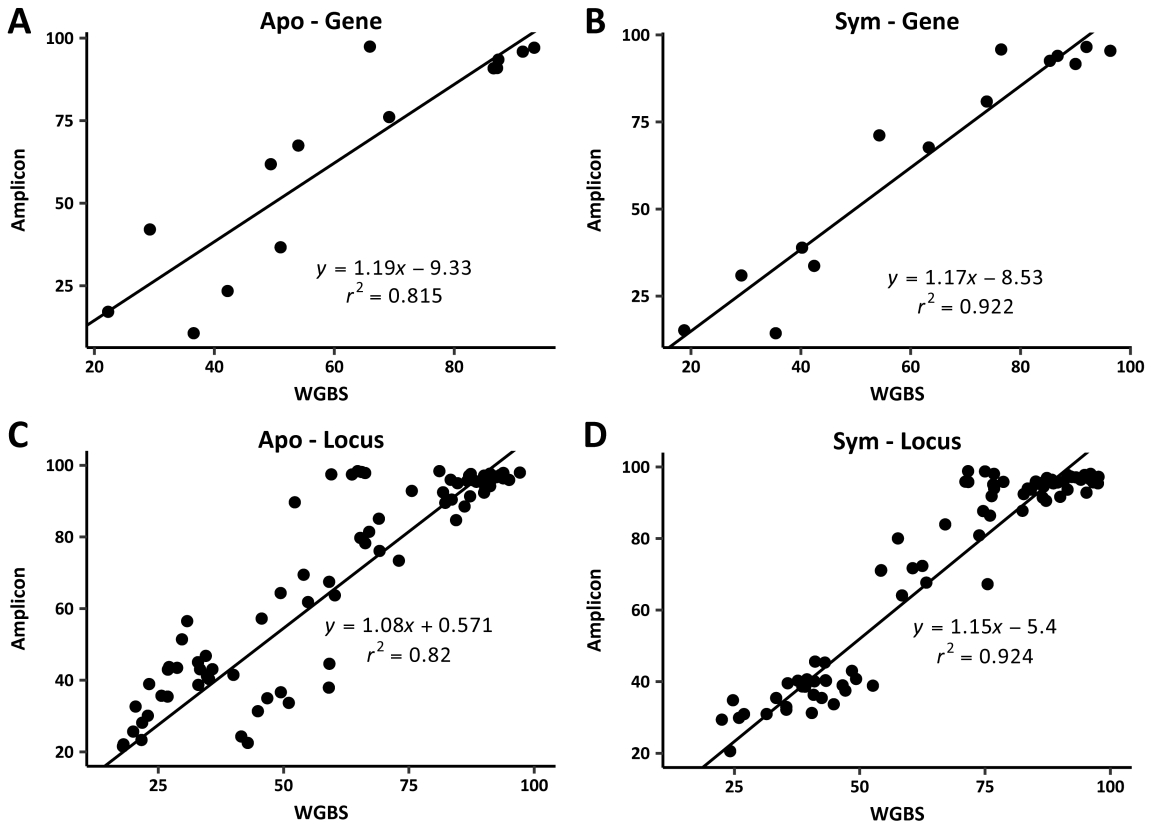
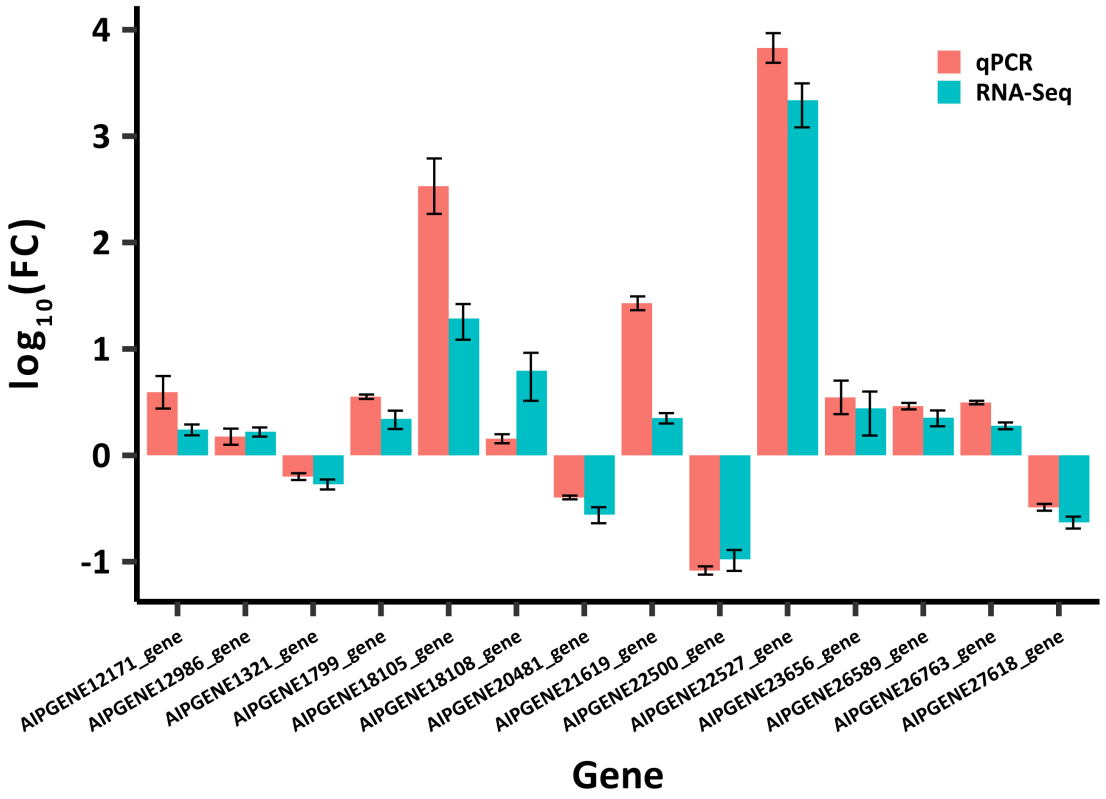


Fig. S5. Validation of methylation level

Validation of methylation level using bisulfite PCR on selected genes. (**A**, **B**) validation of methylation level on genes (median methylation levels of methylated CpGs were used to represent genes). (**C**, **D**) validation of methylation level on locus (methylated CpGs). WGBS: whole genome bisulfite sequencing; Amplicon: MiSeq sequencing results of bisulfite PCR amplicons on selected genes.

720

721



722

723 **Fig. S6. qPCR validation of gene expression levels**

724 Validation of gene expression changes using qPCR. Expression levels are shown as $\log_{10}$(fold

725 change). All genes show similar expression changes as determined by RNA-seq and q-PCR.

726

727

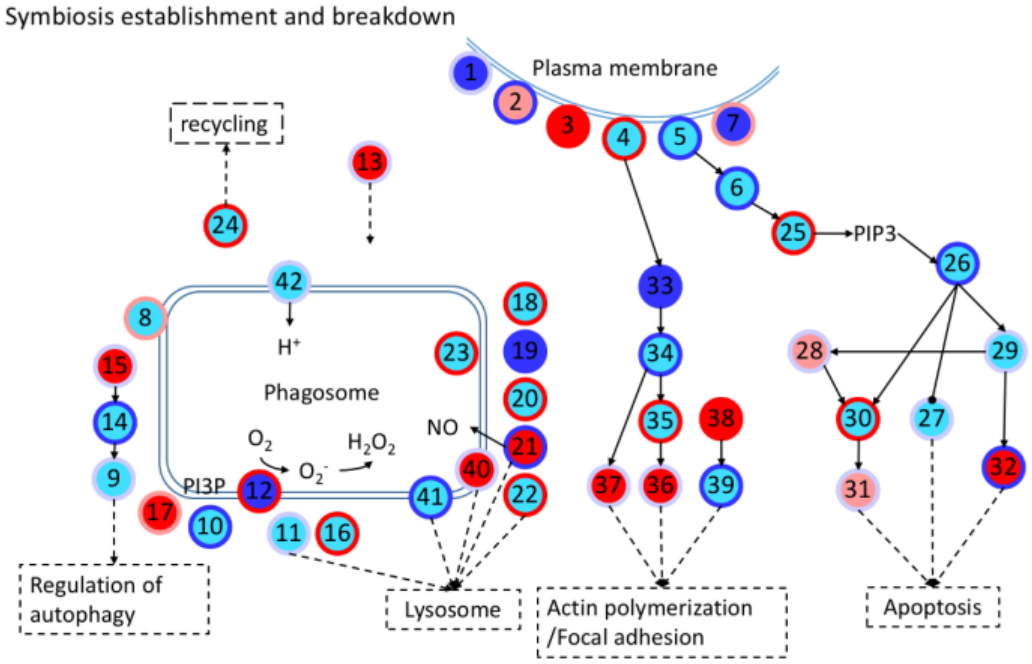


Fig. S7. Schematic diagram of symbiosis establishment and breakdown associated genes. Every node represents a category of genes, and generally has multiple corresponding genes. The inside colors of nodes represent the expression changes of corresponding genes, including non-DEGs (cyan), up-regulated (red), down-regulated (blue) and up- and down-regulated DEGs (light red). The colors of node edges represent the methylation level changes of corresponding genes, including non-DMGs (light blue), hypermethylated (red), hypomethylated (blue) and hyper- and hypo-methylated DMGs (light red). Numbers in circles denote genes/proteins as detailed below.

1. Complement receptor
2. Scavenger receptor
3. C-type lectin
4. Integrin
5. Toll-like receptor
6. Ras-related C3 botulinum toxin substrate 1 - rho family (RAC1)
7. Collagen
8. Vesicle-associated membrane protein (VAMP)
9. Autophagy-related protein 16 (ATG16)
10. Ras-related protein 5 (Rab5)
11. Ras-related protein 7 (Rab7)
12. NADPH oxidase (NOX)
13. Syntaxin 12
14. Autophagy-related protein 5 (ATG5)
15. Autophagy-related protein 10 (ATG10)
16. Programmed cell death 6-interacting protein
17. Sorting nexin (SNX)
18. Cytoplasmic dynein
19. Tubulin alpha chain (TUBA)
20. Tubulin beta chain (TUBB)
21. Nitric oxide synthase (NOS)
22. Lysosome-associated membrane glycoprotein/Cluster of differentiation (LAMP/CD)
23. Cathepsin L
24. Kinesin
25. Phosphatidylinositol 4,5-bisphosphate 3-kinase (PI3K)
26. RAC serine/threonine-protein kinase (AKT)
27. Bcl-2-antagonist of cell death (BAD)
28. TNF receptor-associated factor (TRAF)
29. Nuclear factor of kappa light polypeptide gene (NFKB)
30. Caspase 8 (CASP8)
31. Caspase 7 (CASP7)
32. Apoptosis regulator/Bcl-2 (BCL2)
33. Ras homolog (RHO)
34. Rho-associated protein kinase (ROCK)
35. Phosphatidylinositol 4-phosphate 5-kinase / Phosphatidylinositol 5-phosphate 4-kinase / Phosphatidylinositol 3-phosphate 5-kinase (PI4P5K/ PI5P4K/ PI3P5K)
36. Vinculin
37. Radixin
38. Profilin
39. Actin
40. CD63
41. Lysosomal-associated transmembrane protein
42. V-type proton ATPase

PI3P: phosphatidylinositol-3-phosphate
PIP3: Phosphatidylinositol (3,4,5)-trisphosphate

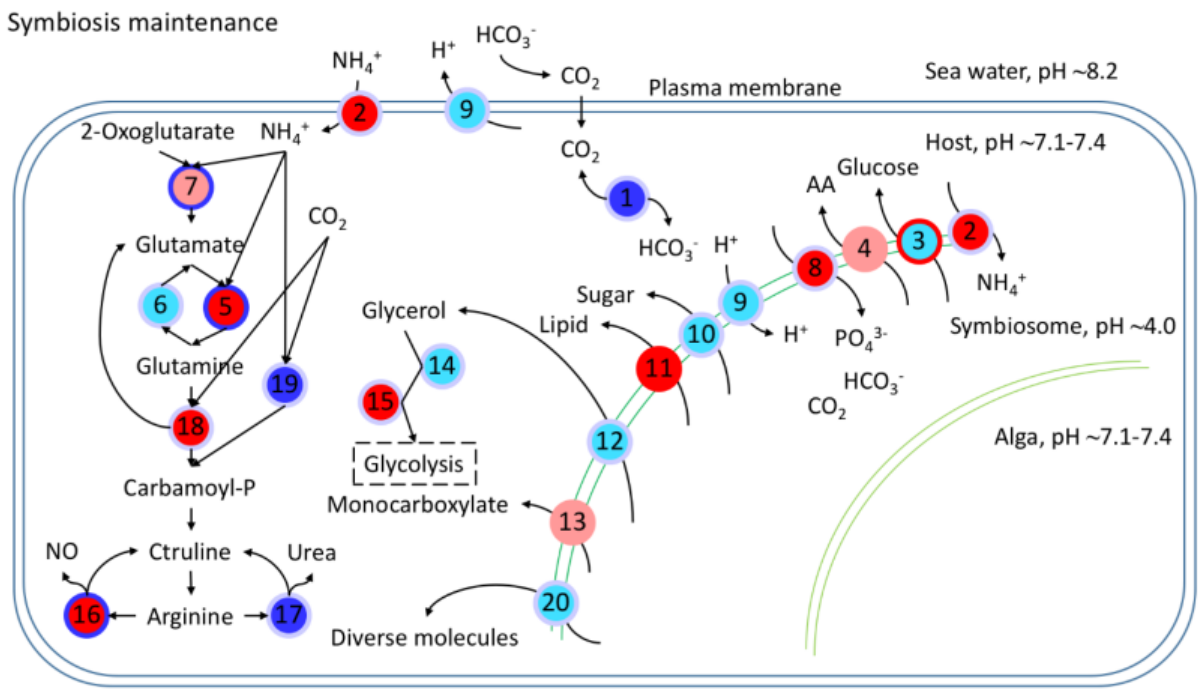


Fig. S8. Schematic diagram of symbiosis maintenance associated genes. Every node represents a category of genes, and generally has multiple corresponding genes. The inside colors of nodes represent the expression changes of corresponding genes, including non-DEGs (cyan), up-regulated (red), down-regulated (blue) and up- and down-regulated DEGs (light red). The colors of node edges represent the methylation level changes of corresponding genes, including non-DMGs (light blue), hypermethylated (red), hypomethylated (blue) and hyper- and hypo-methylated DMGs (light red).

1. Carbonic anhydrase (CA)
2. Ammonium transporter
3. Glucose transporter
4. Amino acid transporter
5. Glutamine synthetase (GS)
6. Glutamate synthase
7. Glutamate dehydrogenase (GDH)
8. Phosphate transporter
9. V-type proton ATPase
10. Sugar transporter
11. Lipid transfer protein
12. Aquaporin 3 (Glycerol transporter)
13. Monocarboxylate transporter
14. Alcohol dehydrogenase
15. Aldehyde dehydrogenase
16. Nitric oxide synthase (NOS)
17. Arginase
18. Carbamoyl-phosphate synthase / Aspartate carbamoyltransferase / Dihydroorotase (CAD)
19. Carbamoyl-phosphate synthase (ammonia) (CPS1)
20. ABC transporter

References

1. Sunagawa S*, et al.* Generation and analysis of transcriptomic resources for a model system on the rise: the sea anemone Aiptasia pallida and its dinoflagellate endosymbiont. *BMC Genomics* **10**, 258 (2009).

2. Baumgarten S*, et al.* The genome of Aiptasia, a sea anemone model for coral symbiosis. *Proc Natl Acad Sci U S A* **112**, 11893-11898 (2015).

3. Xiang T, Hambleton EA, DeNofrio JC, Pringle JR, Grossman AR. Isolation of clonal axenic strains of the symbiotic dinoflagellate Symbiodinium and their growth and host specificity(1). *J Phycol* **49**, 447-458 (2013).

4. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, (2011).

5. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Meth* **9**, 357-359 (2012).

6. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572 (2011).

7. Suzuki MM, Kerr AR, De Sousa D, Bird A. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res* **17**, 625-631 (2007).

8. Wang Y, Leung FC. GC content increased at CpG flanking positions of fish genes compared with sea squirt orthologs as a mechanism for reducing impact of DNA methylation. *PLoS One* **3**, e3612 (2008).

9. Zeng J, Yi SV. DNA Methylation and Genome Evolution in Honeybee: Gene Length, Expression, Functional Enrichment Covary with the Evolutionary Signature of DNA Methylation. *Genome Biology and Evolution* **2**, 770-780 (2010).

10. Hastie TJ, Pregibon D. Generalized linear models. In: *Statistical Models in S* (ed^(eds) (1992).

11. R Core Team. R: A Language and Environment for Statistical Computing. https://wwwr-projectorg/, (2016).

12. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Meth* **12**, 357-360 (2015).

13. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotech* **33**, 290-295 (2015).

14. Grabherr MG*, et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644-652 (2011).

15. Haas BJ*, et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512 (2013).

16. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 1-16 (2011).

17. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).

18. Ashburner M*, et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

19. Adrian Alexa JR. topGO: Enrichment Analysis for Gene Ontology. *R package version 2240*, (2016).

20. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, D457-462 (2016).

21. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

22. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35**, W182-185 (2007).

23. Lehnert EM, Mouchka ME, Burriesci MS, Gallo ND, Schwarz JA, Pringle JR. Extensive Differences in Gene Expression Between Symbiotic and Aposymbiotic Cnidarians. *G3: Genes|Genomes|Genetics* **4**, 277-295 (2014).