Role of gene body methylation in coral acclimatization and adaptation.

Groves Dixon[1], Yi Liao[1], Line K. Bay[2] and Mikhail V. Matz[1]

[1]Department of Integrative Biology, University of Texas at Austin, 2401 Speedway Ave, Austin, TX 78712, USA

[2]Australian Institute of Marine Science, PMB 3, Townsville, Queensland 4810, Australia

Corresponding Author:

Mikhail Matz

matz@utexas.edu

512-992-8086

One-sentence summary:

Genome-wide shifts in gene body methylation predict gene expression and fitness during acclimatization but do not contribute to epigenetic divergence between populations.

Abstract

Gene body methylation (GBM) has been hypothesized to modulate responses to environmental change, including transgenerational plasticity, but the evidence thus far has been lacking. Here we show that coral fragments reciprocally transplanted between two distant reefs respond with genome-wide increase or decrease in GBM disparity among genes. Surprisingly, this simple genome-wide adjustment predicted broad-scale gene expression changes and fragments' fitness in the new environment. This supports GBM's role in acclimatization, which may consist in modulating the expression balance between environmentally-responsive and housekeeping genes. At the same time, constitutive differences in GBM between populations did not align with plastic GBM changes upon transplantation and were mostly observed among F_{ST} outliers, indicating that they arose through genetic divergence rather than through transgenerational inheritance of acquired GBM states.

GBM is a taxonomically widespread epigenetic modification the function of which remains unclear (*1*, *2*). GBM is bimodally distributed among genes: it is high in ubiquitously expressed housekeeping genes and low in inducible genes (*2*, *3*). Only the detrimental effect of GBM is well understood: GBM causes hypermutability in protein-coding regions (*4*). Indeed, in humans, GBM is the primary driver of deleterious parent-age-related mutations (*5*). To merit pervasive evolutionary conservation, the fundamental biological function of GBM must be important enough to outweigh this risk (*1*). Among putative cellular functions of GBM,

41 suppression of intragenic transcription initiation is currently the most studied (*6*, *7*) but still
42 remains controversial (*8*). The ecological genomic literature has long expected that GBM (as well
43 as other epigenetic marks) respond to the environment and assist acclimatization by modulating
44 gene expression, possibly across generations (*9*–*11*). Still, thus far the only well-documented
45 case of GBM responding to the environment and resulting in phenotypic change is caste
46 determination in social insects depending on diet (*12*). Beyond this highly taxon-specific
47 example, GBM response to the environment has been reported in a recent paper on coral
48 acclimatization to acidic conditions (*13*); notably, GBM change did not specifically affect genes
49 expected to be involved in such acclimatization. One way to address the role of GBM in
50 adaptation would be to examine differences between populations, as has been recently done in
51 plants (*14*, *15*); however, this interpretation requires disentangling the effect of environment
52 from the effect of genetic divergence among populations (*16*). As for transgenerational
53 plasticity, for GBM to be involved in this process it must be both responsive to the environment
54 and heritable. This somewhat contradictory combination of properties has not yet been
55 demonstrated for GBM in any study system.

56 Here, we used a reciprocal transplantation framework (*17*) to test for the roles of GBM,
57 gene expression, and genetics in local adaptation and acclimatization in a reef-building coral
58 *Acropora millepora*. In corals, the possibility of clonal replication makes disentangling the
59 effects of genotype and environment very straightforward, solving a major problem of
60 ecological epigenetics (*16*). We compared corals from two reefs, Orpheus Island in the central
61 sector of the Great Barrier Reef (GBR) and Keppel Island in the south of GBR (Fig. 1 A). These

62 reefs are notably different in temperature (Fig. 1 B) as well as several other abiotic parameters
63 (*18*) and host slightly genetically divergent *A. millepora* populations (*19*, *20*). Fifteen coral
64 colonies from each reef were halved and reciprocally transplanted between the sites for three
65 winter months (Fig. 1 A, B). In this way, each individual genotype was simultaneously exposed to
66 two distinct reef conditions. After three months in the field, tissue samples were collected from
67 each fragment and assayed for genome-wide gene expression using TagSeq (*21*) and for DNA
68 methylation using MBD-seq (*22*). These data were analyzed in the context of several fitness
69 proxies and genetic distances between individuals (based on genetic polymorphisms detected in
70 the MBD-seq data). All experimental corals were dominated by *Symbiodinium* clade C, which
71 was established by analyzing *Symbiodinium*-matching MBD-seq reads.

72 For both Keppel- and Orpheus-origin corals fitness proxies were higher at Keppel,
73 possibly due to higher concentration of inorganic nutrients there (*18*, *19*). For Keppel-origin
74 corals the difference in performance was predominantly in weight gain (i.e., skeletal growth)
75 while for Orpheus-origin corals the difference was mostly in internal stores (lipids, proteins and
76 carbohydrates, Fig. 1 C). This indicates that (i) Keppel and Orpheus environments are indeed
77 different from corals' perspective, (ii) both populations are capable of plastic responses to these
78 differences, and (iii) there is also constitutive divergence in physiology between populations.
79

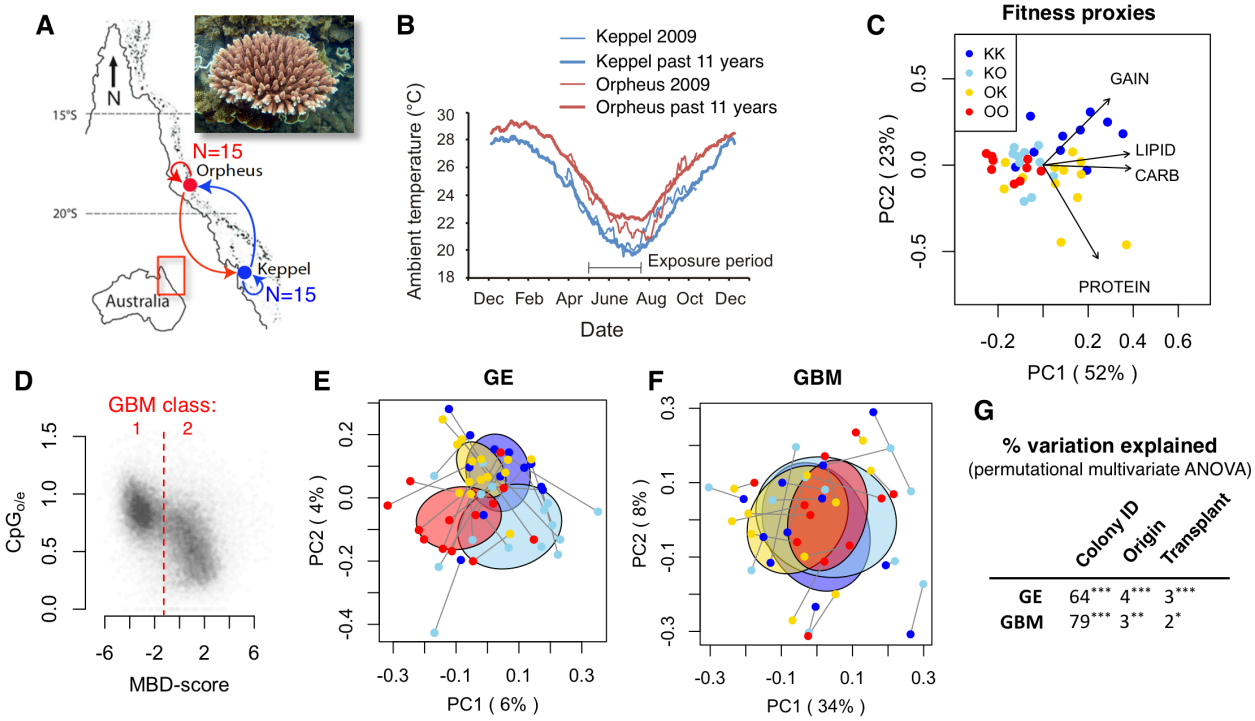


Figure 1. Overview of the experiment. A: Design of reciprocal transplantation. Inset – *Acropora millepora*. B: Temperature profiles at the transplantation sites, historical (thick lines) and during the year of experiment (thin lines). C: Principal component analysis of the four fitness proxies: weight gain, lipid, protein and carbohydrate content. In the sample names, the first letter is the origin location and the second letter is transplantation location. D: Scatter plot of the measure of depletion of CpG dinucleotides (ratio of observed vs. expected number of CpGs, $CpG_{O/E}$) against MBD-score (log_2 of fold-difference in read counts between MBD-captured and flow-through fraction) for each gene. The MBD-score threshold defining the two GBM classes is indicated. E and F: Principal coordinate analysis of gene expression (E) and gene body methylation (F). Lines connect fragments of the same original colony; the color scheme follows the legend on panel C. For GE, PC1 corresponds to the colony's origin and PC2 – to the transplantation site; no such partitioning is visible for GBM. G: Percentage of variation in GE and GBM

93 explained by colony identity (i.e., by similarity among clonal fragments, or broad-sense heritability),
94 colony's origin and transplantation site. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

95 To determine the characteristic level of GBM for each gene we used MBD-seq results for
96 12 samples (three per each experimental group, KK, KO, OK, and OO) for which both MBD-
97 captured and flow-through fraction were sequenced. Logarithm with the base 2 of each gene's
98 abundance in captured vs. flow-through fractions ("MBD-score" (*22*)) exhibited the expected
99 bimodal distribution (Fig. S1 A) and showed strong correlation (Fig. 1 D) with the established
100 proxy of historical methylation level, the ratio of observed to expected numbers of CpG
101 dinucleotides ("$CpG_{O/E}$" (*3*)). In addition to this genome-wide validation of our methylation data,
102 we validated GBM results by bisulfite-sequencing 13 amplicons representing different MBD-
103 score classes (Fig. S1 B, C). Genes were assigned to either low-methylated class (class 1) or high-
104 methylated class (class 2) based on the MBD-score threshold indicated on Fig. 1 D. Relative
105 quantification results based only on MBD-captured read data were nearly identical to those
106 obtained when using both captured and flow-through fractions (Fig. S2). For the remaining 32
107 samples, we sequenced only the MBD-captured fraction.

108 Multivariate analysis of gene expression (GE, Fig. 1 E) and GBM (Fig. 1 F) revealed that
109 GBM is more consistent across fragments of the same original colony than GE, resulting in the
110 estimate of broad-sense heritability of 0.79 compared to 0.64 for GE (Fig. 1G, Fig. S3 and S4). In
111 addition, significant effects of origin and transplantation site were observed for GE and GBM,
112 both more pronounced for GE (Fig. 1 E and G).

113 Surprisingly, the change in GBM in response to transplantation from Orpheus to Keppel
114 consisted mainly in genome-wide reduction of disparity between high- and low-methylated
115 genes: highly-methylated genes became less methylated and low methylated became more
116 methylated (Fig. 2 A). This change was mirrored by less pronounced but clearly reciprocal GBM
117 adjustment in Keppel corals transplanted to Orpheus (Fig. 2 B, D) and recapitulated the
118 difference between fragments planted in their native environment (Fig. 2 C). Both exons and
119 introns underwent this change (Fig. S5). Absolute changes in methylation included both positive
120 and negative shifts (Fig. S6 A, B), indicating that the change in relative GBM levels among genes
121 is not due to biased genome-wide methylation increase or decrease. Methylation levels of
122 intergenic regions and repeated elements remained relatively constant compared to genic
123 regions (Fig. S6 C, D).

124 GE response was also reciprocal between transplantation directions but demonstrated
125 much broader range of fold-changes compared to GBM response (Fig. 2 D). Remarkably, gene
126 expression changes correlated with low-and high-methylated gene classes and were in the
127 opposite direction relative to GBM changes (Fig. 2 F, G). Although at the gene level the
128 correlation between GE and GBM was weak (but still highly significant, Fig. S7), it was very
129 strong when comparing broader functional groups of genes (Gene Ontology categories, Fig. 2 H
130 and Fig. S8).

131 To characterize GBM disparity among genes in each sample, we computed GBM class
132 difference ("GBMcd") as difference in mean $\log_2$(GBM) between high- and low-methylated gene

classes (Fig. 2 I). This measure aligned nearly perfectly with the first principal component of GBM variation for the whole experiment, explaining 33% of total variation (Fig. 1 F and Fig. 2 J, which show the same PCA ordination but are colored differently). For comparison, the next principal component explains only 8% of GBM variation.

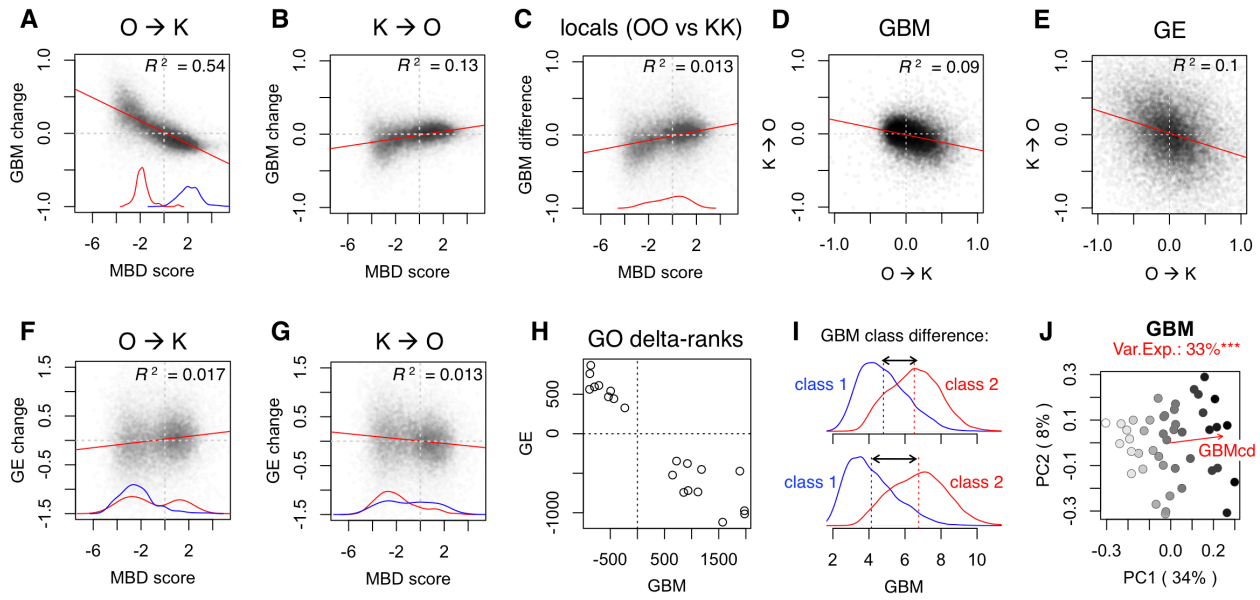
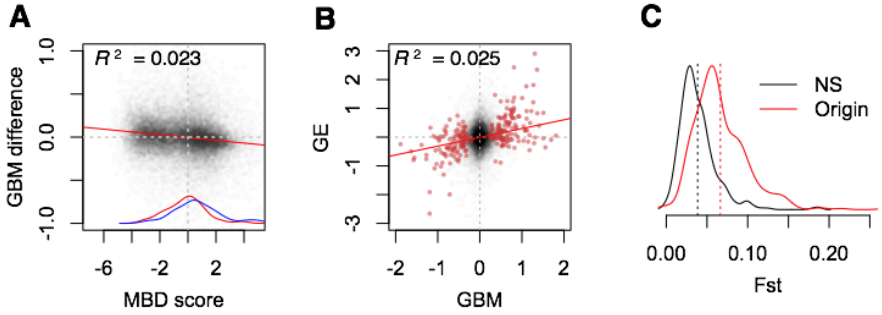


Figure 2. Changes in gene body methylation (GBM) and gene expression (GE) in response to transplantation ("away" relative to "home"). On all panels, GE and GBM measures are $\log_2$ fold-changes. A though C: Differences in GBM in fragments transplanted from Orpheus to Keppel (A), from Keppel to Orpheus (B), and fragments planted in their home environment (C) plotted against MBD-score for each gene. Red and blue curves near the x-axis are density plots of significantly up- and down-regulated genes (there were none for the K→O experiment). D and E: Correlation between GBM (D) and GE (E) changes in reciprocally transplanted fragments. F and G: Change in GE in fragments transplanted from Orpheus to Keppel (F) and from Keppel to Orpheus (G) plotted against MBD-score for each gene (compare to panels A and B). H: Plot of delta-ranks for 20 Gene Ontology categories that were significantly enriched with

147 either up-regulated genes (positive delta-rank) or down-regulated genes (negative delta-rank) in both GE
148 and GBM analyses. I: Calculation of GBM class difference (GBMcd), a proxy of genome-wide GBM
149 disparity among genes. Each panel shows two density plots of GBM levels for the two GBM gene classes
150 (see Fig. 1 D), with the mean for each class indicated by a dashed line. The top panel shows the sample
151 with the lowest GBMcd, bottom panel – the sample with the highest GBMcd. J: Principal coordinate
152 biplot (the same ordination as on Fig. 1 F) colored according to GBMcd. The purple vector shows the
153 direction of correlation of GBMcd with the ordination. All correlations with reported R^2 are significant at
154 the $p<0.0001$ level.

155 Our data do not support transgenerational inheritance of acquired GBM changes. If
156 plastic changes in GBM were transmitted across generations, one would expect constitutive
157 GBM differences between Orpheus- and Keppel-origin corals to parallel the GBM response
158 induced upon transplantation. Instead, constitutive differences between Orpheus and Keppel
159 (differences between all Orpheus-origin and all Keppel-origin fragments, irrespective of
160 transplantation site) were in the opposite direction compared to plastic changes (compare Fig. 3
161 A and Fig. 2 C), did not predominantly affect genes at the opposite ends of MBD-score spectrum
162 (compare density curves on Fig. 2 A and Fig. 3 A), and were positively rather than negatively
163 associated with differences in GE (Fig. 3 B). In addition, genes showing significant constitutive
164 differences in GBM also showed elevated F_{ST} (Fig. 3 C) unlike genes undergoing significant plastic
165 changes (Fig. S9). This indicates that the mechanism giving rise to constitutive GBM changes
166 between populations is unrelated to the mechanism of plasticity and is most likely genetic
167 divergence.

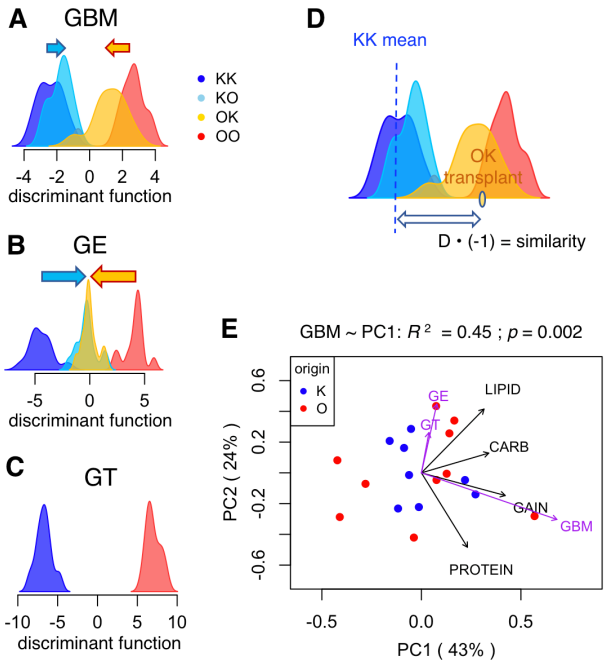


168

169 Figure 3. Constitutive $\log_2$ fold-differences in gene body methylation (GBM) and gene expression (GE)
170 between populations (Orpheus relative to Keppel). A: Difference in GBM plotted against MBD-score for
171 each gene. Red and blue curves by the x-axis are density plots of significantly up- and down-regulated
172 genes. B: Correlation between GBM and GE differences. Red points are genes significantly different in
173 GBM at 10% FDR level. This correlation is positive, while correlations between plastic GBM and GE
174 changes are negative (Fig. 2 H and Fig. S7). E: Density plot of between-population F_{ST} for genes showing
175 significant constitutive difference in GBM (red line) compared to 500 randomly chosen non-significant
176 genes (black line). Dashed vertical lines mark the mean of each dataset. All correlations with reported R^2
177 are significant at $p<0.0001$ level.

178 To see whether plastic GBM changes were related to acclimatization (rather than, for
179 example, overall stress in response to transplantation) we used Differential Analysis of Principal
180 Components (DAPC, Fig. 4 A-C) to test whether GBM similarity to the local population (Fig. 4 D)
181 predicted the transplanted fragment's fitness in the new environment. The resulting "GBM
182 similarity to locals" was highly correlated with the first principal component of measured fitness
183 proxies, the most strongly with the weight gain (Fig. 4 E). This correlation held both before and
184 after (as shown on Fig 4 E) controlling for overall higher fitness of Orpheus corals transplanted

185 to Keppel. This result indicates that a coral's GBM profile, despite being highly variable among
186 individuals (Fig 1 F, G), also reflects adaptation and acclimatization to local environmental
187 conditions. Interestingly, the same analysis performed for GE and genotypes (GT) revealed no
188 relationships with fitness (Fig. 4 E). One explanation of this could be the difference in time scales
189 on which the effects of GE and GT on fitness would be detectable: compared to the three-
190 month length of our experiment, GE might be linked to fitness on shorter time scales, and GT –
191 on longer time scales.



192

193 Figure 4. Relationship between gene body methylation (GBM), gene expression (GE), and genotype (GT)
194 with fitness proxies in corals transplanted into novel environment. A-C: Density plots of discriminant
195 function values for each variable. The functions were developed to distinguish between coral fragments
196 in their native environment (KK vs. OO) and applied to transplanted fragments (KO and OK) to quantify
197 their adaptive plasticity (ability to shift towards local values). Arrows indicate mean plasticity. D:

198 Calculation of “similarity to locals” for each transplanted fragment. E: Principal components biplot of
199 fitness proxies in fragments transplanted to another reef, corrected for the overall higher fitness of
200 Orpheus corals transplanted to Keppel. Fit of “similarity to locals” for GBM, GE and GT onto this
201 ordination is shown as purple vectors scaled by correlation coefficient. Only GBM fit is statistically
202 significant.

203 Association between plastic GBM changes and gene expression changes (Fig. 2 F - H)
204 prompts the question: which one is likely to be the leading cause? The magnitude of GBM
205 changes is much lower than changes in GE (Fig. 2 D, E), which suggests that GE might be the
206 primary driver. However, considering the simplicity of plastic GBM changes - genome-wide
207 exaggeration or mitigation of GBM disparity among genes - it is more parsimonious to assume
208 that this is the leading change driven by the environment, which then translates into broad-
209 scale functional shifts in gene expression (Fig. 2 H and Fig. S8). In this way, genome-wide
210 adjustment of gene activity could be achieved by modulating a single process that governs the
211 disparity in GBM among genes.

212 Our results suggest that GBM disparity controls the balance between expression of two
213 broad gene classes: the low-methylated environmentally-responsive genes and highly-
214 methylated housekeeping genes. In the higher-quality environment of the Keppel island, where
215 corals attain higher fitness (Fig. 1C), GBM disparity change is associated with suppression of the
216 environmentally responsive genes and up-regulation of housekeeping genes (Fig. 2 F and Fig.
217 S8). Conversely, at the low-quality Orpheus Island the balance shifts the other way, from
218 housekeeping to environmental responsiveness (Fig. 2G and Fig. S8). We propose that mediating

219 such coarse adjustment of physiology in response to the environment is the ecological function

220 of GBM.

221

222 **References** (Main paper)

223

224 1. D. Zilberman, An evolutionary case for functional gene body methylation in plants and

225 animals. *Genome Biol.* **18**, 87 (2017).

226 2. A. J. Bewick, R. J. Schmitz, Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.*

227 **36**, 103–110 (2017).

228 3. S. Sarda, J. Zeng, B. G. Hunt, S. V Yi, V. Soojin, The Evolution of Invertebrate Gene Body

229 Methylation. *Mol Biol Evol*. **29**, 1907–1916 (2012).

230 4. A. P. Bird, DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8**,

231 1499–1504 (1980).

232 5. L. C. Francioli *et al.*, Genome-wide patterns and properties of de novo mutations in

233 humans. *Nat. Genet.* **47**, 822–826 (2015).

234 6. A. K. Maunakea *et al.*, Conserved role of intragenic DNA methylation in regulating

235 alternative promoters. *Nature*. **466**, 253–7 (2010).

236 7. F. Neri *et al.*, Intragenic DNA methylation prevents spurious transcription initiation.

237 *Nature*. **543**, 72–77 (2017).

238 8. D. Jjingo *et al.*, On the presence and role of human gene-body DNA methylation.

239 *Oncotarget*. **3**, 462–474 (2012).

240 9. S. B. Roberts, M. R. Gavery, Is there a relationship between DNA methylation and

241 phenotypic plasticity in invertebrates? *Front. Physio.* **2**, 1–5 (2012).

242 10. G. E. Hofmann, Ecological Epigenetics in Marine Metazoans. *Front. Mar. Sci.* **4**, 4 (2017).

243 11. R. Feil, M. F. Fraga, Epigenetics and the environment: emerging patterns and implications.

244 *Nat. Rev. Genet.* **13**, 97 (2012).

245 12. H. Yan *et al.*, DNA Methylation in Social Insects: How Epigenetics Can Control Behavior

246 and Longevity. *Annu. Rev. Entomol.* **60**, 435–452 (2015).

247 13. Y. J. Liew *et al.*, Epigenome-associated phenotypic acclimatization to ocean acidification in
248 a reef-building coral. *Sci. Adv.* **4**, eaar8028 (2018).
249 14. P. F. Gugger, S. Fitz-Gibbon, M. Pellegrini, V. L. Sork, Species-wide patterns of DNA
250 methylation variation in Quercus lobata and their association with climate gradients. *Mol.*
251 *Ecol.* **25**, 1665–1680 (2016).
252 15. T. E. Keller, J. R. Lasky, S. V. Yi, The multivariate association between genomewide DNA
253 methylation and climate across the range of *Arabidopsis thaliana*. *Mol. Ecol.* **25**, 1823–
254 1837 (2016).
255 16. J. Hu, R. D. H. Barrett, Epigenetics in natural animal populations. *J. Evol. Biol.* (2017),
256 doi:10.1111/jeb.13130.
257 17. T. J. Kawecki, D. Ebert, Conceptual issues in local adaptation. *Ecol. Lett.* **7**, 1225–1241
258 (2004).
259 18. C. Lønborg *et al.*, "Marine Monitoring Program: Annual report for inshore water quality
260 monitoring 2014 to 2015" (Australian Institute of Marine Science and JCU TropWATER,
261 Townsville, 2016), (available at http://elibrary.gbrmpa.gov.au/jspui/handle/11017/3050).
262 19. M. V. Matz, E. A. Treml, G. V. Aglyamova, L. K. Bay, Potential and limits for rapid genetic
263 adaptation to warming in a Great Barrier Reef coral. *PLOS Genet.* **14**, e1007220 (2018).
264 20. M. J. H. Van Oppen, L. M. Peplow, S. Kininmonth, R. Berkelmans, Historical and
265 contemporary factors shape the population genetic structure of the broadcast spawning
266 coral, Acropora millepora, on the Great Barrier Reef. *Mol. Ecol.* **20**, 4899–4914 (2011).
267 21. E. Meyer, G. V. Aglyamova, M. V. Matz, Profiling gene expression responses of coral
268 larvae (Acropora millepora) to elevated temperature and settlement inducers using a
269 novel RNA-Seq procedure. *Mol. Ecol.* **20** (2011), doi:10.1111/j.1365-294X.2011.05205.x.
270 22. G. B. . Dixon, L. K. Bay, M. V. Matz, Evolutionary Consequences of DNA Methylation in a
271 Basal Metazoan. *Mol. Biol. Evol.* **33**, 2285–2293 (2016).
272
273
274

275 # Materials and Methods

276 *Sample sizes*

277 The experiment started with 15 colonies from each of the two sites, which were halved
278 and transplanted, resulting in a total of 60 fragments distributed across four study groups 15
279 samples each: two "natives" (KK and OO), and two "transplants" (OK and KO, Fig. 1C). Although
280 not all original samples were successfully analyzed using all approaches (Table S1), in the end
281 each of these four groups included 11 fragments representing unique coral genotypes and
282 paired by genotype between "home" and "away" groups that were analyzed for GBM, gene
283 expression, genotype, and fitness proxies.

284

285 *Reciprocal Transplantation Experiment*

286 Field work was conducted with permission from the Great Barrier Reef Marine Park
287 Authority (Research permit G09/29894.1) as described previously (*23*). Reciprocal
288 transplantations were undertaken between two environmentally distinct study sites (Miall
289 Island in the Keppel Island group: 23°09S 150°54E and Hazard Bay on Orpheus Island 18°37S
290 146°29E) separated by 4.5 degrees of latitude on the Great Barrier Reef (Figure 1A). On the
291 23rd of April (Orpheus) and 4th of May (Keppels) 2010 fifteen colonies were collected from wild
292 populations from each site and split in two. One half of each colony was replaced in its native
293 habitat, while the second half was transplanted to the alternate study site. Samples from all
294 coral fragments were collected at midday after three months (9th July 2010 at Orpheus,

295 14th July at Keppels) frozen in liquid nitrogen, then transferred into RNAlater (Ambion, Austin,
296 TX, USA) for gene expression and DNA methylation profiling.
297
298 *Coral growth rate*
299 Total colony skeletal growth was evaluated using wet buoyant weight as in Jokiel,
300 Maragos, & Franzisket (1978). Measurements from beginning and end of the experiment were
301 expressed as percent daily weight gain.
302
303 *Coral condition*
304 The energetic condition of corals was determined by the analysis of total protein,
305 carbohydrate and lipid content. Coral tissue was removed from an average of 8.6 cm^2 of coral
306 branch (mean $\pm$ SE = 0.18) using an air gun in 12 – 15 ml of 0.2 µM filtered sea-water (FSW). The
307 coral slurry was homogenized and centrifuged at 1500 g for 5 min at 4°C to separate coral and
308 endosymbiotic algal components. The coral slurry was aliquoted for protein, carbohydrate and
309 lipid assays and frozen at -20°C (10 ml for lipid, 0.5 ml for protein and carbohydrate). The algal
310 component was re-suspended in 2.75 ml of 0.2 µM FSW. Coral tissue surface area was
311 determined from cleaned (10% bleach for 24 hrs) skeletons of water-blasted samples using a
312 twice dip paraffin wax method (*25*). Surface area of coral branches was determined by the
313 weight gain of the second wax layer from a session specific standard curve (R^2 were 0.97 – 0.98)
314 using seven nylon cylinders of known surface area (5 – 48 cm^2).
315

316 *Protein analysis*

317 Total protein was extracted from a 0.5 ml aliquot using 0.5 ml 1M NaOH. Samples were
318 incubated at 90°C for one hour and spun at 3000 g for 5 minutes to separate cell-debris from
319 the solution and 500 µL clear supernatant was transferred onto a 96-well format. Protein
320 concentration was quantified in three technical replicates of 50 µl of coral protein extract using
321 a microplate Peterson – Lowry assay following the manufacturer's recommendations (Sigma:
322 TP0300). Blank 0.5 M NaOH samples and BSA protein standards (50, 100, 200, 300 and 400 µg /
323 ml) were run in duplicates on each plate and absorbance was read at 595 nm on a
324 spectrophotometer. Standard curves had R^2 of 0.97 – 0.99. The coefficient of variation (CV) was
325 calculated among technical replicates and deemed acceptable when <10%. For samples with a
326 CV > 10%, single deviant outliers (>1 SD deviant from other two measurements) were identified
327 and omitted. Total protein content per sample was expressed as average content (mg) of the
328 technical replicates per cm^2 of coral branch surface area.

329

330 *Carbohydrate analysis*

331 Total carbohydrate content estimates were obtained from the average of three technical
332 replicates of 50 µL coral slurry using D-glucose as a standard (*26*). Blank water samples and BSA
333 protein standards (20, 50, 100, 200, 500 and 1000 µg/ml) were run in triplicate on each plate
334 and absorbance was read at 485 nm on a spectrophotometer. Standard curves had R^2 of 0.96 –
335 0.99 and total carbohydrate content per sample was expressed in average mg of the technical
336 replicates per cm^2 of coral branch surface area after quality control as described for proteins.

337

338 *Lipid analysis*

339 Lipids were extracted using a modified version of Harland et al. (1993). Ten ml of coral
340 slurry was freeze dried overnight then homogenised by vortexing with 2 x 5 mL of 2:1
341 dichloromethane:methanol (DCM:MeOH), each left for 24 hours at 4°C. Extracts were filtered to
342 remove any debris using a glass filter (Whatman GF/C) and purified once with 5 ml 0.88%
343 KCl:H_2O and thrice with 1:1 MeOH:H_2O. Total lipid content was determined gravimetrically from
344 dried samples (60°C overnight) in pre-weighed acetone washed aluminum trays. Total lipid
345 content was expressed in mg per cm^2 of coral branch surface area.

346

347 *MBD-seq library preparation*

348 DNA was isolated from adult holobiont tissue using dispersion buffer (4 M guanadine
349 thiocyanate, 30 mM sodium citrate, 30 mM β-mercaptoethanol) followed by phenol chloroform
350 purification and a final cleanup with Zymo Genomic DNA Clean and Concentrator-10 kit (Catalog
351 No D4011). Genomic DNA was sheared using a Misonix Sonicator 3000 to a size range of ~200 to
352 800 bp checked by gel electrophoresis. Enrichment reactions were performed using the
353 MethylCap kit (Diagenode Cat. No. C02020010) with an initial input of 2 μg of sheared DNA per
354 reaction. The methylated fraction was eluted from the capture beads in a single step using High
355 Elution Buffer. Library preparation using NEBNext® DNA library Prep Master Mix Set (E6040L)
356 and sequencing on an Illumina HiSeq4000 was performed at the University of Texas Genome
357 Sequencing and Analysis Facility. Our total sample size was N=44 (22 colonies divided in half

358 giving 11 samples per treatment group). For the majority of these we sequenced only the
359 enriched library eluted from the capture beads. Fold coverages from these captured libraries
360 were used to estimate relative differences in GBM between samples. For a subset of 12 of the
361 44 samples, we sequenced both the captured and flow-through fractions. Fold differences
362 between these captured and flow-through libraries were used to estimate absolute levels of
363 methylation across genes. We did this for only a subset of samples because we were primarily
364 interested in relative differences in GBM between groups. As relative differences could be
365 reasonably well assessed without sequencing the flow-through (Figure S2), we chose to focus
366 our sequencing resources on increasing sample size rather than more thorough estimates of
367 absolute methylation levels.
368
369 *Repeated elements annotation*
370 *De novo* identification of repeated elements in *A. digitifera* genome was performed using
371 RepeatModeler (http://www.repeatmasker.org/RepeatModeler/, version 1.0.10).
372 RepeatModeler launches two repeat finding programs, RECON 1.05 (*28*)
373 (http://eddylab.org/software/recon/) and RepeatScout 1.0.5 (*29*)
374 (https://bix.ucsd.edu/repeatscout/), to generate a genomic database and upon which to build,
375 refine and classify consensus models of putative interspersed repeats. This pipeline produced a
376 total of 2,858 non-redundant repeat families in recent *A. digitifera* genome sequence. These
377 sequences were then used as the custom library to search the *A. digitifera* genome using

378 RepeatMasker (http://www.repeatmasker.org/). Approximate 7.33% and 5.47% of the genome
379 was annotated to be composed of retrotransposon and DNA transposon elements, respectively.
380
381 *MBD-seq data processing*
382 Forty four MBD-seq libraries were prepared from reciprocally transplanted coral
383 fragments. Sequencing produced 980 million raw reads with a mean of 16 $\pm$ 0.65 SEM million
384 reads per sample. Raw reads were trimmed of non-template sequence using Cutadapt (*30*) and
385 quality filtered using Fastx toolkit (http://cancan.cshl.edu/labmembers/gordon/fastx_toolkit/).
386 Adapter trimming and quality filtering reduced these the total read count 940 million, mean =
387 15 $\pm$ 0.64 SEM million reads per sample. The reference genome and annotations (version 1.1)
388 for *Acropora digitifera* (*31*) were downloaded from NCBI. Trimmed and filtered reads were
389 mapped to this concatenated reference using Bowtie2 (*32*). To ensure that using a reference
390 from an alternate species did not severely impair mapping, we compared the mean mapping
391 efficiency against the *A. digitifera* reference with that of a draft genome sequence for *A.*
392 *millepora* produced by David Miller and coworkers (James Cook University). Mean mapping
393 efficiency against the *A. digitifera* (78.4 $\pm$ 0.7% SEM) reference was only 5.2% lower than that of
394 *A. millepora* (83.6 $\pm$ 0.8% SEM), hence sequence differences between the two species did not
395 appear to substantially impair read mapping. Following mapping, PCR duplicates were removed
396 using Picard (https://broadinstitute.github.io/picard/). Mean duplication frequency was 14.3 $\pm$
397 0.5% SEM. The BAM files were filtered to retain only highly uniquely aligned reads (mapping
398 quality 30, or 0.1% chance of non-unique alignment) using samtools (*33*), after which the

399 number of reads overlapping various genomic regions - exon, intron, genic (exon + intron),
400 repeated elements longer than 500b, and repeat-free intergenic regions at least 1 kb long and at
401 least 2 kb away from any gene - were counted using BEDtools (*34*).
402
403 *MBD-score and GBM gene classes*
404 For 12 samples (3 per each experimental group) we quantified absolute levels of GBM as
405 the $\log_2$ fold difference in coverage between captured and flow-through DNA fractions while
406 controlling for genotype, as described in Dixon et al. (2016). These values were used as gene-
407 specific "MBD scores" throughout the study. As expected based on previous studies (*3*, *35*),
408 MBD-scores were bimodally distributed (Fig. S1 A) and negatively correlated with $CpG_{o/e}$ (Figure
409 1 D) providing genome-wide validation for our MBD-seq technique. The MBD-score threshold
410 separating the two clouds of high point density in the scatter plot on Figure 1 D was used to
411 delineate the two gene classes for computation of the "GBM class difference" (GBMcd) for each
412 sample (Fig. 2 I, J).
413
414 *Tag-seq data processing*
415 Transcription was quantified using Tag-seq (Meyer et al. 2011; Lohman et al. 2016). The
416 Tag-seq reads were downloaded from the SRA database (accession SRP049522; (*23*)) and
417 mapped against the *A.digitifera* genome using SHRiMP (*37*). Mapped reads overlapping
418 annotated coding sequences were counted using intersection-nonempty method in HTseq

419 version 0.6.1p1 (*38*). Normalization of raw counts and statistical analyses were performed using
420 DESeq2 (*39*).
421
422 *Assessing variation in gene body methylation and transcription*
423 As with gene expression analyses, normalization and statistical analyses of MBD-seq
424 reads were performed with DESeq2 (*39*). To test for origin effects, we performed two tests that
425 compared the two groups placed at Keppel to each other (OK vs KK) and the two groups placed
426 at Orpheus to each other (OO vs KO). These tests were intended to identify effects of origin
427 while controlling for environmental conditions experienced during the experiment. To assess
428 effects of transplantation, we compared groups that originated from Orpheus to each other (OO
429 vs OK) and groups that originated from Keppel to each other (KK vs KO). For these tests, we
430 included an additional parameter of colony identity to identify effects of transplantation while
431 controlling for genotype. For TagSeq, only genes with mean read count ≥3 were considered for
432 analysis (19706 genes), and for MBD-seq we chose genes with mean read count ≥20 (27084
433 genes).
434
435 *Discriminant analysis of principal components*
436 GBM data were further analyzed using discriminant analysis of principal components
437 (DAPC) implemented in the R package adegenet (*40*), following the procedures outlined in
438 Kenkel & Matz (2016). DAPC is a multivariate analysis method designed to identify between-
439 group variation while neglecting within-group variation. We used this method to distill our

440 multivariate MBD-seq dataset into single axes that maximized discrimination between natives
441 from the two experimental sites (KK and OO samples). The result was a discriminant axis
442 contrasting the GBM variation of the two native populations—with one pole designating native
443 Keppel-like GBM patterns and the opposite pole designating native Orpheus-like patterns. The
444 function was then applied to data from the transplanted samples, so that their positions along
445 discriminant axis described their similarity in GBM patterns to the contrasted native
446 populations. It seems plausible that, either due to natural selection or plasticity, native corals
447 would possess patterns optimal for their particular site. If this is true, then transplants with
448 DAPC values closer those of local corals would be expected to show improved fitness proxies
449 (eg: samples transplanted to Keppel with DAPC values closer to the 'Keppel-like' pole of the
450 discriminant axis would be expected to show higher fitness proxies and vice versa for samples
451 transplanted to Orpheus). 'DAPC Similarity' values describing each transplant's proximity along
452 the discriminate axis to natives of their transplantation site (Fig. 4 D) were computed by taking
453 the absolute value of the difference between the transplants' DAPC values and the mean value
454 for natives of their respective transplant sites, converting these distances into z-scores, and
455 multiplied them by -1. The DAPC Similarity value for sample X was:
456
457 $$DAPC\ Similarity_x = -1 \times (|D_x| - \overline{D})/\sigma$$
458
459 where, D_x is its distance along the discriminant axis between sample X and the mean DAPC
460 value for natives of its transplantation site, $\overline{D}$ is the mean absolute distance for all transplants,

461 and σ is the standard deviation of absolute distance for all transplants. Correlations of these
462 values with fitness characteristics of transplanted fragments (Fig. 4 E) were computed using
463 function *envfit* (R package *vegan* (*42*))
464
465 *Validation of MBD-seq results by bisulfite amplicon sequencing*
466 To validate our MBD-seq results, we used targeted bisulfite sequencing. Genomic DNA
467 for this procedure was extracted and purified as described above. It should be noted that as
468 with the MBD-seq and Tag-seq, targeted bisulfite sequencing was performed from separate
469 isolations from the same or closely adjacent branch. We performed bisulfite conversion using an
470 EZ DNA Methylation-Gold kit (Zymo Research; cat. no. D5005). Thermocycler conditions for
471 conversion were 98°C for 10 minutes followed by 53°C for 4 hours. Primers for post-conversion
472 amplification were designed using Bisulfite Primer Seeker (Zymo Research;
473 http://www.zymoresearch.com/tools/bisulfite-primer-seeker). Primer sequences are given in
474 Table S1. We designed 13 primer sets to target coding sequences (excised from the *A. digitifera*
475 reference genome) of 13 separate loci. Primer sequences included 5' tails for downstream
476 addition of barcodes and Illumina adapters by PCR. Target loci were selected based on either
477 strong origin or transplant effects in the DESeq2 analyses. Amplification from bisulfite converted
478 samples was performed using EpiMark Hot Start Taq DNA Polymerase (New England Biolabs;
479 cat. no. M0490S). Thermocycler conditions were 95°C for 30 seconds followed by 35-40 cycles
480 of 95°C for 15 seconds, 53°C for 30 seconds, 68°C for 30 seconds. Barcodes and adapters for
481 multiplexed sequencing were added to the resulting PCR products in a second PCR. The

482 oligonucleotide sequences used for barcoding were the same as those given in the Tag-seq
483 library preparation (Meyer et al. 2011; https://github.com/z0on/tag-based_RNAseq).
484 Sequencing was performed using paired-end 600 cycle runs on the Illumina Miseq. Resulting
485 reads were quality trimmed using cutadapt (*30*). Quantification of CpG methylation was
486 performed using Bismark (*43*) using a fasta file of the 13 exon sequences used to design the
487 primers as a reference. For a given sample, CpG sites represented by fewer than 50 reads were
488 excluded from the analysis (2022 low coverage calls excluded). The difference in means based
489 on origin for all CpGs in each gene correlated with the difference in mean normalized counts
490 from the MBD-seq reads (Figure 1 D).
491
492 **Primers used for targeted bisulfite sequencing**. Each primer includes a 5' tail (bold) for
493 amplification with barcoded primers for multiplex Illumina sequencing included in the Tag-seq
494 library preparation.

Primer	Sequence (5'-3')
LOC107327073_bsfTag_For	**CTACACGACGCTCTTCCGATCT**TTTTTGYGAGGTTGATTTTGTTATTATG
LOC107327073_bsfTag_Rev	**ACGTGTGCTCTTCCGAT**TTTCCAAACATATTCCTTTCCATAACATTC
LOC107356898_bsfTag_For	**CTACACGACGCTCTTCCGATCT**TAGATTTYGTTATAATGTTATTAAGAAGTGAAGG
LOC107356898_bsfTag_Rev	**ACGTGTGCTCTTCCGAT**AATAATATTAAACATTCTCTCTACAAATCTACCAC
LOC107358158_bsfTag_For	**CTACACGACGCTCTTCCGATCT**TGGTYGGATTGTTGAAGAGTTTAAGTAG
LOC107358158_bsfTag_Rev	**ACGTGTGCTCTTCCGAT**ACACCCAAATCACCCATCTCATTAAC
LOC107356899_bsfTag_For	**CTACACGACGCTCTTCCGATCT**GTTTTTAAAATTTGAAGGATTTGGTTTTGTTG
LOC107356899_bsfTag_Rev	**ACGTGTGCTCTTCCGAT**TTACATTATATTTTCCAAACATATTTCATACCATAAC
LOC107358871_bsfTag_For	**CTACACGACGCTCTTCCGATCT**GATATYGGGTTTTTAATAATAATTGTATGTTGGTTG
LOC107358871_bsfTag_Rev	**ACGTGTGCTCTTCCGAT**TTTTCCTTTAAAATTATTTCCACCCAACTCC
LOC107334334_bsfTag_For	**CTACACGACGCTCTTCCGATCT**ATAAGGATATGTAGGGTTTTGGTAAGG
LOC107334334_bsfTag_Rev	**ACGTGTGCTCTTCCGAT**AATTCATAATAACTACCCTACAACAAAAAATCCC
LOC107347512_bsTag_For	**CTACACGACGCTCTTCCGATCT**ATYGATAGATTAAAAGAAGTTGGAGTATTG

LOC107347512_bsTag_Rev	**ACGTGTGCTCTTCCGAT**CCTAAACAATCCATAAAAACCTTCCTACAATTC
LOC107350794_bsTag_For	**CTACACGACGCTCTTCCGATCT**GAYGTGTTTTGTATTTAGTTATTGGATATTTGG
LOC107350794_bsTag_Rev	**ACGTGTGCTCTTCCGAT**ATCTTCCRCAATTAAAAAAACATACTTAAAATCTTCAC
LOC107339795_bsTag_For	**CTACACGACGCTCTTCCGATCT**TGAATATAGTTAAGGTAAATGGATGAGTTATATATGG
LOC107339795_bsTag_Rev	**ACGTGTGCTCTTCCGAT**CAAAAATAAACRAATCAAACAAAACAATCATTAC
LOC107336909_bsTag_For	**CTACACGACGCTCTTCCGATCT**AGAGATGYGGAATAGATATTTTTGGTTGG
LOC107336909_bsTag_Rev	**ACGTGTGCTCTTCCGAT**AATAACRAACATTACATCTTATTTCTCTAAAATAATAC
LOC107352877_bsTag_For	**CTACACGACGCTCTTCCGATCT**TGGTGTYGTGAATTTGTTTAAATAATTTATG
LOC107352877_bsTag_Rev	**ACGTGTGCTCTTCCGAT**CATTCCATCRACACAACTAAATAAAAAAC
LOC107351808_bsTag_For	**CTACACGACGCTCTTCCGATCT**GTYGAGATATGTTAAAATTTGAGGATGTGAGTTTG
LOC107351808_bsTag_Rev	**ACGTGTGCTCTTCCGAT**ATCCACCATTCRCCAACTTACTAATAAATCTCCC
LOC107327285_bsTag_For	**CTACACGACGCTCTTCCGATCT**ATTATGTTAGYGTTTTTGTTATTTTGGATTGTGG
LOC107327285_bsTag_Rev	**ACGTGTGCTCTTCCGAT**CCTCATTCACAAAAACAACCTTTAATC

Analysis of genetic differentiation

The MBD-seq reads were also used to analyze genetic distances among the 22 sequenced colonies. For this procedure, we concatenated reads from clone pairs into single files, and mapped the reads to the *A. digitifera* reference genome (*31*) using Bowtie2 (*32*). For the subset of 12 samples where we sequenced both captured and flow-through fractions, both sets of reads were concatenated for genotyping. Between-individual genetic distances for DAPC analysis were determined as (1 - Identity-By-State) using ANGSD (*44*) with the following filter settings: -uniqueOnly 1 -remove_bads 1 -minMapQ 20 -minQ 30 -baq 1 -minInd 10 -snp_pval 1e-2 -minMaf 0.1 (to summarize, only high-quality and uniquely-mapped reads were taken into account; only sites covered in a minimum of 10 individuals were considered; the cutoff for the p-value for the SNP being true was set at 0.01; and only alleles with frequency equal or exceeding 0.1 were used). This resulted in genotyping information for roughly 507,000 polymorphic sites. The between-population F_{ST} for genes passing 10% FDR cutoff for either for

509 origin or transplant effect and a random subset of 500 non-significant genes (rather than whole
510 genome, to speed up computation) were determined using ANGSD based on site frequency
511 spectra as priors, which were generated for every gene subset without using filters that would
512 affect representation of alleles of different frequencies (-uniqueOnly 1 -remove_bads 1 -
513 minMapQ 20 -minQ 30 -baq 1 -minInd 4).
514
515 *Determination of relative proportions of* Symbiodinium *clades*
516 The MBD-seq reads were mapped to the concatenated reference including A. digitifera
517 genome and four *Symbiodinium* sp. trasncriptomes from four different genotypic groups
518 ("clades"). Transcriptomes for *Symbiodinium* clades A and B were from (*45*) and transcriptomes
519 for clades C and D were from (*46*). We then counted the relative proportions of reads producing
520 highly unique matches (mapping quality 40 or higher) to each *Symbiodinium* transcriptome,
521 using a custom perl script *zooxType.pl*. All corals were found to be dominated by *Symbiodinium*
522 clade C.
523
524 *Data and scripts availability*
525 TagSeq, MBD-seq, and amplicon-bisulfite sequencing datasets are available through NCBI
526 Short Read Archive (Project Accession: SRP049522). Scripts and traits data are available in a
527 GitHub repository dedicated to this paper
528 (https://github.com/grovesdixon/reciprocal_transplant_methylation).
529

530 **References** (Materials and Methods)

531 23. G. B. Dixon, L. K. Bay, M. V Matz, Bimodal signatures of germline methylation are linked
532 with gene expression plasticity in the coral Acropora millepora. *BMC Genomics*. **15**, 1109
533 (2014).
534 24. P. Jokiel, J. Maragos, L. Franzisket, in *Coral reefs: research methods*, D. R. Stoddart, R. E.
535 Johannes, Eds. (UNESCO, Paris, ed. 5, 1978), pp. 529–542.
536 25. J. Stimson, R. A. Kinzie, The temporal pattern and rate of release of zooxanthellae from
537 the reef coral Pocillopora damicornis (Linnaeus) under nitrogen-enrichment and control
538 conditions. *J. Exp. Mar. Bio. Ecol.* **153**, 63–74 (1991).
539 26. T. Masuko *et al.*, Carbohydrate analysis by a phenol-sulfuric acid method in microplate
540 format. *Anal. Biochem.* **339**, 69–72 (2005).
541 27. A. D. Harland, J. C. Navarro, P. S. Davies, L. M. Fixter, Lipids of some Caribbean and Red
542 Sea corals: total lipid, wax esters, triglycerides and fatty acids. *Mar. Biol.* **117**, 113–117
543 (1993).
544 28. Z. Bao, S. R. Eddy, Automated de novo identification of repeat sequence families in
545 sequenced genomes. *Genome Res.* **13**, 1269–1276 (2003).
546 29. A. L. Price, N. C. Jones, P. A. Pevzner, De novo identification of repeat families in large
547 genomes. *Bioinformatics*. **21** (2005), doi:10.1093/bioinformatics/bti1018.
548 30. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing
549 reads. *EMBnetjournal*. **17**, 10–12 (2011).
550 31. C. Shinzato *et al.*, Using the Acropora digitifera genome to understand coral responses to
551 environmental change. *Nature*. **476**, 320–3 (2011).
552 32. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. **9**,
553 357–359 (2012).
554 33. H. Li *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**,
555 2078–9 (2009).
556 34. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic
557 features. *Bioinformatics*. **26**, 841–842 (2010).

558 35. G. B. Dixon, L. K. Bay, M. V. Matz, Evolutionary Consequences of DNA Methylation in a
559 Basal Metazoan. *Mol. Biol. Evol.* **33** (2016), doi:10.1093/molbev/msw100.
560 36. B. K. Lohman, J. N. Weber, D. I. Bolnick, Evaluation of TagSeq, a reliable low-cost
561 alternative for RNAseq. *Mol. Ecol. Resour.* **16**, 1315–1321 (2016).
562 37. S. M. Rumble *et al.*, SHRiMP: Accurate mapping of short color-space reads. *PLoS Comput.*
563 *Biol.* **5**, 1–11 (2009).
564 38. S. Anders, P. T. Pyl, W. Huber, HTSeq-A Python framework to work with high-throughput
565 sequencing data. *Bioinformatics*. **31**, 166–169 (2015).
566 39. M. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for
567 RNA-Seq data with DESeq2. *Genome Biol.* (2014) (available at
568 http://www.biomedcentral.com/content/pdf/s13059-014-0550-8.pdf).
569 40. T. Jombart, S. Devillard, F. Balloux, Discriminant analysis of principal components: a new
570 method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94 (2010).
571 41. C. Kenkel, M. V Matz, Enhanced gene expression plasticity as a mechanism of adaptation
572 to a variable environment in a reef-building coral. *Nat. Ecol. Evol.* **1**, 0014 (2016).
573 42. J. Oksanen *et al.*, vegan: Community Ecology Package. *R Packag. ver. 2.4–3* (2017), ,
574 doi:10.4135/9781412971874.n145.
575 43. F. Krueger, S. R. Andrews, Bismark: A flexible aligner and methylation caller for Bisulfite-
576 Seq applications. *Bioinformatics*. **27**, 1571–1572 (2011).
577 44. T. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation
578 Sequencing Data. *BMC Bioinformatics*. **15**, 356 (2014).
579 45. T. Bayer *et al.*, Symbiodinium transcriptomes: genome insights into the dinoflagellate
580 symbionts of reef-building corals. *PLoS One*. **7**, e35269 (2012).
581 46. J. T. Ladner, D. J. Barshis, S. R. Palumbi, Protein evolution in two co-occurring types of
582 Symbiodinium: an exploration into the genetic basis of thermal tolerance in
583 Symbiodinium clade D. *BMC Evol. Biol.* **12**, 217 (2012).
584 47. C. Kosiol *et al.*, Patterns of positive selection in six mammalian genomes. *Plos Genet.* **4**
585 (2008), doi:10.1371/journal.pgen.1000144.

Supplementary Tables and Figures

Table S1. Sample sizes

Sample group	N (TagSeq)	N (MBD-seq, captured and flow-through fractions)	N (MBD-seq, only captured fraction)
KK	14	3	8
KO	15	3	8
OK	15	3	8
OO	15	3	8

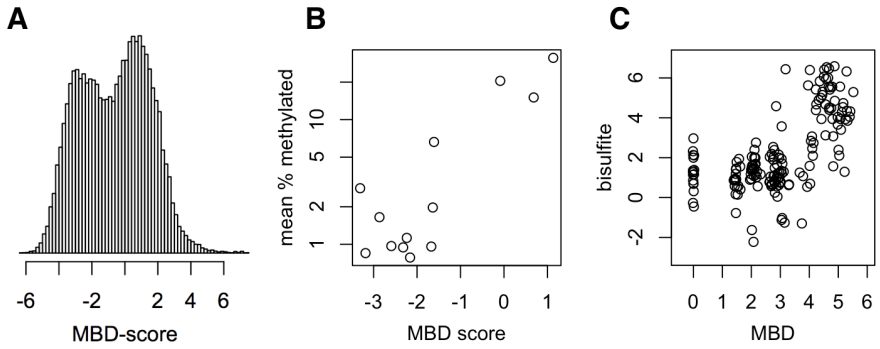
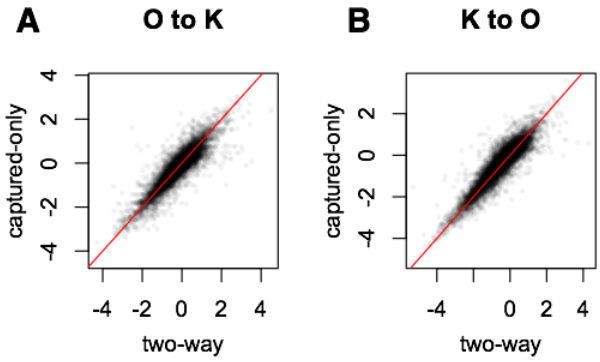


Figure S1. Distribution of MBD-scores (A) and validation by bisulfite sequencing of amplicons (B, C). B: Mean percent methylation was calculated as the proportion of methylated CpG sites within each gene averaged across all samples. C: Correlation between MBD-based and bisulfite-based measurement of the effects of each coral individual on methylation for each of the assayed genes.



596

597 **Figure S2** Correlation between inferred changes in GBM based on comparison of MBD-captured

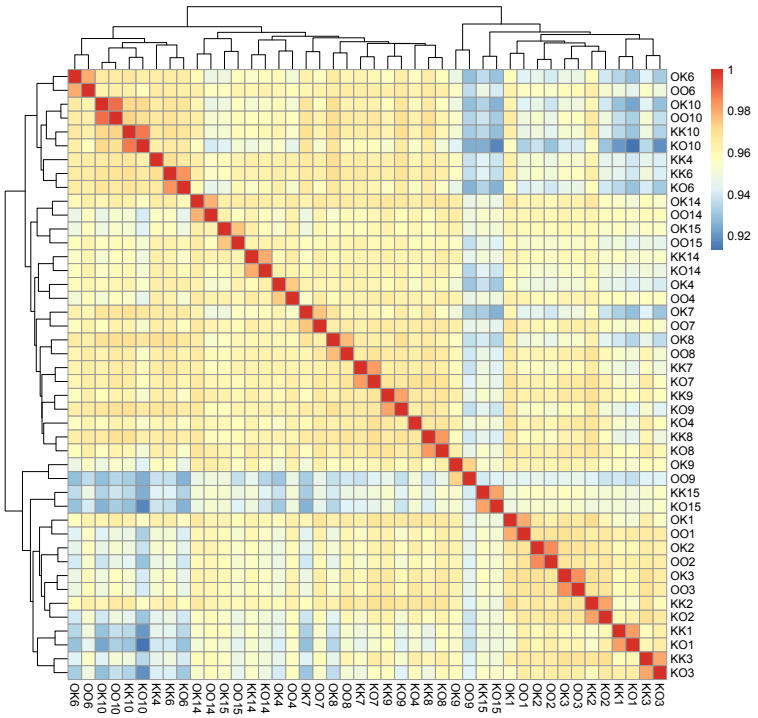
598 and flow-through libraries (x-axis) and only on MBD-captured fraction (y-axis). The red line is 1:1

599 correspondence. (A) Orpheus to Keppel transplantation; (B) Keppel to Orpheus transplantation.

600 These plots were generated for a subset of 12 samples (three for each of the four experimental

601 groups, KK, KO, OK, and OO) for which both the captured and the flow-through fractions were

602 sequenced.



603

604 **Figure S3** Heatmap of GBM correlations among samples, illustrating strong dependence of GBM

605 on genotype. Colors indicate Spearman's rank correlations for normalized MBD-seq read counts

606 across all coding genes (N = 24853); the lowest observed correlation value was 0.9. Samples

607 were clustered by maximum distance method. First letter of sample names indicates sample

608 origin, second letter indicates transplantation site, and number indicates replicate. Samples

609 sharing the same first letter and the same number are clonal fragments from the same colony.

610 All of the 22 clone pairs except one (sample pair KK4 and KO4) clustered together.

611

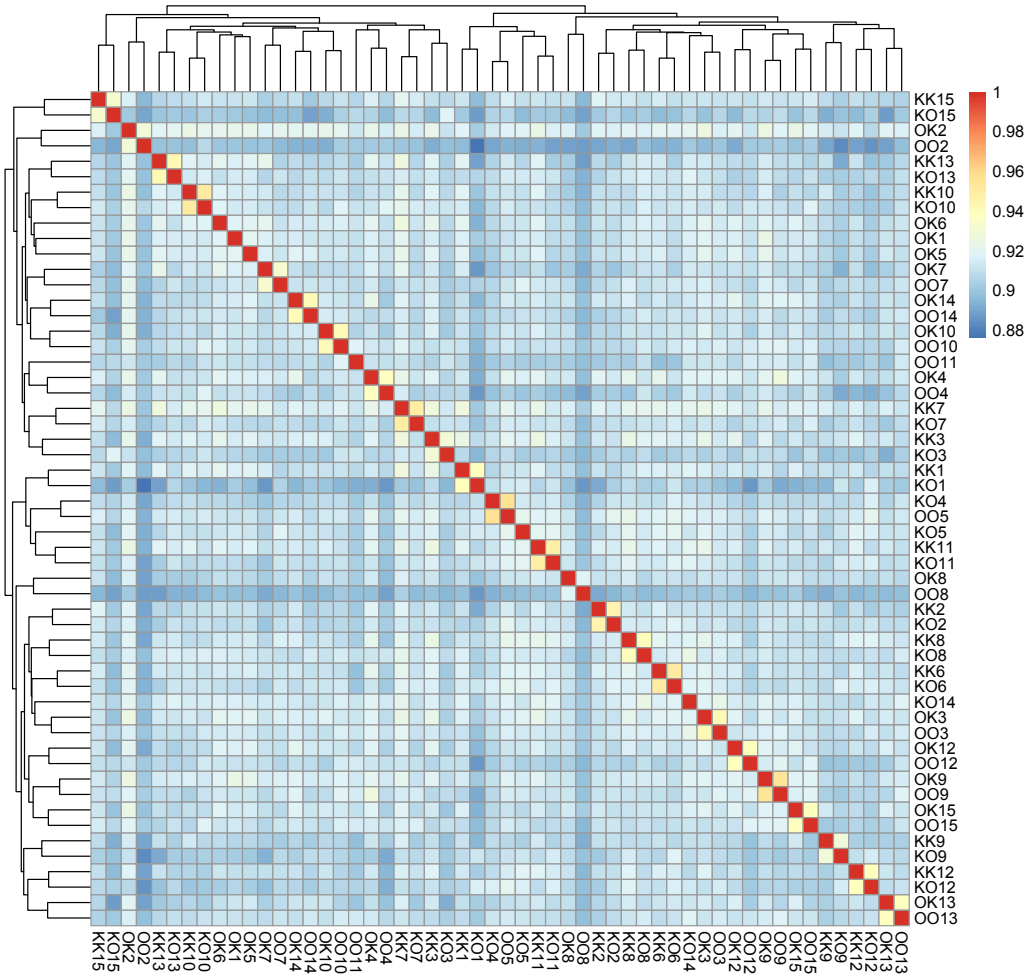
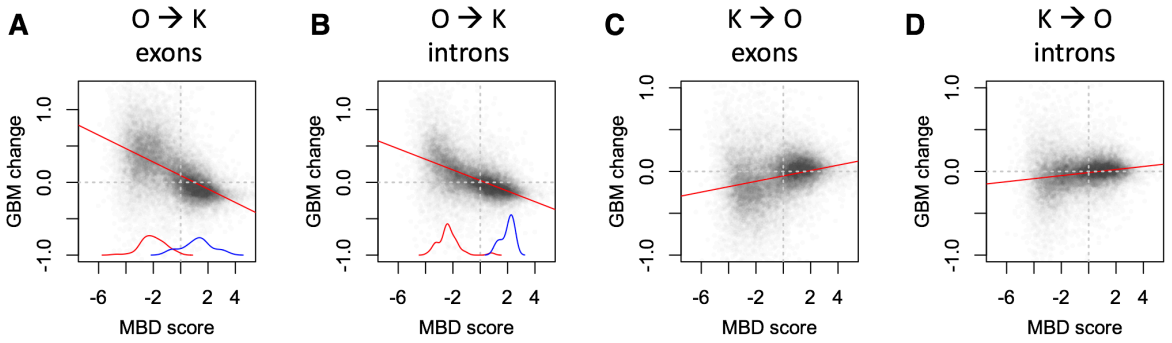


Figure S4 Heatmap of overall correlations in transcription illustrating strong dependence of transcription on genotype. Colors indicate Spearman's correlations for normalized Tag-seq read counts across all coding genes (N=19706). Samples were clustered by maximum distance method. First letter of sample names indicates sample origin, second letter indicates transplantation site, and number indicates replicate. Samples sharing the same first letter and number are clonal fragments from the same colony. All 24 available clone pairs clustered together (six samples lacked data for clone pairs).

621



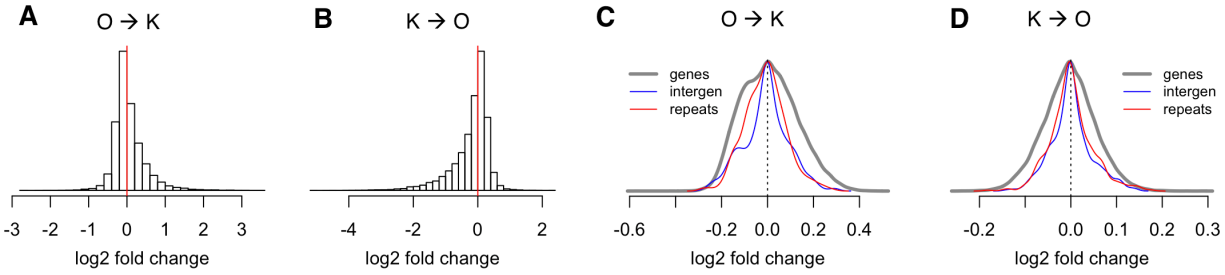
622

623 **Figure S5.** GBM $\log_2$ fold-changes change in response to transplantation in exons and introns

624 plotted against MBD-score for each gene. Red and blue curves near the x-axis are density plots

625 of significantly up- and down-regulated genes.

626



627

628 **Figure S6.** Absolute methylation changes in response to transplantation ("away" vs. "home")

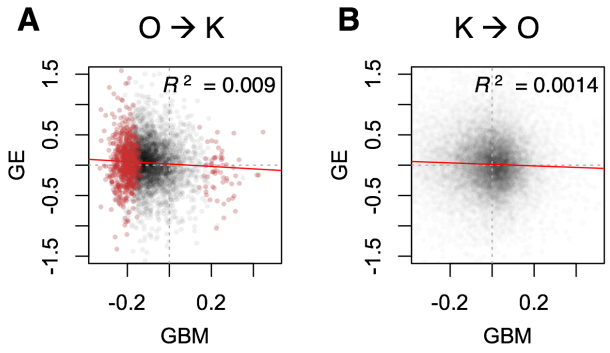
629 can be positive or negative across genes (A, B) and affect mostly protein-coding regions (C, D).

630 Plots on A and B are based on comparison between captured and flow-through fractions for 12

631 samples (three for each of the four experimental groups, KK, KO, OK, and OO). Plots on panels C

632 and D were generated based on captured-only analysis of all 44 MBD-seq samples (Table S1).

633



634

635 **Figure S7**. Gene expression changes in response to transplantation are negatively correlated

636 with GBM changes. Both correlations are highly significant (p<0.0001). Red points on panel A

637 represent genes with statistically significant (FDR<10%) change in GBM.

638

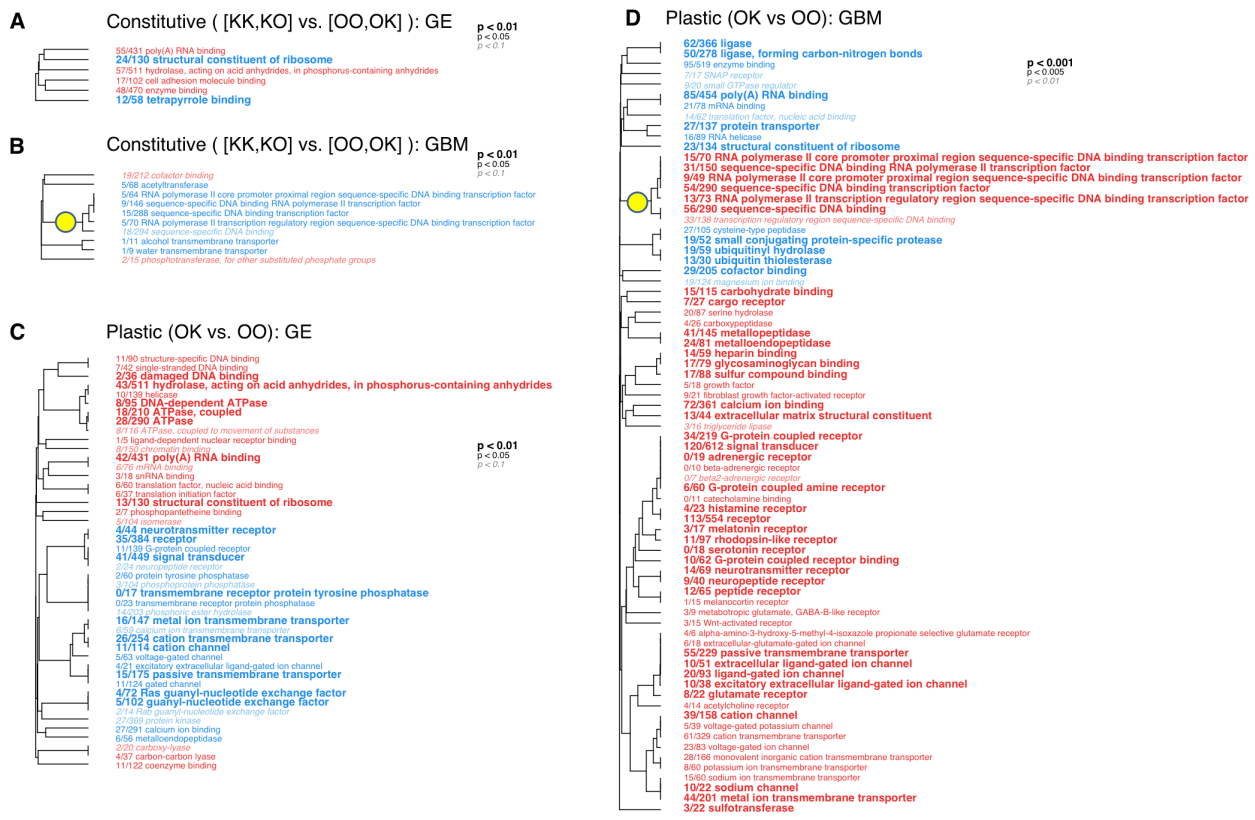
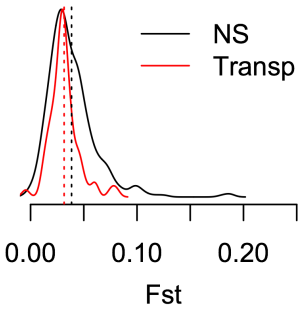


Figure S8. Gene ontology categories significantly enriched with up- or down-regulated genes (red and blue, respectively) in GE and GBM comparisons. The Benjamini-Hochberg corrected significance of enrichment is indicated by the size and type of the font (see legends). Note that for plastic GBM the significance thresholds are ten times more stringent, to fit the GO plot onto one page. The fraction preceding the category name is the number of genes exceeding raw p-value of 0.05 relative to the total number of genes in the category. The dendrograms are hierarchical clustering of GO categories based on proportion of shared genes (*47*). Genes comprising the branch labelled by the yellow circle (transcription factors) show plastic changes (D) that are opposite to the constitutive differences between populations (B).



650

651 **Figure S9**. Density plots of F_{ST} for genes that show significant (FDR <0.1) GBM change in
652 response to transplantation from Orpheus to Keppel ("Transp", red curve) compared to 500
653 randomly picked non-significant genes ("NS", black curve). Dashed vertical lines mark the mean
654 of each dataset.

655