

## Genome-wide DNA methylation landscape of gonads in the deep-sea mussel *G. platifrons*

### Summary of gonadal WGBS

To investigate the overall methylation patterns of the deep-mussel *G. platifrons* and detect those sites associated with sex determination/differentiation, WGBS was performed on the gonad tissues of female and male *G. platifrons*. A total of 336.55 Gb of clean sequencing data average 56.09 Gb per gonad sample were obtained for the whole genome bisulfite of *G. platifrons*. A statistical summary of various sequencing result parameters is reported in Table 1. Reads mapping efficiency for each library is 48.47%, 49.51%, 48.76%, 44.62%, 49.58% and 48.4% for M1, M2, M3, F1, F2 and F3 libraries, respectively. Overall, approximately 3.09 and 3.22% of all genomic C sites were methylated in the female and male *G. platifrons*, respectively. The distribution of the DNA methylation sites in the whole genome of *G. platifrons* were separately analyzed in the female and male individuals and we found methylation exist in three sequence regions: CG, CHG and CHH (H represents base A, C or T) and (Figure 1). On average, the genome of female *G. platifrons* exhibited 23.30% (mCG), 0.18% (mCHG) and 0.68% (mCHH) methylation on the total sequenced CpG, CHG and CHH sites, respectively. According, the male *G. platifrons* showed 24.89, 0.20 and 0.62% methylation in the CG, CHG and CHH regions respectively, with a slight increase in CG and CHG methylation percentage compared to the female mussel (Table 2), but the methylation density was not coincident with methylation level, which female 5mC density was slight higher than male (Table 3). The proportions of mCG, mCHG and mCHH on the total mC sites were analyzed, which reflects the similar rates in the female and male *G. platifrons*, with mC most found at CG sites (>96%) and much less frequently at CHG (<1%) and CHH (~2%) sites (Figure 2). In short, the methylation levels of the female and male gonad shared similar trends, with high levels in CG sites and low levels in CHG and CHH sites.

Table 1 Data summary of WGBS for the female and male gonad of *G. platifrons*

Libr ary	Raw Reads/Gb	Clean Reads/Gb	Q20 (%)	BS Conversion rate (%)	Mapped reads	Mapping rate (%)
<b>F1</b>	186822458 /54.44	186822444 /50.69	98.50	99.91	83360174	44.62
<b>F2</b>	186172849 /54.41	186172842 /50.67	98.60	99.90	92304495	49.58
<b>F3</b>	166278544 /48.47	166278530 /45.13	98.38	99.90	80478808	48.40
<b>M1</b>	208819956 /60.78	208819931 /56.59	98.31	99.91	10121502 0	48.47
<b>M2</b>	214007371 /62.34	214007357 /58.04	98.32	99.91	10595504 2	49.51
<b>M3</b>	192761046 /56.11	192761018 /52.24	98.33	99.91	93990272	48.76

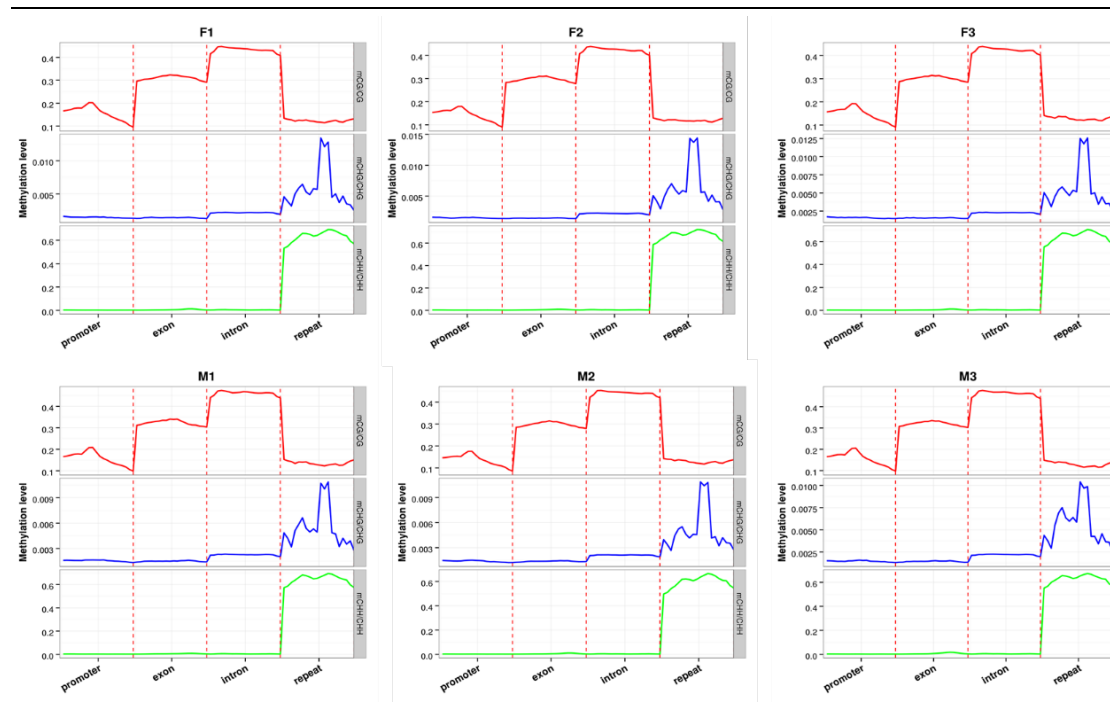


Figure1 Differential methylation level of gonads in different gene regions.

Table 2 Statistical information of methylation at site C

Samples	C	C (Mb)	CG	CHG	CHH	Mean	Mean	Mean	Mean
	coverage		(Mb)	(Mb)	(Mb)	C (%)	CG (%)	CHG	CHH
	Mean							(%)	(%)
<b>F1</b>	3.3	2318.9	255.4	397.3	1666.2	3.14	23.85	0.18	0.67
<b>F2</b>	3.7	2625.2	292.5	454.2	1878.4	3.11	22.92	0.18	0.73
<b>F3</b>	3.2	2244.9	245.4	384	1615.5	3.02	23.14	0.19	0.63
<b>M1</b>	3.9	2791	304.3	477.1	2009.5	3.29	25.59	0.2	0.65
<b>M2</b>	4.3	3076.1	341.3	529.7	2205.2	3.03	23.59	0.19	0.53
<b>M3</b>	3.7	2595.4	287.8	447	1860.6	3.34	25.48	0.2	0.68

C coverage Mean: the average coverage depth of all C loci in the genome; C(Mb): the number of bases mapped to the C sites in the genome; CG(Mb) : the number of cytosine bases mapped on CG region of genome; CHG(Mb): the number of cytosine bases mapped on CHG region of genome; CHH(Mb): the number of cytosine bases mapped on CHH region of genome; Mean C(%): the average methylation level of all C sites in the genome; Mean CG(%): mean methylation level in CG region; Mean CHG(%): mean methylation level in CHG region; Mean CHH(%): mean methylation level in CHH region

Table 3 Statistical information of 5mC density in different sequence contexts

	CG_5mC	CHG_5mC	CHH_5mC	sum
	(%)	(%)	(%)	(%)
<b>F1</b>	36.1195	1.1759	1.211	38.5064
<b>F2</b>	34.8019	1.1553	1.1997	37.1569
<b>F3</b>	35.2119	1.1952	1.2268	37.6339
<b>M1</b>	35.2974	1.2366	1.2485	37.7825
<b>M2</b>	33.6136	1.1041	1.0583	35.776
<b>M3</b>	35.2556	1.1336	1.1288	37.518

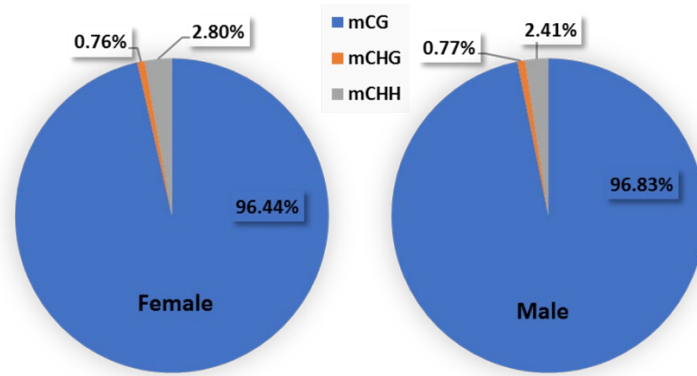


Figure 2 The proportion of mC in the CG, CHG and CHH regions.

### DMRs analysis

The length distribution and the 5mC levels of DMRs were showed in Figure 3. The DMRs length were much accumulation in 0-500bp, and compared to CHG and CHH sequence contexts, much differences of 5mC levels were found in the GC sequence context.

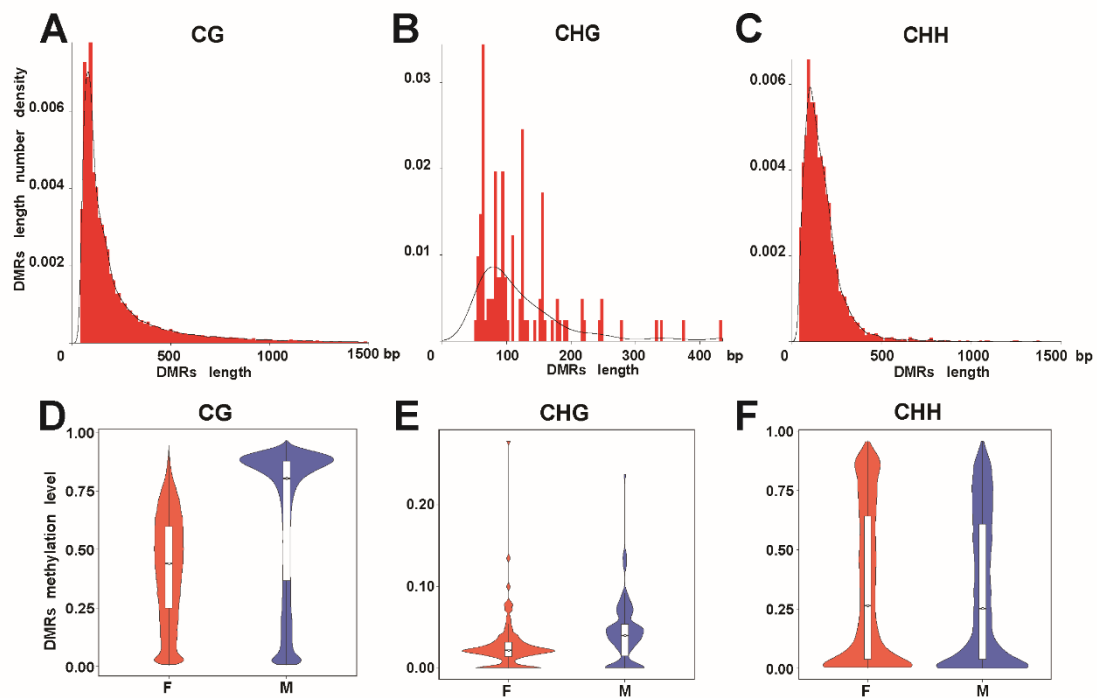


Figure 3 Length and levels distribution of DMRs. Figure A, B, and C represent DMRs length distribution in CG, CHG and CHH sequence contexts respectively, and Figure D, E, and F represent 5mC levels distribution of two sexes in CG, CHG and CHH

respectively.

### Correlation between Sex-related gene expression levels and 5mC levels

To find the roles of methylation in gene expression regulation of sex-related genes, we analyzed the 5mC level in the TSS 2K region (where always including the genes' promoter region) of these genes between female and male. We found a general trend whereby low level of expression level associated with increased 5mC level of corresponding gene for females, however, the trend was reversed for males, similar pattern was also found in the gene body and TTS 2K regions (Figure 4). Spearman correlation between 5mC levels and gene expression were calculated (Table 4 & 5).

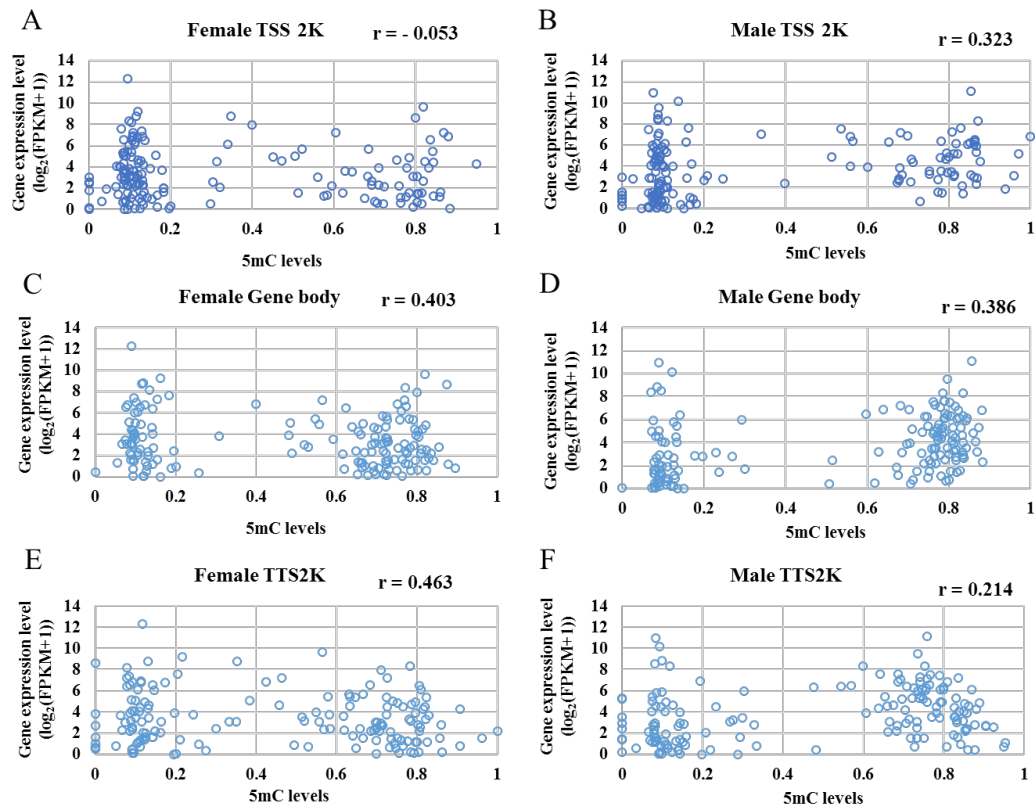


Figure 4 Correlation analysis of sex-related gene expression levels and 5mC levels in the TSS 2K, the gene body, and TTS 2K regions. Spearman correlation r were showed in A to F.

Table4. The correlation between methylation levels of TSS2K, gene body and TTS2K regions and sex-related genes expression in female *G. platifrons*

Spearman r	TSS2K	Gene body	TTS2K
r	-0.05301	0.4028	0.4632
95% confidence interval	-0.2104 to 0.1071	0.2603 to 0.5281	0.3282 to 0.5796
P	0.5042	<0.0001	<0.0001
Significant	No	Yes	Yes

Table5. The correlation between methylation levels of TSS2K, gene body and TTS2K regions and sex-related genes expression in male *G. platifrons*

Spearman r	TSS2K	Gene body	TTS2K
r	0.3232	0.3862	0.2136
95% confidence interval	0.1729 to 0.4588	0.2419 to 0.5138	0.0563 to 0.3605
P	<0.0001	<0.0001	0.0065
Significant	Yes	Yes	Yes