Performance of Bisulfite Sequencing Methods in Invertebrates

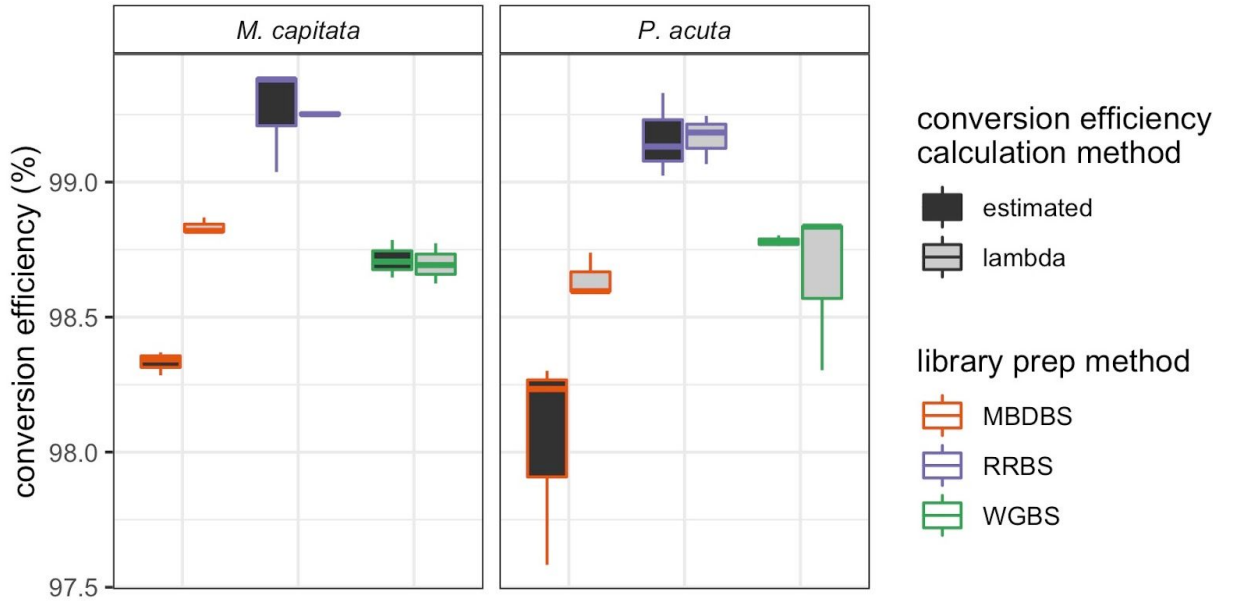
Trigg*[1], YR Venkataraman*[1], Gavery[2], Roberts[1], ***XX*** and HM Putnam[^+]

[1] University of Washington, School of Aquatic and Fisheries Sciences 1122 NE Boat St. Seattle, WA, 98195, USA
[2] NOAA Northwest Fisheries Science Center Montlake 2725 Montlake Blvd E, Seattle, WA, 98112, USA
[^]Department of Biological Sciences, University of Rhode Island, Kingston, RI 02881, USA

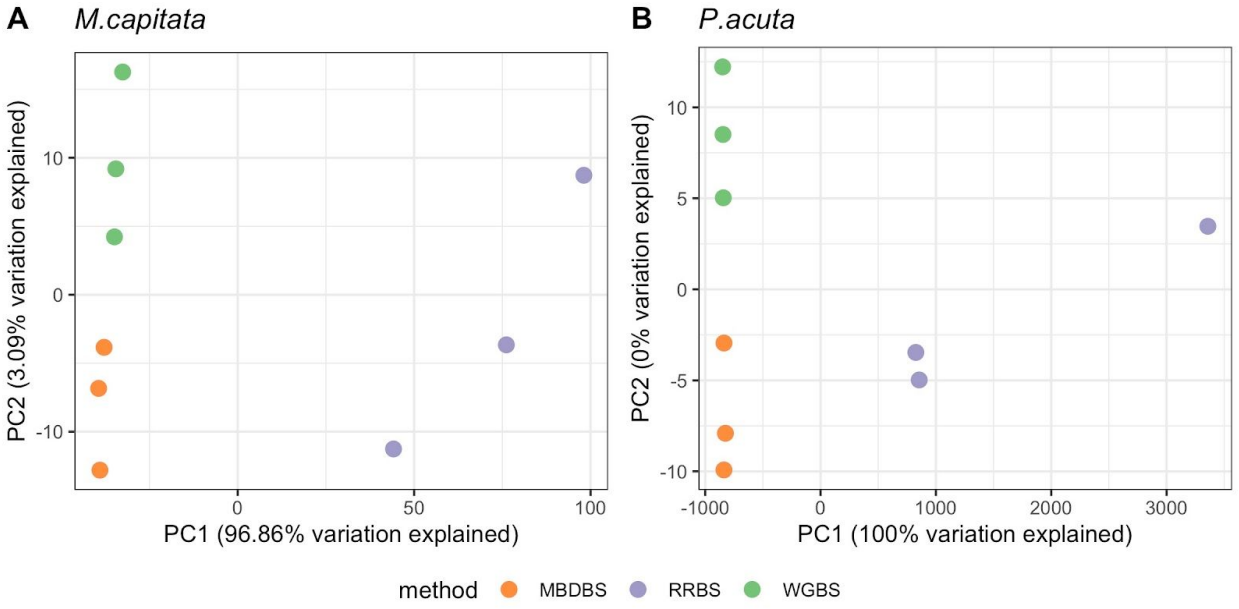
[+]corresponding author: hputnam@uri.edu
* equal contribution

Table of Contents:

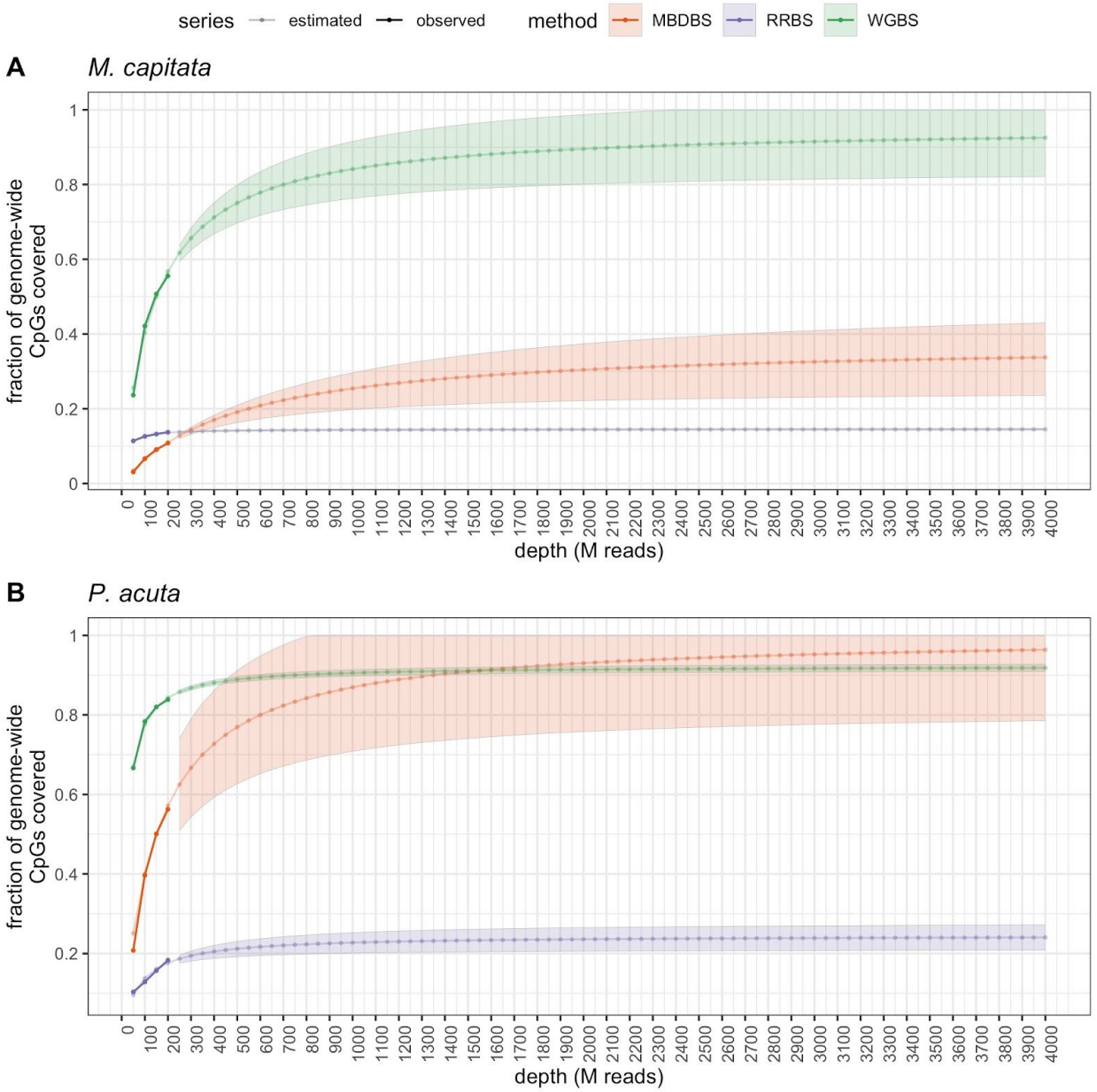
Supplementary Figure 1. Bisulfite conversion efficiency assessment. Bisulfite conversion efficiency is different across library preparation methods whether calculated from lambda alignments or estimated from non-CpG methylation from coral alignments for *M. capitata* libraries and *P. acuta* libraries. Generally, bisulfite conversion efficiency calculation methods are not different.



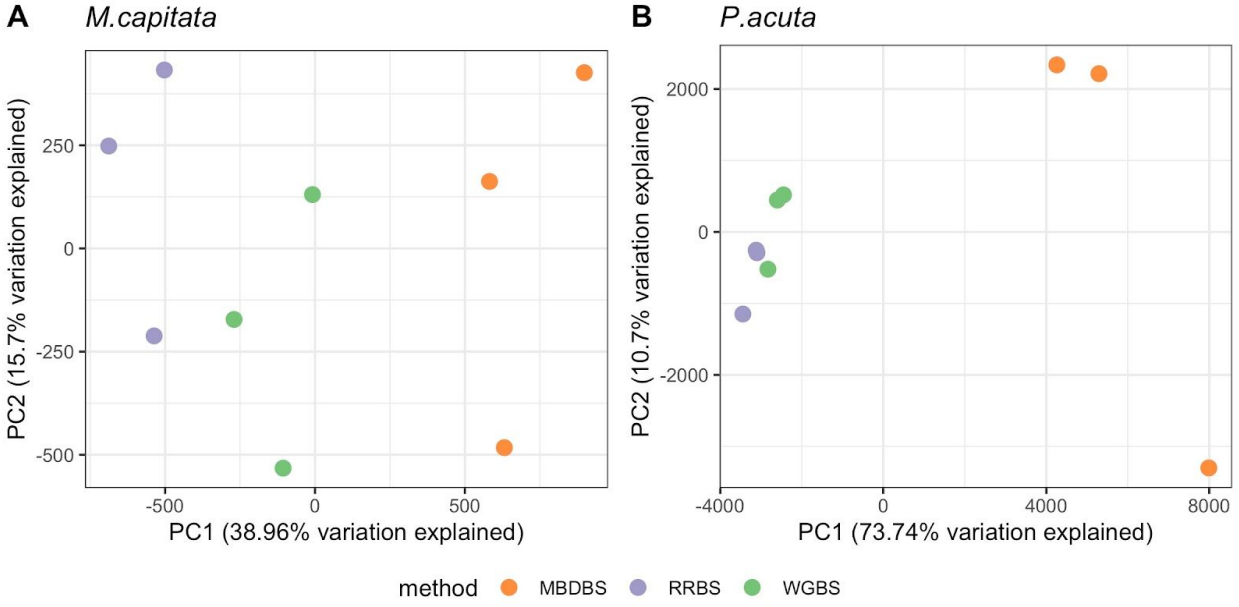
Supplementary Figure 2. PCA of alignment features presented in the summary section of Qualimap MultiBamQC reports for (**A**) *M. capitata* and (**B**) *P. acuta*.



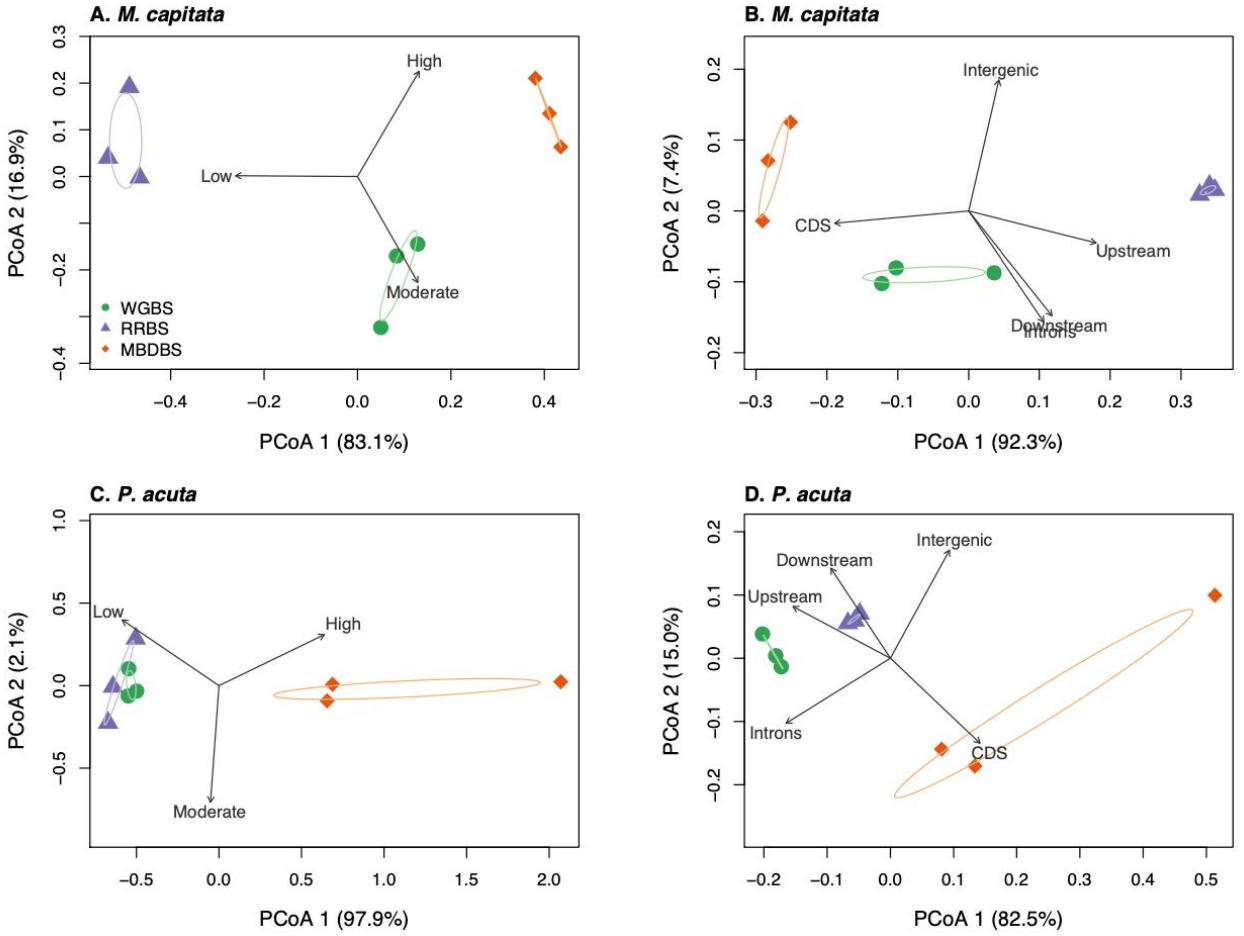
Supplementary Figure 3. Estimated fraction of genome-wide CpG loci covered by at least 5 reads with different sequencing depths for (**A**) *M. capitata* and (**B**) *P. acuta*. 'Observed' (opaque line and dots) denotes the fraction of genome-wide CpG loci covered by at least 5 reads determined from 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.



Supplementary Figure 4. PCA of CpG methylation for loci covered at 5x read depth in all samples for (**A**) *M. capitata* and (**B**) *P. acuta*.



Supplementary Figure 5. Principal Coordinate Analyses associated with perMANOVA and beta-dispersion tests (see **Supplementary Table 6** and **Supplementary Table 8**). Differences in proportion of CpGs that are highly (≥ 50%), moderately (10-50%), or lowly (≤ 10%) methylated in **A**) *M. capitata* and **C**) *P. acuta*. Differences in proportion of CpGs in various genomic locations (CDS, introns, upstream flanks, downstream flanks, and intergenic regions) for **B**) *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.



Supplementary Table 6. Results from perMANOVA and beta-dispersion tests for methylation status. Significant global perMANOVA results were interrogated further using a beta-dispersion model to determine if significant differences were the result of centroid of variance differences. Non-significant beta-dispersion model results (ANOVA *P*-value > 0.05) indicated significant results could be attributed to centroid differences. Pairwise perMANOVA tests were used to investigate differences between two sequencing methods. Beta-dispersion models were only used with pairwise perMANOVA tests if the global beta-dispersion model indicated a significant effect of group variance (ANOVA *P*-value < 0.05) on the global perMANOVA result.

Species	Contrast	perMANOVA R^2	*Adonis* *P*-value	Beta-dispersion model F-statistic	ANOVA *P*-value
M. capitata	Global	0.96	0.004	0.58	0.59
	WGBS vs. RRBS	0.91	0.10	N/A	N/A
	WGBS vs. MBDBS	0.91	0.10	N/A	N/A
	RRBS vs. MBDBS	0.97	0.10	N/A	N/A
P. acuta	Global	0.80	0.02	8.40	0.02
	WGBS vs. RRBS	0.05	0.80	2.46	0.19
	WGBS vs. MBDBS	0.76	0.10	12.53	0.02
	RRBS vs. MBDBS	0.76	0.10	6.21	0.07

Supplementary Table 7. Logs-odd ratio results from post-hoc estimated marginal mean calculations for methylation status. Pairwise generalized linear mixed-effect models were constructed for each methylation status separately (high, moderate, low) to investigate significant global perMANOVA results. Estimated marginal mean calculations were used to obtain contrasts for each pairwise sequencing method comparison. All P-values were corrected using an FDR threshold.

Species	Methylation status	Contrast	t-ratio	*P*-value
M. capitata	High	WGBS vs. RRBS	7.28	0.0019
		WGBS vs. MBDBS	-12.59	0.0034
		RRBS vs. MBDBS	-19.10	0.00013
	Moderate	WGBS vs. RRBS	15.38	0.0002
		WGBS vs. MBDBS	-0.83	0.46
		RRBS vs. MBDBS	-16.12	0.0002
	Low	WGBS vs. RRBS	-25.42	2.13×10^{-5}
		WGBS vs. MBDBS	17.96	5.65×10^{-5}
		RRBS vs. MBDBS	41.97	5.8×10^{-6}
P. acuta	High	WGBS vs. RRBS	0.06	0.96
		WGBS vs. MBDBS	-5.03	0.01
		RRBS vs. MBDBS	-5.13	0.01
	Moderate	WGBS vs. RRBS	0.44	0.69
		WGBS vs. MBDBS	-6.37	0.005
		RRBS vs. MBDBS	-6.76	0.005
	Low	WGBS vs. RRBS	-0.13	0.91
		WGBS vs. MBDBS	3.39	0.04
		RRBS vs. MBDBS	3.44	0.04

Supplementary Table 8. Results from perMANOVA and beta-dispersion tests for genomic location. Significant global perMANOVA results were interrogated further using a beta-dispersion model to determine if significant differences were the result of centroid of variance differences. Non-significant beta-dispersion model results (ANOVA *P*-value > 0.05) indicated significant results could be attributed to centroid differences. Pairwise perMANOVA tests were used to investigate differences between two sequencing methods. Beta-dispersion models were only used with pairwise perMANOVA tests if the global beta-dispersion model indicated a significant effect of group variance (ANOVA *P*-value < 0.05) on the global perMANOVA result.

Species	Contrast	perMANOVA R^2	*Adonis* *P*-value	Beta-dispersion model F-statistic	ANOVA *P*-value
M. capitata	Global	0.96	0.004	4.55	0.06
	WGBS vs. RRBS	0.94	0.10	N/A	N/A
	WGBS vs. MBDBS	0.79	0.10	N/A	N/A
	RRBS vs. MBDBS	0.98	0.10	N/A	N/A
P. acuta	Global	0.67	0.005	13.36	0.006
	WGBS vs. RRBS	0.94	0.10	1.40	0.30
	WGBS vs. MBDBS	0.64	0.10	12.69	0.02
	RRBS vs. MBDBS	0.51	0.10	14.40	0.02

Supplementary Table 9. Logs-odd ratio results from post-hoc estimated marginal mean calculations for genomic location. Pairwise generalized linear mixed-effect models were constructed for each genomic location separately (CDS, introns, upstream flanks, downstream flanks, intergenic regions) to investigate significant global perMANOVA results. Estimated marginal mean calculations were used to obtain contrasts for each pairwise sequencing method comparison. All *P*-values were corrected using an FDR threshold.

Species	Genomic location	Contrast	t-ratio	*P*-value
M. capitata	CDS	WGBS vs. RRBS	13.45	0.0003
		WGBS vs. MBDBS	-6.43	0.003
		RRBS vs. MBDBS	-19.58	0.0001
	Introns	WGBS vs. RRBS	-0.37	0.73
		WGBS vs. MBDBS	9.60	0.001
		RRBS vs. MBDBS	9.97	0.001
	Upstream Flanks	WGBS vs. RRBS	-0.87	0.43
		WGBS vs. MBDBS	2.74	0.08
		RRBS vs. MBDBS	3.61	0.07
	Downstream Flanks	WGBS vs. RRBS	4.15	0.02
		WGBS vs. MBDBS	4.45	0.02
		RRBS vs. MBDBS	0.29	0.78
	Intergenic Regions	WGBS vs. RRBS	-5.24	0.02
		WGBS vs. MBDBS	-4.37	0.02
		RRBS vs. MBDBS	0.87	0.43
P. acuta	CDS	WGBS vs. RRBS	-3.11	0.04
		WGBS vs. MBDBS	-19.80	0.0001
		RRBS vs. MBDBS	-16.73	0.0001

	Introns	WGBS vs. RRBS	1.68	0.17
		WGBS vs. MBDBS	3.41	0.08
		RRBS vs. MBDBS	1.75	0.17
	Upstream Flanks	WGBS vs. RRBS	-0.79	0.48
		WGBS vs. MBDBS	3.56	0.04
		RRBS vs. MBDBS	4.34	0.04
	Downstream Flanks	WGBS vs. RRBS	3.70	0.02
		WGBS vs. MBDBS	8.63	0.003
		RRBS vs. MBDBS	4.94	0.01
	Intergenic Regions	WGBS vs. RRBS	-1.12	0.56
		WGBS vs. MBDBS	-1.01	0.56
		RRBS vs. MBDBS	0.11	0.92