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

Shelly A. Trigg*[1], Yaamini R. Venkataraman*[1], Mackenzie R. Gavery[2], Steven B. Roberts[1], Debashish Bhattacharya[3], Alan Downey-Wall[4], Jose M. Eirin-Lopez[5], Kevin M. Johnson[6], Katie E/ Lotterhos[4], Jonathan B. Puritz[7] and Hollie M. Putnam[7+]

[1] University of Washington, School of Aquatic and Fishery Sciences 1122 NE Boat St. Seattle, WA, 98195, USA
[2] NOAA Northwest Fisheries Science Center Montlake 2725 Montlake Blvd E, Seattle, WA, 98112, USA
[3] Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, NJ 08901 USA
[4] Department of Marine and Environmental Sciences, Northeastern University, 430 Nahant Road, Nahant, MA 01908
[5] Florida International University, Environmental Epigenetics Laboratory, Institute of Environment 3000 NE 151 St. North Miami, FL, 33181, USA
[6]Department of Biological Sciences, University of Rhode Island, Kingston, RI 02881, USA

[+] corresponding author: hputnam@uri.edu

* joint first authors/equal contribution

Table of Contents:

Additional File 5: Table ST4.csv
Lambda conversion efficiency and alignment statistics.
Descriptive statistics of sequencing reads resulting from Bismark alignment to the lambda genome.

Additional File 6: Figure SF2.jpg
Bisulfite conversion efficiency assessment.
Bisulfite conversion efficiency calculated from lambda alignments or estimated from non-CpG methylation from coral alignments for *M. capitata* libraries and *P. acuta* libraries.

Additional File 7: Table ST5.csv
Bisulfite conversion efficiency statistics.
Results from ANOVA test comparing conversion efficiency calculated from lambda alignments and conversion efficiency estimated from non-CpG methylation from coral alignments.

Additional File 8: Figure SF3.jpg
Comparison of CpG methylation status.
Percent of highly methylated (≥ 50%; darkest shade), moderately methylated (10-50%; medium shade), and lowly methylated CpGs (< 10%; lightest shade) detected by each method **A)** for *M. capitata* and **B)** *P. acuta*, based on the number of CpGs captured by each method separately. Principal Coordinate Analyses associated with PERMANOVA and beta-dispersion tests related to Table ST6 that show differences in proportion of CpGs that are highly (≥ 50%), moderately (10-50%), or lowly (≤ 10%) methylated in **C)** *M. capitata* and **D)** *P. acuta*. WGBS is represented by green circles, RRBS by purple triangles, and MBDBS by orange diamonds. Percent 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 $\alpha = 0.05$ level.

Additional File 9: Figure SF4.jpg
Methylation profile comparison of shared loci
PCA of CpG methylation for loci covered at 5x read depth in all samples for **A)** *M. capitata* and **B)** *P. acuta*.

Additional File 10: Figure SF5.jpg
Genome coverage at various sequencing depths.
Estimated fraction of CpG sites in the genome covered by at least 5 reads at different sequencing depths (number of M read pairs) for **A)** *M. capitata* and **B)** *P. acuta* samples for each bisulfite sequencing method. 'Observed' (opaque line and dots) denotes the fraction of genome-wide CpG loci covered by at least 5 reads determined from pooled

data that was subsampled at 50M, 100M, 150M, and 200M reads. 'Estimated' (translucent line and dots) denotes the fraction of genome-wide CpG loci covered by at least 5 reads estimated by michaelis-menten modelling of the 'observed' data with standard error shown by shaded areas. All samples within a bisulfite sequencing method were pooled for the downsampling analyses.

Additional File 11: Figure SF6.jpg
Coverage of orthologs across methods
Mean proportion (n=3 samples per method) of CpGs per gene that have at least 5x coverage in all of the one-to-one orthologous genes, as identified by OrthoFinder (Putnam et al., 2020) for **A)** *M. capitata* and **B)** *P. acuta*.

Additional File 12: Table ST6.csv
Results from PERMANOVA and beta-dispersion tests for genomic location
PERMANOVA R^2 and Adonis *P*-value, and beta dispersion test *F*-statistic and ANOVA *P*-value for obtained by comparing CpG genomic location between methods for *M. capitata* and *P. acuta*.

Additional File 13: Table ST7.csv
Contingency test results for *M. capitata* and *P. acuta.*
Test statistics and *P*-values from chi-squared tests comparing CpG methylation status and genomic location between WGBS, RRBS, and MBDBS for *M. capitata* and *P. acuta*.

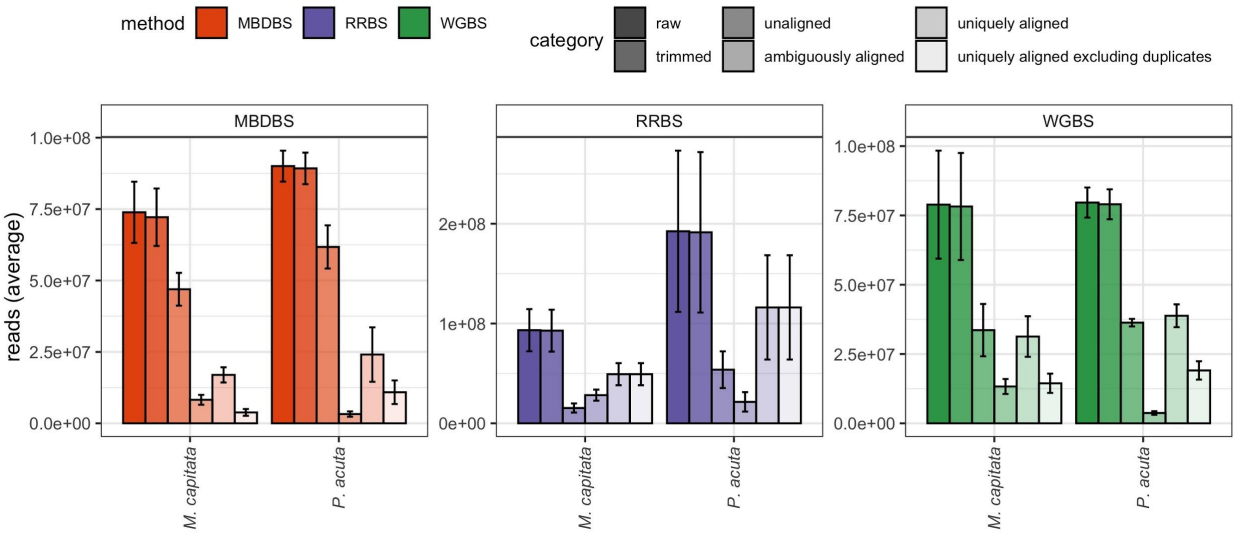

Figure SF1. Summary of sequencing depth and alignments for all libraries. Bars show average number of reads for each method and species and error bars show standard deviation.

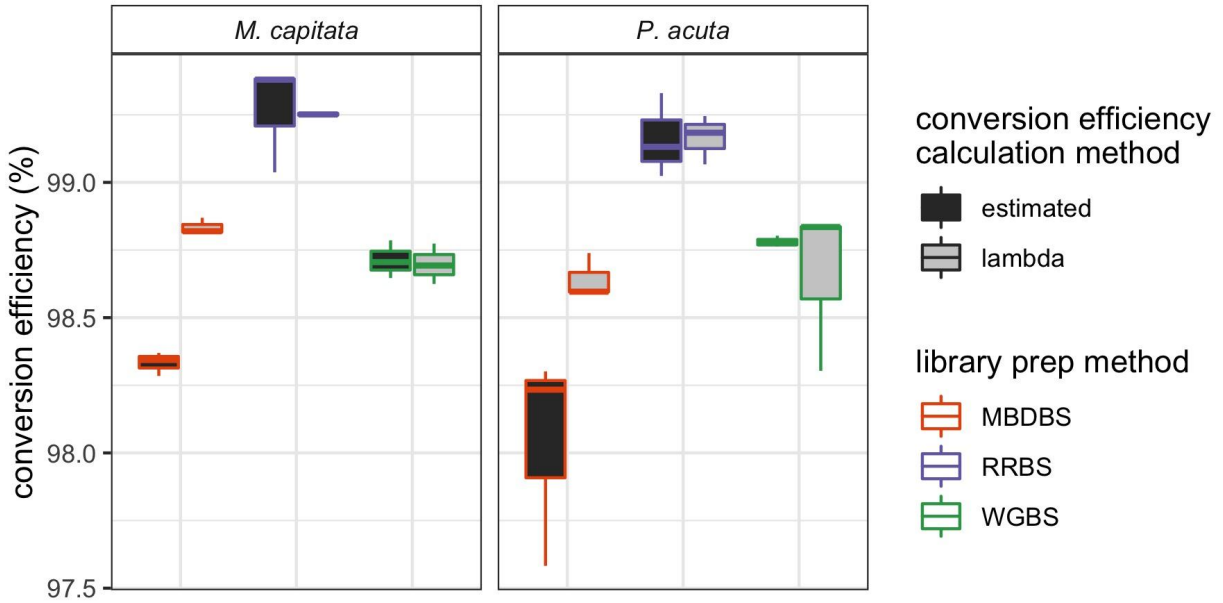

Figure SF2. Bisulfite conversion efficiency assessment. Bisulfite conversion efficiency calculated from lambda alignments or estimated from non-CpG methylation from coral alignments for *M. capitata* libraries and *P. acuta* libraries.

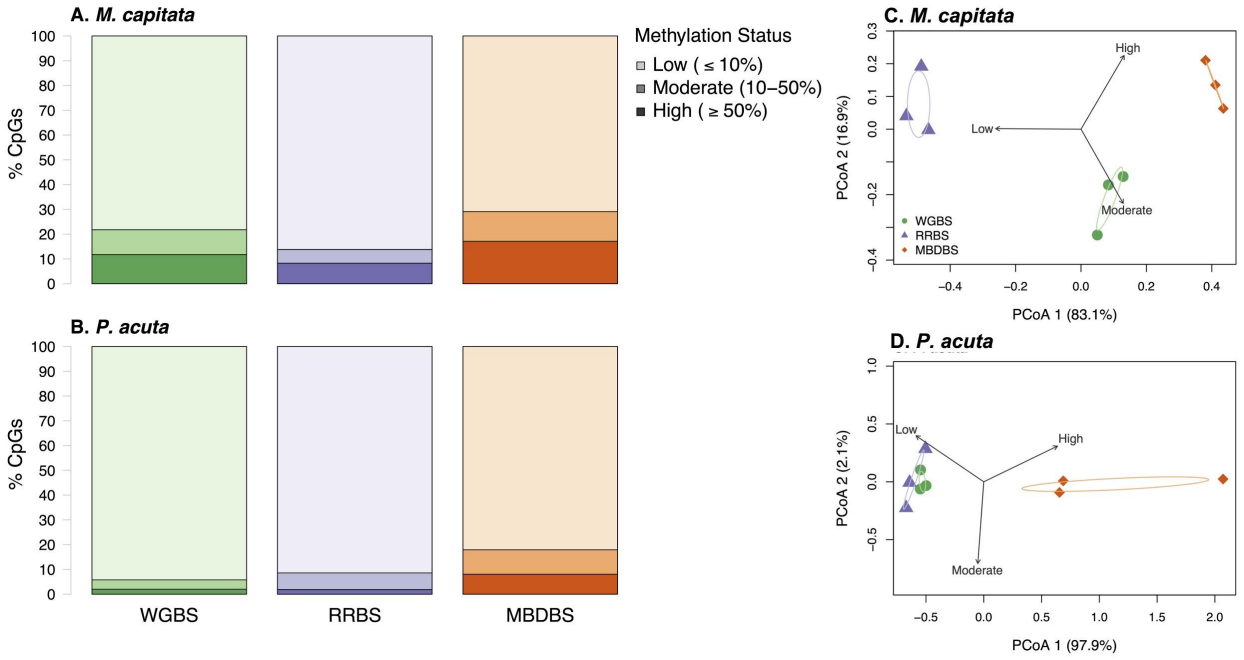

Figure SF3. Percent of highly methylated (≥ 50%; darkest shade), moderately methylated (10-50%; medium shade), and lowly methylated CpGs (< 10%; lightest shade) detected by each method **A**) for *M. capitata* and **B**) *P. acuta*, based on the number of CpGs captured by each method separately. Principal Coordinate Analyses associated with PERMANOVA and beta-dispersion tests related to **Table ST6** that show differences in proportion of CpGs that are highly (≥ 50%), moderately (10-50%), or lowly (≤ 10%) methylated in **C**) *M. capitata* and **D**) *P. acuta*. WGBS is represented by green circles, RRBS by purple triangles, and MBDBS by orange diamonds. Percent 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 α = 0.05 level.

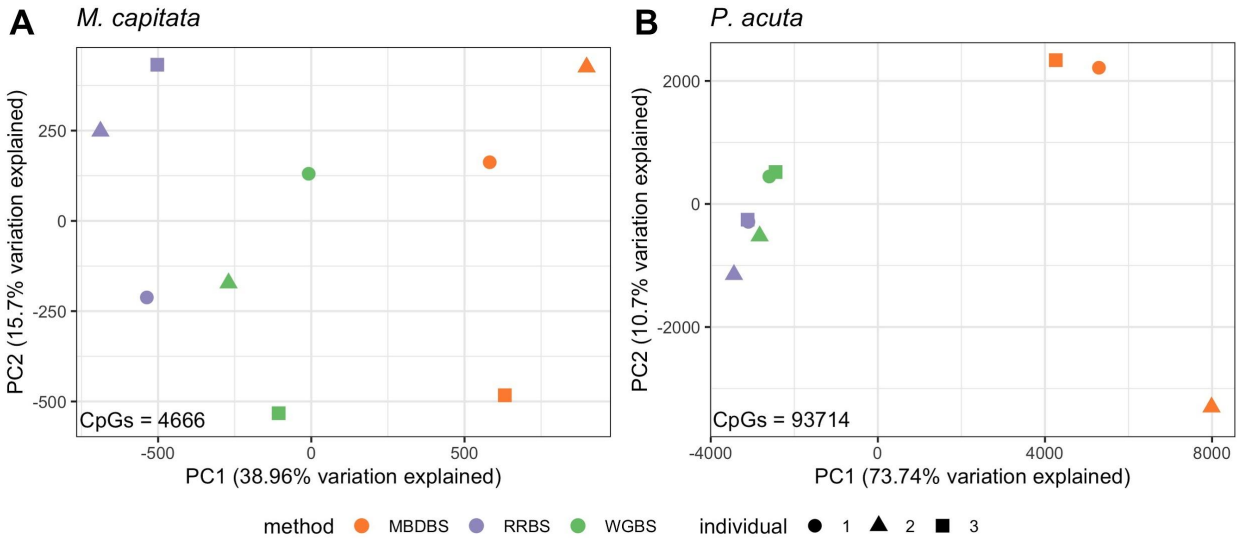

Figure SF4. PCA of CpG methylation for loci covered at 5x read depth in all samples for (**A**) *M. capitata* and (**B**) *P. acuta*.

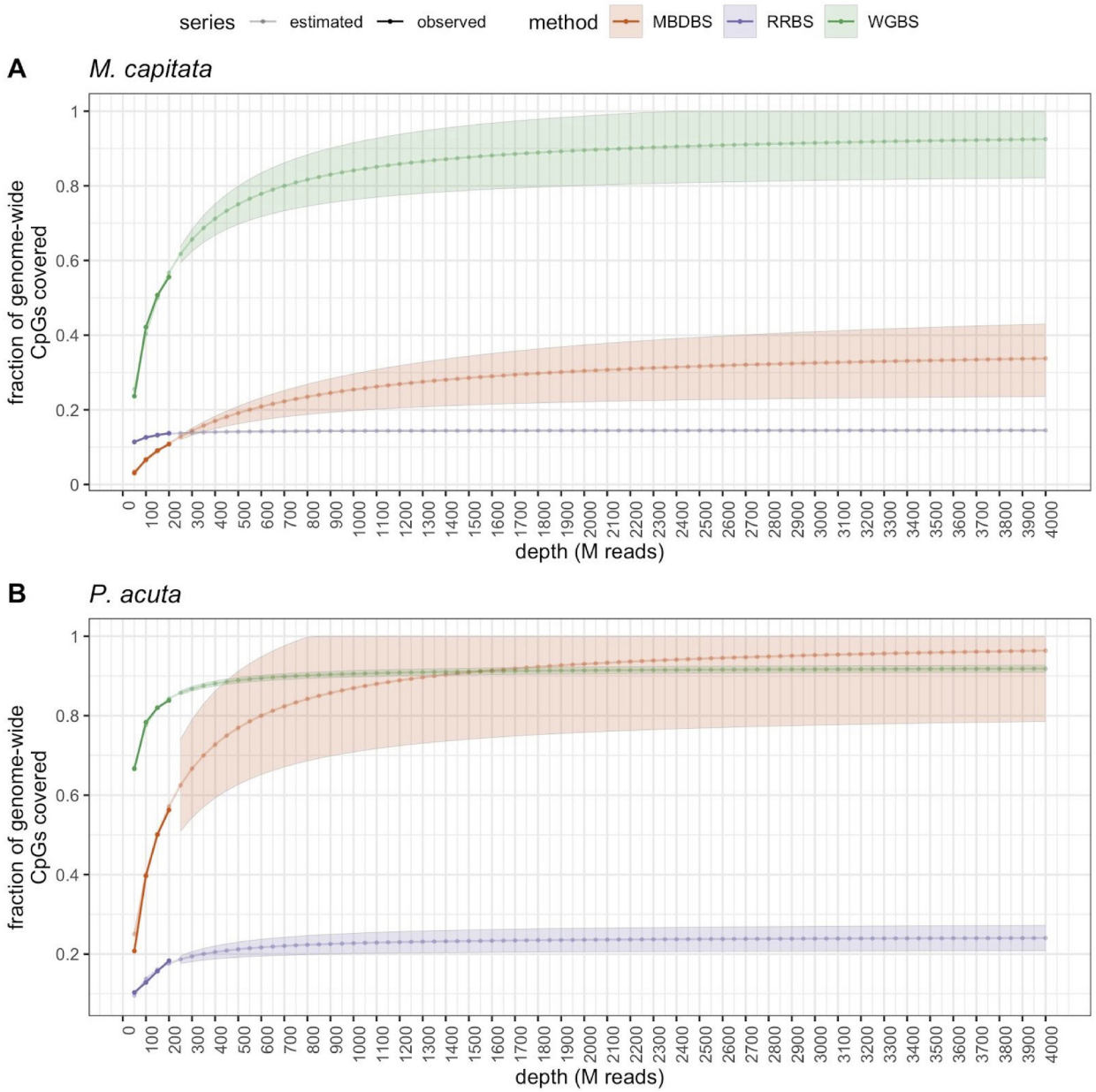

Figure SF5. Estimated fraction of CpG sites in the genome covered by at least 5 reads at different sequencing depths (number of M read pairs) for (**A**) *M. capitata* and (**B**) *P. acuta* samples for each bisulfite sequencing method. 'Observed' (opaque line and dots) denotes the fraction of genome-wide CpG loci covered by at least 5 reads determined from pooled data that was subsampled at 50M, 100M, 150M, and 200M reads. 'Estimated' (translucent line and dots) denotes the fraction of genome-wide CpG loci covered by at least 5 reads estimated by michaelis-menten modelling of the 'observed' data with standard error shown by shaded areas. All samples within a bisulfite sequencing method were pooled for the downsampling analyses.

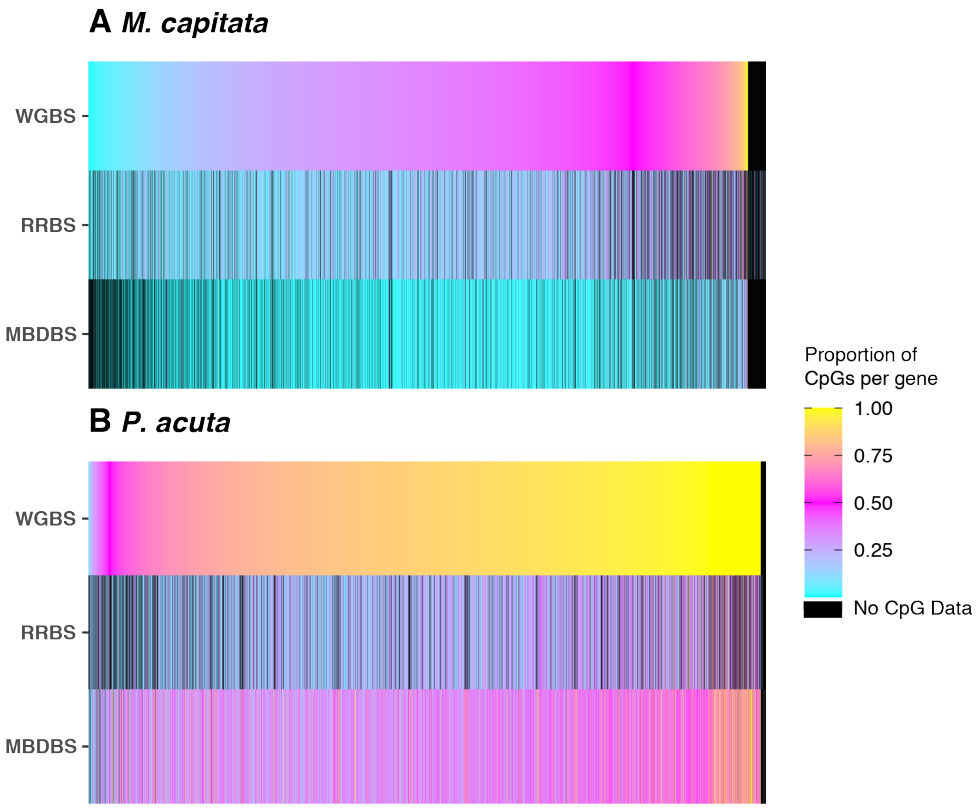

Figure SF6. Mean proportion (n=3 samples per method) of CpGs per gene that have at least 5x coverage in all of the one-to-one orthologous genes, as identified by OrthoFinder (Putnam et al., 2020) for **A**) *M. capitata* and **B**) *P. acuta*.