Gene body methylation is involved in coarse genome-wide adjustment of gene expression during acclimatization.

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

Gene body methylation (GBM) is a taxonomically widespread epigenetic modification of the DNA 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 context-dependent inducible genes [2,3], and it has been hypothesized that changes in GBM might modulate responses to environmental change, including transgenerational plasticity [4,5]. Here, we profiled GBM, gene expression, genotype, and fitness characteristics in clonal fragments of a reef building coral *Acropora millepora* reciprocally transplanted between two distant reefs. We find that genotype-specific GBM is considerably more stable than gene expression and responds to transplantation predominantly by genome-wide increase or decrease in disparity of methylation levels among genes. A proxy of this change, GBM difference between the two gene classes (housekeeping vs. inducible), was the most important determinant of genome-wide GBM variation in our experiment, explaining 33% of it. Surprisingly, despite apparent lack of capacity for environmental specificity, this simple genome-wide GBM adjustment was a good predictor of broad-scale functional shifts in gene expression and of fragments' fitness in the new environment, which supports GBM's role in acclimatization. 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. We propose that during acclimatization GBM acts as a "single-knob equalizer" to rapidly achieve coarse genome-wide adjustment of gene expression, after which further fine-tuning is provided by expression plasticity of individual genes and longer-term genetic adaptation of both GBM and gene expression to local conditions.

Only the detrimental effect of GBM is well understood: GBM causes hypermutability in protein-coding regions [6]. Indeed, in humans, GBM is the primary driver of deleterious parent-age-related mutations [7]. 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 , suppression of intragenic transcription initiation is currently the most studied[8,9] but still remains controversial [10]. The ecological genomic literature has long expected that GBM (as well as other epigenetic marks) respond to the environment and assist acclimatization by modulating gene expression, possibly across generations [4,5,11]. Still, thus far the only well-documented case of GBM responding to the environment and resulting in phenotype change is determination of castes in social insects depending on diet [12]. Beyond this highly taxon-specific example, GBM response to the environment has been reported in a recent paper on coral acclimatization to acidic conditions [13]; notably, GBM change did not specifically affect genes expected to be involved in such acclimatization. One way to address the role of GBM in adaptation would be to examine differences between populations, as has been recently

done in plants [14,15]; however, this interpretation requires disentangling the effect of environment from the effect of genetic divergence among populations [16]. As for transgenerational plasticity, for GBM to be involved in this process it must be both responsive to the environment and heritable. This somewhat contradictory combination of properties has not yet been demonstrated for GBM in any study system.

Here, we used a reciprocal transplantation framework [17] to test for the roles of GBM, gene expression, and genetics in local adaptation and acclimatization in a reef-building coral *Acropora millepora*. In corals, the possibility of clonal replication makes disentangling the effects of genotype and environment very straightforward, solving a major problem of ecological epigenetics [16]. We have compared corals from two reefs, Orpheus Island in the central sector of the Great Barrier Reef (GBR) and Keppel Island in the south of GBR (Fig. 1 A). These reefs are notably different in temperature (Fig. 1 B) as well as several other abiotic parameters [18] and host slightly genetically divergent *A. millepora* populations [19,20]. Fifteen coral colonies from each reef were halved and reciprocally transplanted between the sites for three winter months (Fig. 1 A, B). In this way, each individual genotype was simultaneously exposed to two distinct reef conditions. After three months in the field, tissue samples were collected from each fragment and assayed for genome-wide gene expression using TagSeq [21] and for DNA methylation using MBD-seq [22]. These data were analyzed in the context of several fitness proxies and genetic distances between individuals (based on genetic polymorphisms detected in the MBD-seq data). All experimental corals were dominated by *Symbiodinium* clade C, which was established by analyzing *Symbiodinium*-matching MBD-seq reads.

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

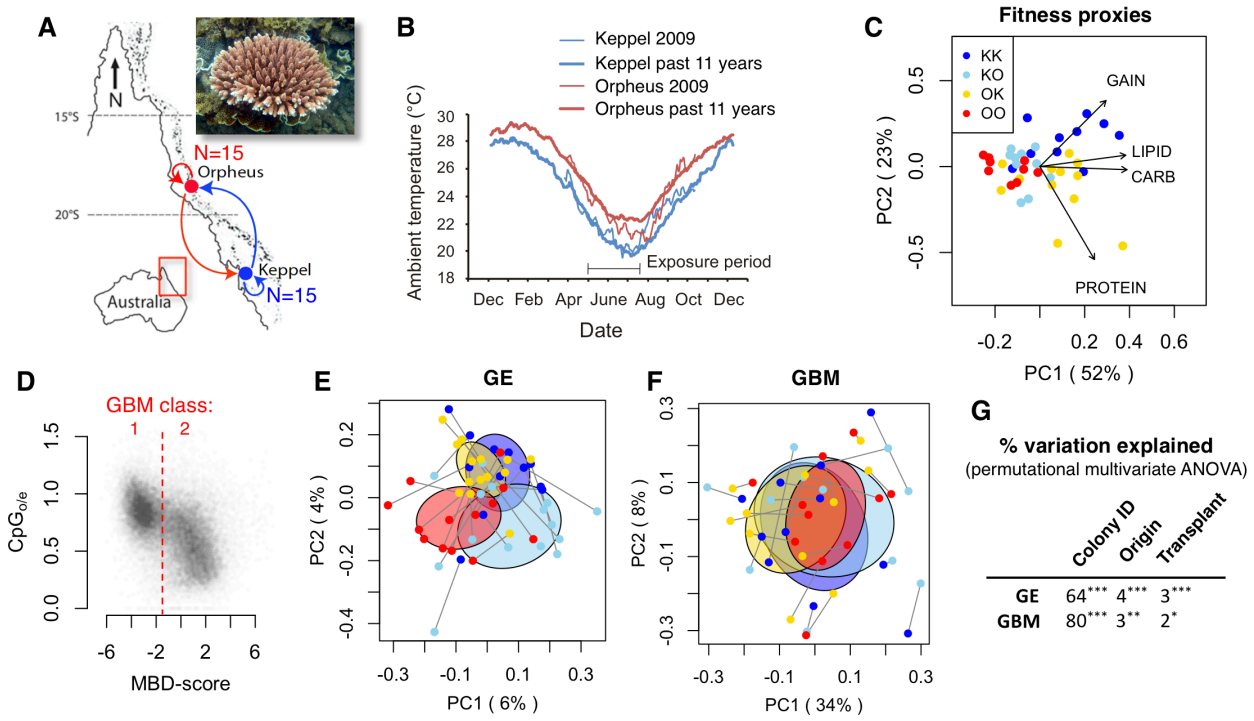

Figure 1. Overview of the experiment. A: Design of reciprocal transplantation. Inset – *Acropora millepora*. B: Temperature profiles at the transplantation sites, historical (thick lines) and during the year of experiment (thin lines). C: Principal component analysis of the four fitness proxies: weight gain, lipid, protein and carbohydrate content. In the sample name, 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 explained by colony identity (i.e., by similarity among clonal fragments, or broad-sense heritability), colony's origin and transplantation site. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

To determine the characteristic level of GBM for each gene we used MBD-seq results for 12 samples (three per each experimental group, KK, KO, OK, and OO) for which both MBD-captured and flow-through fraction were sequenced. Logarithm with the base 2 of each gene's abundance in captured vs. flow-through fractions ("MBD-score" [22]) exhibited the expected bimodal distribution (Fig. S1 A) and showed strong correlation (Fig. 1 D) with the established proxy of historical methylation level, the ratio of observed to expected numbers of CpG dinucleotides ("$CpG_{O/E}$" [3]). In addition to this genome-wide validation of our methylation data, we validated GBM results by bisulfite-sequencing 13 amplicons representing different MBD-

score classes (Fig. S1 B, C). Genes were assigned to either low-methylated class (class 1) or high-methylated class (class 2) based on the MBD-score threshold indicated on Fig. 1 D. Relative quantification results based only on MBD-captured read data were similar to those obtained when using both captured and flow-through fractions, although they were less powerful (Fig. S2). For the remaining 32 coral samples we sequenced only the MBD-captured fraction.

Multivariate analysis of gene expression (GE, Fig. 1 E) and GBM (Fig. 1 F) revealed that GBM is more consistent across fragments of the same original colony than GE (Fig. 1G, Fig. S3 and S4). In addition, significant effects of origin and transplantation site were observed for GE and GBM, both more pronounced for GE (Fig. 1 E and G).

Surprisingly, the change in GBM in response to transplantation from Orpheus to Keppel predominantly consisted in genome-wide reduction of GBM disparity among genes: highly-methylated genes became less methylated and low methylated became more methylated (Fig. 2 A). This change was mirrored by less pronounced but clearly reciprocal GBM adjustment in Keppel corals transplanted to Orpheus (Fig. 2 B, C). In other words corals planted at Keppel showed reduced genome-wide GBM disparity compared to their clonemates planted at Orpheus. Observing this in both Orpheus- and Keppel-origin corals indicates that this is indeed an environment-induced plastic response. Both exons and introns underwent this change (Fig. S5). Absolute gene-specific changes in methylation inferred from comparison of MBD-captured and flow-through fractions included both positive and negative shifts (Fig. S6 A, B), indicating that the change in relative GBM levels among genes is not due to biased genome-wide methylation increase or decrease. Methylation levels of intergenic regions and repeated elements remained relatively constant compared to genic regions (Fig. S6 C, D). Like GBM, GE response was also reciprocal between transplantation directions but was much more pronounced than GBM response, indicated by much broader range of fold-changes (Fig. 2 D). Notably, there was highly significant genome-wide negative correlation between changes in GBM and GE. Although it was weak on gene-by-gene basis (Fig. 2 F, G), it was very strong when comparing broader functional groups of genes (Gene Ontology categories) implicated in both GE and GBM analyses (Fig. 2 H and Fig. S7). To characterize GBM disparity among genes in each sample, we computed GBM class 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.

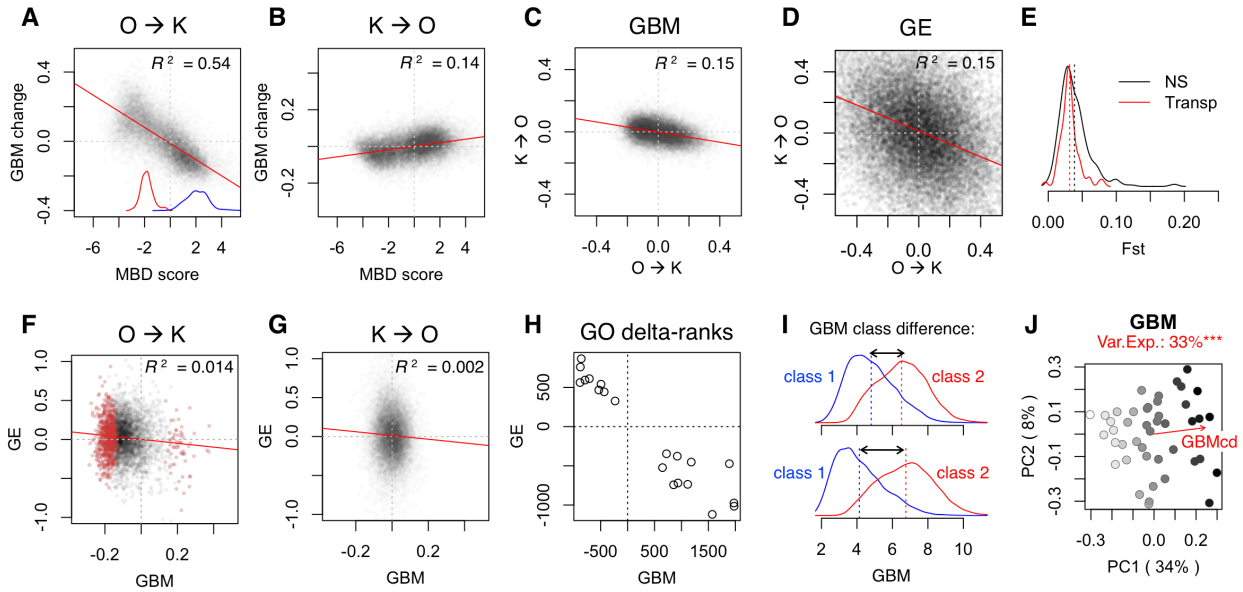

Figure 2. Changes in gene body methylation (GBM) and gene expression (GE) in response to transplantation ("away" relative to "home"). On all panels, GE and GBM measures are $\log_2$ fold-changes. A and B: Change in GBM in fragments transplanted from Orpheus to Keppel (A) and from Keppel to Orpheus (B) plotted against MBD-score for each gene (Fig. 1 D). Red and blue curves near the x-axis are density plots of significantly up- and down-regulated genes. C and D: Correlation between GBM (C) and GE (D) changes in reciprocally transplanted fragments. E: Density plot of between-population F_{ST} for genes showing significant change in GBM with respect to transplantation site (red line) compared to 500 randomly picked non-significant genes (black line). F and G: Correlation between changes in GE and GBM in fragments transplanted from Orpheus to Keppel (F) and from Keppel to Orpheus (G). Red points on panel F are genes significantly different in GBM at 10% FDR level. H: Plot of delta-ranks for 20 Gene Ontology categories that were significantly enriched with either up-regulated genes (positive delta-rank) or down-regulated genes (negative delta-rank) in both GE and GBM analyses. I: Calculation of GBM class difference (GBMcd), a proxy of genome-wide GBM disparity among genes. Each panel shows two density plots of GBM levels for the two GBM gene classes (see Fig. 1 D), with the mean for each class indicated by a dashed line. The top panel shows the sample with the least GBMcd, bottom panel – the sample with the most GBMcd. J: Principal coordinate biplot (the same ordination as on Fig. 1 F) colored according to GBMcd. The purple vector shows the direction of correlation of GBMcd with the ordination. All correlations with reported R^2 are significant at the $p<0.0001$ level.

If plastic changes in GBM were transmitted across generations, one would expect constitutive GBM differences between Orpheus- and Keppel-origin corals to parallel the GBM response induced upon transplantation. This was not the case: constitutive differences between Orpheus and Keppel (Fig. 3 A) tended to be in the opposite direction compared to plastic changes (Fig. 2 A), did not predominantly affect genes at the opposite ends of MBD-score

spectrum (compare density curves on Fig. 2 A and Fig. 3 A), and were positively rather than negatively associated with differences in GE (Fig. 3 B, in contrast to Fig. 2 F). In addition, genes showing significant constitutive differences in GBM tended to be F_{ST} outliers (Fig. 3 C) unlike genes undergoing significant plastic changes (Fig. 2 E). This indicates that the mechanism giving rise to constitutive GBM changes between populations is unrelated to the mechanism of plasticity and is most likely genetic divergence.

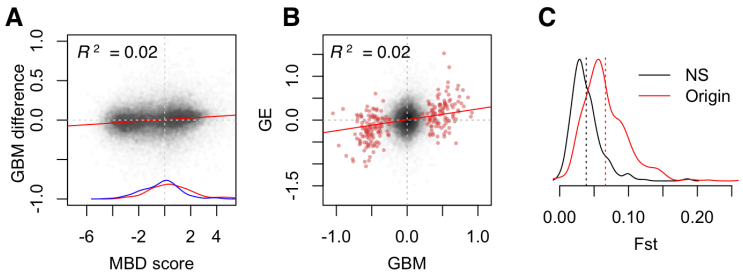

Figure 3. Constitutive $\log_2$ fold-differences in gene body methylation (GBM) and gene expression (GE) between populations (Orpheus relative to Keppel). A: Difference in GBM plotted against MBD-score for each gene. Red and blue curves by the x-axis are density plots of significantly up- and down-regulated genes. B: Correlation between GBM and GE differences. Red points are genes significantly different in GBM at 10% FDR level. Note that, unlike correlations among transplantation-induced changes (Fig. 2 F), this correlation is positive. E: Density plot of between-population F_{ST} for genes showing significant constitutive difference in GBM with respect to the place of coral's origin (red line) compared to 500 randomly chosen non-significant genes (black line). All correlations with reported R^2 are significant at $p<0.0001$ level.

To see whether plastic GBM changes were related to acclimatization (rather than, for example, overall stress in response to transplantation) we used Differential Analysis of Principal Components (DAPC, Fig. 4 A-C) to test whether GBM similarity to the local population (Fig. 4 D) predicted the transplanted fragment's fitness in the new environment. Plastic GBM changes generally followed the differences between locals (Fig. S8) and the resulting "GBM similarity to locals" was highly correlated with the first principal component of measured fitness proxies, the most strongly with the weight gain (Fig. 4 E). This correlation held both before and after (as shown on Fig 4 E) controlling for overall higher fitness of Orpheus corals transplanted to Keppel. This result indicates that a coral's GBM profile, despite being highly variable among individuals (Fig 1 F, G), also reflects adaptation and acclimatization to local environmental conditions. Interestingly, the same analysis performed for GE and genotypes (GT) failed to reveal their relationship with fitness (Fig. 4 E). One explanation of this could be the difference in time scales on which the effects of GE and GT on fitness would be detectable: compared to the three-

month length of our experiment, GE might be linked to fitness on shorter time scales, and GT – on longer time scales.

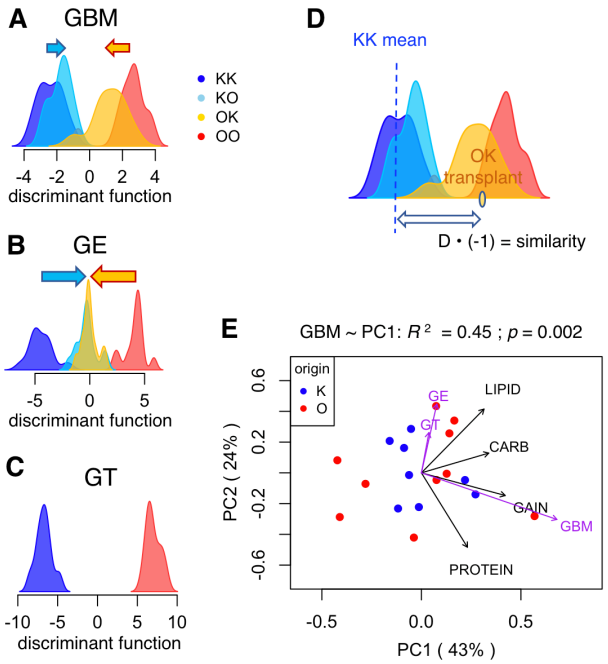

Figure 4. Relationship between gene body methylation (GBM), gene expression (GE), and genotype (GT) with fitness proxies in corals transplanted into novel environment. A-C: Density plots of discriminant function values for each variable. The functions were developed to distinguish between coral fragments in their native environment (KK vs. OO) and applied to transplanted fragments (KO and OK) to quantify their adaptive plasticity (ability to shift towards local values). Arrows indicate mean plasticity. D: Calculation of "similarity to locals" for each transplanted fragment: it is the inverse of the standardized distance between the fragment's discriminant function value and the locals' mean. E: Principal components biplot of fitness proxies in fragments transplanted to another reef, corrected for the overall higher fitness of Orpheus corals transplanted to Keppel. Fit of "similarity to locals" for GBM, GE and GT onto this ordination is shown as purple vectors scaled by correlation coefficient. Only GBM fit is statistically significant.

Correlation between plastic GBM changes and gene expression changes (Fig. 2 F - H) prompts the question: which one is likely to be the leading cause? The magnitude of GBM changes is considerably lower than changes in GE (Fig. 2 C, D), which suggests that GE might be the primary driver. However, considering the simplicity of plastic GBM changes - genome-wide exaggeration or mitigation of GBM disparity among genes - it is more parsimonious to assume that this is the leading change driven by the environment, which then translates into broad-scale functional shifts in gene expression (Fig. 2 H and Fig. S7). In this way, genome-wide

adjustment of gene activity could be achieved by modulating a single process that governs the disparity in GBM among genes, which could be likened to coarsely adjusting a complex multi-frequency audio output using a single-knob equalizer. We propose that mediating such coarse adjustment of genome-wide gene expression in response to the environment is the ecological function of GBM.

References

1. Zilberman, D. An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol.* **18,** 87 (2017).
2. Bewick, A. J. & Schmitz, R. J. Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.* **36,** 103–110 (2017).
3. Sarda, S., Zeng, J., Hunt, B. G., Yi, S. V & Soojin, V. The Evolution of Invertebrate Gene Body Methylation. *Mol Biol Evol* **29,** 1907–1916 (2012).
4. Roberts, S. B. & Gavery, M. R. Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? *Front. Physio.* **2,** 1–5 (2012).
5. Hofmann, G. E. Ecological Epigenetics in Marine Metazoans. *Front. Mar. Sci.* **4,** 4 (2017).
6. Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8,** 1499–1504 (1980).
7. Francioli, L. C. *et al.* Genome-wide patterns and properties of de novo mutations in humans. *Nat. Genet.* **47,** 822–826 (2015).
8. Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466,** 253–7 (2010).
9. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543,** 72–77 (2017).
10. Jjingo, D. *et al.* On the presence and role of human gene-body DNA methylation. *Oncotarget* **3,** 462–474 (2012).
11. Feil, R. & Fraga, M. F. Epigenetics and the environment: emerging patterns and implications. *Nat. Rev. Genet.* **13,** 97 (2012).
12. Yan, H. *et al.* DNA Methylation in Social Insects: How Epigenetics Can Control Behavior and Longevity. *Annu. Rev. Entomol.* **60,** 435–452 (2015).
13. Liew, Y. J. *et al.* Epigenome-associated phenotypic acclimatization to ocean acidification in a reef-building coral. *Sci. Adv.* **4,** eaar8028 (2018).
14. Gugger, P. F., Fitz-Gibbon, S., Pellegrini, M. & Sork, V. L. Species-wide patterns of DNA methylation variation in Quercus lobata and their association with climate gradients. *Mol. Ecol.* **25,** 1665–1680 (2016).
15. Keller, T. E., Lasky, J. R. & Yi, S. V. The multivariate association between genomewide DNA methylation and climate across the range of *Arabidopsis thaliana*. *Mol. Ecol.* **25,** 1823–1837 (2016).
16. Hu, J. & Barrett, R. D. H. Epigenetics in natural animal populations. *J. Evol. Biol.* (2017). doi:10.1111/jeb.13130

17. Kawecki, T. J. & Ebert, D. Conceptual issues in local adaptation. *Ecol. Lett.* **7,** 1225–1241 (2004).
18. Lønborg, C. *et al. Marine Monitoring Program: Annual report for inshore water quality monitoring 2014 to 2015*. (Australian Institute of Marine Science and JCU TropWATER, 2016). at <http://elibrary.gbrmpa.gov.au/jspui/handle/11017/3050>
19. Matz, M. V., Treml, E. A., Aglyamova, G. V. & Bay, L. K. Potential and limits for rapid genetic adaptation to warming in a Great Barrier Reef coral. *PLOS Genet.* **14,** e1007220 (2018).
20. Van Oppen, M. J. H., Peplow, L. M., Kininmonth, S. & Berkelmans, R. Historical and contemporary factors shape the population genetic structure of the broadcast spawning coral, Acropora millepora, on the Great Barrier Reef. *Mol. Ecol.* **20,** 4899–4914 (2011).
21. Meyer, E., Aglyamova, G. V. & Matz, M. V. Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Mol. Ecol.* **20,** (2011).
22. Dixon, G. B. ., Bay, L. K. & Matz, M. V. Evolutionary Consequences of DNA Methylation in a Basal Metazoan. *Mol. Biol. Evol.* **33,** 2285–2293 (2016).

Materials and Methods

Sample sizes

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

Reciprocal Transplantation Experiment

Field work was conducted with permission from the Great Barrier Reef Marine Park Authority (Research permit G09/29894.1) as described previously [1]. Reciprocal transplantations were undertaken between two environmentally distinct study sites (Miall Island in the Keppel Island group: 23°09S 150°54E and Hazard Bay on Orpheus Island 18°37S 146°29E) separated by 4.5 degrees of latitude on the Great Barrier Reef (Figure 1A). On the 23rd of April (Orpheus) and 4th of May (Keppels) 2010 fifteen colonies were collected from wild populations from each site and split in two. One half of each colony was replaced in its native habitat, while the second half was transplanted to the alternate study site. Samples from all coral fragments were collected at midday after three months (9th July 2010 at Orpheus, 14th July at Keppels) frozen in liquid nitrogen, then transferred into RNAlater (Ambion, Austin, TX, USA) for gene expression and DNA methylation profiling.

Coral growth rate

Total colony skeletal growth was evaluated using wet buoyant weight as in Jokiel, Maragos, & Franzisket (1978). Measurements from beginning and end of the experiment were expressed as percent daily weight gain.

Coral condition

The energetic condition of corals was determined by the analysis of total protein, carbohydrate and lipid content. Coral tissue was removed from an average of 8.6 cm^2 of coral branch (mean $\pm$ SE = 0.18) using an air gun in 12 – 15 ml of 0.2 µM filtered sea-water (FSW). The coral slurry was homogenized and centrifuged at 1500 g for 5 min at 4°C to separate coral and endosymbiotic algal components. The coral slurry was aliquoted for protein, carbohydrate and lipid assays and frozen at -20°C (10 ml for lipid, 0.5 ml for protein and carbohydrate). The algal

component was re-suspended in 2.75 ml of 0.2 µM FSW. Coral tissue surface area was determined from cleaned (10% bleach for 24 hrs) skeletons of water-blasted samples using a twice dip paraffin wax method [3]. Surface area of coral branches was determined by the weight gain of the second wax layer from a session specific standard curve (R^2 were 0.97 – 0.98) using seven nylon cylinders of known surface area (5 – 48 cm^2).

Protein analysis

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

Carbohydrate analysis

Total carbohydrate content estimates were obtained from the average of three technical replicates of 50 µL coral slurry using D-glucose as a standard [4]. Blank water samples and BSA protein standards (20, 50, 100, 200, 500 and 1000 µg/ml) were run in triplicate on each plate and absorbance was read at 485 nm on a spectrophotometer. Standard curves had R^2 of 0.96 – 0.99 and total carbohydrate content per sample was expressed in average mg of the technical replicates per cm^2 of coral branch surface area after quality control as described for proteins.

Lipid analysis

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

MBD-seq library preparation

DNA was isolated from adult holobiont tissue using dispersion buffer (4 M guanadine thiocyanate, 30 mM sodium citrate, 30 mM β-mercaptoethanol) followed by phenol chloroform purification and a final cleanup with Zymo Genomic DNA Clean and Concentrator-10 kit (Catalog No D4011). Genomic DNA was sheared using a Misonix Sonicator 3000 to a size range of ~200 to 800 bp checked by gel electrophoresis. Enrichment reactions were performed using the MethylCap kit (Diagenode Cat. No. C02020010) with an initial input of 2 μg of sheared DNA per reaction. The methylated fraction was eluted from the capture beads in a single step using High Elution Buffer. Library preparation using NEBNext® DNA library Prep Master Mix Set (E6040L) and sequencing on an Illumina HiSeq4000 was performed at the University of Texas Genome Sequencing and Analysis Facility. Our total sample size was N=44 (22 colonies divided in half giving 11 samples per treatment group). For the majority of these we sequenced only the enriched library eluted from the capture beads. Fold coverages from these captured libraries were used to estimate relative differences in GBM between samples. For a subset of 12 of the 44 samples, we sequenced both the captured and flow-through fractions. Fold differences between these captured and flow-through libraries were used to estimate absolute levels of methylation across genes. We did this for only a subset of samples because we were primarily interested in relative differences in GBM between groups. As relative differences could be reasonably well assessed without sequencing the flow-through (Figure S2), we chose to focus our sequencing resources on increasing sample size rather than more thorough estimates of absolute methylation levels.

Repeated elements annotation

De novo identification of repeated elements in *A. digitifera* genome was performed using RepeatModeler (http://www.repeatmasker.org/RepeatModeler/, version 1.0.10). RepeatModeler launches two repeat finding programs, RECON 1.05 [6] (http://eddylab.org/software/recon/) and RepeatScout 1.0.5 [7] (https://bix.ucsd.edu/repeatscout/), to generate a genomic database and upon which to build, refine and classify consensus models of putative interspersed repeats. This pipeline produced a total of 2,858 non-redundant repeat families in recent *A. digitifera* genome sequence. These sequences were then used as the custom library to search the *A. digitifera* genome using RepeatMasker (http://www.repeatmasker.org/). Approximate 7.33% and 5.47% of the genome was annotated to be composed of retrotransposon and DNA transposon elements, respectively.

MBD-seq data processing

Forty four MBD-seq libraries were prepared from reciprocally transplanted coral fragments. Sequencing produced 980 million raw reads with a mean of 16 ± 0.65 SEM million reads per sample. Raw reads were trimmed of non-template sequence using Cutadapt [8] and quality filtered using Fastx toolkit (http://cancan.cshl.edu/labmembers/gordon/fastx_toolkit/). Adapter trimming and quality filtering reduced these the total read count 940 million, mean = 15 ± 0.64 SEM million reads per sample. The reference genome and annotations (version 1.1) for *Acropora digitifera* [9] were downloaded from NCBI. Trimmed and filtered reads were mapped to this concatenated reference using Bowtie2 [10]. To ensure that using a reference from an alternate species did not severely impair mapping, we compared the mean mapping efficiency against the *A. digitifera* reference with that of a draft genome sequence for *A. millepora* produced by David Miller and coworkers (James Cook University). Mean mapping efficiency against the *A. digitifera* (78.4 ± 0.7% SEM) reference was only 5.2% lower than that of *A. millepora* (83.6 ± 0.8% SEM), hence sequence differences between the two species did not appear to substantially impair read mapping. Following mapping, PCR duplicates were removed using Picard (https://broadinstitute.github.io/picard/). Mean duplication frequency was 14.3 ± 0.5% SEM. The BAM files were filtered to retain only highly uniquely aligned reads (mapping quality 30, or 0.1% chance of non-unique alignment) using samtools [11], after which the number of reads overlapping various genomic regions - exon, intron, genic (exon + intron), repeated elements longer than 500b, and repeat-free intergenic regions at least 1 kb long and at least 2 kb away from any gene - were counted using BEDtools [12].

MBD-score and GBM gene classes

For 12 samples (3 per each experimental group) we quantified absolute levels of GBM as the $\log_2$ fold difference in coverage between captured and flow-through DNA fractions while controlling for genotype, as described in Dixon et al. (2016). These values were used as gene-specific "MBD scores" throughout the study. As expected based on previous studies [14,15], MBD-scores were bimodally distributed (Fig. S1 A) and negatively correlated with $CpG_{o/e}$ (Figure 1 D) providing genome-wide validation for our MBD-seq technique. The MBD-score threshold separating the two clouds of high point density in the scatter plot on Figure 1 D was used to delineate the two gene classes for computation of the "GBM class difference" (GBMcd) for each sample (Fig. 2 I, J).

Tag-seq data processing

Transcription was quantified using Tag-seq (Meyer et al. 2011; Lohman et al. 2016). The Tag-seq reads were downloaded from the SRA database (accession SRP049522; [1]) and mapped

against the *A.digitifera* genome using SHRiMP [18]. Mapped reads overlapping annotated coding sequences were counted using intersection-nonempty method in HTseq version 0.6.1p1 [19]. Normalization of raw counts and statistical analyses were performed using DESeq2 [20].

Assessing variation in gene body methylation and transcription

As with gene expression analyses, normalization and statistical analyses of MBD-seq reads were performed with DESeq2 [20]. To test for origin effects, we performed two tests that compared the two groups placed at Keppel to each other (OK vs KK) and the two groups placed at Orpheus to each other (OO vs KO). These tests were intended to identify effects of origin while controlling for environmental conditions experienced during the experiment. To assess effects of transplantation, we compared groups that originated from Orpheus to each other (OO vs OK) and groups that originated from Keppel to each other (KK vs KO). For these tests, we included an additional parameter of colony identity to identify effects of transplantation while controlling for genotype. Results from these tests were used to identify genes showing evidence of plasticity that were input into downstream analyses. For TagSeq, only genes with mean read count ≥3 were considered for analysis (19706 genes), and for MBD-seq we chose genes with mean read count ≥20 (27084 genes).

Discriminant analysis of principal components

GBM data were further analyzed using discriminant analysis of principal components (DAPC) implemented in the R package adegenet [21], following the procedures outlined in Kenkel & Matz (2016). DAPC is a multivariate analysis method designed to identify between-group variation while neglecting within-group variation. We used this method to distill our highly multivariate MBD-seq dataset into single axes that maximized discrimination between natives from the two experimental sites (KK and OO samples). The result was a discriminant axis contrasting the GBM variation of the two native populations—with one pole designating native Keppel-like GBM patterns and the opposite pole designating native Orpheus-like patterns. The function was then applied to data from the transplanted samples, so that their positions along discriminant axis described their similarity in GBM patterns to the contrasted native populations. It seems plausible that, either due to natural selection or plasticity, native corals would possess patterns optimal for their particular site. If this is true, then transplants with DAPC values closer those of local corals would be expected to show improved fitness proxies (eg: samples transplanted to Keppel with DAPC values closer to the 'Keppel-like' pole of the discriminant axis would be expected to show higher fitness proxies and vice versa for samples transplanted to Orpheus). 'DAPC Similarity' values describing each transplant's proximity along the discriminate axis to natives of their transplantation site (Fig. 4 D) were computed by taking the absolute value of the difference between the transplants' DAPC values and the mean value

for natives of their respective transplant sites, converting these distances into z-scores, and multiplied them by -1. The DAPC Similarity value for sample X was:

$$DAPC\ Similarity_x = -1 \times (|D_x| - \overline{D})/\sigma$$

where, D_x is its distance along the discriminant axis between sample X and the mean DAPC value for natives of its transplantation site, $\overline{D}$ is the mean absolute distance for all transplants, and σ is the standard deviation of absolute distance for all transplants. Correlations of these values with fitness characteristics of transplanted fragments (Fig. 4 E) were computed using function *envfit* (R package *vegan* [23])

Validation of MBD-seq results by bisulfite amplicon sequencing

To validate our MBD-seq results, we used targeted bisulfite sequencing. Genomic DNA for this procedure was extracted and purified as described above. It should be noted that as with the MBD-seq and Tag-seq, targeted bisulfite sequencing was performed from separate isolations from the same or closely adjacent branch. We performed bisulfite conversion using an EZ DNA Methylation-Gold kit (Zymo Research; cat. no. D5005). Thermocycler conditions for conversion were 98°C for 10 minutes followed by 53°C for 4 hours. Primers for post-conversion amplification were designed using Bisulfite Primer Seeker (Zymo Research; http://www.zymoresearch.com/tools/bisulfite-primer-seeker). Primer sequences are given in Table S1. We designed 13 primer sets to target coding sequences (excised from the *A. digitifera* reference genome) of 13 separate loci. Primer sequences included 5' tails for downstream addition of barcodes and Illumina adapters by PCR. Target loci were selected based on either strong origin or transplant effects in the DESeq2 analyses. Amplification from bisulfite converted samples was performed using EpiMark Hot Start Taq DNA Polymerase (New England Biolabs; cat. no. M0490S). Thermocycler conditions were 95°C for 30 seconds followed by 35-40 cycles of 95°C for 15 seconds, 53°C for 30 seconds, 68°C for 30 seconds. Barcodes and adapters for multiplexed sequencing were added to the resulting PCR products in a second PCR. The oligonucleotide sequences used for barcoding were the same as those given in the Tag-seq library preparation (Meyer et al. 2011; https://github.com/z0on/tag-based_RNAseq). Sequencing was performed using paired-end 600 cycle runs on the Illumina Miseq. Resulting reads were quality trimmed using cutadapt [8]. Quantification of CpG methylation was performed using Bismark [24] using a fasta file of the 13 exon sequences used to design the primers as a reference. For a given sample, CpG sites represented by fewer than 50 reads were excluded from the analysis (2022 low coverage calls excluded). The difference in means based on origin

for all CpGs in each gene correlated with the difference in mean normalized counts from the MBD-seq reads (Figure 1 D).

Primers used for targeted bisulfite sequencing. Each primer includes a 5' tail (bold) for amplification with barcoded primers for multiplex Illumina sequencing included in the Tag-seq 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 [9] using Bowtie2 [10]. 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 [25] 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 origin or transplant effect and a random subset of 500 non-significant genes (rather than whole genome, to speed up computation) were determined using ANGSD based on site frequency spectra as priors, which were generated for every gene subset without using filters that would affect representation of alleles of different frequencies (-uniqueOnly 1 -remove_bads 1 -minMapQ 20 -minQ 30 -baq 1 -minInd 4).

Determination of relative proportions of Symbiodinium *clades*

The MBD-seq reads were mapped to the concatenated reference including A. digitifera genome and four *Symbiodinium* sp. trasncriptomes from four different genotypic groups ("clades"). Transcriptomes for *Symbiodinium* clades A and B were from [26] and transcriptomes for clades C and D were from [27]. We then counted the relative proportions of reads producing highly unique matches (mapping quality 40 or higher) to each *Symbiodinium* transcriptome, using a custom perl script *zooxType.pl*. All corals were found to be dominated by *Symbiodinium* clade C.

Data and scripts availability

TagSeq, MBD-seq, and amplicon-bisulfite sequencing datasets are available through NCBI Short Read Archive (Project Accession: SRP049522). Scripts and traits data are available in a GitHub repository dedicated to this paper (https://github.com/grovesdixon/reciprocal_transplant_methylation).

References

1. Dixon, G. B., Bay, L. K. & Matz, M. V. Bimodal signatures of germline methylation are linked with gene expression plasticity in the coral Acropora millepora. *BMC Genomics* **15,** 1109 (2014).
2. Jokiel, P., Maragos, J. & Franzisket, L. in *Coral reefs: research methods* (eds. Stoddart, D. R.

& Johannes, R. E.) 529–542 (UNESCO, 1978).
3. Stimson, J. & Kinzie, R. A. The temporal pattern and rate of release of zooxanthellae from the reef coral Pocillopora damicornis (Linnaeus) under nitrogen-enrichment and control conditions. *J. Exp. Mar. Bio. Ecol.* **153,** 63–74 (1991).
4. Masuko, T. *et al.* Carbohydrate analysis by a phenol-sulfuric acid method in microplate format. *Anal. Biochem.* **339,** 69–72 (2005).
5. Harland, A. D., Navarro, J. C., Davies, P. S. & Fixter, L. M. Lipids of some Caribbean and Red Sea corals: total lipid, wax esters, triglycerides and fatty acids. *Mar. Biol.* **117,** 113–117 (1993).
6. Bao, Z. & Eddy, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* **13,** 1269–1276 (2003).
7. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21,** (2005).
8. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17,** 10–12 (2011).
9. Shinzato, C. *et al.* Using the Acropora digitifera genome to understand coral responses to environmental change. *Nature* **476,** 320–3 (2011).
10. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
11. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–9 (2009).
12. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).
13. Dixon, G. B. ., Bay, L. K. & Matz, M. V. Evolutionary Consequences of DNA Methylation in a Basal Metazoan. *Mol. Biol. Evol.* **33,** 2285–2293 (2016).
14. Sarda, S., Zeng, J., Hunt, B. G., Yi, S. V & Soojin, V. The Evolution of Invertebrate Gene Body Methylation. *Mol Biol Evol* **29,** 1907–1916 (2012).
15. Dixon, G. B., Bay, L. K. & Matz, M. V. Evolutionary Consequences of DNA Methylation in a Basal Metazoan. *Mol. Biol. Evol.* **33,** (2016).
16. Meyer, E., Aglyamova, G. V. & Matz, M. V. Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. *Mol. Ecol.* **20,** (2011).
17. Lohman, B. K., Weber, J. N. & Bolnick, D. I. Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Mol. Ecol. Resour.* **16,** 1315–1321 (2016).
18. Rumble, S. M. *et al.* SHRiMP: Accurate mapping of short color-space reads. *PLoS Comput. Biol.* **5,** 1–11 (2009).
19. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).
20. Love, M., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* (2014). at <http://www.biomedcentral.com/content/pdf/s13059-014-0550-8.pdf>
21. Jombart, T., Devillard, S. & Balloux, F. Discriminant analysis of principal components: a

new method for the analysis of genetically structured populations. *BMC Genet.* **11,** 94 (2010).
22. Kenkel, C. & Matz, M. V. Enhanced gene expression plasticity as a mechanism of adaptation to a variable environment in a reef-building coral. *Nat. Ecol. Evol.* **1,** 0014 (2016).
23. Oksanen, J. *et al.* vegan: Community Ecology Package. *R package ver. 2.4–3* (2017). doi:10.4135/9781412971874.n145
24. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27,** 1571–1572 (2011).
25. Korneliussen, T., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15,** 356 (2014).
26. Bayer, T. *et al.* Symbiodinium transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PLoS One* **7,** e35269 (2012).
27. Ladner, J. T., Barshis, D. J. & Palumbi, S. R. Protein evolution in two co-occurring types of Symbiodinium: an exploration into the genetic basis of thermal tolerance in Symbiodinium clade D. *BMC Evol. Biol.* **12,** 217 (2012).

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

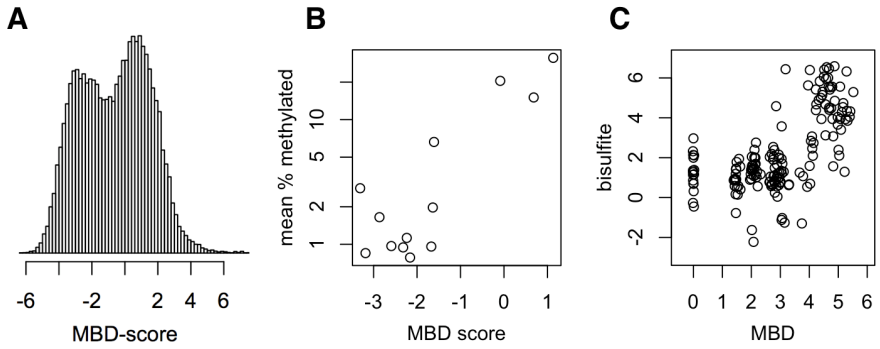

Figure S1. Distribution of MBD-scores (A) and validation by bisulfite sequencing of amplicons (B, C). B: Mean percent methylation was calculated as the proportion of methylated CpG sites within each gene averaged across all samples. C: Correlation between MBD-based and bisulfite-based measurement of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

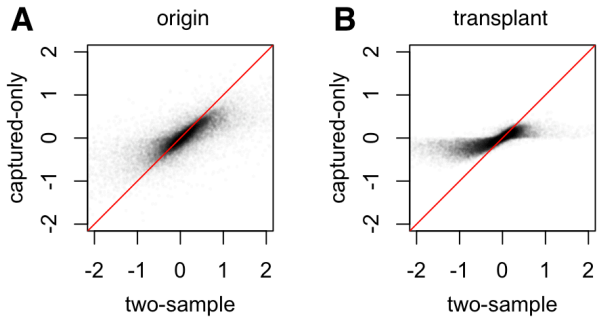

Figure S2 Correlation between inferred differences in GBM depending on origin (A) and transplant site (B) based on comparison of MBD-captured and flow-through libraries (x-axis) and only on MBD-captured fraction (y-axis). These plots were generated for a subset of 12 samples (three for each of the four experimental groups, KK, KO, OK, and OO) for which both the captured and the flow-through fractions were sequenced.

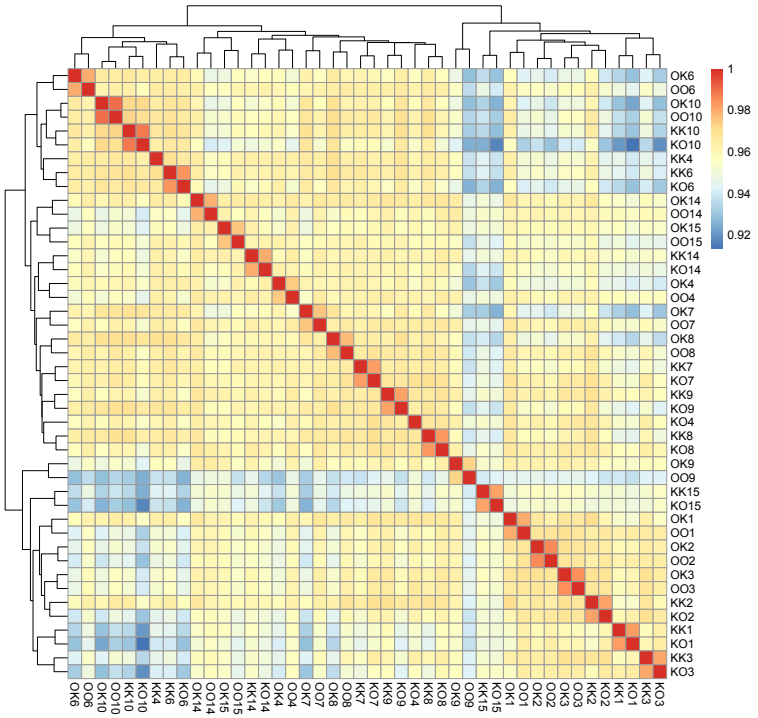

Figure S3 Heatmap of GBM correlations among samples, illustrating strong dependence of GBM on genotype. Colors indicate Spearman's rank correlations for normalized MBD-seq read counts across all coding genes (N = 24853); the lowest observed correlation value was 0.9. Samples were clustered by maximum 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 the same number are clonal fragments from the same colony. All of the 22 clone pairs except one (sample pair KK4 and KO4) clustered together.

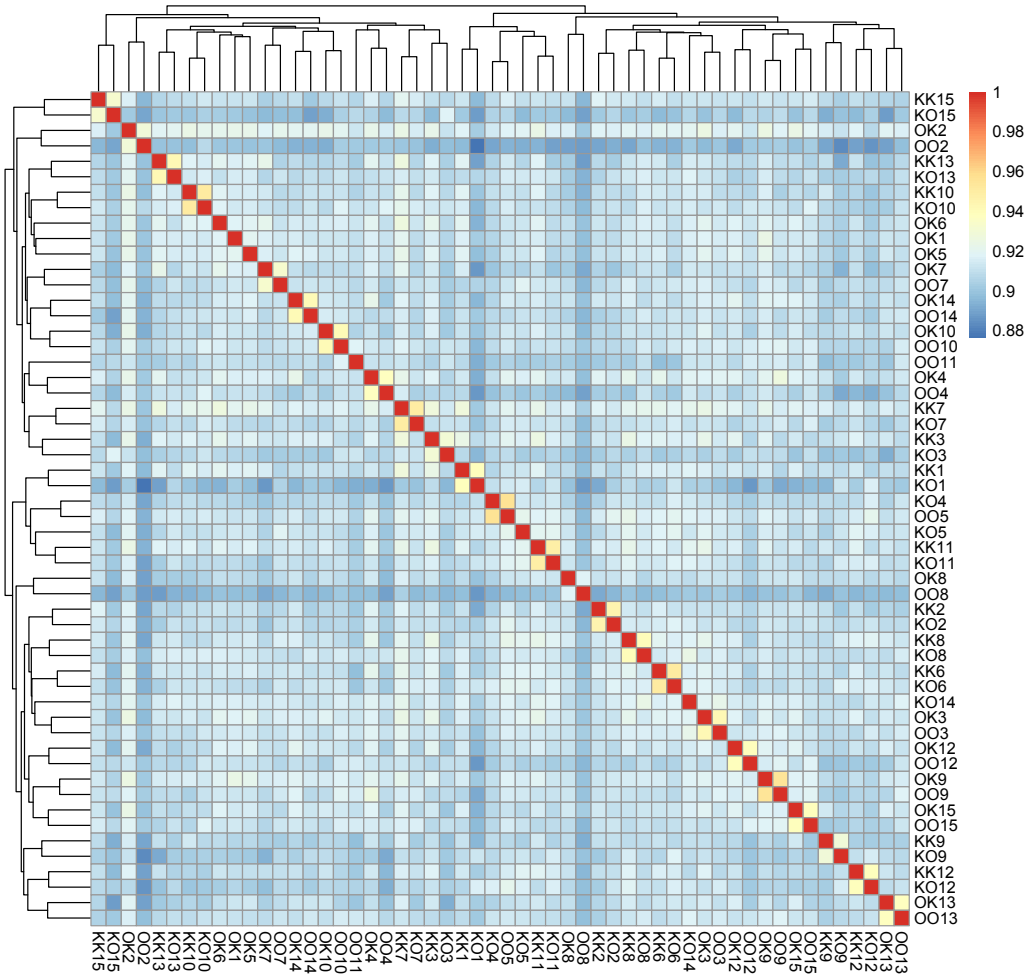

Figure S4 Heatmap of overall correlations in transcription illustrating strong dependence of transcription on genotype. Colors indicate Spearman's correlations for normalized Tag-seq read counts across all coding genes (N=19706). Samples were clustered by maximum distance method. First letter of sample 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).

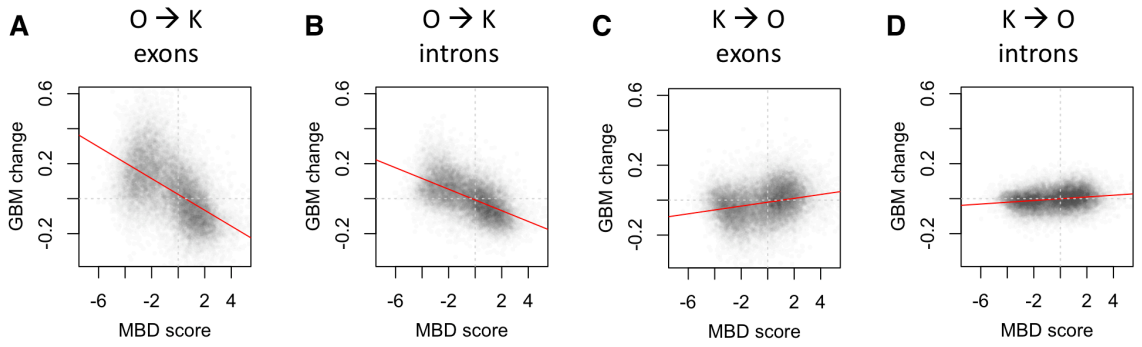

Figure S5. GBM change in response to transplantation in exons and introns.

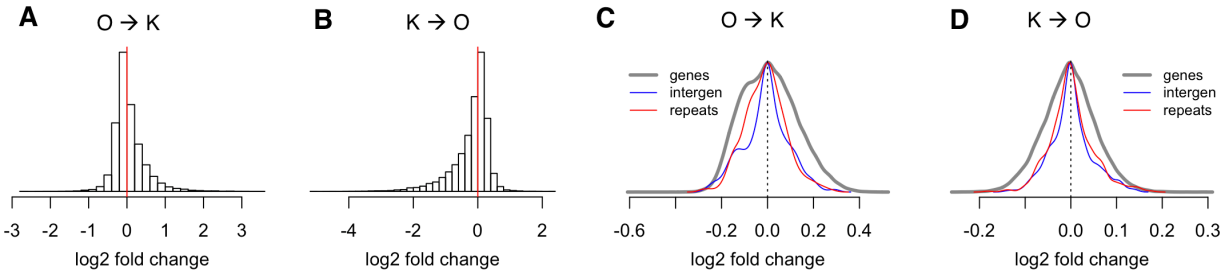

Figure S6. Absolute methylation changes in response to transplantation ("away" vs. "home") can be positive or negative across genes (A, B) and affect mostly protein-coding regions (C, D). Plots on A and B are based on comparison between captured and flow-through fractions for 12 samples (three for each of the four experimental groups, KK, KO, OK, and OO). Plots on panels C and D were generated based on captured-only analysis of all 44 MBD-seq samples (Table S1).

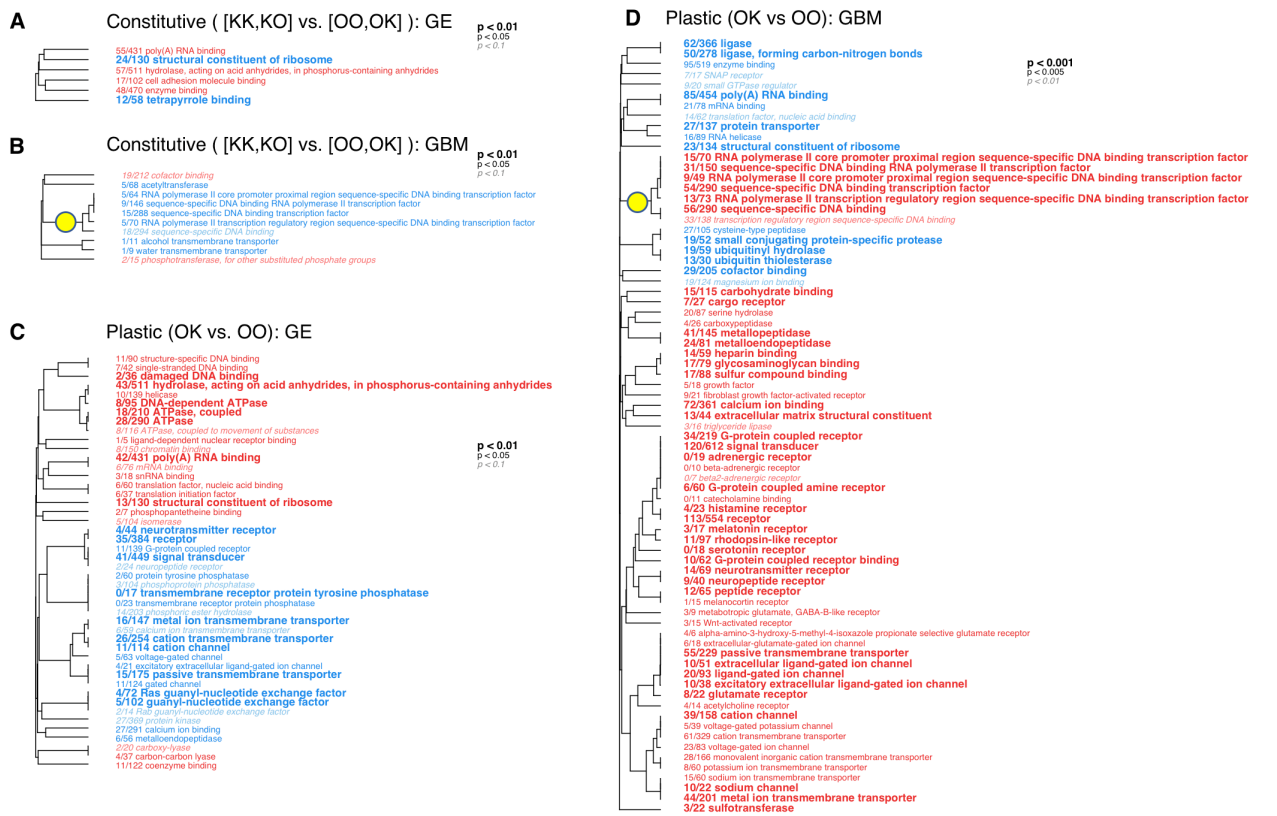

Figure S7. Gene ontology categories significantly enriched with up- or down-regulated genes (red and blue, respectively) in GE and GBM comparisons. The Benjamini-Hochberg corrected significance of enrichment is indicated by the size and type of the font (see legends). Note that for plastic GBM the significance 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[1]. 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).

1. Kosiol, C. *et al.* Patterns of positive selection in six mammalian genomes. *Plos Genet.* **4,** (2008).

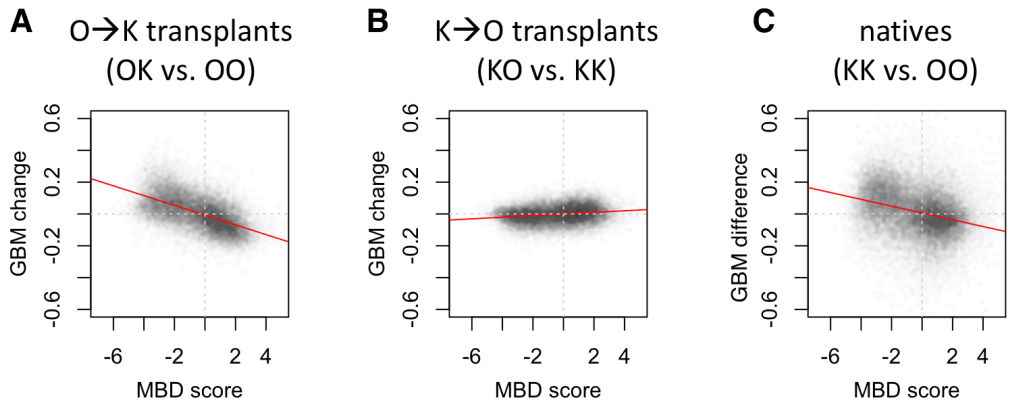

Figure S8. Plastic GBM changes in response to transplantation (A, B) broadly recapitulate the differences between native corals at the two sites (C). Panels A and B use the same data as Fig. 2 A and B, rescaled to facilitate comparison with differences between "natives" (C).