








RESOURCE ARTICLE

Invertebrate methylomes provide insight into mechanisms of environmental tolerance and reveal methodological biases

Shelly A. Trigg¹ | Yaamini R. Venkataraman^{1,2}  | Mackenzie R. Gavery³  |
 Steven B. Roberts¹ | Debashish Bhattacharya⁴  | Alan Downey-Wall⁵ |
 Jose M. Eirin-Lopez⁶  | Kevin M. Johnson^{7,8} | Katie E. Lotterhos⁵  |
 Jonathan B. Puritz⁹  | Hollie M. Putnam⁹ 

¹School of Aquatic and Fishery Sciences, University of Washington, Seattle, Washington, USA

²Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA

³Environmental and Fisheries Sciences Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, Washington, USA

⁴Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, New Jersey, USA

⁵Department of Marine and Environmental Sciences, Northeastern University, Nahant, Massachusetts, USA

⁶Environmental Epigenetics Laboratory, Institute of Environment, Florida International University, North Miami, Florida, USA

⁷Center for Coastal Marine Sciences, California Polytechnic State University, San Luis Obispo, California, USA

⁸California Sea Grant, University of California San Diego, La Jolla, California, USA

⁹Department of Biological Sciences, University of Rhode Island, Kingston, Rhode Island, USA

Correspondence

Hollie M. Putnam, Department of Biological Sciences, University of Rhode Island, Kingston, RI, USA
 Email: hputnam@uri.edu

Funding information

This work was partially supported by the USDA: National Institute of Food and Agriculture, Hatch accession number 1017848 to H.M.P., NJ01180 to D.B., and the NRSP-8 National Animal Genome Research Program to H.M.P. and S.B.R. This work was also supported by funding from the National Science Foundation: 1756623 and 1921465 to H.M.P., 1756616 to D.B., 1921149 to S.B.R., 1921402 to J.E.L., and 1635423 to K.E.L.

Abstract

There is a growing focus on the role of DNA methylation in the ability of marine invertebrates to rapidly respond to changing environmental factors and anthropogenic impacts. However, genome-wide DNA methylation studies in nonmodel organisms are currently hampered by a limited understanding of methodological biases. Here, we compare three methods for quantifying DNA methylation at single base-pair resolution—whole genome bisulfite sequencing (WGBS), reduced representation bisulfite sequencing (RRBS), and methyl-CpG binding domain bisulfite sequencing (MBDBS)—using multiple individuals from two reef-building coral species with contrasting environmental sensitivity. All methods reveal substantially greater methylation in *Montipora capitata* (11.4%) than the more sensitive *Pocillopora acuta* (2.9%). The majority of CpG methylation in both species occurs in gene bodies and flanking regions. In both species, MBDBS has the greatest capacity for detecting CpGs in coding regions at our sequencing depth, but MBDBS may be influenced by intrasample methylation heterogeneity. RRBS yields robust information for specific loci albeit without enrichment of any particular genome feature and with significantly reduced genome coverage. Relative genome size strongly influences the number and location of CpGs detected by each method when sequencing depth is limited, illuminating nuances in cross-species comparisons. As genome-wide methylation differences, supported by data across bisulfite sequencing methods, may contribute to environmental sensitivity

Shelly A. Trigg and Yaamini R. Venkataraman contributed equally.

phenotypes in critical marine invertebrate taxa, these data provide a genomic resource for investigating the functional role of DNA methylation in environmental tolerance.

KEYWORDS

bisulfite sequencing, coral, epigenetics, marine invertebrate

1 | INTRODUCTION

Environmental stimuli interact with genomic content to drive variation in gene and protein expression, resulting in phenotypic plasticity. This plasticity has the potential to buffer against mortality under environmental change (Baldwin, 1902), or conversely be maladaptive (Velotta et al., 2018). Furthermore, plasticity may enhance or diminish evolutionary rates (Ghalambor et al., 2007), which is particularly relevant to plasticity–evolution feedbacks (Ghalambor et al., 2007, 2015; Kronholm & Collins, 2016). This is of particular concern in the Anthropocene (Lewis & Maslin, 2015), as global change exacerbates the mismatch between phenotype and a rapidly changing environment.

The increase in negative global climate change consequences have prompted an intensification of research into phenotypic plasticity, gene regulation, and epigenetic mechanisms in nonmodel marine invertebrates (as reviewed by Eirin-Lopez & Putnam, 2019; Hofmann, 2017; Roberts & Gavery, 2012). Specifically, carryover effects and cross- and multigenerational plasticity in response to climate change (Byrne et al., 2020) may be generated by epigenetic regulation of gene expression (Dixon et al., 2018; Liew et al., 2018, 2020). As epigenetic research has increased, there has been a focus on DNA methylation: the addition of a methyl group on the cytosine residues in the genome often in the cytosine phosphate guanine (CpG) context (Zemach et al., 2010). DNA methylation has gene expression regulation capacity through the interaction of base modification with transcriptional elements. Early bulk enzyme-based and fingerprinting methods for quantifying DNA methylation in marine invertebrates provided initial insights into DNA methylation and organismal phenotypic plasticity in response to environmental changes (Dimond et al., 2017; Gavery & Roberts, 2010; Gonzalez-Romero et al., 2017; Putnam et al., 2016; Riviere et al., 2013; Rodriguez-Casariago et al., 2018; Suarez-Ulloa et al., 2018).

Nonsequencing approaches that quantify global or bulk DNA methylation, such as colorimetric or fluorescence ELISAs (enzyme-linked immunosorbent assays; Dimond et al., 2017; Gavery & Roberts, 2010; Putnam et al., 2016; Riviere et al., 2013; Rodriguez-Casariago et al., 2018) are low-cost, rapidly applied, and do not require genomic resources to generate information on the responsiveness of the methylome. These global estimates do not, however, fully capture local changes in DNA methylation across different genome regions. Specifically, differences in the location and amount of methylation in two samples or treatments could lead to biased conclusions when based on average per cent methylation at the bulk level. Consequently, nonsequencing methods are limited in their

ability to elucidate specific mechanisms of expression regulation and thus are unable to fully address the functional implications of methylation-driven regulation within the genome. In contrast, the use of genome-wide approaches that provide single base-pair resolution allow the testing of hypotheses regarding spurious transcription, alternative splicing and exon skipping (Roberts & Gavery, 2012). For example, the use of whole genome bisulfite sequencing (WGBS) to investigate the role of DNA methylation in regulating genes involved in caste specification in honeybees identified differential methylation in an exon of the *anaplastic lymphoma kinase* (*ALK*) gene; this exon was differentially retained in a splice variant between queens and workers (Foret et al., 2012). Thus, there is a clear need for single base-pair assessment of DNA methylomes facilitated by next generation sequencing to more fully elucidate the relationship of DNA methylation and gene expression in nonmodel invertebrates.

Genome-wide levels of DNA methylation can be estimated by several bisulfite conversion and sequencing approaches. Bisulfite conversion of DNA results in the deamination of unmethylated cytosine to uracil, which leaves a base change signature in the DNA that can be tracked via sequence comparison between bisulfite-converted samples and reference genomes. While the number of bisulfite sequencing approaches are expanding (e.g., epiGBS; van Gurp et al., 2016), the widely used approaches are WGBS, reduced representation bisulfite sequencing (RRBS) and, more recently, methyl-CpG binding domain bisulfite sequencing (MBDBS). WGBS is considered to be the gold-standard of bisulfite sequencing because it provides full coverage of the genome (given deep sequencing coverage) and the capacity to detect the entire methylome at single base-pair resolution.

While providing a comprehensive approach, the high cost of WGBS is juxtaposed against the often very small fraction of methylated DNA in invertebrate genomes (Tweedie et al., 1997). Alternatively, approaches such as RRBS also use bisulfite conversion to quantitatively assess DNA methylation with base-pair resolution. RRBS incorporates a restriction digestion of the genome to enrich for CpG-rich regions, and was designed to enrich for promoters and other genomic regions containing CpG islands because they have important regulatory functions in mammals (Meissner et al., 2008); however, applications of RRBS in more basal vertebrates, such as fish, report this method is less biased toward CpG islands (Chatterjee et al., 2013). This is a more cost-effective approach as it only sequences a small portion of the genome, but requires restriction enzyme recognition sites near other CpGs to gather high-resolution data. Since DNA methylation in invertebrates is primarily limited

to coding regions (Dixon et al., 2018; Flores et al., 2012; Roberts & Gavery, 2012), it is less clear whether enrichment of CG-rich DNA using RRBS will enrich for informative or regulatory regions of invertebrate genomes.

In contrast to the CpG-rich, region-specific targeting of RRBS, MBDBS uses methyl binding domain proteins to target and enrich methylated CpGs, then employs bisulfite conversion to provide single base-pair resolution of DNA fragments enriched for methylated regions. Many marine invertebrate genomes consist of highly methylated regions that are distributed in a mosaic pattern throughout predominantly unmethylated DNA (Suzuki et al., 2007). When invertebrate methylomes have been characterized, these highly methylated regions overlap with gene bodies and have been shown to play a role in gene expression activity (Roberts & Gavery, 2012). Therefore, using an enrichment approach such as MBDBS to isolate gene body methylation can be a cost-effective and gene body-focused alternative to WGBS or RRBS (Gavery & Roberts, 2013; Venkataraman et al., 2020). The base-pair resolution and ability to quantify loci methylation offered by the combination of MBD enrichment and BS conversion is an advantage compared to MBD sequencing alone (Dixon & Matz, 2021), as the latter assumes that methylation level is proportional to read depth. In contrast to WGBS or RRBS, the quantification and interpretation of MBDBS data could be complicated by individual variation in methylation levels (e.g., one individual that has high methylation in a particular region would have data showing enrichment of that region, whereas another individual that lacks methylation in that region would have missing data there).

Given the need to assess plasticity mechanisms and the acclimatization potential of a variety of marine taxa, it is critical to compare the potential of different approaches to detect, quantify and assess DNA methylation with respect to specific biological hypotheses of interest. To this end, we studied three DNA methylation quantification approaches that provide single base-pair resolution data using bisulfite conversion and sequencing: WGBS, RRBS and MBDBS. We applied these methods to two reef-building corals, *Montipora capitata* and *Pocillopora acuta*, which have different environmental sensitivity, phenotypic plasticity, inducible DNA methylation (Putnam et al., 2016), and genome sizes (Shumaker et al., 2019; Vidal-Dupiol et al., 2019). We assessed species-specific differences in genome-wide methylation and contrasted per cent methylation of common loci, gene coverage and orthologous genes across methods. Then, we compared the coverage and genomic location of CpG data generated from the three methods. Compared to WGBS, both MBDBS and RRBS have advantages and potential limitations associated with biology, genome characteristics and experimental design, highlighting the need to fully consider these aspects when evaluating DNA methylation for particular hypotheses of methylation function in invertebrates. As part of this effort, we characterized DNA methylation differences in two coral species, providing valuable insights into the epigenetic underpinnings of phenotypic plasticity in nonmodel marine invertebrates.

2 | MATERIALS AND METHODS

2.1 | Sample collection

The reef-building scleractinian coral species *M. capitata* and *P. acuta* were collected from 1 to 2 m depth on the patch reefs of Kane'ohe Bay Hawai'i under SAP 2019-60 between September 4 and 7, 2018. Corals were transported to the Hawai'i Institute of Marine Biology where they were held in outdoor tanks under ambient conditions for 15 days, then snap-frozen in liquid nitrogen and stored at -80°C until nucleic acid extraction was performed. For each of the two coral species, fragments were collected from three different individuals in ambient conditions.

2.2 | Nucleic acid extraction

Samples were removed from -80°C and small tissue fragments were clipped directly into a tube containing RNA/DNA shield (1 ml) and glass beads (0.5 mm). The tissue clippings consisted of all coral cell types and their symbionts. Samples were homogenized on a vortexer for 1 min for the thin tissue imperforate coral *P. acuta* and 2 min for the thick tissue perforate coral *M. capitata* at maximum speed to ensure tissue extraction of all cell types. The supernatant was removed and DNA was extracted using the Zymo Quick-DNA/RNA Miniprep Plus Kit and subsequently checked for quality using gel electrophoresis on an Agilent 4200 TapeStation and quantified using a Qubit. One DNA preparation was made from each of the three individuals per coral species and was subsequently divided into three aliquots for each of the three bisulfite sequencing methods (WGBS, MBDBS and RRBS) to yield a total of 18 libraries (Figure 1).

2.3 | Genome information

Previously sequenced and assembled coral genomes were used for mapping *M. capitata* (Shumaker et al., 2019) and *P. acuta* (Vidal-Dupiol et al., 2019) DNA methylation data. Both of the coral genomes have a high and similar number of predicted genes (63,227 in *M. capitata* and 64,558 in *P. acuta*). However, *P. acuta* is much smaller in size (~ 352 vs. ~ 886 Mb in *M. capitata*), has less repetition, a greater number of scaffolds (25,553 in *P. acuta* vs. 3043 in *M. capitata*), and lower genome assembly continuity (N50 is 171,375 in *P. acuta* and 540,623 in *M. capitata*).

Genome feature tracks for *M. capitata* and *P. acuta* were derived directly from the published genomes for use in DNA methylation analyses. The *M. capitata* genome annotation yielded gene (a combination of AUGUSTUS and GEMOMA predictions), coding sequence, and intron tracks (Shumaker et al., 2019). Similarly, gene (AUGUSTUS predictions), coding sequence, and intron information was obtained from the *P. acuta* genome (Vidal-Dupiol et al., 2019). Flanking regions 1000 bp upstream and downstream of annotated genes were generated with BEDTOOLS version 2.29.2 (flankBED) for each

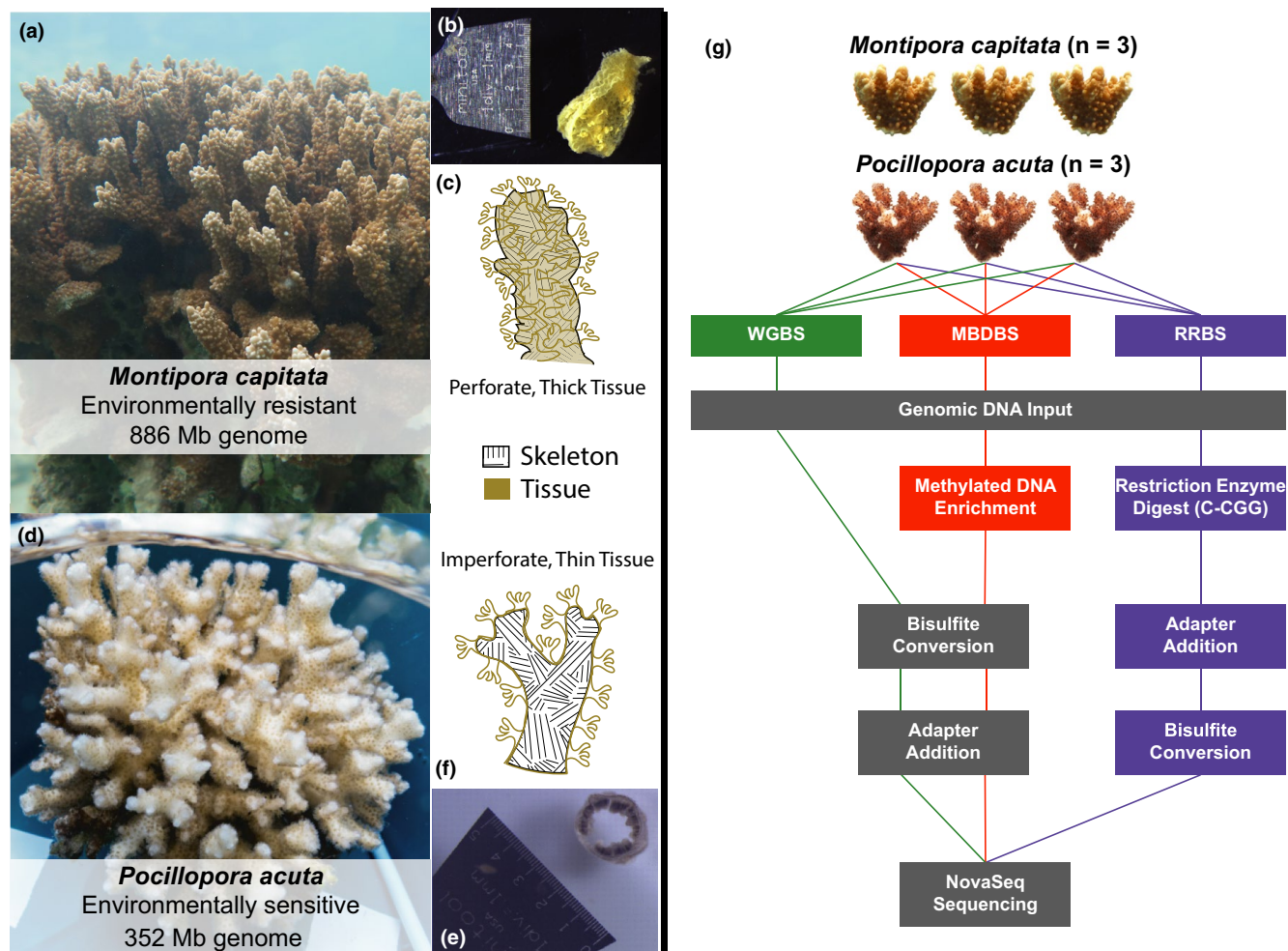


FIGURE 1 Experimental design. Three biological replicate coral samples were obtained from both coral species. (a) *Montipora capitata*, where (b) a cross-section of a decalcified fragment reveals thick tissue, and (c) a perforate tissue skeletal interaction. In contrast in (d) *Pocillopora acuta*, (e) a cross-section of a decalcified fragment reveals thin tissue, and (f) an imperforate tissue skeletal interaction. (g) DNA was extracted from each coral sample and split for use in whole genome bisulfite sequencing (WGBS), reduced representation bisulfite sequencing (RRBS), and methyl-CpG binding domain bisulfite sequencing (MBDBS) library preparation methods. Three libraries were generated for each of the three methods, yielding nine libraries for each species and 18 libraries total. Tissue photo credit: Ariana Huffmyer. Colony photo credit: Hollie Putnam and Danielle Claar

genome separately (Quinlan & Hall, 2010). Overlaps between genes and flanks were removed from up- or downstream flanking region tracks using subtractBED. Similarly, an intergenic region track was created by finding the complement of genes with complementBED, then removing any overlaps with flanking regions using subtractBED. All tracks were verified with the Integrative Genomics Viewer (Thorvaldsdóttir et al., 2013). Feature track files generated for both species are available in the project large file repository (Putnam et al., 2020).

2.4 | MBD enrichment

Before enrichment, DNA (1 μ g) in 80 μ l Tris HCl (pH 8.0) was sheared to 500 bp using a QSonica Q800R3. Samples were sonicated for

90 s, with 15 s on and 15 s off intervals at 25% amplitude. Fragment length was checked using a D5000 TapeStation System (Agilent Technologies) and samples were sonicated for an extra 15 s to shear DNA from 600 bp to 500 bp as needed.

The MethylMiner kit (Invitrogen; Cat. no. ME10025) was used to enrich for methylated DNA prior to MBDBS library generation, with 1 μ g of input DNA. The manufacturer's instructions were adhered to with the following modifications: The capture reaction containing the fragmented DNA and MBD beads was incubated with mixing at 4°C overnight, and enriched DNA was obtained with a single fraction elution using 2 M NaCl. Following ethanol addition, samples were centrifuged at 14,000 RCF (relative centrifugal force) at 1°C for 5 min. Pellets were resuspended in 25 μ l ultrapure water. Captured DNA was quantified using a Qubit dsDNA HS Kit (Invitrogen).

2.5 | MBDBS and WGBS library preparation

WGBS and MBDBS libraries were prepared using the Pico Methyl-Seq Library Prep Kit (ZymoResearch Cat. no. D5456). The manufacturer's instructions were followed with the following modifications: For each sample, 1 ng of coral DNA and 0.05 ng of *Escherichia coli* nonmethylated genomic DNA (ZymoResearch Cat. no. D5016) were used. Samples were always centrifuged at 12,000 RCF for 30 s with the exception of a 90 s centrifugation at 12,000 RCF after the second 200 μ l addition of M Wash Buffer. Warmed elution buffer (56°C) was added to each sample to increase DNA elution yield. During the second amplification cycle, 0.5 μ l of PreAmp Polymerase was added. After initial clean-up with the DNA Clean and Concentrator kit (ZymoResearch Cat. no. D4013), the first amplification step was run for eight cycles. For amplification with i5 and i7 index primers, 1 μ l of each primer (10 μ M) was used to improve amplification. The volume of the 2 \times LibraryAmp Master Mix was increased to 14 μ l to match the increase in index primer volume.

To remove excess primers from WGBS and MBDBS preparations, samples were cleaned with 11 μ l of KAPA pure beads (1 \times) (KAPA Cat. no. KK8000) and 80% ethanol. Cleaned samples were resuspended in 12 μ l of room-temperature DNA elution buffer from the Pico Methyl-Seq Library Prep Kit. Samples were re-amplified with either two or four cycles, depending on DNA concentration. Re-amplification was conducted with only 0.5 μ l of each i5 and i7 index primer (10 μ M). After re-amplification, 26 μ l of KAPA pure beads (1 \times) and 80% ethanol were used for clean-up. Final samples were resuspended in 14 μ l of room-temperature elution buffer. Primer removal and library size were confirmed by running samples on a D5000 TapeStation System.

2.6 | RRBS library preparation

Reduced representation bisulfite sequencing libraries were prepared with the EZ DNA RRBS Library Prep Kit (ZymoResearch Cat. no. D5460). The manufacturer's instructions were used with the following modifications: for *MspI* digestion, 300 ng of input DNA and 15 ng of *E. coli* nonmethylated genomic DNA spike were used. Digestions were carried out at 37°C for 4 hr. Adapter ligation was performed overnight, with samples held at 4°C once cycling was completed. Similar to WGBS and MBDBS library preparation, samples were always centrifuged at 12,000 RCF for 30 s, with the exception of a 90 s centrifugation after the second 200 μ l addition of M Wash Buffer. Warmed elution buffer (65°C) was added to each sample to increase DNA elution yield. Adapter sequences were added via PCR using index primers following the recommended thermocycling protocol with 11 cycles. Samples were cleaned using 50 μ l of KAPA pure beads (1 \times) and 80% ethanol, then resuspended in 16 μ l of the elution buffer. Primer removal and library size were confirmed by running samples on a D5000 TapeStation System.

2.7 | DNA sequence alignment

All libraries ($n = 18$) were pooled in equimolar amounts and loaded at 250 pM onto a single Illumina NovaSeq S4 flow cell lane for 2 \times 150 bp sequencing at Genewiz. This was estimated to yield 111–138 million reads per library and 99–123 \times coverage of the *P. acuta* genome (3.3 Mbp) and 38–47 \times coverage of the *M. capitata* genome (8.8 Mbp), assuming 100% even coverage (e.g., 150 bp read \times 2 pairs \times 111 million reads/336,684,533 bp for *P. acuta*).

Sequence quality was checked by FASTQC version 0.11.8 and adapters from paired-end sequences were trimmed using TRIMGALORE! version 0.4.5 (Krueger, 2012). Following recommendations for methylation sequence analysis from the manufacturer's protocol and from the Bismark User Guide, 10 bp were hard trimmed from the 5' and 3' end of each read for WGBS and MBDBS samples, and RRBS samples were trimmed with --non_directional and --rrbs options. Bisulfite-converted genomes were created *in silico* with BOWTIE 2-2.3.4 (Linux version x84_64; Langmead & Salzberg, 2012) using bismark_genome_preparation through BISMARK version 0.21.0 (Krueger & Andrews, 2011). Trimmed reads were aligned to the BS-converted *P. acuta* genome (Vidal-Dupiol et al., 2019) and the BS-converted *M. capitata* genome (Shumaker et al., 2019) with BISMARK version 0.21.0 with alignment stringency set by -score_min L,0,-0.6 and the default MAPQ score threshold of 20. To check mapping rates for endosymbionts and quantify per cent methylation, trimmed reads from *P. actua* libraries were also aligned to the *Cladicopium goreau* genome (type C1, previously *Symbiodinium goreau*; Liu et al., 2018) using the same settings as specified above. Reads that mapped ambiguously were excluded and alignment files containing uniquely mapped reads were deduplicated with deduplicate_bismark for WGBS and MBDBS samples only. Methylation calls were extracted from sorted deduplicated alignment files using bismark_methylation_extractor. Cytosine coverage reports were generated using coverage2cytosine with the --merge_CpG option to combine methylation data from both strands. Resulting files include BEDGRAPHS and BISMARK coverage files (Putnam et al., 2020). MULTIQC version 1.8 (Ewels et al., 2016) was run on the trimmed reads, FASTQC output, and BISMARK reports to assess quality and summarize results.

2.8 | Bisulfite conversion efficiency assessment

Trimmed sequence reads were aligned to the genome of *E. coli* strain K-12 MG1655 (Riley et al., 2006) using BISMARK version 0.21.0 with the --non_directional option and alignment stringency set by -score_min L,0,-0.6. Bisulfite conversion efficiency was also estimated from coral alignments as the ratio of the sum of unmethylated cytosines in CHG and CHH context to the sum of methylated and unmethylated cytosines in CHG and CHH. ANOVA was used to test for an effect of library preparation method on conversion efficiency within each species (conversion efficiency ~library preparation method) for both coral data estimated and *E. coli* alignment calculated conversion efficiencies. A two-sample *t* test was used to test if conversion

efficiency calculated from *E. coli* alignments was the same as estimated conversion efficiency for each library preparation method within each coral species.

2.9 | Genome-wide methylation

General *M. capitata* and *P. acuta* methylation was characterized to describe species-specific patterns. This was carried out by combining BEDgraphs derived from all methods for each species using unionBedGraphs. Per cent methylation for every CpG locus with at least 5x coverage was averaged, irrespective of how many samples had coverage for that locus. Loci with no data within a method were excluded from downstream analysis. CpGs were classified as being either highly methylated ($\geq 50\%$ methylation), moderately methylated ($>10\%$ and $<50\%$) or lowly methylated ($\leq 10\%$ methylation).

2.10 | Per cent methylation of shared CpG loci

Comparisons of per cent DNA methylation at CpG loci analysed by more than one method were performed using the R package METHYLKIT (Akalin et al., 2012). A minimum of 5x coverage was required across all samples for a CpG locus to be considered in the analyses. The unite function in METHYLKIT was used to identify CpG loci that were covered across all nine samples (three individuals per method) per species. Scatterplots and Pearson correlation coefficients were calculated using the function getCorrelation. Additionally, differential methylation tests were performed on pairwise comparisons between methods (WGBS vs. RRBS, WGBS vs. MBDBS, and RRBS vs. MBDBS). Discordant methylation was quantified using a logistic regression model on CpG loci that were covered across all six samples (three samples from each method compared) in each pairwise comparison using the calculateDiffMeth function with default parameters.

2.11 | CpG coverage

To assess average genome-wide CpG coverage, the number of cytosines passing different read depth thresholds (5x, 10x, 15x, 20x, 25x, 30x, 40x, and 50x) were totalled from the CpG coverage reports output by the BISMARK coverage2cytosine function (detailed above) for each sample. These totalled CpGs were then relativized to the number of CpGs in their respective genomes (*M. capitata*, 28,684,519 CpGs; *P. acuta*, 9,155,620 CpGs). Next, average and standard deviation of genome-wide CpG fractions were calculated for each method within each species ($n = 3$), and these were plotted across different read depth thresholds using ggplot2 (Gómez-Rubio, 2017).

To estimate overall genome-wide CpG coverage, a downsampling analysis was performed by pooling all sample reads within a method and species. Briefly, trimmed fastq files were concatenated for each method and species, then randomly subsampled to 50, 100, 150, and 200 million reads. Next, alignment and methylation calling

were carried out as described above on each subset, and the number of cytosines with five or more reads were totalled from CpG coverage reports from each subset. Sequencing saturation was estimated from a Michaelis–Menten model with the drm function from the R package DRC (Ritz et al., 2015) using CpG coverage reports from subsampled data as input. Both observed CpG coverage from subsampled data and estimated CpG coverage were plotted using the R package GGLOT2 (Gómez-Rubio, 2017).

2.12 | Gene coverage

To compare the differences in genes with methylation data by method, we identified the CpGs with 5x coverage that intersected with gene regions using intersectBED (BEDTOOLS version 2.30.0). The proportion of genes with methylation data was calculated by identifying genes in each method that had 5x CpG data and dividing by the total number of genes in the reference genome. We assessed library preparation method bias on functional information by considering gene ontology (GO) terms associated with genes containing CpG data (at least one CpG per gene with 5x coverage in any library). For the set of genes with CpG coverage we performed enrichment analysis to determine if these genes resulted in significant enrichment of particular GO terms using goseq (Young et al., 2010) and accounting for gene length.

2.13 | Proportion of detected CpGs for orthologues

To describe the differences in DNA methylation detected by each method at a more functional level, and given the connection of gene body methylation and gene expression in invertebrates (Roberts & Gavery, 2012) and corals specifically (Liew et al., 2018), the presence of CpG data within all genes was calculated for each species, by method. First, a CpG gff track was generated using EMBOS (Rice et al., 2011) with the fuzznuc command searching for the pattern CG. For each sample, intersectBED was used to identify CpGs with 5x coverage that intersected with gene regions. This was also done for the reference genome CpG gff track. CpG counts per gene were compiled for each sample and the mean taken per method. The proportion of CpGs per orthologous gene was calculated by dividing the mean number of CpGs with 5x coverage from the three samples per method and dividing that by the number of CpGs possible summed per gene from the reference genome CpG gff track. The proportion of CpG data in a gene was then visualized in heatmaps for all genes of *M. capitata* and *P. acuta*.

2.14 | Genomic location of CpGs

For both *M. capitata* and *P. acuta*, the overlap between genome feature tracks and species-specific CpG data at 5x coverage was characterized with BEDTOOLS version 2.29.2 to assess the presence

of CpGs in various regions by method (Quinlan & Hall, 2010). Since only gene, coding sequence, intron, flanking regions, and intergenic region tracks were common between species, these were the tracks used in downstream analyses. A combination of principal coordinates analysis (PCoA), permutational multivariate analysis of variance (PERMANOVA) and beta-dispersion tests, and chi-squared contingency tests were used to determine if the library preparation method influenced the proportion of CpGs detected in a specific genomic feature. A separate contingency test was used for each genomic feature.

Code for all calculations can be found in Putnam et al. (2020).

3 | RESULTS

To compare the performance of bisulfite sequencing methods in the reef-building scleractinian corals *M. capitata* and *P. acuta*, we isolated DNA and generated WGBS, RRBS, and MBDBS libraries for three individuals from each species to yield a total of 18 libraries (Figure 1).

3.1 | DNA sequence alignment

Sequencing of all 18 libraries resulted in 1.82×10^9 read pairs, of which 99.1% remained after QC and trimming (Table S1). Individual libraries were generally sequenced to the same depth ($8.1 \pm 0.34 \times 10^7$ reads; mean \pm SE) across library preparation methods and species, with the exception of *P. acuta* RRBS libraries, which were sequenced 2- to 4-fold deeper ($19.2 \pm 4.67 \times 10^7$ reads; mean \pm SE). The average mapping efficiencies for all *P. acuta* and all *M. capitata* libraries were $45.5 \pm 5.2\%$ and $38.9 \pm 4.3\%$ of reads, respectively (Table S2). In comparison to other methods, MBDBS libraries had a larger proportion of reads ($73.1 \pm 9.9\%$) that did not align to the coral genomes (Figure S1). To investigate this we aligned *P. acuta* libraries to a known symbiont *Cladocopium goreaui* genome (C1; Liu et al., 2018) for which the genome sequence was available at the time of analysis. We found a sizable proportion of the MBDBS reads mapped to the symbiont genome ($23.6 \pm 10.6\%$), while a much smaller proportion of RRBS and WGBS reads mapped to the symbiont genome ($5.04 \pm 0.22\%$ and $1.92 \pm 0.3\%$ respectively) (Table S3).

3.2 | Bisulfite conversion efficiency assessment

Bisulfite conversion efficiency calculated from alignments of the *Escherichia coli* non-methylated genomic DNA spike-in ranged from 98.6% to 99.3% in *M. capitata* and from 98.3% to 99.1% in *P. acuta* (Table S4), and this differed by library preparation method for both *M. capitata* ($F_{2,6} = 114.22$, $p = 1.676 \times 10^{-5}$) and *P. acuta* ($F_{2,6} = 7.24$, $p = .025$) libraries. In general, conversion efficiency calculated from the *E. coli* alignments did not differ from conversion efficiency estimates from CHG and CHH methylation (under the assumption that non-CpG methylation does not occur in corals; see also Liew et al.,

2018) from coral alignments in *M. capitata* and *P. acuta* (Figure S2 and Table S5).

3.3 | Methylation characterization

For each species, the general methylation landscape was characterized for CpG loci with 5 \times coverage identified in any method. The *M. capitata* genome was more methylated than that of *P. acuta* (Figure 2). Using a cutoff of $\geq 50\%$ methylation to define methylated CpGs, of the 13,340,268 CpGs covered by the *M. capitata* data, 11.4% were methylated. In contrast, only 2.9% of the 7,326,297 CpGs in *P. acuta* were methylated. Both genomes were predominantly lowly methylated ($\leq 10\%$ methylated): 79.6% CpGs in *M. capitata* and 91.3% CpGs in *P. acuta* were lowly methylated. The remaining 9.0% of CpGs in *M. capitata* and 5.8% of CpGs in *P. acuta* were moderately methylated (10%–50% methylation). The different methods captured varying proportions of highly, moderately, and lowly methylated CpGs (Figure S3).

3.4 | Correlation of methylation among common CpG loci

For quantitative comparison of method performance, we reduced the data set to loci covered at 5 \times read depth across all methods and samples for each species, referred to here as "shared loci." The number of shared loci was 4666 CpGs for *M. capitata* and 93,714 CpGs for *P. acuta*. A principal components analysis of CpG methylation for loci covered at 5 \times read depth showed that libraries tended to cluster in PC space by preparation method, rather than by individual (Figure S4). Variation in methylation levels of the shared loci across all *M. capitata* samples was lower within a method than between methods (Figure S4A). For *P. acuta*, RRBS and WGBS methods showed similar methylation levels of shared loci, but these were different from the methylation level of loci identified in MBDBS (Figure S4B). To further explore the variation in methylation observed by method, we directly correlated quantitative methylation calls for the shared loci (Figure 3). Correlations among biological replicates within a method were higher and less variable for *M. capitata* compared to *P. acuta*. Correlations between pairs of methods for *M. capitata* ranged on average from .75 to .82, whereas correlations for *P. acuta* ranged from .40 to .64. For *M. capitata*, WGBS versus MBDBS had the highest correlation. For *P. acuta*, WGBS versus RRBS had the highest correlation.

Discordance in methylation quantification between methods was evaluated by identifying the number of CpG loci with large differences ($>50\%$) in methylation for each species. WGBS versus RRBS showed the lowest discordance in both species (0.4% for *M. capitata* and 0.5% for *P. acuta*). The highest discordance in methylation was found in comparisons with MBDBS for *P. acuta*, with 11% and 15% of CpG sites being called at least 50% different for comparisons with WGBS and RRBS, respectively. In contrast, only 0.4% and 5%

M. capitata

P. acuta

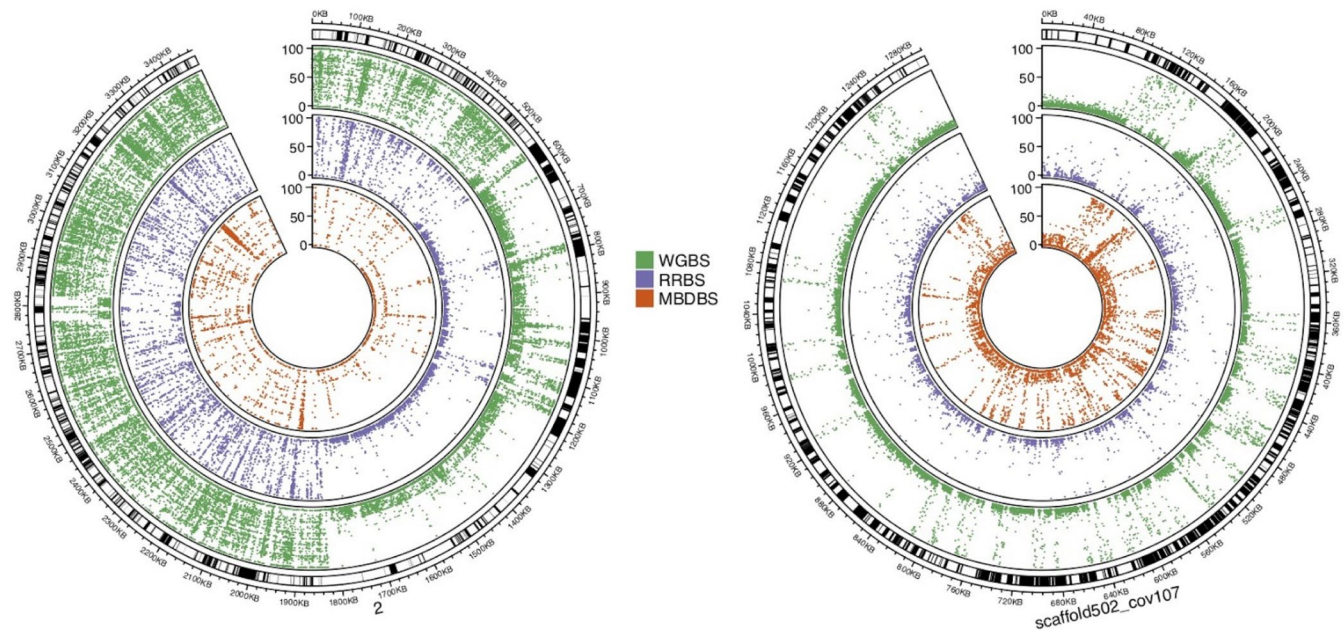


FIGURE 2 Mean per cent methylation of CpGs. Data are presented for CpGs with 5x coverage for each method on the largest scaffolds of each genome. The outer track shows the scaffold locations and dots indicate the per cent methylation as indicated by the y-axes from 0% to 100% for each of the inner tracks

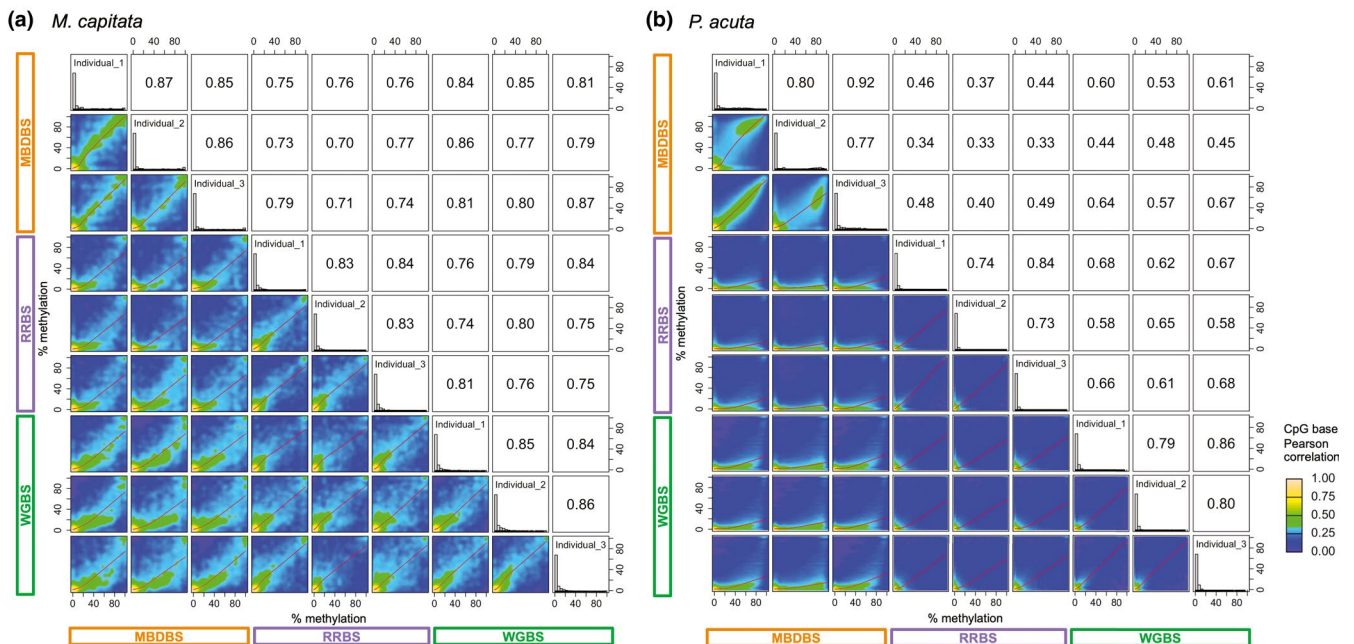


FIGURE 3 Matrix of pairwise scatter plots for shared CpG loci. Data are presented for CpG covered at $\geq 5\times$ across all samples) for (a) *Montipora capitata* ($n = 4666$ common loci) and (b) *Pocillopora acuta* ($n = 93,714$ common loci). The red lines represent linear regression fits and the green lines are polynomial regression fits. Pearson correlation coefficients for each pairwise comparison are presented in the upper right boxes. Methods are colour coded on the x- and y-axes (WGBS = green, MBDBS = orange, RRBS = purple) and replicate samples are indicated on the diagonal along with histograms of per cent CpG methylation

of common CpG sites were at least 50% different between MBDBS versus WGBS and MBDBS versus RRBS, respectively, for *M. capitata*. A majority of the discordance was due to higher methylation calls in MBDBS compared to WGBS or RRBS (Figure 3b).

3.5 | CpG coverage

Consistent with what would be expected based on genome size, *P. acuta* libraries have higher genome-wide CpG coverage than

M. capitata regardless of library preparation method (Figure 4). For both species, WGBS and MBDBS libraries covered more CpGs than RRBS libraries, whereas RRBS libraries tended to show greater read depth for the CpGs that it did cover. In other words, at $>20\times$ read depth, RRBS libraries covered more CpGs than either WGBS or MBDBS (Figure 4 insets). Modelling increased sequencing depth for RRBS or MBDBS libraries but showed little impact on the fraction of genome-wide CpGs covered in *M. capitata*, while increasing sequencing depth from 50 million to 200 million for WGBS libraries in both species and for MBDBS in *P. acuta* showed a substantially larger fraction of CpGs covered (Figure S5).

3.6 | Gene coverage

For each species, WGBS provided CpG data for the highest proportion of genes in the genome: 84.5% and 94.3% for *M. capitata* and *P. acuta*, respectively. RRBS generated CpG data for 54.7% and 63.5% of the genes for *M. capitata* and *P. acuta*, respectively. MBDBS provided the most divergent coverage of genes with 44.1% and 86.9% for *M. capitata* and *P. acuta*, respectively. When performing functional gene enrichment on the gene sets that contained CpG data for each method, there were multiple enriched GO categories for all methods and both species (Figures S6 and S7). For *M. capitata*, there were 223, 401 and 253 enriched molecular function (MF) terms (Table S6) and 508, 1123 and 853 biological process (BP) terms for WGBS, RRBS and MBDBS, respectively (Table S7). For *P. acuta* there were 164, 282 and 287 terms enriched for MF (Table S8) and 313, 749 and 682 terms enriched for BP for WGBS, RRBS and MBDBS, respectively (Table S9).

3.7 | CpG coverage within orthologous genes

To assess the potential for cross-species comparisons using an equivalent data set, we quantified CpG data available across one-to-one

orthologous genes. For *M. capitata*, WGBS yielded the highest proportion of CpGs, followed by RRBS, and then MBDBS (Figure S8A). This differed in *P. acuta* with WGBS yielding the highest proportion of CpGs on average across orthologues, followed by MBDBS, and then RRBS (Figure S8B).

3.8 | Genomic location of CpGs

To compare locations of CpG data between genomic features for each species and method, all CpGs with $5\times$ coverage were characterized based on genomic feature location (Figure 5). Global PERMANOVA tests found significant differences between library preparation methods for CpG coverage in various genome features for *M. capitata* and *P. acuta* (Table S10). Although *post hoc* pairwise PERMANOVA tests did not reveal differences between sequencing methods, power for these was probably low (due to low sample size). Pairwise chi-squared tests indicated there are differences in CpG location for both species. In particular, CpGs in gene bodies were significantly enriched over other genomic features with MBDBS (Figure 5; Table S11). Visual inspection of PCoA also revealed the proportion of CpGs captured in coding sequences (CDS) drove differences between MBDBS and the other methods in both species (Figure 5c,d).

4 | DISCUSSION

We evaluated the performance of three approaches that use bisulfite-treated DNA for library preparation to enable single base-pair resolution quantification of DNA methylation in corals. Our results demonstrate that the methylation landscape can vary significantly across species, which is a critical consideration for both interpreting environmental response capacity, and therefore for experimental design. Whereas WGBS is the gold standard for studying methylation, it comes at a high cost. MBDBS enriches for

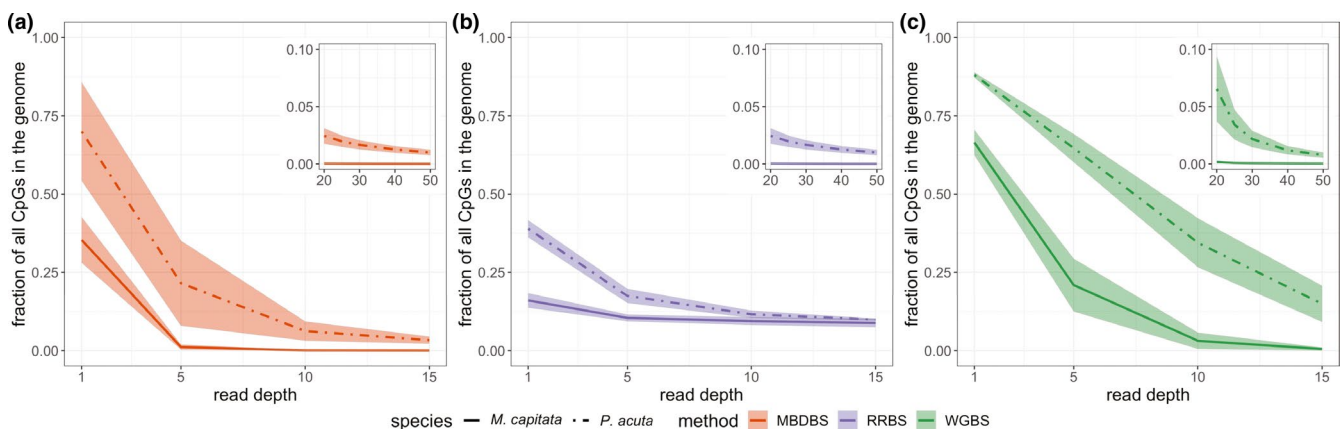


FIGURE 4 CpG site coverage across library preparation methods. Mean fraction of CpG sites in the genome covered at different sequencing depths (read depths) by (a) MBDBS libraries, (b) RRBS libraries and (c) WGBS libraries with standard deviations shown by shaded areas (see Table S2 for number of reads in each sample)

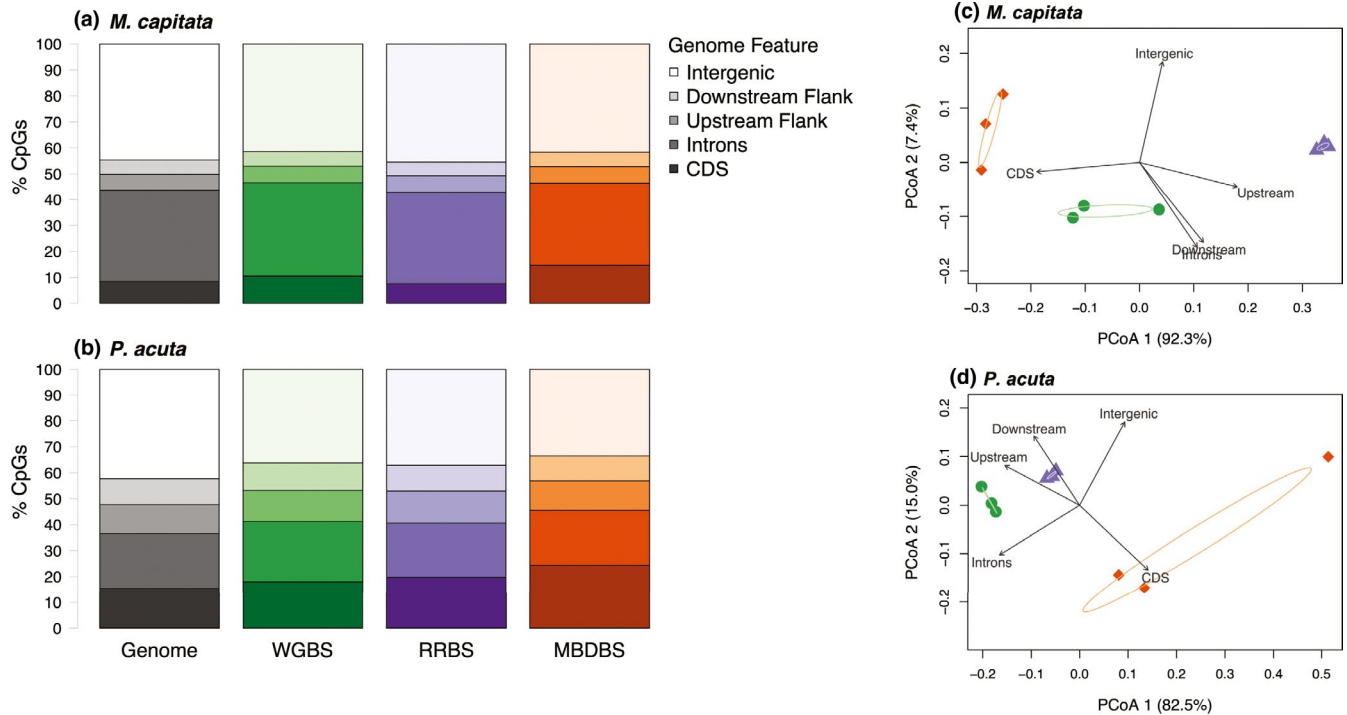


FIGURE 5 Per cent of CpGs detected by sequencing methods in genome features (a) for *Montipora capitata* and (b) *Pocillopora acuta*. Genome features considered were coding sequences (CDS), introns, 1-kb flanking regions upstream (Upstream Flank) or downstream of genes (Downstream Flank), and intergenic regions. Each bar corresponds to all the CpGs in the genome (Genome), or each method (WGBS, RRBS or MBDBS). The order of the genome features depicted in each bar is identical to what is displayed in the key, with the darkest shade representing CDS, and the lightest shade representing intergenic regions. Principal coordinate analyses associated with PERMANOVA and beta-dispersion tests related to Table S6 that show differences in proportion of CpGs in various genomic locations (CDS, introns, upstream flanks, downstream flanks and intergenic regions) for (c) *M. capitata* and (d) *P. acuta*. WGBS is represented by green circles, RRBS by purple triangles, and MBDBS by orange diamonds. Per cent variation explained by each PCoA axis is included in the axis label. Ellipses depict 95% confidence intervals for each sequencing method. All eigenvectors are significant at the $\alpha = .05$ level

gene regions, which may be useful for taxa with gene body methylation. On the other hand, RRBS provides greater coverage depth for a smaller fraction of the genome, but lacks specificity for genomic features or DNA methylation. Taken together, our findings indicate biology, genome architecture, regions of interest, and depth of coverage are critical considerations when choosing methods for high-resolution quantification of DNA methylation profiles in invertebrates.

Montipora capitata has a relatively high environmental tolerance (Bahr et al., 2016; Gibbin et al., 2015; Putnam et al., 2016), which has previously been attributed to its symbiont composition (Cunning et al., 2016), genome characteristics (Shumaker et al., 2019), perforate tissue-skeletal architecture and tissue thickness, and heterotrophic capacity (Rodrigues & Grottoli, 2007). Of particular relevance to DNA methylation are genomic aspects such as gene family duplication and high repeat content in *M. capitata* (Shumaker et al., 2019). We found overall DNA methylation was higher in *M. capitata* than in *P. acuta*, supporting early bulk analyses of DNA methylation in these species (Putnam et al., 2016). While the predicted number of genes is similar, the genome size of *M. capitata* is over twice that of *P. acuta* (Shumaker et al., 2019; Vidal-Dupiol et al., 2019). One explanation for the higher

methylation in *M. capitata* is that with greater energy availability—through translocation from high-density Symbiodiniaceae populations and energy stores in perforate tissues—there is greater capacity for methyltransferase to maintain high methylation, and thus reduce gene expression variability and spurious expression (Li et al., 2018; Liew et al., 2018). High constitutive methylation could allow “frontloading” of stress response genes (e.g., Barshis et al., 2013), providing greater stress tolerance. Another possible explanation is that the higher level of methylation contributes to the silencing of repeated genetic elements. In contrast, with a small and nonrepetitive genome, imperforate thin tissues, and low energy reserves, *P. acuta* may be more energetically limited. Thus, *P. acuta* may be expected to show lower DNA methylation across the genome as we demonstrate here, as well as a higher propensity for inducible methylation in the presence of stressors (Putnam et al., 2016).

Another striking contrast in DNA methylation in these species is the lack of concordance in the per cent methylation values for *P. acuta* among methods compared to *M. capitata* (Figure 3). The potential for chimerism in corals (Oury et al., 2020; Schweinsberg et al., 2015) and differences in tissue structure (e.g., perforate or imperforate) between species could contribute to differences in

concordance across methods for quantifying DNA methylation. One possibility is that Pocilloporids are chimeric and multiple genotypes exist (Oury et al., 2020; Schweinsberg et al., 2015). Although per cent DNA methylation concordance across methods was generally high, in *P. acuta* there was an ~10% higher level of discordance in per cent methylation quantification when comparing WGBS to RRBS or MBDBS (Figure 3). This discordance could have resulted from differences in *P. acuta* and *M. capitata* tissue structure. There is the potential to homogenize and extract DNA from all cell types from the thin, imperforate tissues of *P. acuta*, as opposed to the thick, perforate tissues in *M. capitata* (Putnam et al., 2017), probably contributing to a greater number of cell types, and thus methylation differences, captured in our *P. acuta* samples. Furthermore, the microhabitats created in the tissues of these two species probably differ substantially spatially (Putnam et al., 2017), creating cell-to-cell variability in methylation content. Since the likelihood of capturing multiple cell types in bulk DNA extractions varies with tissue structure, future studies should consider methods such as fluorescence-activated cell sorting (Hu et al., 2020; Rosental et al., 2017), or laser microdissection (e.g., Liew et al., 2018), to target specific tissues or cell types and reduce cell-to-cell methylation variability. Whereas this does not necessarily indicate a bias in our methods, it highlights the need to account for the biological characteristics of a species when designing an experiment and evaluating results. Also when comparing across species, given genetic–epigenetic correlations, particularly in the case of DNA methylation and the requirement for a CpG sequence target site (Dimond & Roberts, 2020; Johnson et al., 2020), variation in genome architecture, gene number and content will impact the presence and use of DNA methylation as a mechanism of gene expression regulation.

The gold standard for bisulfite sequencing, WGBS, can be cost prohibitive particularly if comparing multiple species and treatments. As expected, we found that WGBS performed well, particularly for *P. acuta* which has a smaller genome. We found GO enrichment of the genes covered by CpG data was affected by library preparation method, probably attributed to genome characteristics and relatively higher methylation in *M. capitata* affecting the genes covered by reduced representation methods (e.g. 44.1% of genes in *M. capitata* and 86.9% of genes in *P. acuta* covered by MBDBS). There may be preferential enrichment of hypermethylated genes in *M. capitata* by MBDBS and thus MBDBS in *M. capitata* may not capture the breadth of genes found in the less methylated *P. acuta*. Therefore, MBDBS approaches could benefit from species-specific protocol optimization (Aberg et al., 2018). GO enrichment analysis identified changes in the significantly enriched GO terms as the proportion of genes with data decreases (Figures S7 and S8) and these GO terms were found across many BP and MF terms, not limited to a particular set (Tables S6–S9). Focusing on direct comparison of gene orthologues, WGBS performed the best in terms of data for CpGs per gene. Based on the gene orthologue comparisons, MBDBS provided more information than RRBS for *P. acuta*, but the opposite

held true for *M. capitata*. Collectively, these differences are probably attributable to the different genome size, assembly quality and/or inherent differences in methylation that result in differential enrichment.

For both species, WGBS and MBDBS libraries covered more CpGs than RRBS libraries; however, RRBS libraries showed greater read depth for CpGs. This is because RRBS subsampled a specific, smaller portion of the genome than MBDBS or WGBS, allowing more read coverage. Hence, CpG coverage did not largely increase when deeper sequencing was modelled using RRBS data (Figure S5). RRBS was designed to enrich for CpG islands, short stretches of DNA with higher levels of CpGs (~1 CpG per 10 bp), which are typically found in mammalian promoters and enhancer regions and thought to play a role in gene regulation (Gu et al., 2011). We found RRBS yielded a well-covered reduced representation of the genome, which is important for bisulfite data where high read depth is desired, and locus methylation levels were concordant with WGBS for both species. However, RRBS did not enrich for promoters or other particular genomic regions compared to the other bisulfite sequencing methods (Figure 5), and in fact tended to identify unmethylated regions. For this reason, RRBS is not the best choice for gene-focused methylation studies in corals and other invertebrate taxa with gene body methylation.

A critical consideration in deciding to perform MBDBS in corals is the amount of DNA methylation present from any symbiont. If any nontarget organisms have substantially more DNA methylation than the target organism, MBDBS data could become saturated by methylated DNA from nontarget organisms, lowering sampling of the target species. We observed this in *P. acuta*, for which we had the genome of its Symbiodiniaceae which has ~90% genome-wide methylation (Lohuis & Miller, 1998; de Mendoza et al., 2018). When compared to RRBS and WGBS data, we found a 4- to 10-fold enrichment of Symbiodiniaceae DNA in *P. acuta* MBDBS data. Separation of host and symbionts is therefore recommended to obtain the greatest read counts for the organism of interest, but this comes at the cost of not being able to obtain RNA from the same nucleic acid pool. For example, physical separation of the host and symbiont in living cells impacts gene expression, and attempts at physical separation after freezing can degrade the host RNA. Simultaneous extraction of holobiont RNA and DNA from the same nucleic acid pool provides the optimal approach for detecting interactions between DNA methylation and epigenetic regulation of gene expression. This comes at the cost of generating excess reads to overcome highly methylated Symbiodiniaceae DNA.

MBDBS can enrich for gene regions in species where methylation is primarily found in gene bodies such as in corals (reviewed by Eirin-Lopez & Putnam, 2019), and can thus provide insight into mechanisms underlying physiological or organismal responses. We found that MBDBS significantly enriched for gene bodies, specifically CDS and introns, when compared to RRBS and WGBS in both *M. capitata* and *P. acuta* (Figure 5). While MBDBS may be a good choice to examine gene body methylation at a reduced cost,

species differences in CpG coverage within orthologous genes with MBDBS (Figure S6) may complicate cross-species comparisons by reducing the amount of data available for analysis. Additionally, we found high discordance between MBDBS and nonenrichment methods, WGBS and RRBS, for *P. acuta*. MBDBS is the only method we evaluated that can nonrandomly subsample genomes present in a DNA sample through preferential pull-down of methylated DNA. Differences in methylation across the sampled genomes could result from cell-to-cell heterogeneity in methylation or cell type (e.g., methylation of calcifying cells may differ from symbiont hosting cells). In other words, MBDBS data may represent only a subpopulation of highly methylated cells, while WGBS and RRBS represent the average methylation across all cells in the sample. Using a consistent tissue type is important to limit potential methylation heterogeneity, and caution should be taken when comparing MBDBS data directly to that of nonenrichment bisulfite sequencing approaches.

Although MBDBS did enrich for methylated regions of the genome, 80% of CpGs in *M. capitata* and 82% of CpGs in *P. acuta* interrogated with MBDBS were lowly methylated (<10% methylated) (Figure S3). This is expected and is consistent with previous reports applying MBDBS in other marine invertebrates where unmethylated CpGs actually represent the highest proportion of loci in the data, attributable to the nature of the methylation landscape and enrichment protocol (e.g. Gavery & Roberts, 2013; Venkataraman et al., 2020). The base-pair resolution of methylation revealed by MBDBS is a benefit over MBD sequencing (MBD-seq) alone because it enables a fine-scale examination of specific genomic features (e.g. exon-intron boundaries) that may not be possible with the regional resolution of MBD-seq. Without complete knowledge of the relative importance of a single locus compared to a region, it is difficult to compare trade-offs between MBDBS and MBD-seq. However, bisulfite sequencing requires significant coverage to quantify DNA methylation.

Methyl-CpG binding domain bisulfite sequencing may have potential biases that should be considered when interpreting results. If a treatment comparison, population comparison, or developmental change results in a given region (~500 bp) going from being highly methylated to fully unmethylated, then it is likely that this region would not be interrogated by MBDBS, due to an absence of data in the unmethylated condition. This is a potential source of bias in MBDBS data and may contribute to important differentially methylated regions being overlooked: for example if one treatment results in high methylation and is captured by MBDBS and another treatment results in no methylation and is not captured by MBDBS, this region would be filtered out of the analysis because of missing data in some individuals. Further, the potential of MBDBS to provide limited information for unmethylated genes may introduce bias in studies that seek to draw relationships between methylation level and gene expression. Just as with many interpretations of key findings we present, a more complete understanding of the mechanistic functional role DNA methylation plays in genome regulation in the species of interest is needed.

There is a greater capacity to gain mechanistic insight when using methods that have single base-pair resolution of methylation data compared to methylation enrichment without bisulfite treatment or to bulk per cent methylation approaches. For example, hypotheses such as the linkage between DNA methylation and alternative splicing (Roberts & Gavery, 2012) are more accurately tested with bisulfite sequencing approaches. We acknowledge the cost of generating genomic resources and bisulfite sequencing data can be higher than other approaches. While WGBS is supported here as the gold standard for DNA methylation quantification, consideration should be given to specific study hypotheses in light of the pros and cons of the enrichment or reduced representation approaches presented here and in other comparative works (Dixon & Matz, 2021). Our results suggest that it would be unwise to use multiple different library preparation methods for comparing individuals within a study, especially for studies in which familial relationships are to be compared. As technology advances, it would be ideal to move away from harsh bisulfite conversion to assess DNA methylation with single base-pair resolution across whole genomes in the absence of DNA treatment (e.g., Oxford Nanopore).

Our results provide a quantitative comparative assessment that can be used to inform the choice of sequencing DNA methylation in corals and other nonmodel invertebrates. Together these metrics enable comparative capacity for three common methods in two coral taxa that vary in their phylogeny, genome size, symbiotic unions and environmental performance, and thus provide the community with a more comprehensive foundation upon which to build laboratory and statistical analyses of DNA methylation, plasticity and acclimatization.

ACKNOWLEDGEMENTS

We thank Putnam Laboratory members E. Strand and M. E. Schedl for their technical support. Data analyses were facilitated through the use of advanced computational, storage and networking infrastructure provided by the Hyak supercomputer system at the University of Washington.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS

H.M.P. designed the study and collected the samples. Y.R.V., S.A.T., M.R.G., H.M.P. and S.B.R. performed data analyses and drafted the manuscript. D.B., A.D.-W., J.M.E.-L., K.M.J., K.E.L. and J.B.P. provided analytical insight and manuscript revisions. All authors read and approved the final manuscript.

OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at <http://doi.org/10.17605/OSF>.

IO/X5WAZ. All raw data can be accessed under NCBI Bioproject PRJNA691891.

DATA AVAILABILITY STATEMENT

The data sets supporting the conclusions of this article are available in the Coral Methylation Methods Comparison repository, <http://doi.org/10.17605/OSF.IO/X5WAZ>, and included within the article and its additional files. All raw data can be accessed under NCBI Bioproject PRJNA691891.

ORCID

Yaamini R. Venkataraman  <https://orcid.org/0000-0002-0364-1829>

Mackenzie R. Gavery  <https://orcid.org/0000-0001-5010-3399>

Debashish Bhattacharya  <https://orcid.org/0000-0003-0611-1273>

Jose M. Eirin-Lopez  <https://orcid.org/0000-0002-8041-9770>

Katie E. Lotterhos  <https://orcid.org/0000-0001-7529-2771>

Jonathan B. Puritz  <https://orcid.org/0000-0003-1404-4680>

Hollie M. Putnam  <https://orcid.org/0000-0003-2322-3269>

REFERENCES

- Aberg, K. A., Chan, R. F., Xie, L., Shabalin, A. A., & van den Oord, E. J. C. G. (2018). Methyl-CpG-binding domain sequencing: MBD-seq. In J. Tost (Ed.), *DNA methylation protocols* (pp. 171–189). Springer.
- Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, 13(10), R87. <https://doi.org/10.1186/gb-2012-13-10-r87>
- Bahr, K. D., Jokiel, P. L., & Rodgers, K. S. (2016). Relative sensitivity of five Hawaiian coral species to high temperature under high-pCO₂ conditions. *Coral Reefs*, 35(2), 729–738. <https://doi.org/10.1007/s00338-016-1405-4>
- Baldwin, J. M. (1902). *Development and evolution, including psychophysical evolution, evolution by orthoplasy, and the theory of genetic modes*. MacMillan Co.
- Barshis, D. J., Ladner, J. T., Oliver, T. A., Seneca, F. O., Traylor-Knowles, N., & Palumbi, S. R. (2013). Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 110(4), 1387–1392. <https://doi.org/10.1073/pnas.1210224110>
- Byrne, M., Foo, S. A., Ross, P. M., & Putnam, H. M. (2020). Limitations of cross-and multigenerational plasticity for marine invertebrates faced with global climate change. *Global Change Biology*, 26(1), 80–102. <https://doi.org/10.1111/gcb.14882>
- Chatterjee, A., Ozaki, Y., Stockwell, P. A., Horsfield, J. A., Morison, I. M., & Nakagawa, S. (2013). Mapping the zebrafish brain methylome using reduced representation bisulfite sequencing. *Epigenetics*, 8(9), 979–989. <https://doi.org/10.4161/epi.25797>
- Cunning, R., Ritson-Williams, R., & Gates, R. D. (2016). Patterns of bleaching and recovery of *Montipora capitata* in Kāne 'ohe Bay, Hawai'i, USA. *Marine Ecology Progress Series*, 551, 131–139. <https://doi.org/10.3354/meps11733>
- de Mendoza, A., Bonnet, A., Vargas-Landin, D. B., Ji, N., Li, H., Yang, F., Li, L., Hori, K., Pflueger, J., Buckberry, S., Ohta, H., Rosic, N., Lesage, P., Lin, S., & Lister, R. (2018). Recurrent acquisition of cytosine methyltransferases into eukaryotic retrotransposons. *Nature Communications*, 9(1), 1–11. <https://doi.org/10.1038/s41467-018-03724-9>
- Dimond, J. L., Gamblewood, S. K., & Roberts, S. B. (2017). Genetic and epigenetic insight into morphospecies in a reef coral. *Molecular Ecology*, 26(19), 5031–5042. <https://doi.org/10.1111/mec.14252>
- Dimond, J. L., & Roberts, S. B. (2020). Convergence of DNA methylation profiles of the reef coral *Porites astreoides* in a Novel Environment. *Frontiers in Marine Science*, 6, 792. <https://doi.org/10.3389/fmars.2019.00792>
- Dixon, G., Liao, Y., Bay, L. K., & Matz, M. V. (2018). Role of gene body methylation in acclimatization and adaptation in a basal metazoan. *Proceedings of the National Academy of Sciences of the United States of America*, 115(52), 13342–13346. <https://doi.org/10.1073/pnas.1813749115>
- Dixon, G., & Matz, M. (2021). Benchmarking DNA methylation assays in a reef-building coral. *Molecular Ecology Resources*, 21(2), 464–477. <https://doi.org/10.1111/1755-0998.13282>
- Eirin-Lopez, J. M., & Putnam, H. M. (2019). Marine environmental epigenetics. *Annual Review of Marine Science*, 11, 335–368. <https://doi.org/10.1146/annurev-marine-010318-095114>
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
- Flores, K., Wolschin, F., Corneveaux, J. J., Allen, A. N., Huentelman, M. J., & Amdam, G. V. (2012). Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics*, 13, 480. <https://doi.org/10.1186/1471-2164-13-480>
- Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S. E., Robinson, G. E., & Maleszka, R. (2012). DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. *Proceedings of the National Academy of Sciences of the United States of America*, 109(13), 4968–4973. <https://doi.org/10.1073/pnas.1202392109>
- Gavery, M. R., & Roberts, S. B. (2010). DNA methylation patterns provide insight into epigenetic regulation in the Pacific oyster (*Crassostrea gigas*). *BMC Genomics*, 11(1), 483. <https://doi.org/10.1186/1471-2164-11-483>
- Gavery, M. R., & Roberts, S. B. (2013). Predominant intragenic methylation is associated with gene expression characteristics in a bivalve mollusc. *PeerJ*, 1, e215. <https://doi.org/10.7717/peerj.215>
- Ghalambor, C. K., Hoke, K. L., Ruell, E. W., Fischer, E. K., Reznick, D. N., & Hughes, K. A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, 525(7569), 372–375. <https://doi.org/10.1038/nature15256>
- Ghalambor, C. K., McKay, J. K., Carroll, S. P., & Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, 21(3), 394–407. <https://doi.org/10.1111/j.1365-2435.2007.01283.x>
- Gibbin, E. M., Putnam, H. M., Gates, R. D., Nitschke, M. R., & Davy, S. K. (2015). Species-specific differences in thermal tolerance may define susceptibility to intracellular acidosis in reef corals. *Marine Biology*, 162(3), 717–723. <https://doi.org/10.1007/s00227-015-2617-9>
- Gómez-Rubio, V. (2017). *ggplot2 - Elegant graphics for data analysis* (2nd ed.). In *Journal of Statistical Software* (Vol. 77, Issue Book Review 2). <https://doi.org/10.18637/jss.v077.b02>
- Gonzalez-Romero, R., Suarez-Ulloa, V., Rodriguez-Casariago, J., Garcia-Souto, D., Diaz, G., Smith, A., Pasantes, J. J., Rand, G., & Eirin-Lopez, J. M. (2017). Effects of Florida Red Tides on histone variant expression and DNA methylation in the Eastern oyster *Crassostrea virginica*. *Aquatic Toxicology*, 186, 196–204. <https://doi.org/10.1016/j.aquatox.2017.03.006>
- Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., & Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols*, 6(4), 468–481. <https://doi.org/10.1038/nprot.2010.190>

- Hofmann, G. E. (2017). Ecological epigenetics in marine metazoans. *Frontiers in Marine Science*, 4, 4. <https://doi.org/10.3389/fmars.2017.00004>
- Hu, M., Zheng, X., Fan, C.-M., & Zheng, Y. (2020). Lineage dynamics of the endosymbiotic cell type in the soft coral *Xenia*. *Nature*, 582(7813), 534–538. <https://doi.org/10.1038/s41586-020-2385-7>
- Johnson, K. M., Sirovy, K. A., Casas, S. M., La Peyre, J. F., & Kelly, M. W. (2020). Characterizing the epigenetic and transcriptomic responses to *Perkinsus marinus* Infection in the Eastern Oyster *Crassostrea virginica*. *Frontiers in Marine Science*, 7, 598. <https://doi.org/10.3389/fmars.2020.00598>
- Kronholm, I., & Collins, S. (2016). Epigenetic mutations can both help and hinder adaptive evolution. *Molecular Ecology*, 25(8), 1856–1868. <https://doi.org/10.1111/mec.13296>
- Krueger, F. (2012). *Trim Galore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisulfite-Seq) libraries*. http://www.Bioinformatics.Babraham.Ac.Uk/projects/trim_galore/. Accessed 28 April 2016.
- Krueger, F., & Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27(11), 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lewis, S. L., & Maslin, M. A. (2015). Defining the anthropocene. *Nature*, 519(7542), 171–180. <https://doi.org/10.1038/nature14258>
- Li, Y., Liew, Y. J., Cui, G., Czielski, M. J., Zahran, N., Michell, C. T., Voolstra, C. R., & Aranda Lastra, M. (2018). DNA methylation regulates transcriptional homeostasis of algal endosymbiosis in the coral model *Aiptasia*. *Science Advances*, 4(8), eaat2142. <https://doi.org/10.1126/sciadv.aat2142>
- Liew, Y. J., Howells, E. J., Wang, X., Michell, C. T., Burt, J. A., Idaghdour, Y., & Aranda, M. (2020). Intergenerational epigenetic inheritance in reef-building corals. *Nature Climate Change*, 10(3), 254–259. <https://doi.org/10.1038/s41558-019-0687-2>
- Liew, Y. J., Zoccola, D., Li, Y., Tambutté, E., Venn, A. A., Michell, C. T., Cui, G., Deutekom, E. S., Kaandorp, J. A., Voolstra, C. R., Forêt, S., Allemand, D., Tambutté, S., & Aranda, M. (2018). Epigenome-associated phenotypic acclimatization to ocean acidification in a reef-building coral. *Science Advances*, 4(6), 188227. <https://doi.org/10.1126/sciadv.aar8028>
- Liu, H., Stephens, T. G., González-Pech, R. A., Beltran, V. H., Lapeyre, B., Bongaerts, P., Cooke, I., Aranda, M., Bourne, D. G., Forêt, S., Miller, D. J., van Oppen, M. J. H., Voolstra, C. R., Ragan, M. A., & Chan, C. X. (2018). Symbiodinium genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. *Communications Biology*, 1, 95. <https://doi.org/10.1038/s42003-018-0098-3>
- Lohuis, M. R., & Miller, D. J. (1998). Hypermethylation at CpG-motifs in the dinoflagellates *Amphidinium carterae* (dinophyceae) and *Symbiodinium microadriaticum* (dinophyceae): Evidence from restriction analyses, 5-azacytidine and ethionine treatment. *Journal of Phycology*, 34(1), 152–159. <https://doi.org/10.1046/j.1529-8817.1998.340152.x>
- Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205), 766–770. <https://doi.org/10.1038/nature07107>
- Oury, N., Gélín, P., & Magalon, H. (2020). Together stronger: Intracolony genetic variability occurrence in *Pocillopora* corals suggests potential benefits. *Ecology and Evolution*, 10(12), 5208–5218. <https://doi.org/10.1002/ece3.5807>
- Putnam, H. M., Barott, K. L., Ainsworth, T. D., & Gates, R. D. (2017). The vulnerability and resilience of reef-building corals. *Current Biology*, 27(11), R528–R540. <https://doi.org/10.1016/j.cub.2017.04.047>
- Putnam, H. M., Davidson, J. M., & Gates, R. D. (2016). Ocean acidification influences host DNA methylation and phenotypic plasticity in environmentally susceptible corals. *Evolutionary Applications*, 9(9), 1165–1178. <https://doi.org/10.1111/eva.12408>
- Putnam, H. M., Roberts, S. B., & Venkataraman, Y. (2020). Coral methylation methods comparison. Open Science Framework. <https://doi.org/10.17605/OSF.IO/X5WAZ>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Rice, P. M., Bleasby, A. J., & Ison, J. C. (2011). *EMBOSS user's guide: Practical bioinformatics with EMBOSS*. Cambridge University Press. <https://play.google.com/store/books/details?id=zC5WGJxDNtC>
- Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K. B., Blattner, F. R., Chaudhuri, R. R., Glasner, J. D., Horiuchi, T., Keseler, I. M., Kosuge, T., Mori, H., Perna, N. T., Plunkett, G. 3rd, Rudd, K. E., Serres, M. H., Thomas, G. H., Thomson, N. R., Wishart, D., & Wanner, B. L. (2006). *Escherichia coli* K-12: A cooperatively developed annotation snapshot-2005. *Nucleic Acids Research*, 34(1), 1–9. <https://doi.org/10.1093/nar/gkj405>
- Ritz, C., Baty, F., Streibig, J. C., & Gerhard, D. (2015). Dose-response analysis using R. *PLoS One*, 10(12), e0146021. <https://doi.org/10.1371/journal.pone.0146021>
- Riviere, G., Wu, G.-C., Fellous, A., Goux, D., Sourdain, P., & Favrel, P. (2013). DNA methylation is crucial for the early development in the Oyster *C. gigas*. *Marine Biotechnology*, 15(6), 739–753. <https://doi.org/10.1007/s10126-013-9523-2>
- Roberts, S. B., & Gavery, M. R. (2012). Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? *Frontiers in Physiology*, 2, 116. <https://doi.org/10.3389/fphys.2011.00116>
- Rodrigues, L. J., & Grotto, A. G. (2007). Energy reserves and metabolism as indicators of coral recovery from bleaching. *Limnology and Oceanography*, 52(5), 1874–1882. <https://doi.org/10.4319/lo.2007.52.5.1874>
- Rodriguez-Casariello, J. A., Ladd, M. C., Shantz, A. A., Lopes, C., Cheema, M. S., Kim, B., Roberts, S. B., Fourqurean, J. W., Ausio, J., Burkepille, D. E., & Eirin-Lopez, J. M. (2018). Coral epigenetic responses to nutrient stress: Histone H2A.X phosphorylation dynamics and DNA methylation in the staghorn coral *Acropora cervicornis*. *Ecology and Evolution*, 8(23), 12193–12207. <https://doi.org/10.1002/ece3.4678>
- Rosental, B., Kozhekbaeva, Z., Fernhoff, N., Tsai, J. M., & Taylor-Knowles, N. (2017). Coral cell separation and isolation by fluorescence-activated cell sorting (FACS). *BMC Cell Biology*, 18(1), 30. <https://doi.org/10.1186/s12860-017-0146-8>
- Schweinsberg, M., Weiss, L. C., Striewski, S., Tollrian, R., & Lampert, K. P. (2015). More than one genotype: How common is intracolony genetic variability in scleractinian corals? *Molecular Ecology*, 24(11), 2673–2685. <https://doi.org/10.1111/mec.13200>
- Shumaker, A., Putnam, H. M., Qiu, H., Price, D. C., Ehud, Z., Harel, A., Wagner, N. E., Gates, R. D., Yoon, H. S., & Bhattacharya, D. (2019). Genome analysis of the rice coral *Montipora capitata*. *Scientific Reports*, 9(2571), <https://doi.org/10.1038/s41598-019-39274-3>
- Suarez-Ulloa, V., Rivera-Casas, C., Michel, M., & Eirin-Lopez, J. M. (2018). Seasonal variation and epigenetic changes in the flat tree oyster *Isognomon alatus* during a 2-year study of a mangrove ecosystem in North Biscayne Bay, Florida. *Journal of Shellfish Research*, 38(1), 79–88.
- Suzuki, M. M., Kerr, A. R. W., De Sousa, D., & Bird, A. (2007). CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Research*, 17(5), 625–631. <https://doi.org/10.1101/gr.6163007>
- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. <https://doi.org/10.1093/bib/bbs017>

- Tweedie, S., Charlton, J., Clark, V., & Bird, A. (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Molecular and Cellular Biology*, 17(3), 1469–1475. <https://doi.org/10.1128/MCB.17.3.1469>
- van Gurp, T. P., Wagemaker, N. C. A. M., Wouters, B., Vergeer, P., Ouborg, J. N. J., & Verhoeven, K. J. F. (2016). epiGBS: Reference-free reduced representation bisulfite sequencing. *Nature Methods*, 13(4), 322–324. <https://doi.org/10.1038/nmeth.3763>
- Velotta, J. P., Ivy, C. M., Wolf, C. J., Scott, G. R., & Cheviron, Z. A. (2018). Maladaptive phenotypic plasticity in cardiac muscle growth is suppressed in high-altitude deer mice. *Evolution*, 72(12), 2712–2727. <https://doi.org/10.1111/evo.13626>
- Venkataraman, Y. R., Downey-Wall, A. M., Ries, J., Westfield, I., White, S. J., Roberts, S. B., & Lotterhos, K. E. (2020). General DNA methylation patterns and environmentally-induced differential methylation in the eastern oyster (*Crassostrea virginica*). *Frontiers in Marine Science*, 225. <https://doi.org/10.3389/fmars.2020.00225>
- Vidal-Dupiol, J., Chaparro, C., Pralong, M., Pontaroti, P., Grunau, C., & Mitta, G. (2019). Sequencing, de novo assembly and annotation of the genome of the scleractinian coral, *Pocillopora acuta*. *bioRxiv*, (p. 698688). <https://doi.org/10.1101/698688>
- Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology*, 11(2), R14. <https://doi.org/10.1186/gb-2010-11-2-r14>
- Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, 328(5980), 916–919. <https://doi.org/10.1126/science.1186366>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Trigg, S. A., Venkataraman, Y. R., Gavery, M. R., Roberts, S. B., Bhattacharya, D., Downey-Wall, A., Eirin-Lopez, J. M., Johnson, K. M., Lotterhos, K. E., Puritz, J. B., & Putnam, H. M. (2021). Invertebrate methylomes provide insight into mechanisms of environmental tolerance and reveal methodological biases. *Molecular Ecology Resources*, 00, 1–15. <https://doi.org/10.1111/1755-0998.13542>