

PROF. THOMAS MOCK (Orcid ID : 0000-0001-9604-0362)

Article type : Regular Manuscript

Single-base methylome profiling of the giant kelp *Saccharina japonica* reveals significant differences in DNA methylation to microalgae and plants

Authors:

Xiao Fan^{1,2,3}, Wentao Han¹, Linhong Teng^{1,4,5}, Peng Jiang⁶, Xiaowen Zhang¹, Dong Xu¹, Chang Li⁷, Matteo Pellegrini⁸, Chunhui Wu⁶, Yitao Wang¹, Michelle Joyce Slade Kaczurowski⁹, Xin Lin¹⁰, Leila Tirichine¹¹, Thomas Mock^{12*}, Naihao Ye^{1,2*}

Affiliations:

¹Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, China

²Function Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao, China

³Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306, China

⁴College of Life Science, Dezhou University, Dezhou 253023, China

⁵Shandong Key Laboratory of Biophysics, Dezhou University, Dezhou 253023, China

⁶Chinese Academy of Sciences, Qingdao 266071, China.

⁷BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, China

⁸Department of Molecular, Cell and Developmental Biology, and Institute for Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA

⁹Biological Sciences, Flinders University, GPO Box 2100, Adelaide, SA 5001, Australia

¹⁰State Key Laboratory of Marine Environmental Science, College of Ocean & Earth Sciences, Xiamen University, China

¹¹CNRS UMR 6286, Faculté des Sciences et des Techniques, Université de Nantes, 2 rue de la Houssinière, 44322, Nantes, France

¹²School of Environmental Sciences, University of East Anglia, Norwich Research Park, NR4 7TJ Norwich, UK

Corresponding authors:*

Naihao Ye: Email: yenh@ysfri.ac.cn Tel: +86-532-85830360

Thomas Mock: Email: t.mock@uea.ac.uk Tel: +44(0) 1603 592566

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/nph.16125

This article is protected by copyright. All rights reserved.

Received: 27 July 2019

Accepted: 6 August 2019

ORCID IDs:

Xiao Fan: <https://orcid.org/0000-0002-7093-8868>

Wentao Han: <http://orcid.org/0000-0002-8294-1044>

Linhong Teng: <https://orcid.org/0000-0002-2245-9563>

Xiaowen Zhang: <https://orcid.org/0000-0001-5338-1837>

Chang Li: <https://orcid.org/0000-0002-7823-3467>

Matteo Pellegrini: <https://orcid.org/0000-0001-9355-9564>

Xin Lin: <https://orcid.org/0000-0001-8323-5056>

Naihao Ye: <http://orcid.org/0000-0002-8294-1044>

Thomas Mock: <https://orcid.org/0000-0001-9604-0362>

Summary

- Brown algae have convergently evolved plant-like body plans and reproductive cycles, which in plants are controlled by differential DNA methylation. Here we provide the first single-base methylome profiles of haploid gametophytes and diploid sporophytes of a multicellular alga.
- Although only *c.* 1.4% of cytosines in *Saccharina japonica* were methylated mainly at CHH sites and characterised by 5-methylcytosine (5mC), there were significant differences between life-cycle stages. DNA methyltransferase 2 (DNMT2), known to efficiently catalyze tRNA methylation, is assumed to methylate the genome of *S. japonica* in the structural context of tRNAs as the genome does not encode any other DNA methyltransferases. Circular and long non-coding RNA genes were the most strongly methylated regulatory elements in *S. japonica*.
- Differential expression of genes was negatively correlated with DNA methylation with the highest methylation levels measured in both haploid gametophytes. Hypomethylated and highly expressed genes in diploid sporophytes included genes involved in morphogenesis and halogen metabolism.

- Our data give evidence that cytosine methylation, although occurring at a low level, is significantly contributing to the formation of different life-cycle stages, tissue differentiation, and metabolism in brown algae.

Key Words: DNA methylation, WGBS-seq, MeDIP-seq, BS-PCR, brown algae, life-cycle stages, gene expression, DNMT2

Introduction

Cytosine DNA methylation is a common epigenetic mark essential for genomic imprinting, X-chromosome inactivation, silencing of transposable elements as well as regulation of gene expression in many species (Meissner *et al.*, 2008; Feng *et al.*, 2010; Law & Jacobsen, 2010; Stelzer *et al.*, 2015). Genome-wide methylation studies with many different plant and animal species but also unicellular eukaryotes (e.g. fungi, algae) over the last few years have revealed significant inter- and intra-specific variations in cytosine DNA methylation (Cokus, *et al.*, 2008; Lee *et al.*, 2010; Molaro *et al.*, 2011; Greaves *et al.*, 2012; Ziller *et al.*, 2013). For example, vertebrate genomes are characterised by significant levels of cytosine methylation whereas genomes of invertebrates, plants, and fungi are characterized by sparse DNA methylation (Sturgill *et al.*, 2007; Feng *et al.*, 2010). Generally, the degree of DNA methylation is positively correlated with the complexity of organisms, i.e. prokaryotes and unicellular eukaryotes have much lower levels of DNA methylation compared to organisms with complex life cycles, developmental stages and cell-type differentiation (Molaro *et al.*, 2011; Lopez *et al.*, 2015). Furthermore, there are significant differences in establishing, maintaining and modifying DNA methylation in plants vs animals. For instance, in plants, DNA methylation has been observed for GC, CHG and CHH contexts with H being any nucleotide but G. DNA methylation in plants predominantly occurs on repetitive DNA elements (e.g. transposons) regardless of their life cycle and developmental stages (Lopez *et al.*, 2015; Takuno *et al.*, 2016). Similar DNA methylation patterns have been observed for unicellular photosynthetic eukaryotes (e.g. microalgae) (Veluchamy *et al.*, 2013a). In contrast, DNA methylation in mammals mostly occurs in the GC context throughout the genome except in clusters near promoters (CpG islands) (Molaro *et al.*, 2011; Smith *et al.*, 2012; Hon *et al.*, 2013; Kundaje *et al.*, 2015). Generally, DNA methylation in plants and animals is dynamic with genome-wide reduction during both male and female gametogenesis as well as development (Molaro *et al.*, 2011; Shao *et al.*, 2014; Lopez *et al.*, 2015). However, in plants, many DNA methylation patterns seem to be inherited over many generations whereas

transgenerational DNA methylation in animals is much more variable (Hon *et al.*, 2013; Lister *et al.*, 2013; Shao *et al.*, 2014).

Brown algae, although only distantly related to plants and animals, have convergently evolved plant-like body plans and reproductive cycles including male and female gametogenesis (Charrier *et al.*, 2012). Furthermore, many brown algae have an alternative life cycle with two life stages, termed gametophytes (1N) and sporophytes (2N) (Charrier *et al.*, 2012; Cock *et al.*, 2014). Gametophytes from some kelp species, such as *Saccharina japonica*, can even develop into larger multicellular organisms but rarely reach tissue differentiation (Ye *et al.*, 2015). Some other brown algae, such as *Ectocarpus siliculosus*, have isomorphic life-cycle stages that are filamentous and phenotypically not distinguishable from one another (Luthringer *et al.*, 2014).

Brown algae comprise a group of *c.* 2000 species, possessing a large variety of phenotypes including the largest multicellular photosynthetic organisms in the ocean with distinct and specialized tissue differentiation, such as holdfast, blade and stripe. Despite their convergently evolved plant-like body plans, their ecological (e.g. main primary producers of temperate and polar rocky shores) and commercial significance (e.g. alginate, fucoidan), genomes of only three brown algal species are available to date, namely *E. siliculosus* (Cock *et al.*, 2010), *S. japonica* (Ye *et al.*, 2015) and *Cladosiphon okamuranus* (Nishitsuji *et al.*, 2016). A comparative study based on the genomes of *E. siliculosus* and *S. japonica* revealed that they share 4,309 gene families, which comprise 17,379 genes in *S. japonica* and 14,136 genes in *E. siliculosus*, covering 93% and 86% of the gene content of each genome, respectively. About 40% of the assembled *S. japonica* genome is comprised of repetitive elements, which is nearly twice as much as for the *E. siliculosus* genome (*c.* 23%) (Ye *et al.*, 2015).

DNA methylation in brown algal genomes was considered to be negligible based on preliminary HPLC analyses of deoxycytosine methylation (5mdC) of hydrolysed DNA from *E. siliculosus* (Cock *et al.*, 2010). These data indicated that the percentage of 5mdC in *E. siliculosus* is below 0.035%. However, our study using single-base DNA methylome profiling of a fully developed *S. japonica* sporophyte (SP), female (FG) and male (MG) gametophyte revealed that *c.* 1.4% of all cytosines in *S. japonica* were methylated in GC, CHG and CHH contexts potentially mediated by a DNA methyltransferase 2 (DNMT2) as there is no other DNA methyltransferase encoded in the genome of *S. japonica*, which is

different to most plants, animals and unicellular algae such as the green alga *Chlamydomonas reinhardtii* (Lopez *et al.*, 2015) and even the more closely related diatom *Phaeodactylum tricornutum* (Takuno *et al.*, 2016; Tirichine *et al.*, 2017). Among the three life-cycle stages of *S. japonica*, we found the highest level of DNA methylation in both gametophytes.

Furthermore, the highest methylated elements of any life cycle stage were found to be genes encoding non-coding RNAs (circular and long non-coding), which is different to other photosynthetic eukaryotes as they mostly methylate repetitive elements (e.g. transposons) (Sleutels *et al.*, 2002; Dinger *et al.*, 2008; Law & Jacobsen, 2010). Differentially methylated genes in MG and in FG were significantly enriched for cellular processes, cell-wall organization and cell-cell junctions whereas genes in SP were more enriched for rRNA modification and RNA methylation. However, for all life-cycle stages, we found an overall negative correlation between DNA methylation and gene expression. Thus, despite an overall low level of cytosine methylation in the genome of *S. japonica*, there is evidence that it may play a significant role for the development of life-cycle stages and regulation of metabolism via the control of gene expression and non-coding RNAs, which is different to what has been observed so far in plants, animals and many microbes.

Material and Methods

Strain selection, DNA and RNA extraction and purification

The *S. japonica* strain Ye-c12 was used for methylated DNA immunoprecipitation (MeDIP), bisulfite and RNA sequencing. The haploid male (MG) and female (FG) gametophytes were collected from a blade of the diploid Ye-c12 sporophyte (SP) and expanded via the application of gametophyte cloning technology (Wang *et al.*, 2013). The diploid sporophyte of Ye-c12 was obtained from sexual reproduction of FG and MG, and collected when the blade length was *c.* 3cm. Genomic DNA from haploid gametophytes and diploid sporophytes was extracted using the Plant Genomic DNA kit (Tiangen, China), and RNA was removed by incubating the DNA solution at 37°C with a DNase-free RNase A (Tiangen, China) (Table S1) for 30 minutes. DNA integrity was assessed using agarose gels and a NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). DNA was quantified with a Qubit® 2.0 Fluorometer (Life Technologies, CA, USA) using the Qubit® DNA Assay Kit.

Total RNA was extracted using the RNeasy Pure Plant Kit (Tiangen, China) and DNA was removed using an RNase-Free DNase I treatment according to the instructions by the manufacturer (Tiangen, China). RNA integrity was assessed on 1% agarose gels and by using

the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). RNA was quantified with a Qubit® 2.0. Fluorometer (Life Technologies, CA, USA). A total amount of 3 µg RNA per sample was used as input material. Ribosomal RNA was depleted by Epicentre Ribo-zero™ rRNA Removal Kit (Epicentre, USA), and the rRNA-free residue was cleaned up by ethanol precipitation. Subsequently, libraries were generated using the rRNA-depleted RNA by NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (NEB, USA). The Illumina HiSeq 2500 paired-end platform was used to sequence the libraries with a read length of 125 bps.

Identification of genes encoding circRNAs, lncRNAs and tRNAs

Unmapped RNA-seq reads were kept and 20-mers from 5' and 3' end of these reads were extracted and aligned independently to reference sequences by Bowtie v2.0.6. Anchor sequences were extended by nd_circ such that the complete reads aligned and the breakpoints were anchored by GU/AG splice sites. The back-spliced reads with at least two supporting reads were annotated as circRNAs. We used CNCI (Coding-Non-Coding-Index) (v2) with default parameters to effectively distinguish protein-coding and non-coding sequences independent of known annotations. We used CPC (Coding Potential Calculator) (0.9-r2) mainly to assess the extent and quality of the ORF in a transcript and search the sequences with the NCBI eukaryotes' protein database with the e-value of '1e-10'. We translated each transcript in all three possible reading frames and used Pfam Scan (v1.3) to identify occurrence of any of the known protein family domains documented in the Pfam database (Pfam A and Pfam B), by default parameters of -E 0.001 --domE 0.001. Transcripts predicted with coding potential by either / all of the three tools above were filtered out, and those without coding potential were our candidate set of lncRNAs. Genome wide tRNA genes were predicted using the tRNAscan-SE web server (Lowe & Chan, 2016). The genomic loci for circRNA, lncRNA and tRNA were deposited at CNGB Nucleotide Sequence Archive (<https://db.cngb.org/cnsa/>, project CNP0000364).

Determining the level of heterozygosity between gametophytes and the reference genome.

Both parental haploid gametophytes were used to perform whole-genome sequencing (WGS) employing a Illumina Hiseq 2500 platform, and the reads of each parent were mapping to the reference *S. japonica* genome using BWA (Li & Durbin, 2010). The heterozygosity sites and indel sites were called using SAMTOOLS (Li *et al.*, 2009), which were also used to calculate the heterozygosity between individual haplotypes and the reference genome.

Methylated DNA immunoprecipitation (MeDIP).

Extracted genomic DNA was sonicated to obtain fragments from 100 to 500 bp. The fragmented DNA was end repaired, subjected to A-tailing, and PE adapters were ligated. Subsequently, the treated DNA fragments were immunoprecipitated with an antibody that specifically recognizes 5-methylcytosine using the MagMeDIP Kit (Diagenode, Liege, Belgium). The specificity of the enrichment was confirmed by quantitative real-time PCR (qRT-PCR). After PCR amplification of the enriched fragments, we quantified them with the Agilent 2100 Analyzer (Agilent Technologies, Palo Alto, CA, USA). Sequencing libraries of MeDIP fragments were constructed by adopting the Illumina paired-end protocol. The negative control (input) MeDIP experiment was performed using the same procedure except for the absence of DNA immunoprecipitation with the 5-methylcytosine recognizing antibody.

MeDIP-Seq sequence alignments and data analysis.

Raw sequencing data were filtered by removing the adaptors and discarding low-quality reads. Clean reads were aligned to the reference genome of *S. japonica* with Bowtie's (Langmead & Salzberg, 2012) best mode using default parameters. Read depths of each sample were normalized to eliminate the influence via differences of total read numbers between samples (Fig. 1A). After using input reads to eliminate the background noise in 3 MG samples, peaks were called using "MACS2 callpeak" (Zhang *et al.*, 2008) (Fig. S2A). According to GTF annotation, the normalized absolute read depth (NARD) between genomic components were calculated, including GENOME (genome wide), TSSUP (transcript start site), TESDOWN (transcript stop site), exon, intron, transposable elements (TEs), genes encoding for long non-coding RNAs (lncRNAs), circular RNAs (circRNAs) and tRNAs. Absolute read depth were normalized to genome read depth be comparable between any two classes. Absolute read depth (ARD) for each class of genomic components was calculated as following: $ARD = (\text{absolute read counts (ARC)} * 150) / \text{total length of this class}$. Finally, ARDs of each class were normalized to the ARD of genome, termed as NARD (Fig. S2B). The NARD across the genomic components were represented via meta-plot (Fig. S2C). The regions 2 kb upstream and downstream of the components were split into 20 non-overlapping bins, whereas genes were split into 40 equal windows.

Library preparation and quantification of whole-genome bisulfite sequencing (WGBS).

A total amount of 5.2 μ g genomic DNA spiked with 26 ng lambda DNA were fragmented to 200-300bp by sonication (Covaris S220), followed by end-repair and adenylation.

Cytosine-methylated barcodes were ligated to sonicated DNA. DNA fragments were treated

twice with bisulfite using EZ DNA Methylation-Gold™ Kit (Zymo Research), before the resulting single-stranded DNA fragments were PCR amplified using KAPA HiFi HotStart Uracil + ReadyMix (2X). Libraries were quantified by Qubit® 2.0 Fluorometer (Life Technologies, CA, USA) and quantitative RT-PCR, and the insert size was assessed on the Agilent Bioanalyzer 2100 system.

Clustering and data analysis of WGBS.

Clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq 2500 platform and 125bp paired-end reads were generated. Image analysis and base calling were performed with the Illumina CASAVA pipeline, and finally 125bp paired-end reads were generated. Bismark software (version 0.16.1) (Krueger & Andrews, 2011) was used to perform alignments of bisulfite-treated reads to a reference genome using default parameters. Results of the methylation extractor were transformed to bigWig format for visualization using the IGV browser (Thorvaldsdóttir *et al.*, 2012). The bisulfite non-conversion rate was calculated as the percentage of cytosines sequenced at cytosine reference positions in the lambda genome.

Calculation of methylation levels of mCs using WGBS.

The methylation level of methylated cytosines (**ML_{mc}**) was calculated by the read count of methylated cytosines mapped to their genomic locus divided by the read count of all cytosines mapped to the same locus:

$$ML(C) = \frac{\text{reads}(mC)}{\text{reads}(mC) + \text{reads}(C)}$$

The methylation level of a specific genomic fragment (**ML_{gf}**) (including every single gene, intron, circRNA, lncRNA, TSSUP, TESDOWN, CGI, TE, tRNA and 10k bin of chromosomes or the entire chromosome) was calculated based on the methylation levels of all mCs of this genomic fragment divided by the sum of cytosines and guanines of the same fragment. (Fig. 2F, Fig. 3) (Xiang *et al.*, 2010).

To identify genomic components that were preferentially methylated, the methylation level of a class of genomic component (**ML_{gc} such as tRNAs, lncRNAs**) was calculated as

following: the sum of MLmc of all mCs belonging to a specific component (e.g. tRNAs, lncRNAs) divided by the sum of cytosines and guanines of this component. The approach is also used to calculate the methylation level of the whole genome (**MLwg**, Fig. 2B). To remove experimental artifacts, noise and any batch effects, MLgc were normalized according to MLwg, which is defined as the normalized methylation level of genomic components (**NMLgc**, Fig. 2C) (Xiang *et al.*, 2010).

Differentially methylated sites (DMSs: a DMS corresponds to a single nucleotide), regions (DMRs) and promoters (DMPs) of WGBS.

DMRs were identified using the swDMR software (<http://122.228.158.106/swDMR/>), which uses a sliding-window approach. The window size of 1000bp using a step length of 100bp was used. To avoid overlap of neighbouring DMRs, the sliding-window always started downstream of already identified DMRs. Fisher's exact test was implemented to detect DMRs. The areaStat value that combines both the p-values and fold change between two experimental groups were introduced to identify the DMR (see details for areaStat: http://www.bioconductor.org/packages/release/bioc/vignettes/DSS/inst/doc/DSS.html#11_background). A larger absolute value of the areaStat is more likely to be a DMR. The minimal absolute value of areaStat in our study is 40, resulting in a more than 6 fold difference in methylation level between two experimental groups (N = 3) with a p-value < 0.05. Customised scripts were used to identify DMSs using Fisher's exact test with FDR multiple test correction. DMSs with p-values < 0.05 and a fold change of more than 2 in methylation level between two experimental groups (N = 3) were considered candidate DMSs. The identification of DMPs was conducted for 2000bps upstream of each gene using the same significance criteria as for DMRs.

Validation of DNA methylation by bisulfite sequencing PCR (BS-PCR).

MLgf of specific genomic regions were validated by BS-PCR. Genomic DNA (500ng) of MG, FG and SP were converted using EpiTect Fast Bisulfite Conversion Kit (Qiagen) according to the manufacture's instruction. CpG islands were identified using the MethPrimer (<http://www.urogene.org/methprimer/>), which was also used to design specific BSP methylation primers. PCRs were carried out with the following program: initial denaturation at 95°C for 30s followed by 40 cycles of 95°C for 5s, 60°C for 10s, and 72°C for 40s. The PCR products were purified and recovered by a DNA Universal DNA Purification Kit

(Tiangen, Beijing, China). Next, purified PCR products were cloned into the pMD18-T vector (Takara, Beijing, China) and transformed into *E. coli* DH5 α -competent cells (Yixin, Shanghai, China) for further replication. Positive clones were screened and sequenced (Sunny, Shanghai, China). Up to 30 randomly picked clones for each amplified locus were sequenced using the ABI3730xl platform.

Phylogenetic analysis.

Given the large diversity of DNMTs, only the DNA methylase motif region (PF00145) was used for constructing phylogenetic trees after sequence alignment using muscle 3.8.31 (Edgar, 2004). Maximum-likelihood (ML) phylogenetic trees were constructed using MEGA 6.0 with the LG + G + F model, predicted as best model by the 'Find best DNA/protein models' module of MEGA 6.0 (Tamura, 2013). Motif structures of genes were visualised using the Interactive Tree of Life (iTOL) (Ivica & Peer, 2016). 3D structures of DNMT proteins were visualised using X-ray structure models from SWISS-MODEL <https://www.swissmodel.expasy.org/>.

Quantitative real-time PCR.

Total RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA, USA) according to the user manual. Purified RNA was dissolved in diethylpyrocarbonate (DEPC)-treated water. cDNA was synthesized from the total RNA using Moloney murine leukemia virus reverse transcriptase (Promega Biotech Co., Madison, Wisconsin, USA). Quantitative real-time PCR (qPCR) reactions were performed with the ABI StepOne Plus Real-Time PCR System (Applied Biosystems, USA) using SYBR Green fluorescence (TaKaRa) according to the manufactures instruction. The 18S rDNA gene was used as reference. For each of the selected genes, three biological replicates were assayed independently. The qPCR amplifications were carried out in a total volume of 20 μ L containing 10 μ L of 26 SYBR Premix Ex TaqTM II (TaKaRa Biotech Co., Dalian, China), 0.6 μ L (10 mM) of each primer, 2.0 μ L of the diluted cDNA mix, and 6.8 μ L de-ionized water. The qPCR protocol was as follows: initial denaturation at 95°C for 30s followed by 40 cycles of 95°C for 5s, 60°C for 10s, and 72°C for 40s. The $2^{-\Delta\Delta CT}$ method was used to analyze the qPCR data.

Results

Assessment of DNA methylation based on an improved genome assembly.

To obtain evidence of cytosine methylation in the genome of *S. japonica*, we performed methylated DNA immunoprecipitation (MeDIP) using a mixture of DNA from all three life stages of the *S. japonica* strain Ye-c12 (SP, FG, MG). Methylated DNA was recognized by a 5-methylcytosine antibody. After PCR amplification, we selected 220-520 bp fragments to construct the sequencing libraries. The libraries revealed a positive signal of the expected length (220-520 bp, Fig. S1A), indicating the presence of 5-methylcytosines in the DNA of *S. japonica*. However, the positive control (Bos DNA) gave *c.* 6-fold stronger signals (Fig. S1B), suggesting that the concentration of 5-methylcytosines was significantly higher in Bos DNA (average MLwg of 8.43%) compared to *S. japonica* (Dechow & Liu, 2018).

To provide first insights into the methylome landscape on a chromosome-wide scale, we constructed pseudo-chromosomes by combining a genetic linkage map (Zhang *et al.*, 2015) with scaffold information. A total of 1,576 scaffolds were anchored to 31 linkages of the genetic map (Fig. 1 and Methods S1 Part 1), accounting for 64.69 % (352,93 Mb) of the assembled kelp scaffolds. The remaining scaffolds were concatenated to construct an artificial chromosome (Fig. S2, S3, Supporting Information Methods S1 Part 1).

Cytosine methylome landscapes based on MeDIP-seq.

To obtain insights into the global methylation landscape between life stages, we performed independent MeDIP experiments for FG, MG and SP with three biological replicates each. However, only MG produced enough immunoprecipitated DNA after PCR amplification. The negative results for FG and SP were verified by additional MeDIP experiments, suggesting reduced levels of methylation in FG and SP. We then performed WGB-seq for 3 MG replicates and one negative control (also named as input). A total of 20 giga bp paired-end reads were generated from three MG MeDIP-seq libraries and one input library (Table S1). The MeDIP-seq reads were mapped across chromosomal regions in 3 MG and one control sample (Fig. S2A). We obtained 8,824 peaks in a total length (3 replicates) of 8,764,952 bp, covering 1.61% of the genome (Table S2). The distribution of MeDIP-seq reads (normalized absolute read depth = NARD) showed that the reads were distributed across all the genomic feature regions (Fig. S2B). Interestingly, lncRNAs genes and circRNAs genes had the highest level of NARDs (p -value < 0.05) while exons and genes encoding tRNAs had the lowest level of methylation (p -value < 0.05). The distribution of NARDs across lncRNAs genes and CGIs had a higher level of methylation than regions 2 kb up- and downstream (Fig. S2C).

Cytosine methylome landscapes based on whole-genome bisulfite sequencing (WGBS).

Overall, only *c.* 1.4% of cytosines in the genome of *S. japonica* were methylated based on WGBS. At such low levels of DNA methylation, a few sequencing errors can significantly impact the results. Thus, to minimize false positives and negatives in our dataset, we sequenced three biological replicates for each life-cycle stage of *S. japonica* using whole-genome bisulfite sequencing (WGBS) (Fig. 1, Table S1). By calculating the percentage of cytosines sequenced at cytosine reference positions in the lambda genome, we identified a final conversion rate of > 99.95% for unmethylated cytosines. By combining nine individual methylome profiles, we were able to obtain a high-quality, high-resolution kelp methylome with a final sequence yield of 92.4 Gb covering 86% of all cytosines in the genome (Table S1). To minimize the number of false positives, the lower threshold for identification of methylated cytosines was set at 4 sequence reads per methylated cytosine (Table S1, Datasets S1-S9).

Most of the chromosomes had the highest MLgf in MG with additional hyper-methylated chromosomes 4, 17 and regions on chromosomes 9, 12, making MG the most methylated life-cycle stage, followed by FG and SP (Fig. 1, Fig. S4 and S5B, Table S3). Although chromosomes 7 and 29 had a higher MGgf in FG compared to MG (Fig. 1, Fig. S4 and S5B), sex-determining genes were not found to be localised on these chromosomes (Lipinska *et al.* 2017) (Tables S3-4). Approximately 57% of the methylated cytosines in the genome were methylated in a CHH context, whereas 19% and 24% were in CHG or CpG contexts, respectively (Fig. 2A, Table S4). However, from the intra-sample comparing the median MLmc is higher in the CpG context than the other two contexts ($P < 0.05$, Fig. S6).

Interestingly, MLgfs in 10,000 bps bins of chromosomes show a significantly negative correlation with the CpG O/E (Observed / expected value), CpG ratio (nucleotide counts of all CpGs in bin / bin length) and CGI (CpG Island) ratio (nucleotide counts of all CGIs in bin / bin length), and a positive correlation with the CHH ratio (nucleotide counts of all CHHs in bin / bin length) (Figs. S7-12), indicating the CHH sites are preferentially methylated.

Among all life-cycle stages, SP had the lowest counts of methylated cytosines and the lowest MLgw of whole genome and MLgc for all class of components (CpG, CHG and CHH) ($p < 0.05$, Fig. 2A and 2B; Fig. S6; Fig. S13). However, NMLgc fluctuated significantly across the genome (Fig. 2C), reflecting a mosaic methylation pattern where non-methylated regions were interspersed with methylated regions. Notably, the most highly methylated genetic

elements in the genome of *S. japonica* were loci encoding for circular (circRNA) and long non-coding RNAs (lncRNAs) (Fig. 2C, Fig. S14, Tables S5-S15) with 81% and 66% of all elements methylated, respectively (Fig. 2F, Tables S4-S14). Coding genes and their regulatory parts (transcriptional start (TSSup) and stop (TESdown) sites) were significantly lower methylated compared to non-coding circRNA and lncRNA genes (Fig. 2C, $p < 0.05$). Exons had the NMLgc, even they had the highest CG frequency (Fig. 2D, $p\text{-value} < 0.05$). Most of the exons were methylated in the CHH context with a concave methylation pattern. Introns, however, showed a convex methylation pattern (Fig. 2E, Fig. S15). Convex methylation patterns were also found in the core regions of genes encoding for lncRNAs and circRNAs (Fig. S15).

We predicted 24, 29, and 757 tRNA genes for the mitochondrion, chloroplast and nuclear genomes of *S. japonica*, respectively. On average, 29.1% of all tRNA genes were methylated (Fig. 2F; Table S14). In plants and algae, repetitive elements such as transposable elements (TEs) are often most highly methylated (Cokus, Shawn J. *et al.*, 2008; Law & Jacobsen, 2010; Su, C. *et al.*, 2014; Takuno *et al.*, 2016). However, in *S. japonica*, methylation of TEs does not seem to play a significant role (Fig. 2C, Fig. S12-S13), as only 47.8% TEs methylated, which is lower compared to genes (93.1%), introns (81.3%), circRNAs (80.9%), lncRNAs (66.4%), TSSUPs (90.3%), TESDOWNs (89.6%) and CGIs (60.5%), but just a little higher than exons (42.4%) and tRNA genes (29.1%) (Fig. 2F; Tables S5-14). Although only 42% (Table S7) of exons were methylated, methylated exons had above average MLgf (Fig. 2F).

Functional analysis of differentially methylated regions (DMRs) between life-cycle stages.

DMRs were identified using a sliding-window approach with a window size of 1000bp and a step length of 100bp (see the Materials and Methods section). Despite the overall low level of DNA methylation in *S. japonica*, the number of DMRs varied considerably between life-cycle stages. By systematically surveying the regions that were differentially methylated between all life-cycle stages, we found 12 hypermethylated DMRs in sporophytes (Table S15-a and d), 646 hypermethylated DMRs in female (Table S15-b and e) and 790 hypermethylated DMRs in male gametophytes (Table S15-c and f). DMRs were identified to overlap with exons, introns, TEs, TSSUPs, TESDOWNs, circRNA genes, lncRNA genes and intergenic DNA, and the length of overlapping fragments ranged from 10 to 500 bp (Fig. 3A). By clustering the overlapped elements with DMRs according to their MLgf for each life-cycle stage, we found they are most highly methylated in MG and FG (Fig. 3A),

reflecting the genome wide trend (see above). GO and KEGG enrichments were conducted for all DMRs and all life-cycle stages representing protein-coding genes (Table S16). Differentially methylated genes (genes overlapped with DMRs) in SP were enriched in rRNA modification (GO:0000154), RNA methylation (GO:0001510), generation of precursor metabolites and energy (GO:0006091), methyltransferase activity (GO:0008168), transferase activity (GO:0016758, GO:0016758) (table S15). Differentially methylated genes in MG were enriched for the following GO terms: cellular process (GO:0009987), cell wall organization (GO:0071555), cellular metabolic process (GO:0044237), metabolic process (GO:0008152) and Pentose phosphate pathway (ko00030) (Fig. S16-18, tables S16-17). FG specific differentially methylated genes were enriched for GO terms of cell-cell junctions (GO:0005911), transition metal ion binding (GO:0046914), cellular developmental processes (GO:0048869) and plant hormone signal transduction pathways (ko04075) (Fig. S16-18, tables S16-17).

DNA methylation and gene expression.

We identified more expressed genes in sporophytes than in gametophytes (RPKM >10). Generally, there was a negative correlation between gene expression and MLgfs (Fig. 3B, Fig. S19). In SP, 3884 genes and 3490 TSSUP were significantly hypomethylated and higher expressed than in FG whereas 4132 genes and 4062 TSSUP were significantly hypomethylated and higher expressed than in MG (T-test, $p < 0.05$, Fig. 3C, Tables S17-19). Only 51 genes in SP were hypermethylated with significantly lower expression than in either FM or MG (T-test, $p < 0.05$, Fig. 3C, Tables S18-20). Especially some of the hypomethylated and highly expressed genes in SP with their respective GO term categories in comparison to FG and MG revealed some insights into processes that might be important to develop fully grown kelp sporophytes (Tables S21-22). For instance, enriched GO terms for this category in SP vs FG included structural molecule activity (GO:0005198, $p\text{-value} < 0.01$) and carbohydrate derivative binding (GO:0097367, $p\text{-value} < 0.05$). The same category for SP vs MG include the GO term membrane protein complex (GO:0098796, $p\text{-value} < 0.01$). Interestingly, four genes that were reported to have moved from sex-determining chromosomal regions to auto chromosomes (Lipinska, Agnieszka P. *et al.*, 2017) were more highly expressed in SP and had significantly reduced MLgf in all life-cycle stages (Fig. 3D, Table S23, Fig. S20, Methods S1 Part 2). Other hypomethylated and strongly expressed

genes in SP included a cellulose synthase, a mannuronate C5-epimerases (MC5E) and an iodoperoxidase (vIPO) (Fig. S21).

Although there was no significant correlation between MLgf and expression levels of lncRNAs and circRNAs genes (Fig. S22-23) ($|\text{Pearson correlation index}| < 0.01$, $p > 0.05$), lncRNAs genes with expression levels in the top 10% had the lowest level of MLgf, and lncRNAs with MLgf in the top 10% had the lowest expression levels. The similar trends were also found in circRNA. However, there is no significant correlation between MLgf and the expression of TEs (Fig. S24).

A comparison between MeDIP-seq, BS-PCR and WGBS.

MeDIP-seq and WGBS give evidence for DNA methylation in *S. japonica* although at a relatively low level. Furthermore, there is a high consistency between results from both methods (Pearson correlation index > 0.5 , $p\text{-value} < 0.001$, Fig. S25). For example, both methods independently revealed that methylation preferentially took place at genes encoding lncRNAs and circRNAs. These results were confirmed by BS-PCR with subsequent cloning and Sanger sequencing. We assessed 10 genes (Fig. 3D and Fig. S21) and obtained a good agreement between the WGBS and BS-PCR (Fig. S26-35). Thus, the independent validation of DNA methylation based on 3 different methods suggests that the DNA of *S. japonica* indeed appears to be methylated albeit a low level.

Genome-wide heterozygosity and its influence on estimating the level of DNA methylation.

The level of heterozygosity between the reference genome and the DNA obtained from the individuals used for the experiments can have an influence on estimating methylated cytosines. This is particularly critical if the overall level of methylation is low at a considerable level of heterozygosity between the reference genome and the genome of the individuals used for the experiments. To address this potential issue, we performed Illumina WGS for both male (MG) and female (FG) parents to determine differences in the level of heterozygosity. After mapping *c.* 50Gbp of FG and MG to the reference genome, we calculated that the overall level of heterozygosity was not higher than 0.14% (Table S24). Of all identified heterozygosity sites, C-T / G-A polymorphic sites (168,560 for FG and 118,421 for MG) overlapped with only 0.02% ~ 0.04% of all mCs site for FG and MG samples (Table S24-b), indicating that heterozygosity was not impacting more than 0.04% of all identified mCs in FG and MG.

DNA methyltransferases and demethylases in *S. japonica*.

DNA methyltransferases (DNMTs) and demethylases (DNDMTs) are involved in the establishment of tissue and cell-type-specific methylation patterns during developmental processes in most multicellular organisms (Meissner *et al.*, 2008; Kohli & Zhang, 2013; Satgé *et al.*, 2016; Iurlaro *et al.*, 2017). As the overall DNA methylation was relatively low in *S. japonica*, we did a homology-based search in the *S. japonica* genome to identify all possible DNMTs and DNDMTs (Figs. S36-40). Although the DNA-methylase domain PF00145 is common in all DNMTs, effective DNA methylation requires the following additional domains, which only occur in the DNMT1/3/4/5/6 family: ADD (IPR025766), PWWP (PF00855), PHD (IPR001965) and a bromo-domain (PF00439) (Ponger & Li, 2005). All of the 6 DNMTs found in *S. japonica* had the highest similarity to members of the DNMT2 family without any of the domains required for efficient DNA methylation (Fig. 4A and 4B, Fig. S38). There were no homologs found with significant similarity to any of the other DNMTs such as 1, 3, 4, 5 or 6 (Fig. S37). Interestingly, we found ADD, PWWP, PHD and bromo-domains at other loci in the genome of *S. japonica* but none of them were found to be DNMT2 (Fig. S39-40).

To identify proteins involved in active DNA demethylation (DNDMTs), we searched for RRM-fold domains (PF15628), DNA glycosylase domains (IPR011257), DME, DML, ROS, TET and MBD domains, all of which were previously reported to catalyse the removal of the 5mC base (Meissner *et al.*, 2008; Kohli, & Zhang, 2013; Su *et al.*, 2014; Lipinska, *et al.*, 2017; Teng *et al.*, 2017). We only found one potential candidate in the genome, which included a DNA glycosylase and the methyl-CpG-binding domain protein 2 (MBD2, PF01429) (Fig. S41-42). The MBD2 domain was reported to catalyze the removal of a methyl group (Wu & Zhang, 2014).

To have 6 genes from the DNMT2 family encoded in the *S. japonica* genome with all other DNMTs missing may indicate that DNMT2 methylates DNA in the structural context of tRNAs (Kaiser *et al.*, 2017). Thus, we analysed the MLgf of tRNA genes encoded in the mitochondrial, chloroplast and nuclear genomes of *S. japonica* for all three life stages individually (Fig. S43, Table S25). Overall, the sporophyte had the lowest MLgf of tRNA genes for all three genomes. However, tRNA genes of the nuclear genome of the sporophyte were significantly more methylated than in the organellar genomes, which was not noticeable in the other life stages as tRNA genes in their organellar genomes were much more

Accepted Article
methylated than their counterparts in sporophytes. The ratio between unmethylated to methylated tRNA genes is highest for SP (Fig. S43). The MLgf of tRNA genes in the nucleus was positively correlated with the MLgf of nuclear genes only for MG. These results suggest possible DNMT2 mediated DNA methylation, with RNAs as natural substrates of DNMT2 acting as guides as shown previously (Kaiser *et al.*, 2017).

Discussion

Differential DNA methylation is the hallmark of cell-type specific development and life cycle regulation in multicellular organisms (Law & Jacobsen, 2010; Takuno *et al.*, 2016). For most multicellular organisms, DNMT2 is not considered to be the enzyme for DNA methylation as it primarily catalyzes tRNA methylation and only has very weak DNA methyltransferase activity (Kunert *et al.*, 2003; Goll *et al.*, 2006; Shanmugam *et al.*, 2014; Francesca *et al.*, 2015; Kaiser *et al.*, 2017). However, recent data (Kaiser *et al.*, 2017) have provided first *in vitro* evidence that DNMT2 can efficiently methylate DNA when DNA fragments are presented as covalent DNA-RNA hybrids in the structural context of tRNAs. Thus, methylated tRNAs can serve as substrates for methylating DNA presented as covalent hybrids with RNA. Although our data do not provide direct evidence for this process to be responsible for DNA methylation *in vivo* in *S. japonica*, it is intriguing to see that there are significantly more methylated tRNA genes in both gametophytes which might serve as a substrate to methylate genes via DNMT2. In MG, a significantly positive correlation between the MLgf of tRNA genes and both the expression level of DNMT2s (Fig. 4C and 4D) and the MLgf of genes corroborates this assumption (Fig. S43). However, if DNMT2 does indeed methylate the DNA of *S. japonica in vivo*, it is being done inefficiently as the overall DNA methylation level (MLmc and MLwg) is still very low even though six different DNMT2s are encoded in the genome and expressed. DNMT2 appears to have evolved from a DNA methyltransferase precursor, which is considered an evolutionary ‘relict’ in terms of biochemical catalysis (Jurkowski & Jeltsch, 2011; Raddatz *et al.*, 2013) and therefore might explain its inefficiency compared to other DNMT enzymes such as 1 and 3. Although it is possible to have missed DNMTs because of gaps in the genome assembly, the fact that multiple DNMT2 genes are encoded in the *S. japonica* genome, and the positive correlation of their expression with an elevated level of C5 DNA methylation (Pearson correlation test, $p\text{-value} < 0.05$) makes DNMT2s likely candidates for C5 DNA methylation. Support for the role of DNMT2 in C5 DNA methylation comes from reverse-genetics studies with

Drosophila melanogaster where knockdown and overexpression of DNMT2 resulted in lower and higher levels of 5 methylcytosine in embryos, respectively (Kunert *et al.*, 2003).

Likewise, human DNMT2 was shown to have a residual DNA C5 methyltransferase activity in a particular sequence context (Hermann *et al.*, 2003).

Although the overall DNA methylation level in *S. japonica* is lower than observed for most multicellular organisms with complex life cycles (Takuno *et al.*, 2016), DNA methylation still appears to contribute to the regulation of gene expression and particularly non-coding regulatory RNAs. Furthermore, significant differences between life-cycle stages suggest that differential DNA methylation is involved in their formation, which is similar to other multicellular organisms although most of them have much higher and more dynamic levels of DNA methylation (Lopez *et al.*, 2015; Willing *et al.*, 2015; Gaunt *et al.*, 2016; Kawakatsu *et al.*, 2016; Mayasich *et al.*, 2016; Panikar *et al.*, 2017).

Another striking difference is that genome-wide patterns and mechanisms of cytosine methylation in *S. japonica* neither are more similar to plants nor animals, and they are not even similar to other stramenopile species for which single-base methylome data are available such as *P. tricornutum* (Veluchamy *et al.*, 2013). In many photosynthetic organisms including plants, green algae and diatoms, TEs are the main target of cytosine methylation whereas animals methylate genes and TEs more equally (Cokus *et al.*, 2008; Law & Jacobsen, 2010; Takuno *et al.*, 2016; Tirichine *et al.*, 2017). However, in *S. japonica*, the percentage of methylated TEs is < 48% (Table S5), which is lower than the percentage of methylated genes (93% Table S6). Surprisingly, the highest methylated elements in *S. japonica* were found to be genes encoding circular and long non-coding RNAs (p-value < 0.05 in FG and MG, Fig. 2C, Fig. S14). Significant differences in methylation of both groups of genes encoding non-coding RNAs between SP, FG and MG indicate that they appear to have a role in regulating life cycle stages of *S. japonica*, which, to the best of our knowledge, has not been observed before in any organism (Fig. S22-23). Further differences to plants and animals have been revealed by comparing mechanisms of DNA methylation. Animals mainly methylate cytosines in the CpG context, whereas plants methylate CHG and CHH sites and many algae methylate in CpG context (Cokus *et al.*, 2008; Lopez *et al.*, 2015). *S. japonica* methylates cytosines mainly in the CHH context with much lower but almost equal methylation of CHG and CpG sites.

The fact that a significant number of genes is methylated in *S. japonica*, and that the level of their methylation is negatively correlated with gene expression is in common with many plants and animals. Hence, there is evidence of methylation-regulated gene expression. Furthermore, methylated genes in MG and FG were negatively correlated with gene expression (Fig. 3B). GO enrichment for these genes revealed that many were involved in transport, membrane structure and function (Table S16 and S17), which suggests that a) these processes are under epigenetic control in MG and b) that they are downregulated in comparison to SP and FG. Interestingly, methylated genes in SP and FG were enriched for *e.g.* small molecule metabolic process and organo-nitrogen compound metabolic processes. Hypomethylated and highly expressed genes involved in morphogenesis (*e.g.* MC5E, cellulose synthase) and halogen metabolism (*e.g.* vIPO) were found in SP (fig. S21), which suggests that cytosine methylation contributes to regulating fundamental processes underpinning cell differentiation, growth and stress response (Pear *et al.*, 1996; Colin *et al.*, 2005; Ye *et al.*, 2015; Inoue *et al.*, 2016). Genes of the sex-determination system do not show a negative correlation between their level of methylation and gene expression (Fig. S20), suggesting that 5mC DNA methylation might not significantly contribute to their regulation. Interestingly, the four genes that are characterised by different protein domains (MEMO, glycosyltransferase, RING-type zinc finger, Rab-GTPase-TBC) and which moved from the sex-determination locus in the genome to the autosomal loci were hypomethylated and highly expressed in SP (Lipinska, *et al.*, 2017) (Fig. S20).

The first single-base methylome study of a brown alga has revealed that cytosine methylation might be mediated by DNMT2. Although the overall level of DNA methylation is low, it appears to play a significant role in these multicellular algae mainly for controlling regulatory elements (non-coding RNAs) and the expression of genes. Significant differences in overall DNA methylation and methylation-mediated gene expression between life-cycle stages suggests that cytosine methylation is involved in regulating growth, tissue differentiation, and several different physiological responses. However, it remains to be seen if DNMT2 indeed is responsible for DNA methylation in *S. japonica*.

Data availability

Datasets S1-S9 and their source Illumina clean data for methylome have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE117191 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE117191>). The Illumina raw data for transcriptomes has been deposited at the NCBI database (the SRA accession numbers are SRR5860561-SRR5860568). The Illumina data of the MeDIP-seq experiments and the WGS data were deposited at CNGB Nucleotide Sequence Archive (<https://db.cngb.org/cnsa/>, with project accession number CNP0000364), where we also deposited the reassembled genome and genomic loci for TEs, CGIs, genes, circRNAs, lncRNAs, tRNAs and the detailed methylation data for genome-wide visualization (visualization data is also deposited in GEO Series accession number GSE117191). The methylation data based on WGBS and the gene expression based on RNA-seq can be visualized using the online genome browser at the OrcAE database (<https://bioinformatics.psb.ugent.be/orcae/>).

Acknowledgements

We are grateful to Jim Mitchell for general advice. This work was supported by National key research and development program of China (2018YFD0900703, 2018YFD0901503-8, 2016YFC1402102), National Natural Science Foundation of China (41676145); Shandong key Research and Development Plan (2018GHY115010); Qingdao Municipal Science and Technology plan project (17-1-1-96-jch); Special Scientific Research Funds for Central Non-Profit Institutes, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences (20603022019006, 20603022016001); Youth Talent Program Supported by Laboratory for Marine Fisheries Science and Food Production Processes, Pilot National Laboratory for Marine Science and Technology (Qingdao) (2018-MFS-01); China Agriculture Research System (CARS-50); Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (NO. 2018SDKJ0406-3); Financial Fund of the Ministry of Agriculture and Rural Affairs, P. R. of China (NFZX2018); Taishan Scholars Funding; Talent Projects of Distinguished Scientific Scholars in Agriculture. TM acknowledges the School of Environmental Sciences at the University of East Anglia for partial support. Lt acknowledges support from Region Pays de la Loire and Nantes Métropole.

Author contributions

N.H.Y. designed the project with contributions from T.M., L.T., X.F., W.T.H., L.H.T., P.J., X.W.Z., D.X., C.H.W., C.L. and Y.T.W. performed the research. X.F., N.H.Y., W.T.H., T.M., L.T., C.L. and L.H.T. analyzed the data. T.M. wrote the manuscript with contributions from X.F., N.H.Y., and L.T.. M.P., M.J.S.K. and X.L. revised and edited the manuscript. All authors read and approved the manuscript before submission. The authors declare no conflict of interest.

References

- Charrier B, Le BA, De RB. 2012.** Plant Proteus: brown algal morphological plasticity and underlying developmental mechanisms. *Trends in Plant Science* **17**(8): 468-477.
- Cock JM, Godfroy O, Macaisne N, Peters AF, Coelho SM. 2014.** Evolution and regulation of complex life cycles: a brown algal perspective. *Current Opinion in Plant Biology* **17**(1): 1-6.
- Cock JM, Sterck L, Rouzé P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury JM, Badger JH. 2010.** The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature* **465**(7298): 617.
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008.** Shotgun bisulfite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**(7184): 215-219.
- Colin C, Leblanc C, Michel G, Wagner E, Leizewagner E, Van DA, Potin P. 2005.** Vanadium-dependent iodoperoxidases in *Laminaria digitata*, a novel biochemical function diverging from brown algal bromoperoxidases. *Journal of Biological Inorganic Chemistry Jbic A Publication of the Society of Biological Inorganic Chemistry* **10**(2): 156.
- Dechow CD, Liu WS. 2018.** DNA methylation patterns in peripheral blood mononuclear cells from Holstein cattle with variable milk yield. *BMC Genomics*.
- Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, Askarian-Amiri ME, Ru K, Solda G, Simons C, Sunkin SM, Crowe ML, Grimmond SM, Pekins AC, Mattick JS. 2008.** Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Research* **18**(9): 1433.
- Edgar RC 2004.** MUSCLE: Multiple Sequence Alignment with Improved Accuracy and Speed. *Computational Systems Bioinformatics Conference*.
- Feng S, Cokus SJ, Zhang X, Chen P, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME. 2010.** Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America* **107**(19): 8689-8694.
- Francesca T, Friederike H, Nader A, Sebastian B, Oliver P, Giuseppina F, Sonja R, Reinhard L, Georg S, Hermann-Josef G. 2015.** The tRNA methyltransferase *Dnmt2* is required for accurate polypeptide synthesis during haematopoiesis. *Embo Journal* **34**(18): 2350-2362.
- Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, Zheng J, Duggirala A, Mcardle WL, Ho K. 2016.** Systematic identification of genetic influences on methylation across the human life course. *Genome Biology* **17**(1): 61.

- Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jakobsen SE, Bestor TH. 2006.** Methylation of tRNA by the DNA methyltransferase homolog *Dnmt2*. *Science* **311**(5759): 395-398.
- Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES. 2012.** Trans Chromosomal Methylation in *Arabidopsis* hybrids. *Proceedings of the National Academy of Sciences of the United States of America* **109**(9): 3570-3575.
- Hermann A, Schmitt S, Jeltsch A. 2003.** The Human *Dnmt2* Has Residual DNA-(Cytosine-C5) Methyltransferase Activity. *Journal of Biological Chemistry* **278**(34): 31717-31721.
- Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, Ren B. 2013.** Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature Genetics* **45**(10): 1198-1206.
- Inoue A, Satoh A, Morishita M, Tokunaga Y, Miyakawa T, Tanokura M, Ojima T. 2016.** Functional heterologous expression and characterization of mannuronan C5-epimerase from the brown alga *Saccharina japonica*. *Algal Research* **16**: 282-291.
- Iurlaro M, Von MF, Reik W. 2017.** DNA methylation homeostasis in human and mouse development. *Current Opinion in Genetics & Development* **43**: 101-109.
- Ivica L, Peer B. 2016.** Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* **44**(Web Server issue): W242-W245.
- Jurkowski TP, Jeltsch A. 2011.** On the Evolutionary Origin of Eukaryotic DNA Methyltransferases and *Dnmt2*. *Plos One* **6**(11): e28104.
- Kaiser S, Jurkowski TP, Kellner S, Schneider D, Jeltsch A, Helm M. 2017.** The RNA methyltransferase *Dnmt2* methylates DNA in the structural context of a tRNA. *Rna Biology* **14**(9): 1241-1251.
- Kawakatsu T, Valdes M, Breakfield N, Schmitz RJ, Nery JR, Urich MA, Han X, Lister R, Benfey PN. 2016.** Unique cell-type-specific patterns of DNA methylation in the root meristem. *Nature Plants* **2**(5): 16058-16058.
- Kohli RM, Zhang Y. 2013.** *TET* enzymes, *TDG* and the dynamics of DNA demethylation. *Nature* **502**(7472): 472-479.
- Krueger F, Andrews SR. 2011.** Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**(11): 1571-1572.
- Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravimoussavi A, Kheradpour P, Zhang Z, Wang J. 2015.** Integrative analysis of 111 reference human epigenomes. *Nature* **518**(7539): 317-330.
- Kunert N, Marhold J, Stanke J, Stach D, Lyko F. 2003.** A *Dnmt2-like* protein mediates DNA methylation in *Drosophila*. *Development* **130**(21): 5083.
- Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**(4): 357-359.
- Law JA, Jacobsen SE. 2010.** Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**(3): 204-220.
- Lee T, Zhai J, Meyers BC. 2010.** Conservation and divergence in eukaryotic DNA methylation. *Proc Natl Acad Sci U S A* **107**(20): 9027-9028.
- Li H, Durbin R. 2010.** Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**(5): 589-595.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N. 2009.** The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.
- Lipinska AP, Toda NRT, Heesch S, Peters AF, Cock JM, Coelho SM. 2017.** Multiple gene movements into and out of haploid sex chromosomes. *Genome Biology* **18**(1): 104.

- Lister R, Mukamel EA, Nery JR, Urich M, Puddifoot CA, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD. 2013.** Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* **341**(6146): 629.
- Lopez DA, Hamaji T, Kropat J, De HP, Morselli M, Rubbi L, Fitzgibbon ST, Gallaher SD, Merchant SS, Umen JG. 2015.** Dynamic changes in the transcriptome and methylome of *Chlamydomonas reinhardtii* throughout its life cycle. *Plant Physiology* **169**(4): 2730.
- Lowe TM, Chan PP. 2016.** tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research* **44**(Web Server issue): W54-W57.
- Luthringer R, Cormier A, Ahmed S, Peters AF, Cock JM, Coelho SM. 2014.** Sexual dimorphism in the brown algae. *Perspectives in Phycology* **1**(1): 11-25.
- Mayasich SA, Bemis LT, Clarke BL. 2016.** DNA methylation in a sea lamprey vasotocin receptor gene promoter correlates with tissue- and life-stage-specific mRNA expression. *Comparative Biochemistry & Physiology Part B Biochemistry & Molecular Biology* **202**: 56-66.
- Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB. 2008.** Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**(7205): 766-770.
- Molaro A, Hodges E, Fang F, Song Q, McCombie WR, Hannon GJ, Smith AD. 2011.** Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**(6): 1029-1041.
- Nishitsuji K, Arimoto A, Iwai K, Sudo Y, Hisata K, Fujie M, Arakaki N, Kushiro T, Konishi T, Shinzato C. 2016.** A draft genome of the brown alga, *Cladosiphon okamuranus*, S-strain: a platform for future studies of 'mozuku' biology. *Dna Research* **23**(6): 561-570.
- Panikar CS, Paingankar MS, Deshmukh S, Abhyankar V, Deobagkar DD. 2017.** DNA methylation changes in a gene-specific manner in different developmental stages of *Drosophila melanogaster*. *Current Science* **112**(6): 1165-1175.
- Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM. 1996.** Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. *Proceedings of the National Academy of Sciences of the United States of America* **93**(22): 12637-12642.
- Ponger L, Li WH. 2005.** Evolutionary Diversification of DNA Methyltransferases in Eukaryotic Genomes. *Molecular Biology & Evolution* **22**(4): 1119-1128.
- Raddatz G, Guzzardo PM, Olova N, Fantappiè MR, Rampp M, Schaefer M, Reik W, Hannon GJ, Lyko F. 2013.** *Dnmt2*-dependent methylomes lack defined DNA methylation patterns. *Proceedings of the National Academy of Sciences of the United States of America* **110**(21): 8627-8631.
- Satgé C, Moreau S, Sallet E, Lefort G, Auriac MC, Remblière C, Cottret L, Gallardo K, Noirot C, Jardinaud MF. 2016.** Reprogramming of DNA methylation is critical for nodule development in *Medicago truncatula*. *Nat Plants* **2**(11): 16166.
- Shanmugam R, Aklujkar M, Schäfer M, Reinhardt R, Nickel O, Reuter G, Lovley DR, Ehrenhofermurray A, Nellen W, Ankri S. 2014.** The *Dnmt2* RNA methyltransferase homolog of *Geobacter sulfurreducens* specifically methylates tRNA-Glu. *Nucleic Acids Research* **42**(10): 6487.
- Shao C, Li Q, Chen S, Zhang P, Lian J, Hu Q, Sun B, Jin L, Liu S, Wang Z. 2014.** Epigenetic modification and inheritance in sexual reversal of fish. *Genome Research* **24**(4): 604.

- Sleutels F, Zwart R, Barlow DP. 2002.** The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**(6873): 810-813.
- Smith ZD, Chan MM, Mikkelsen TS, Gu H, Gnirke A, Regev A, Meissner A. 2012.** A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**(7394): 339-344.
- Stelzer Y, Shivalila CS, Soldner F, Markoulaki S, Jaenisch R. 2015.** Tracing dynamic changes of DNA methylation at single-cell resolution. *Cell* **163**(1): 218-229.
- Sturgill D, Zhang Y, Parisi M, Oliver B. 2007.** Demasculinization of X chromosomes in the *Drosophila* genus. *Nature* **450**(7167): 238.
- Su C, Wang C, He L, Yang C, Wang Y. 2014.** Shotgun bisulfite sequencing of the *Betula platyphylla* genome reveals the tree's DNA methylation patterning. *International Journal of Molecular Sciences* **15**(12): 22874-22886.
- Takuno S, Ran JH, Gaut BS. 2016.** Evolutionary patterns of genic DNA methylation vary across land plants. *Nat Plants* **2**(2): 15222.
- Tamura K. 2013.** MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology & Evolution* **30**(12): 2725-2729.
- Teng L, Fan X, Xu D, Zhang X, Mock T, Ye N. 2017.** Identification of Genes under Positive Selection Reveals Differences in Evolutionary Adaptation between Brown-Algal Species. *Frontiers in Plant Science* **8**: 1429.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013.** Integrated Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**(2): 178-192.
- Tirichine L, Rastogi A, Bowler C. 2017.** Recent progress in diatom genomics and epigenomics. *Current Opinion in Plant Biology* **36**: 46-55.
- Veluchamy A, Lin X, Maumus F, Rivarola M, Bhavsar J, Creasy T, Brien KO, Sengamalay N, Tallon LJ, Smith AD. 2013a.** Insights into the role of DNA methylation in diatoms by genome-wide profiling in *Phaeodactylum tricornutum*. *Nature Communications* **4**(1): 2091.
- Wang Y, Xu D, Fan X, Zhang X, Wang W, Mao Y, Mou S, Cao S. 2013.** Variation of photosynthetic performance, nutrient uptake, and elemental composition of different generations and different thallus parts of *Saccharina japonica*. *Journal of Applied Phycology* **25**(2): 631-637.
- Willing EM, Rawat V, Mandã kT, Maumus F, James GV, Nordstrã M KJ, Becker C, Warthmann N, Chica C, Szarzynska B. 2015.** Genome expansion of *Arabidopsis alpina* linked with retrotransposition and reduced symmetric DNA methylation. *Nat Plants* **1**(2): 14023.
- Wu H, Zhang Y. 2014.** Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**(1-2): 45.
- Xiang H, Zhu J, Chen Q, Dai F, Li X, Li M, Zhang H, Zhang G, Li D, Dong Y. 2010.** Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nature Biotechnology* **28**(5): 516.
- Ye N, Zhang X, Miao M, Xiao F, Zheng Y, Xu D, Wang J, Zhou L, Wang D, Gao Y. 2015.** *Saccharina* genomes provide novel insight into kelp biology. *Nature Communications* **6**: 6986.
- Zhang N, Zhang L, Tao Y, Guo L, Sun J, Li X, Zhao N, Peng J, Li X, Zeng L. 2015.** Construction of a high density SNP linkage map of kelp (*Saccharina japonica*) by sequencing Taq I site associated DNA and mapping of a sex determining locus. *BMC Genomics*, 16,1(2015-03-15) **16**(1): 1-11.

Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**(9): R137.

Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**(7463): 477.

The following Supporting Information is available for this article:

Methods S1 1, Constructing the pseudo-chromosomes using a genetic linkage map; 2, methylation profiles of sex-determining regions and their gene expression; 3, DNA methyltransferases evolution and their putative function; 4, other methods in this article.

Fig. S1: MeDIP-seq libraries.

Fig. S2: Methylation profiles of MG based on MeDIP-seq.

Fig. S3: Profiles of SP, MG, FG methylomes and transcriptomes of Chromosome 0.

Fig. S4: Chromosome-specific profiles for SP, MG, FG methylomes and transcriptomes.

Fig. S5: GC characteristics and methylation levels of chromosomes.

Fig. S6: Comparison between methylation levels of individual cytosines.

Fig. S7: Correlation between CpGo/e and the methylation level.

Fig. S8: Correlation between C+G percentage and the methylation level.

Fig. S9: Correlation between CpG percentage and the methylation level.

Fig. S10: Correlation between CHG percentage and the methylation level.

Fig. S11: Correlation between CHH percentage and the methylation level.

Fig. S12: Correlation between CpG islands ratio and the methylation level.

Fig. S13: Difference in methylation levels for different types of genomic elements between life-cycle stages.

Fig. S14: Normalized methylation level for genomic elements.

Fig. S15: Methylation for all methylation contexts of lncRNA genes and circRNA genes.

Fig. S16: Most enriched (top 20) KEGG pathways for methylated genes in FG.

Fig. S17: Most enriched (top 20) KEGG pathways for methylated genes in MG.

Fig. S18: Most enriched (top 20) KEGG pathways for methylated genes in SP.

Fig. S19: Genome-wide gene expression profiling and DNA methylation profiles.

Fig. S20: Methylation profiles of the sex-determining region.

Fig. S21: Base-specific methylation and transcriptional coverage of selected metabolic genes.

Fig. S22: Correlation between MLgf and gene expression for lncRNA genes.

Fig. S23: Correlation between MLgf and gene expression for circRNA genes.

Fig. S24: The correlation ship between MLgf of TEs and their expression level.

Fig. S25: Consistency assessment of MeDIP-seq vs WGBS data.

Fig. S26: BS-PCR for SJ18945.

Fig. S27: BS-PCR for SJ15797.

Fig. S28: BS-PCR for SJ15874.

Fig. S29: BS-PCR for SJ13722.

Fig. S30: BS-PCR for SJ21435.

Fig. S31: BS-PCR for SJ22030.

Fig. S32: BS-PCR for SJ15673

Fig. S33: BS-PCR for SJ19510.

Fig. S34: BS-PCR for SJ22283.

Fig. S35: BS-PCR for SJ20815.

Fig. S36: Phylogeny of methyltransferases based on Pfam PF00145.

Fig. S37: Domain architecture of DNMT4/5/6.

Fig. S38: DNA methylase domain PF000145 and their expression profile.

Fig. S39: Gene structure, expression patterns and chromosomal locations of MBT, PWWD, DNMT1, Zinc finger PHD-type.

Fig. S40: Genes structure, expression patterns and chromosomal locations of ADD, Pre_SET, SET, SNF, helicase and AP2/ERF.

Fig. S41: Genes structure and phylogeny of demethyltransferases.

Fig. S42: Genes structure of MBD domains and their phylogeny.

Fig. S43: Methylation levels of tRNA genes.

table S1: Summary of methylation and RNA sequencing data and mapping results.

table S2: Peak-specific information.

table S3: Chromosome-specific gene methylation and expression.

table S4: Features of methylated sites for all methylation contexts and life-cycle stages.

table S5: Features of TE methylation.

table S6: Features of gene methylation.

table S7: Features of exon methylation.

table S8: Features of intro methylation.

table S9: Features of lncRNA gene methylation.

table S10: Features of circle RNA gene methylation.

table S11: Features of TSSUP methylation.

table S12: Features of TESDOWN methylation.

table S13: Features of CGI methylation.

table S14: Features of tRNA gene methylation.

table S15: Information of DMRs.

table S16: GO enrichment of genes located in DMRs.

table S17: KEGG enrichment of the genes located in DMRs.

table S18: Expression and statistical tests of all genes and life-cycle stages.

table S19: Methylation level and statistical tests of all TSSUPs and life-cycle stages.

table S20: Methylation level and statistical tests of all genes and life-cycle stages.

table S21: KEGG enrichment of genes and TSSUPs for Fig. 3C.

table S22: GO enrichment of genes and TSSUPs for Fig. 3C.

table S23: Methylation and expression of genes located in the sex-determining-region for all life-cycle stages.

table S24: Heterozygosity between FG and MG vs the reference genome.

table S25: Methylation levels of tRNA genes.

Figure legends

Figure 1. Methyloome and transcriptome profiles for 31 chromosomes of the *Saccharina japonica* sporophyte (SP), male (MG) and female (FG) gametophyte genomes. All calculations were conducted by applying a 10kb sliding window. The sliding step length was set at 10kb ensuring no region was accounted for twice. The CpGI ratio (grey track), CpG_{O/e} (Observed value / Expected value, dark red track), C+G Ratio (yellow track), CpG ratio (green track), CHH ratio (light red track) and the CHG ratio (blue track) are displayed by the 6 outermost circles in the graph. Differential transcription (normalised read counts) between the different life-cycle stages in relation to DNA methylation (ML_{gf}) is shown by the 6 innermost circles.

Figure 2. Genomewide methylation landscapes and NML_{gc} of *Saccharina japonica* sporophyte (SP), male (MG) and female (FG) gametophytes. A) The counts of methylated cytosines (mCs) in CpG, CHG, CHH contexts for SP, FG and MG. The three dots in each column are results of replicates. B) Distribution of ML_{wg} in CpG, CHG, CHH contexts for SP, FG and MG. See the Materials and Methods section for the calculation of ML_{wg}. The

The data of MLmc for each mC and its context attribute is obtained from the direct result Bismark (Datasets S1-S9). The white spots of the violin plots stand for the median value. C) MLgc of TSSUP (transcript start site), TESDOWN (transcript stop site), exon, intron, transposable elements (TEs), genes encoding for long non-coding RNAs (lncRNAs), circular RNAs (circRNAs) and tRNAs. To eliminate the variability among samples all MLgc are normalized by the intra-sample MLwg (GENOME). Error bars indicate \pm standard deviation. D) The distribution of CG ratios (number of G and C residues/total number of residues) of all elements for all life stages. Upper panel: density plot lines of different types (in different colors) show the distribution of CG ratios of all element types. The density command was used with parameter $bw=0.01$ to draw these plots. Lower panel: Scatter plots show the distribution of the lengths of individual elements (Y-axis) and their CG ratios (X-axis). E) Meta-plots of MLgf across genomic elements of SP, FG, MG in three contexts (CpG, CHG, CHH). Each single element is divided into 20 equal bins and numbered them by one's order of position, the MLgf for each bin is calculated in three contexts (CpG, CHG, CHH) respectively. The MLgf of bins in the same order in the same genomic class are averaged. F) Methylation levels of methylated genomic elements (MLgf). The pie charts on top describe the percentage of methylated (purple) and unmethylated (orange) elements whereas box plots display their MLgf (regardless of unmethylated genomic elements). Statistics were based on Duncans and Student t-tests, the lowercase letters notations associated with plots in Fig. 2A, B, C, F indicate the level of significance among different groups.

Figure 3. Genomic elements overlapped with differentially methylated regions (DMRs) for all *Saccharina japonica* life-cycle stages (Sporophyte (SP), male (MG) and female (FG) gametophytes) and their relation to gene expression. A) Methylation differences of genomic elements (different colours) that overlapped with DMRs (see the Materials and Methods section) for SP, FG and MG. Ternary plots were drawn according to the MLgf of genomic elements. B) MLgf of genes for SP, FG and MG including a 2kb-up- (TSSUP) and down-stream (TESDOWN) flanking regions. All 18,733 genes were classified into four quantiles according to their expression levels (high, medium, low and none). Two-kilobase regions up- and down-stream of genes were divided into 100-bp intervals for screening. Each gene was divided into 20 intervals (5% of total coding length per interval). Plots show the MLgf of each interval for both coding and non-coding regions up- and down-stream of genes. C) Heat maps of genes and their promoter regions (TSSUP) with significant negative

correlation (p -value < 0.05) between their MLgf and expression levels for pairwise comparisons between all life-cycle stages. Heat maps are colored according to MLgf whereas gene expression levels are colored according to $\log(\text{RPKM}+1)$. Abbreviations used in the figure: methy-level (MLgf), gene-exp (gene expression). D) Significant negative correlation (p -value < 0.05) between the MLgf and the expression of genes (including their promoter regions) which moved from the sex determination region to autosomal loci. Upper row of panels: transcript read counts of four genes (SJ18945: glycosyl transferase, SJ15975: villin, SJ15874: GTPase activating protein, SJ13722: memo-like protein) from SP, FG, MG. Middle row of panels: RT-PCR results of the same four genes using SP as control. Lower row of panels: Methylation levels of the four genes and their promoters. $N=3$; Statistics were based on Duncans and Student t-tests. Error bars indicate \pm standard deviation.

Figure 4. DNA methyltransferases (DNMTs) and their evolution and expression in *Saccharina japonica* life-cycle stages (Sporophyte (SP), male (MG) and female (FG) gametophytes) A) Phylogeny of DNMTs including characteristic 3-D structures for members of each clade. DNMTs of *S. japonica* were marked using red pentagrams in the tree. Green ellipse marked gene SJ00150 were identified to maintain chloroplast targeting signal peptide. To compare the 3D structure of DNMTs 1-3, we used homologous X-ray structure models in SWISS-MODEL (<https://www.swissmodel.expasy.org/>). Seven X-ray structure models (PDB accession numbers 4wxx.1, 4ft4.1, 1g55.1, 4da4.1, 4h0n.1, 5yx2.1, 2i9k.1) were used as they had the highest coverage and similarity to 7 genes marked by red points in the tree. B) Scaffold locations of DNMT2s in *S. japonica*. Red marks represent genes in scaffolds, and the black triangle indicates the scaffold location of DNMTs. Red numbers indicate the start and end sites of the scaffolds. C) Heat-map showing the expression levels of DNMT2s for each life-cycle stage (SP, FG, MG) and biological replicates ($N=3$). A significant positive correlation (p -value < 0.05) between the level of DNA methylation and the expression of DNMT2s was shown for FG and MG (blue frame). D) qRT-PCR result of the six DNMT2s for SP, MG and FG. $N=3$; Statistics were based on Duncans and Student t-tests, error bars indicate \pm standard deviation, the lowercase letters notations associated with plots in Fig. 4D indicate the level of significance among different groups.







