DNA Research, 2018, 25(5), 511–520
doi: 10.1093/dnares/dsy021
Advance Access Publication Date: 29 June 2018
Full Paper

OXFORD

Full Paper

Genome-wide characterization of DNA methylation, small RNA expression, and histone H3 lysine nine di-methylation in *Brassica rapa* L.

Satoshi Takahashi[1,†], Kenji Osabe[2,†], Naoki Fukushima[3,†], Shohei Takuno[4], Naomi Miyaji[3], Motoki Shimizu[5], Takeshi Takasaki-Yasuda[3], Yutaka Suzuki[6], Elizabeth S. Dennis[7,8], Motoaki Seki[1,9,10], and Ryo Fujimoto[3,*]

[1]RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 230-0045, Japan, [2]Plant Epigenetics Unit, Okinawa Institute of Science and Technology Graduate University, Onna-son, Okinawa 904-0495, Japan, [3]Graduate School of Agricultural Science, Kobe University, Rokkodai, Nada-ku, Kobe 657-8501, Japan, [4]Department of Evolutionary Studies of Biosystems, SOKENDAI (The Graduate University for Advanced Studies), Hayama, Kanagawa 240-0193, Japan, [5]Iwate Biotechnology Research Center, Narita Kitakami, Iwate 024-0003, Japan, [6]Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-0882, Japan, [7]CSIRO Agriculture and Food, Canberra, ACT 2601, Australia, [8]University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Australia, [9]Core Research for Evolutional Science and Technology, Japan Science and Technology, Kawaguchi, Saitama 332-0012, Japan, and [10]RIKEN Cluster for Pioneering Research, Wako, Saitama 351-0198, Japan

*To whom correspondence should be addressed. Tel. +81 78 803 5827. Fax. +81 78 803 5827.
Email: leo@people.kobe-u.ac.jp

[†]These authors equally contributed to this work.

Edited by Dr. Satoshi Tabata

Received 8 March 2018; Editorial decision 27 May 2018; Accepted 30 May 2018

Abstract

Epigenetic gene regulation is crucial to plant life and can involve dynamic interactions between various histone modifications, DNA methylation, and small RNAs. Detailed analysis of epigenome information is anticipated to reveal how the DNA sequence of the genome is translated into the plant's phenotype. The aim of this study was to map the DNA methylation state at the whole genome level and to clarify the relationship between DNA methylation and transcription, small RNA expression, and histone H3 lysine 9 di-methylation (H3K9me2) in *Brassica rapa*. We performed whole genome bisulfite sequencing, small RNA sequencing, and chromatin immunoprecipitation sequencing using H3K9me2 antibody in a Chinese cabbage inbred line, RJKB-T24, and examined the impact of epigenetic states on transcription. Cytosine methylation in DNA was analysed in different sequence contexts (CG, CHG, and CHH) (where H could be A, C, or T) and position (promoter, exon, intron, terminator, interspersed repeat regions), and the H3K9me2 and 24 nucleotide small interfering RNAs (24 nt-siRNA) were overlaid onto the *B. rapa* reference genome. The epigenome was compared with that of *Arabidopsis thaliana* and the relationship between the position of DNA methylation and gene expression, and the involvement of 24 nt siRNAs and H3K9me2 are discussed.

Key words: DNA methylation, histone H3 lysine 9 di-methylation, small RNA, epigenetic, Brassica

1. Introduction

Genetic variation can cause phenotypic variation in organisms; due to the changes in the structure of an encoded protein or gene expression level by mutations of the DNA sequence. In addition to genetic variation, gene expression can be controlled by changes in the structure of chromatin without changing the DNA sequence, and this phenomenon is termed 'epigenetic' control. DNA methylation and histone modification are well-known epigenetic modifications.[1,2]

Nucleosomes are formed by a histone octamer containing two of the core histones H2A, H2B, H3, and H4, and 147 bp of DNA is wrapped around this core. The N-terminal regions of histone proteins are subject to various chemical modifications, such as methylation or acetylation, and these histone modifications are associated with gene transcription. Different histone marks can be controlled by different histone methyltransferase and can lead to different effects on gene regulation.[3,4] In *Arabidopsis thaliana*, histone H3 lysine 9 di-methylation (H3K9me2) is associated with heterochromatic regions,[5] and histone lysine methyltransferases, KRYPTONITE (KYP)/SU(VAR)3-9 HOMOLOG 4 (SUVH4), SUVH5, and SUVH6, catalyze addition of H3K9me2.[6]

In plants, cytosines in all contexts, CG, CHG, and CHH (H is A, T, or C), can be methylated. Spontaneous epimutation is defined as heritable stochastic changes in the methylation states at CG, CHG, and CHH sites, and the rate of epimutation is overwhelmingly higher than the rate of genetic mutations in *A. thaliana*.[7,8] Sometimes, epimutation becomes the driving force of phenotypic variation.[1,2] DNA methylation has an important role in the regulation of gene expression, silencing of repeat sequences and transposons, and genome imprinting.[2,9] DNA methylation in CG contexts is largely maintained by METHYLTRANSFERASE I (MET1), and those in CHG contexts are largely maintained by CHROMOMETHYLASE 3 (CMT3)-associated with H3K9me2.[2,6,9] CHH site methylation is maintained by CMT2 or DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2).[10,11] The *de novo* methylation in all contexts is catalyzed by DRM2 and is triggered by 24 nucleotide small-interfering RNAs (24 nt-siRNAs) produced by the RNA interference (RNAi) pathway, termed RNA-directed DNA methylation (RdDM). Two plant-specific RNA polymerases, Polymerase IV (Pol IV) and Pol V, RNA-DEPENDENT RNA POLYMERASE 2 (RDR2), DICER-LIKE 3 (DCL3), and ARGONAUTE 4 (AGO4) proteins function in this RNAi pathway.[2,12]

Genome-wide profiles of epigenetic information (the epigenome) are available in plants using technologies such as tiling arrays or high-throughput sequencing. High-resolution maps of epigenetic features have been obtained from whole genome bisulfite sequencing (WGBS) (bisulfite converted DNA is directly sequenced) or a combination of chromatin immunoprecipitation (ChIP) technology and high-throughput sequencing (ChIP-seq).[4,13] Using these technologies, variations of DNA methylation states between accessions in a specific species or between species have been shown at the whole genome level.[14–17] In *A. thaliana*, DNA methylation is enriched in heterochromatic regions of genomes, which consist of repetitive sequences and transposable elements (TEs).[18,19] DNA methylation is also observed in the euchromatic regions, and genes having DNA methylation at CG sites in transcribed regions, termed gene-body methylation (gbM), showed high expression level.[18,19]

Brassica rapa L. comprises commercially important vegetable crops consumed worldwide including Chinese cabbage (var. *pekinensis*), pak choi (var. *chinensis*), and komatsuna (var. *perviridis*), root vegetables including turnip (var. *rapa*), and oilseed (var. *oleifera*). Most cultivars of *B. rapa* are self-incompatible, preventing self-fertilization, and are F_1 hybrid cultivars, which have increased yields relative to their parents (heterosis/hybrid vigour).[20,21] Chinese cabbage forms a head with large pale-green-coloured leaves and wide white midribs and is an important vegetable in Asia. The draft genome sequence of Chinese cabbage (Chiifu-401-42) has been released and 199 accessions in *B. rapa* representing various morphotypes have been re-sequenced.[22,23] The *B. rapa* genome has undergone a whole-genome triplication (WGT), which results in multiple copies of paralogous genes. Three subgenomes, the least fractioned (LF) subgenome and two more fractionated subgenomes (MF1 and MF2), were found in *B. rapa* genome, and biased gene fractionation and dominant gene expression pattern among the subgenomes were shown.[22,24,25]

The aim of this study was to identify the DNA methylation states at the whole genome levels and to clarify the relationship between DNA methylation and transcription, small RNA expression, and H3K9me2 modification in *B. rapa*. *B. rapa* is related to the model plant *A. thaliana*, which enables us to compare the conservation and diversification of epigenetic states between orthologous genes of these species. In addition, *B. rapa* has more repetitive sequences than *A. thaliana* and has experienced WGT after speciation, which allows us to examine the effect of differences in genome composition on epigenetic states. Furthermore, the knowledge of the DNA methylation state, small RNA expression, and H3K9me2 placement at the whole genome level will be useful for examining natural variation, change of epigenetic states by abiotic or biotic stress, and understanding epigenetic states related to agronomically important traits using segregation of loci.

2. Materials and methods

2.1. Plant materials and growth conditions

A Chinese cabbage inbred line developed in a previous study, RJKB-T24, was used as plant material.[26] Plants were grown on agar-solidified Murashige and Skoog (MS) plates with 1.0% (w/v) sucrose under long day condition (16 h light) at 21°C. For WGBS, small RNA sequencing (small RNA-seq), and ChIP analyses, first and second leaves were harvested from plants at 14 days after sowing.

2.2. Whole genome bisulfite sequencing (WGBS)

Genomic DNAs from 14 days first and second leaves were isolated using DNeasy Plant Mini Kit (Qiagen). Genomic DNAs were fragmented to 100–300 bp length by sonication. DNA-end repair, 3′-dA overhang, ligation of methylated sequence adaptors, and bisulfite treatment by ZYMO EZ DNA Methylation-Gold kit (ZYMO RESEARCH) were performed for constructing bisulfite sequence library. The library was sequenced using Illumina Hiseq2000 (100 bp paired end).

The reads of WGBS were mapped to the *B. rapa* reference genome v.1.5 (http://brassicadb.org/brad/) or to RJKB-T24 re-sequenced genome, whose SNPs were replaced in the reference genome, using Bowtie2 version 2.2.3 and Bismark v0.14.3.[27] In order to estimate methylation levels of CG, CHG, and CHH contexts, the numbers of methylated and unmethylated reads were extracted on each cytosine position using bismark_methylation_extractor script with the paired-end parameter. The methylation level at each cytosine position was calculated by dividing the number of methylated cytosines (mC) reads by the total reads.

2.3. Classification of gene-body methylation (gbM)

We classified cytosine sites into methylated and unmethylated cytosines by employing a binomial test, following Lister *et al.*[18] We estimated the sequence context-specific error rates of bisulfite conversion (0.0173 for CG, 0.0178 for CHG, and 0.0215 for CHH sites) from the mapping result of the unmethylated phage genome, and used these values for the binomial test. We applied the test only to the cytosine site with >2 WGBS coverage, and used a significant cut-off of two-tailed *P*-value < 0.01 to detect mC.

We quantified the level of DNA methylation for each protein-coding region that was defined as the region from start to stop codons, which includes both exons and introns. We screened genes for which ≥60% of all cytosine residues and ≥20 cytosines in the CG context have >2 aligned WGBS reads, and calculated CG, CHG, and CHH methylation levels for these genes. We further classified genes into body- and under-methylated genes, following Takuno and Gaut[28] with a slight modification. We calculated the number of cytosine residues with >2 coverage at CG sites (n_{CG}) and the number of mC residues at CG sites (m_{CG}) for each gene. We calculated *P*-values (P_{CG}) for the departure of CG methylation levels from the average CG methylation level across all genes (p_{CG}):

$$P_{CG} = \sum_{i=m_{CG}}^{n_{CG}} \binom{n_{CG}}{i} p_{CG}^{i} (1 - p_{CG})^{n_{CG}-i}$$

The smaller P_{CG} is, the higher CG methylation level is. We defined body- and under-methylated genes as genes with $P_{CG} < 0.05$ and $P_{CG} > 0.95$, respectively. Using the same rationale, we calculated P_{CHG} and P_{CHH} for CHG and CHH methylation, respectively.

2.4. Small RNA sequencing (small RNA-seq)

Total RNA was isolated from 14 days first and second leaves of *B. rapa* using TRIzol Reagent (Thermo Fisher) with DNase I treatments. Small RNAs of 17–30 nucleotide (nt) were purified. The sequence library was prepared using Oligonucleotide sequences for TruSeq™ Small RNA Sample Prep kits (Illumina), and small RNA-seq was performed using Illumina Hiseq2000 (36 bp single end).

The reads of small RNA-seq were purged from low quality reads or adapter sequences using cutadapt version 1.7.1 and Trim Galore! version 0.3.7. Then the reads were mapped to the *B. rapa* reference genome v.1.5 using Bowtie2 version 2.2.3. To examine the size distribution of mapped small RNAs, we classified the alignment reads by length. The 24 nt aligned reads were used to identify 24 nt-siRNA clusters by clustering adjacent reads using bedtools 2.15.0 with the 'merge –d 200 (reads within 200 nt from each other were merged in one cluster)' parameter. Clusters that contain at least five reads were picked up.

2.5. Chromatin immunoprecipitation sequencing (ChIP-seq)

ChIP experiments were performed as described by Kawanabe *et al.*[29] One gram of 14 days first and second leaves of *B. rapa* was used for ChIP analysis and anti-H3K9me2 (ACTIVE MOTIF, 39753) antibody was used. Before the ChIP-seq, we validated the enrichment of purified immunoprecipitated DNAs by qPCR using the positive and negative control primer sets of H3K9me2 previously developed.[29] Purified immunoprecipitated DNAs and input DNA were sequenced by Hiseq2000 (36 bp single end). The reads of ChIP-seq were purged from low quality reads or adapter sequences using cutadapt version 1.7.1 and Trim Galore! version 0.3.7. Then the reads were mapped to the *B. rapa* reference genome v.1.5 using Bowtie2 version 2.2.3. We performed peak calling on alignment results using model-based analysis for ChIP-seq (MACS) 2 2.1.0 and identified the regions having H3K9me2 as peaks. Primer sequences used in this study are shown in Supplementary Table S1.

2.6. Bisulfite sequencing

Genomic DNAs from 14 days first and second leaves was isolated using DNeasy Plant Mini Kit (Qiagen). A total of 500 ng DNA was fragmented by sonication and fragments were approximately 300–800 bp in length. MethylCode Bisulfite Conversion Kit (Thermo Fisher) was used for chemical bisulfite reaction and PCR was performed using bisulfite treated DNAs as templates. PCR conditions were 95 °C for 2 min followed by 40 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s. Amplified PCR fragments were gel purified using GENECLEAN III Kit (MP Biomedicals) and cloned into pGEM-T Easy vector (Promega). Ten independent clones were sequenced. Primers used for bisulfite sequencing are listed in Supplementary Table S1.

3. Results

3.1. Whole genome DNA methylation states of RJKB-T24

We performed WGBS using 14-day old first and second leaves of the Chinese cabbage (*B. rapa*) inbred line, RJKB-T24. Approximately 192 M clear reads were obtained and 70 M reads were uniquely mapped on the *B. rapa* reference genome (from chromosome 1 to 10) using Bismark, about 27 times the depth of the sequenced genome. Approximately 86% of cytosine sites in the regions from A01 to A10 were covered (Supplementary Table S2). Most of the methylated CG sites were heavily methylated, while methylated CHG and CHH sites had low methylation levels (Fig. 1A). A total of 70% of methyl-CG sites were fully methylated (Fig. 1A), which is higher than that in *A. thaliana* (20–30%).[18] The average of methylation levels of the whole genome (total) for CG, CHG, and CHH sites were 36.5%, 13.4%, and 5.3%, respectively, and interspersed repeat regions (IRRs) were heavily methylated (CG, 73.7%; CHG, 33.8%; CHH, 13.0%) (Fig. 1B, Supplementary Fig. S1). These methylation levels were slightly lower but similar to those seen using the reference genome with scaffold sequences or RJKB-T24 re-sequenced genome as a template of mapped reads (Supplementary Table S3), indicating that sequence polymorphisms between RJKB-T24 and the reference line did not greatly affect the quantification of the DNA methylation levels. We validated the WGBS results in the five selected regions by bisulfite sequencing (Supplementary Fig. S2).

Previously, we examined DNA methylation states by methylated DNA immunoprecipitation sequencing (MeDIP-seq) using the same plant material and tissues but independently harvested.[30] We compared the results from the two methods, WGBS and MeDIP-seq, using genes having DNA methylation in the genic regions. More genes were detected by WGBS than by MeDIP-seq, and about 90% of genes having MeDIP-peaks in the 2-kb upstream, exon, intron, and 2-kb downstream regions overlapped with genes having DNA methylation detected by WGBS (Supplementary Fig. S3). The correlation coefficient between the methylation levels of all cytosine sites detected by WGBS and methylation levels detected by MeDIP-seq [log 10 score of reads per kilobase of exon per million mapped reads (RPKM), RPKM > 0.01] in genic regions (200-bp upstream/exon/

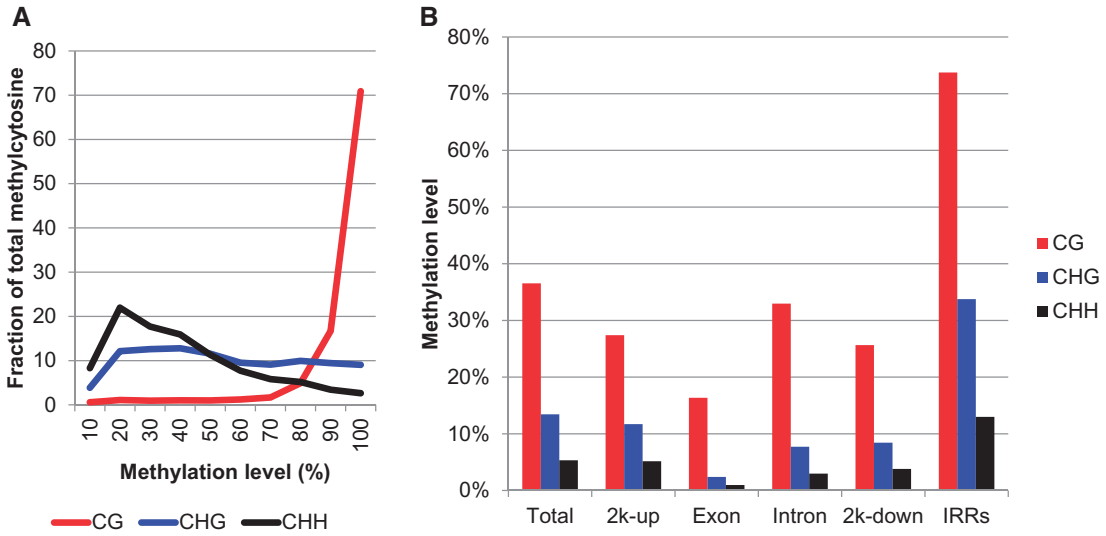

Figure 1. DNA methylation level detected by WGBS. (A) Distribution of the percentage of methylation site at each context. (B) Average of DNA methylation levels in the 2-kb upstream (2k-up), exon, intron, and 2-kb downstream (2k-down) regions, and interspersed repeats regions (IRRs). 'Total' represents the average of DNA methylation levels in all regions of the genome.

Table 1. Correlation coefficients between DNA methylation levels and expression levels

	CG		CHG		CHH	
	r	*P*-values	*r*	*P*-values	*r*	*P*-values
2-kb upstream	−0.035	<1.0e-5	−0.002	NS	0.016	<0.05
500-bp upstream	−0.081	<1.0e-5	−0.043	<1.0e-5	−0.009	NS
200-bp upstream	−0.103	<1.0e-5	−0.087	<1.0e-5	−0.057	<1.0e-5
First exon	−0.123	<1.0e-5	−0.108	<1.0e-5	−0.078	<1.0e-5
Exon	−0.012	NS	−0.137	<1.0e-5	−0.129	<1.0e-5
First intron	−0.091	<1.0e-5	−0.109	<1.0e-5	−0.081	<1.0e-5
Intron	−0.035	<1.0e-5	−0.133	<1.0e-5	−0.103	<1.0e-5
200-bp downstream	−0.079	<1.0e-5	−0.103	<1.0e-5	−0.081	<1.0e-5
500-bp downstream	−0.100	<1.0e-5	−0.095	<1.0e-5	−0.057	<1.0e-5
2-kb downstream	−0.058	<1.0e-5	−0.049	<1.0e-5	−0.034	<1.0e-5

r, correlation coefficient.
Expression levels is log 2 score of FPKM (FPKM > 0.01).
NS, not significantly.
P-values are calculated by permutation test.

intron/200 bp-downstream regions) was high ($r = 0.76$, $P < 10e\text{-}6$), indicating that WGBS data is consistent with the MeDIP-seq data. These results indicate that MeDIP-seq detected highly methylated regions, while WGBS can examine all DNA methylation contexts including the regions detected by MeDIP-seq analysis.

3.2. Genes having DNA methylation and their expression levels

We examined the impact of DNA methylation on transcription by comparing the DNA methylation data to previous RNA sequencing (RNA-seq) data with two replicates generated from samples using the same tissue and stage, but independently harvested.[30,31] The correlation coefficient between the CG, CHG, and CHH methylation levels at specific genic regions (2-kb, 500-bp, and 200-bp upstream and downstream regions, first exon, exon, first intron, and intron) and the expression levels (log 2 score of FPKM, FPKM > 0.01) was examined. There was a negative correlation between methylation levels and expression levels, except for CG methylation in exon regions, CHG methylation in the 2-kb upstream region, and CHH methylation in the 500-bp and 2-kb upstream regions (Table 1). CG and CHG methylation levels in the 200-bp upstream and 500-bp downstream regions and CG, CHG, and CHH methylation levels in the 200-bp downstream regions correlated with lower expression levels. CG, CHG, and CHH methylation levels in the first exon and first intron also showed negative correlation with expression levels. In summary, there were negative correlations between CHG and CHH methylation levels of the exon and intron regions and expression levels, whereas no negative correlation was found between the expression level and CG methylation levels of the exon and intron region (Table 1).

3.3. Comparison of DNA methylation states of *B. rapa* genes with orthologous genes of *A. thaliana*

We first identified gbM genes (methylated only in the CG context), under-methylated genes (under-methylated at all three types of sites),

and genes having methylation at all three types of sites, in *B. rapa* by a probabilistic approach (see Section 2.3).[28] A total of 3,521 of 29,612 genes (11.9%) were classified as gbM genes. The proportion of gbM genes in *B. rapa* is lower than that in *A. thaliana* (23.3%; using the same calculation methods on the data from Lister *et al.*[18]) (Supplementary Table S4). A total of 3,141 of 29,612 genes (10.6%) had methylation both in the CG, CHG, and CHH contexts, which was higher than that in *A. thaliana* (3.6%), while 14,583 of 29,612 genes (49.2%) were categorized into under-methylated genes in *B. rapa*, similar to that in *A. thaliana* (44.1%) (Supplementary Table S4).

Next, we compared the DNA methylation states between the orthologous genes of *B. rapa* and *A. thaliana*. Because *B. rapa* experienced a WGT event,[22,24,25] there are from one to three *A. thaliana* orthologous genes in *B. rapa*. The previous study provided the list of orthologous genes based on the similarity of protein sequences and syntenic information.[24] The orthologous genes were found in *A. thaliana* genes for 3,448 of 3,521 gbM genes in *B. rapa*, 2,976 of which were one-to-one orthologous genes between the two species (Table 2). A total of 1,904 of these 2,976 genes (64.0%) had gbM in *A. thaliana*. A total of 14,147 of 14,583 under-methylated genes in *B. rapa* had the orthologous genes in *A. thaliana*, and this was reduced to 10,333 one-to-one orthologous genes. Only 1,842 of these 10,333 genes (17.8%) had gbM in the *A. thaliana* genome (Table 2). These results suggest gbM tended to be conserved between orthologous genes in *B. rapa* and *A. thaliana*.

We further compared the level of gbM (level of CG methylation) between orthologous genes in *B. rapa* and *A. thaliana*. Again, we found significant correlations between them in all three subgenome domains of *B. rapa*, LF, MF1, and MF2 ($r = 0.433$ for LF; $r = 0.393$ for MF1; $r = 0.402$ for MF2; $P < 10^{-5}$ by a permutation test in all three cases; Fig. 2).[24] We also compared the levels of gbM between paralogous genes in *B. rapa* that were derived in WGT event, and found significant correlations between them ($r = 0.467$ for LF vs. MF1; $r = 0.494$ for LF vs. MF2; $r = 0.442$ for MF1 vs. MF2; $P < 10^{-5}$ by a permutation test in all three cases; Supplementary Fig. S4), indicating that gbM was conserved between paralogous genes in *B. rapa*.

3.4. Small RNA sequencing

24 nt siRNAs are associated with targeting DNA methylation to complementary sequences by RdDM.[12] To examine the relationship between the distribution of 24 nt siRNAs and DNA methylation, we performed small RNA-seq. A total of 47.4 M clean reads were obtained, and 6.6 M reads (14.0%) were uniquely mapped on the *B. rapa* genome. A total of 3.7 of 6.6 M reads (55.3%) were mapped on the IRRs (Supplementary Table S5). The majority of small RNAs were 24 nt in length. The second and third most abundant small RNAs were 23 nt and 21 nt in length, respectively (Supplementary Fig. S5). We identified 65,142 clusters of 24 nt siRNAs in total, and 11,354, 1,678, 2,057, and 8,795 genes had more than one 24 nt-siRNA cluster in the 2-kb upstream, exon, intron, and 2-kb downstream regions, respectively (Supplementary Table S6).

We compared 24 nt-siRNA level (RPKM) and DNA methylation level of each sliding window of per 100 kb, and 24 nt-siRNA levels were most associated with CHH methylation levels (Fig. 3A). We compared the regions having 24 nt-siRNA clusters and regions having DNA methylation detected by WGBS. About 70–80% of genes having 24 nt-siRNA clusters in the 2-kb upstream, exon, intron, and 2-kb downstream regions overlapped with genes having DNA methylation detected by WGBS (Supplementary Fig. S6). About 48.0%, 39.0%, and 60.0% of total 24 nt-siRNA clusters overlapped with CG, CHG, and CHH methylation, respectively, and

Table 2. Comparison of gbM states between orthologous genes in *B. rapa* and *A. thaliana*

	B. rapa			*A. thaliana*		
	Number	Orthologous genes	Percentage (%)	Orthlogous genes[a]	gbM genes	Percentage (%)
gbM genes	3,521	3,448	97.9	2,976	1,904	64.0
Under-methylated genes	14,583	14,147	97.0	10,333	1,842	17.8
Heavily methylated genes	911	424	46.5	395	91	23.0

gbM, gene-body methylation.
[a]*B. rapa* genes having sequence homology to *A. thaliana* genes with removal where more than two *B. rapa* genes in *A. thaliana*.

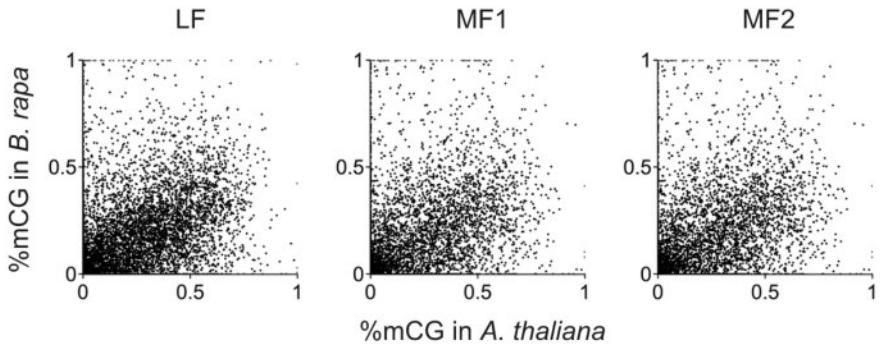

Figure 2. The scatter plots for the level of CG methylation between *A. thaliana* and *B. rapa* orthologous genes in the three subgenomes. Correlation of CG methylation levels between *A. thaliana* and *B. rapa* orthologous genes in LF, MF1, and MF2 were 0.433 ($n = 8,040$; *P*-value $< 10^{-5}$), 0.393 ($n = 5,223$; *P*-value $< 10^{-5}$), and 0.402 ($n = 4,580$; *P*-value $< 10^{-5}$), respectively. *P*-values were calculated by a permutation test.

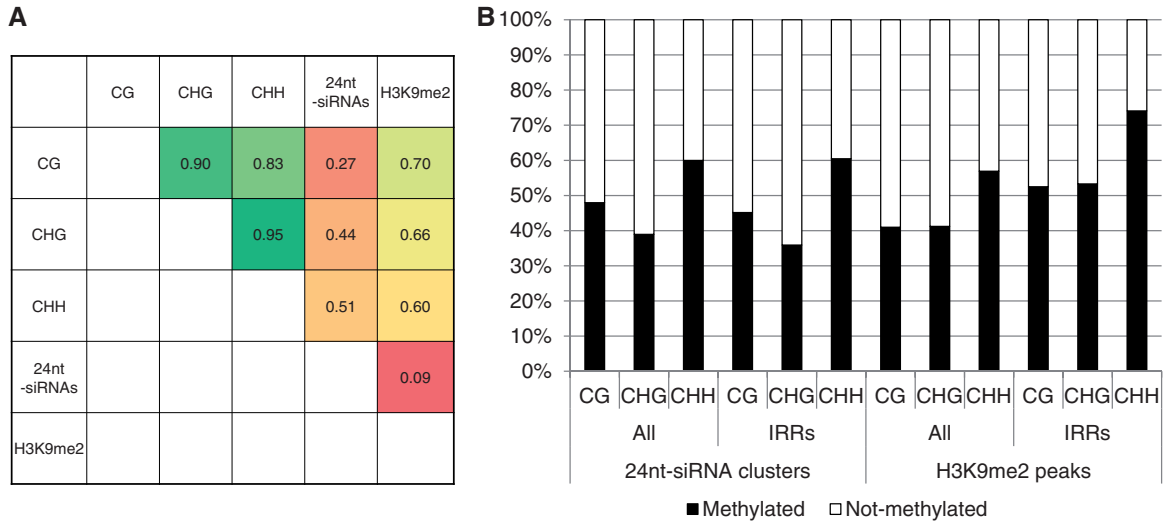

Figure 3. The comparison between 24 nt-siRNA or H3K9me2 level and DNA methylation level. (A) Comparison of the 24 nt-siRNA or H3K9me2 level of the RPKM and DNA methylation level in each sliding window per 100 kb. (B) Percentage of genes both having 24 nt-siRNAs clusters and CG, CHG, and CHH methylation or both H3K9me2 peaks and CG, CHG, and CHH methylation. 'All' and 'IRRs' indicate the average of DNA methylation levels in all regions and interspersed repeats regions, respectively, with 24 nt-siRNA clusters or H3K9me2 peaks.

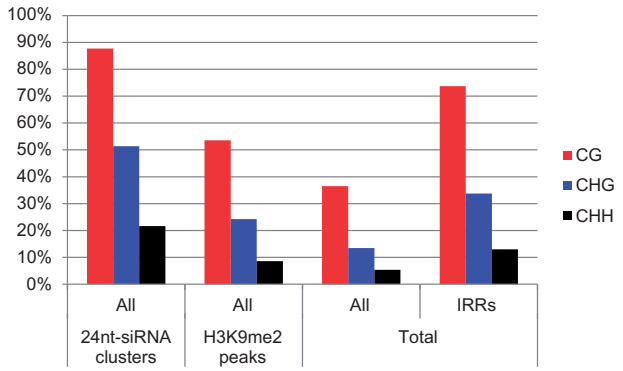

Figure 4. Average of DNA methylation levels in the regions covering 24 nt-siRNA clusters or H3K9me2 peaks. 'All' and 'IRRs' indicate the average of DNA methylation levels in all regions and interspersed repeats regions, respectively.

45.2%, 35.9%, and 60.5% of 24 nt-siRNA clusters in IRRs overlapped with CG, CHG, and CHH methylation, respectively (Fig. 3B). This indicates that regions having 24 nt-siRNA clusters tended to have CHH methylation rather than CG and CHG methylation. The average methylation levels for CG, CHG, and CHH sites in the regions overlapping 24 nt-siRNA clusters were 87.7%, 51.4%, and 21.6%, respectively (Fig. 4, Supplementary Fig. S1). DNA methylation levels overlapping the 24 nt-siRNA clusters were high regardless of where the IRRs were located within the genome (Supplementary Figs S1 and S7).

We examined the relationship between the presence of 24 nt-siRNA clusters and expression levels (log 2 score of FPKM, FPKM > 0.01). Genes having 24 nt-siRNAs in exon regions tended to show a wide range of expression (Fig. 5). Average expression level of genes having 24 nt-siRNAs in exons was higher than that of total genes, and it was also significantly higher than that of genes not having 24 nt-siRNAs in an exon (permutation test, $P < 0.001$) (Fig. 5). Of genes having 24 nt-siRNAs in exons, genes that were more lowly expressed tended to have CG, CHG, and CHH methylation, while more highly expressed genes tended to be under-methylated (Supplementary Fig. S8).

3.5. Chromatin immunoprecipitation sequencing using H3K9me2 antibody

To examine the relationship between DNA methylation and H3K9me2 in the *B. rapa* genome, we performed ChIP-seq analysis using an antibody against H3K9me2. A total of 37.8 M and 40.0 M clean reads were obtained from input-DNA-seq and ChIP-seq, respectively. A total of 18.2 M (input-DNA-seq, 48.2%) and 11.8 M reads (ChIP-seq, 29.5%) were uniquely mapped on the *B. rapa* genome, respectively (Supplementary Table S5). In total, 8.4 M ChIP-seq reads were mapped on the genic regions including exon, intron, and 2-kb upstream and downstream regions. The proportion of mapped reads of ChIP-seq in each of the four regions was similar to those of Input-DNA-seq (Supplementary Table S7). In the ChIP-seq data, 5.6 M (47.4%) reads mapped on the IRRs was higher than the percentage of mapped reads on the IRRs using input-DNA-seq data (39.1%) (Supplementary Table S5), but lower than using the previous MeDIP-seq data (63.9%; 36 bp single end).[30] A total of 9,920 H3K9me2 peaks were identified in RJKB-T24. A total of 1,293, 1,654, 449, and 895 genes had more than one H3K9me2 peak within 2-kb upstream, exon, intron, and 2-kb downstream regions, respectively (Supplementary Table S8). A total of 7,268 of 9,920 H3K9me2 peaks (73.3%) were found in the IRRs (Supplementary Table S8).

We examined the relationship between having H3K9me2 marks and expression levels (log 2 score of FPKM, FPKM > 0.01). The average expression level in genes having H3K9me2 peaks in all four regions, especially in the exon and intron regions, was lower than that of total genes, and the average expression level of genes having H3K9me2 in the exon and intron regions was significantly lower than genes that do not have H3K9me2 peaks (permutation test, $P < 0.001$) (Fig. 5).

We compared H3K9me2 levels (RPKM) and DNA methylation level or 24 nt-siRNA levels (RPKM) of each sliding window of per

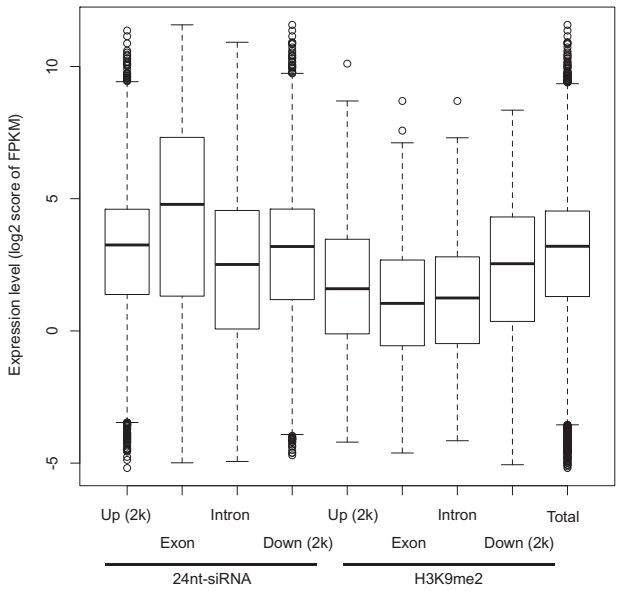

Figure 5. Box plots of the expression levels of log 2 score of FPKM in genes having 24 nt-siRNA clusters or H3K9me2 peaks in the 2-kb upstream regions Up (2k), exon, intron, or 2-kb downstream regions Down (2k) of RJKB-T24. Total indicates the log 2 score of FPKM in all genes (FPKM < 0.01).

100 kb. H3K9me2 levels were associated with DNA methylation levels, but there was no difference among the three contexts. In contrast, the association between H3K9me2 levels and 24 nt-siRNA levels was low (Fig. 3A). In genes having H3K9me2 peaks, about 50% of them overlapped with genes having DNA methylation detected by WGBS in each of the 2-kb upstream, intron, and 2-kb downstream regions, while about 20% of genes overlapped with genes having DNA methylation in exon regions (Supplementary Fig. S9). The regions having H3K9me2 had CHH methylation, especially in IRRs, rather than CG or CHG methylation (Fig. 3B). The average of methylation levels for CG, CHG, and CHH sites in the regions overlapping H3K9me2 peaks were 53.6%, 24.2%, and 8.6%, respectively (Fig. 4, Supplementary Fig. S1), indicating increased DNA methylation levels in the region covering the H3K9me2 peaks. In the regions overlapping H3K9me2 peaks, DNA methylation levels were higher in IRRs (CG; 77.5%, CHG; 31.2%, CHH; 10.4%) in comparison to all regions (Supplementary Fig. S1). However, 6.2% of H3K9me2 peaks overlapped with 24 nt-siRNA clusters, and 4.3% of H3K9me2 peaks overlapped with 24 nt-siRNA clusters in IRRs, indicating that majority of H3K9me2 accumulation was not associated with 24 nt-siRNAs (Supplementary Table S9).

3.6. Genetic and DNA methylation states variation between *B. rapa* and *A. thaliana*

We compared DNA methylation states at specific *B. rapa* genes against the orthologous genes of *A. thaliana*. A total of 1,514 and 1,472 *B. rapa* genes of MeDIP-peaks had CG, CHG, and CHH methylation (heavily methylated genes) in the exon and intron regions, respectively (Supplementary Fig. S10), and 911 genes were in common between methylated exons and introns and distributed evenly over all chromosomes (Supplementary Fig. S11). About 90% of these 911 heavily methylated genes had IRRs in exon and/or intron, and 99% of these IRRs were methylated (Supplementary Table S10), indicating that heavy methylation in exon and intron regions of these 911 genes is due to the presence of IRRs targeted for DNA methylation.

A total of 424 of the 911 genes from *B. rapa* had putative orthologous genes in *A. thaliana* including paralogous genes. This was reduced to 395 genes in *A. thaliana* following the removal of genes where two or three *B. rapa* genes showed sequence similarity to one *A. thaliana* gene (Table 2). Of the 395 genes, 217 (54.9%) were categorized into under-methylated genes and 91 (23.0%) into gbM genes (Table 2). The frequency of under-methylated and gbM genes was similar to that of the total genes (under-methylated genes, 64.7%; gbM genes, 17.7%), suggesting that heavily methylated states were not conserved between orthologous genes between *B. rapa* and *A. thaliana*.

Of the 424 genes in *B. rapa*, 234 genes (55.2%) showed no expression, categorized into Group-0 (Supplementary Fig. S12). In *A. thaliana*, we have divided the level of gene expression into six categories using microarray data; the first and sixth groups had the lowest and highest levels of expression, respectively.[32] Of 395 *A. thaliana* genes with sequence homology to heavily methylated *B. rapa*, 325 genes were on the microarray, and the proportion classified from first to sixth groups were 11% (first, lowest), 11% (second), 14% (third), 16% (fourth), 23% (fifth), and 25% (sixth, highest), respectively (Supplementary Fig. S12), indicating that there was no tendency of transcription suppression. These results indicate that *A. thaliana* genes, which showed sequence similarity to heavily methylated genes in *B. rapa*, had neither heavy DNA methylation nor low expression levels.

4. Discussion

We examined the DNA methylation states and distribution of 24 nt-siRNAs and H3K9me2 of *B. rapa* at the whole genome level. DNA methylation and H3K9me2 are classified as repressive epigenetic marks and 24 nt-siRNAs are involved in RdDM. We showed how these three factors interact with each other and affect gene expression levels in *B. rapa*.

4.1. DNA methylation states at the whole genome levels in *B. rapa*

We performed WGBS using an inbred line of Chinese cabbage, RJKB-T24, and these data were consistent with our previous MeDIP-seq data.[30] The WGBS allowed us to assess different DNA methylation sequence contexts and was more sensitive. The average methylation levels for CG, CHG, and CHH sites in chromosome A01–A10 were 36.5%, 13.4%, and 5.3%, respectively, in RJKB-T24. Previously, WGBS was performed in *B. rapa* and the average methylation levels for CG, CHG, and CHH sites were 52.4%, 31.8%, and 8.3%, respectively,[33] and 37.2%, 17.3%, and 4.4%, respectively.[16] Our results were similar to the latter report, though DNA methylation levels varied between lines or tissues. Most CG sites were highly methylated unlike in *A. thaliana*, but similar to previous reports of *B. rapa*.[16,33] High levels of DNA methylation in the IRRs including TEs or repetitive sequences were detected, which is a common observation not only in *B. rapa* but also in other plant species.[17,33–35]

4.2. Heavy methylation in genic regions results in silencing of transcription

We examined how DNA methylation affects gene expression levels using WGBS and RNA-seq data of identical tissues and stages.

Promoter methylated genes tended to have lower expression levels in *A. thaliana*.[36] Promoter methylation also repressed gene expression in rice, and gene repression by DNA methylation in the transcriptional termination regions was stronger than the effect of DNA methylation in the promoter region.[37] In our study, DNA methylation in the upstream and downstream regions of genes was negatively associated with expression levels, especially DNA methylation in the 200-bp upstream and downstream regions.

There is a strong negative relationship between DNA methylation at all sequence contexts in the first exon/intron and expression in *A. thaliana*.[38] DNA methylation in the region of the first exon is much more tightly linked to transcriptional silencing than is methylation in the upstream promoter region in humans.[39] Our results also showed methylation at CG, CHG, or CHH in the first exon or first intron associated with gene silencing. In soybean, there is positive correlation between CG gbM and expression levels, while CHG or CHH methylation in gene bodies was negatively associated with expression levels.[40] In our study, CHG and CHH methylation in exon or intron regions results in lower expression levels, while there was no negative association between CG methylation in exons (except for in the first exon) and expression levels. This indicates that CHG and CHH methylation in exon or intron regions are associated with gene silencing. In contrast, the frequency of moderately expressed genes (Group-4) having only CG methylation or gbM was higher than in the total genes (Supplementary Fig. S13), indicating that genes having gbM showed higher expression levels, which is consistent with gbM genes in other plant species.[17]

4.3. Conservation and diversification of DNA methylation states of genic regions between *B. rapa* and *A. thaliana*

We categorized genes into three groups by DNA methylation types, gbM, heavily methylated, and under-methylated genes (Table 2). Of 911 heavily methylated genes, 424 (46.5%) had sequence homology to genes in *A. thaliana*, while more than 97% of gbM genes and under-methylated genes had sequence homology to genes in *A. thaliana* (Table 2), indicating that heavily methylated genes tend to lose orthologous genes in *B. rapa*. *A. thaliana* genes, which showed sequence similarity to heavily methylated genes in *B. rapa*, did not tend to have DNA methylation. In *B. rapa*, most heavily methylated genes contained IRRs in the exon/intron regions that were heavily methylated. These results suggest heavy methylation in genes is not conserved between species and heavy methylation occurs after speciation and is caused by transposon insertion that can drive variation of DNA methylation between species.[41,42]

Conservation of gbM in orthologous genes between species, especially in closely related species, has been found.[16,41,43,44] In our study, there was a significant correlation in gbM between orthologous genes in *B. rapa* and *A. thaliana*. Significant correlation in gbM between paralogous genes was also found in *B. rapa*. These results suggest that gbM is conserved not only after speciation but also after WGT. Although gbM is conserved between *A. thaliana* and *B. rapa*, about 14% of genes had gbM in *A. thaliana*, while only 0.5% of genes had gbM in *B. rapa*.[16] In this study, about 23% of genes had gbM in *A. thaliana*, and 12% of genes had gbM in *B. rapa*. Although the percentages are different because of the different analysis methods, the lower frequency of gbM genes in *B. rapa* than in *A. thaliana* is consistent. In *Eutrema salsugineum* and *Conringia planisiliqua*, loss of CMT3 function resulted in loss of gbM,[16,45] suggesting that the lower numbers of gbM genes in *B. rapa* than in *A. thaliana* may be difference of CMT3 function.

4.4. siRNAs are associated with DNA methylation

24 nt-siRNAs guide methylation of cytosine in all classes of sequence contexts, termed RdDM, and the most abundant targets of RdDM are repetitive sequences or transposons, especially smaller and younger ones in euchromatic regions in *A. thaliana*.[10,12] We identified 65,142 clusters of 24 nt-siRNAs using uniquely mapped reads and 56% of the clusters were detected in IRRs. This frequency of the 24 nt-siRNA clusters in IRRs is lower than that of H3K9me2 peaks (73.3%) and MeDIP-peaks (86.5%) in IRRs.[30] The 24 nt-siRNA levels or the 24 nt-siRNA clusters were more associated with CHH methylation than CG and CHG methylation in *B. rapa*, suggesting that this CHH methylation was via RdDM. Surprisingly, the average of methylation levels for CG, CHG, and CHH sites in the regions overlapping 24 nt-siRNA clusters was quite high even in the non-IRRs, indicating that 24 nt-siRNA clusters are strongly associated with DNA methylation.

We examined whether 24 nt-siRNA clusters nearby or within genes affect their expression. Genes having 24 nt-siRNA clusters in the 2-kb upstream or 2-kb downstream did not affect gene expression levels. In contrast, genes having 24 nt-siRNA clusters in exons showed a broad range of expression levels and genes having 24 nt-siRNA clusters in the exon regions tended to show higher expression levels. There is a tendency for negative correlation between the DNA methylation and expression level in genes having 24 nt-siRNA clusters in the exon regions, however, 24 nt-siRNA clusters in exons were not associated with DNA methylation in some genes, especially genes showing higher expression.

4.5. Moderate association between H3K9me2 and DNA methylation

The level of H3K9me2 associated with DNA methylation levels and methylation levels at CG, CHG, and CHH sites in the H3K9me2 regions was higher than the genome average levels, similar to the results in *A. thaliana*.[5] The greatest fraction of CHG methylation overlapped with H3K9me2 positive regions in *A. thaliana*,[5] while there was no tendency for H3K9me2 to be associated with CHG methylation in *B. rapa*. A greater fraction of CHH methylation overlapped with H3K9me2 in the IRRs. In *A. thaliana*, self-reinforcing mechanisms between CMT2/3 and KYP (SUVH4)/SUVH5/SUVH6 have been proposed, and CMT2 and CMT3 add CHH and CHG methylation, respectively.[11,46,47] The genome sequences of most euchromatic regions were determined in *B. rapa*, which is approximately half of the total genome size, but the sequences of the heterochromatic regions including centromere or pericentromere covering TEs and repetitive sequences were not determined.[22] This may result in the weak correlation in *B. rapa* because association between H3K9me2 and CHG methylation levels are mostly high in TEs.[5] In maize there is no orthologous gene of AtCMT2, and there is no strong correlation between CHG methylation and H3K9me2.[48] Although the putative orthologous genes of *AtCMT2*, *AtCMT3*, *AtKYP/AtSUVH4*, *AtSUVH5*, *AtSUVH6* were confirmed in the *B. rapa* genome,[22,49,50] the basis of no strong association between CHG and H3K9me2 is that the interaction between cytosine methyltransferase and H3K9 methyltransferase in *B. rapa* may be different from that in *A. thaliana*, which may explain the smaller number of gbM genes in *B. rapa*.

In *A. thaliana*, H3K9me2 methylation is associated with gene silencing, especially in pericentromeric regions.[5] Although the number

of genes having H3K9me2 peaks in exon and intron regions was small, the average expression levels in genes having H3K9me2 in the exon and intron regions were lower than the total genes. However, the proportion of non-expressed genes having H3K9me2 was lower than genes having CG, CHG, and CHH methylation or MeDIP-peaks,[30] suggesting that in *B. rapa* some genes that were expressed at a lower level but not completely silenced, had H3K9me2.

In *A. thaliana*, TEs located in pericentromeric regions are associated more with high levels of H3K9me2 than in euchromatic arms. In this study, reads mapped to IRRs from ChIP-seq using H3K9me2 antibody (47%) was higher than that of Input-DNA-seq (39%), but this percentage is lower than that of MeDIP-seq (64%),[30] suggesting that H3K9me2 shows moderate association with IRRs in *B. rapa*. In *A. thaliana*, siRNA clusters were associated with H3K9me2 in pericentromeric regions, but generally not in euchromatic arms.[5] In maize, H3K9me2 levels at RdDM loci were similar to genic loci and 2-fold lower than non-RdDM loci.[48] In *B. rapa*, an association between H3K9me2 and 24 nt-siRNA clusters was not detected. We could not examine the relationship between H3K9me2 and 24 nt-siRNA clusters or IRRs in pericentromere or centromere regions due to the limitation of available reference genome sequence information. At least in the euchromatic regions of *B. rapa*, H3K9me2 is moderately associated with IRRs, but the 24 nt-siRNA clusters did not associate with H3K9me2. However, immunostaining using H3K9me2 antibody in interphase nuclei revealed H3K9me2 was present in heterochromatin of chromocenters in *B. rapa*, similar to the results in *A. thaliana*,[51,52] suggesting that H3K9me2 might be associated with IRRs or 24 nt-siRNA clusters in pericentromere or centromere regions in *B. rapa*.

Acknowledgements

We thank Etsuko Itabashi, Kiyomi Imamura, Makiko Tosaka, Taiji Kikuchi, and Terumi Horiuchi for their technical assistance. This work was supported by MEXT KAKENHI (No. 221S0002), by an Open Partnership Joint Projects of JSPS Bilateral Joint Research Projects (14544567), and by Grant-in-Aid for Young Scientists (B) (2478002), Scientific Research on Innovative Areas (24113509), and Grant-in-Aid Young Scientists (A) (15H05614) (JSPS) to R. Fujimoto, and grants from RIKEN and Japan Science and Technology Agency (JST) [Core Research for Evolutionary Science and Technology (JPMJCR13B4) (CREST)] to M.S.

Accession numbers

All sequence data have been deposited with DDBJ under DRA003120, DRA003121, and DRA003122.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at *DNARES* online.

References

1. Fujimoto, R., Sasaki, T., Ishikawa, R., Osabe, K., Kawanabe, T. and Dennis, E. S. 2012, Molecular mechanisms of epigenetic variation in plants, *Int. J. Mol. Sci.*, **13**, 9900–22.
2. Quadrana, L. and Colot, V. 2016, Plant transgenerational epigenetics, *Annu. Rev. Genet.*, **50**, 467–91.
3. Fuchs, J., Demidov, D., Houben, A. and Schubert, I. 2006, Chromosomal histone modification patterns-from conservation to diversity, *Trends Plant Sci.*, **11**, 199–208.
4. Xiao, J., Lee, U. S. and Wagner, D. 2016, Tug of war: adding and removing histone lysine methylation in *Arabidopsis*, *Curr. Opin. Plant Biol.*, **34**, 41–53.
5. Bernatavichute, Y. V., Zhang, X., Cokus, S., Pellegrini, M. and Jacobsen, S. E. 2008, Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*, *PLoS One*, **3**, e3156.
6. Du, J., Johnson, L. M., Jacobsen, S. E. and Patel, D. J. 2015, DNA methylation pathways and their crosstalk with histone methylation, *Nat. Rev. Mol. Cell Biol.*, **16**, 519–32.
7. Becker, C., Hagmann, J., Müller, J., et al. 2011, Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome, *Nature*, **480**, 245–9.
8. Schmitz, R. J., Schultz, M. D., Lewsey, M. G., et al. 2011, Transgenerational epigenetic instability is a source of novel methylation variants, *Science*, **334**, 369–73.
9. Osabe, K., Sasaki, T., Ishikawa, R. and Fujimoto, R. 2012, The role of DNA methylation in plants. In: Tatarinova, T.V., Sablok, G. (eds) *DNA Methylation: Principles, Mechanisms and Challenges*, NOVA Publishers: U.S.A, New York, Hauppauge, pp. 35–66.
10. Zemach, A., Kim, M. Y., Hsieh, P. H., et al. 2013, The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin, *Cell*, **153**, 193–205.
11. Stroud, H., Do, T., Du, J., et al. 2014, Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*, *Nat. Struct. Mol. Biol.*, **21**, 64–72.
12. Matzke, M. A. and Mosher, R. A. 2014, RNA-directed DNA methylation: an epigenetic pathway of increasing complexity, *Nat. Rev. Genet.*, **15**, 394–408.
13. Laird, P. W. 2010, Principles and challenges of genome-wide DNA methylation analysis, *Nat. Rev. Genet.*, **11**, 191–203.
14. Schmitz, R. J., Schultz, M. D., Urich, M. A., et al. 2013, Patterns of population epigenomic diversity, *Nature*, **495**, 193–8.
15. Kawakatsu, T., Huang, S. C., Jupe, F., et al. 2016, Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions, *Cell*, **166**, 492–505.
16. Niederhuth, C. E., Bewick, A. J., Ji, L., et al. 2016, Widespread natural variation of DNA methylation within angiosperms, *Genome Biol.*, **17**, 194.
17. Vidalis, A., Živković, D., Wardenaar, R., Roquis, D., Tellier, A. and Johannes, F. 2016, Methylome evolution in plants, *Genome Biol.*, **17**, 264.
18. Lister, R., O'Malley, R. C., Tonti-Filippini, J., et al. 2008, Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*, *Cell*, **133**, 523–36.
19. Cokus, S. J., Feng, S., Zhang, X., et al. 2008, Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning, *Nature*, **452**, 215–9.
20. Fujimoto, R. and Nishio, T. 2007, Self-incompatibility, *Adv. Bot. Res.*, **45**, 139–54.
21. Saeki, N., Kawanabe, T., Ying, H., et al. 2016, Molecular and cellular characteristics of hybrid vigour in a commercial hybrid of Chinese cabbage, *BMC Plant Biol.*, **16**, 45.
22. Wang, X., Wang, H., Wang, J., et al. 2011, The genome of the mesopolyploid crop species *Brassica rapa*, *Nat. Genet.*, **43**, 1035–9.
23. Cheng, F., Sun, R., Hou, X., et al. 2016, Subgenome parallel selection is associated with morphotype diversification and convergent crop domestication in *Brassica rapa* and *Brassica oleracea*, *Nat. Genet*, **48**, 1218–24.
24. Cheng, F., Wu, J., Fang, L., et al. 2012, Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*, *PLoS One*, **7**, e36442.
25. Tang, H., Woodhouse, M. R., Cheng, F., et al. 2012, Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy, *Genetics*, **190**, 1563–74.
26. Kawamura, K., Kawanabe, T., Shimizu, M., et al. 2016, Genetic distance of inbred lines of Chinese cabbage and its relationship to heterosis, *Plant Gene*, **5**, 1–7.

27. Krueger, F. and Andrews, S. R. 2011, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications, *Bioinformatics*, **27**, 1571–2.
28. Takuno, S. and Gaut, B. S. 2012, Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly, *Mol. Biol. Evol*., **29**, 219–27.
29. Kawanabe, T., Osabe, K., Itabashi, E., Okazaki, K., Dennis, E. S. and Fujimoto, R. 2016, Development of primer sets that can verify the enrichment of histone modifications, and their application to examining vernalization-mediated chromatin changes in *Brassica rapa* L, *Genes Genet. Syst*., **91**, 1–10.
30. Takahashi, S., Fukushima, N., Osabe, K., et al. 2018, Identification of DNA methylated regions by using methylated DNA immunoprecipitation sequencing in *Brassica rapa*, *Crop Pasture Sci*., **69**, 107–20.
31. Shimizu, M., Fujimoto, R., Ying, H., et al. 2014, Identification of candidate genes for fusarium yellows resistance in Chinese cabbage by differential expression analysis, *Plant Mol. Biol*., **85**, 247–57.
32. Kawanabe, T., Fujimoto, R., Sasaki, T., Taylor, J. M. and Dennis, E. S. 2012, A comparison of transcriptome and epigenetic status between closely related species in the genus *Arabidopsis*, *Gene*, **506**, 301–9.
33. Chen, X., Ge, X., Wang, J., Tan, C., King, G. J. and Liu, K. 2015, Genome-wide DNA methylation profiling by modified reduced representation bisulfite sequencing in *Brassica rapa* suggests that epigenetic modifications play a key role in polyploid genome evolution, *Front. Plant Sci*., **6**, 836.
34. Fujimoto, R., Sasaki, T., Inoue, H. and Nishio, T. 2008, Hypomethylation and transcriptional reactivation of retrotransposon-like sequences in *ddm1* transgenic plants of *Brassica rapa*, *Plant Mol. Biol*., **66**, 463–73.
35. Sasaki, T., Fujimoto, R., Kishitani, S. and Nishio, T. 2011, Analysis of target sequences of DDM1s in *Brassica rapa* by MSAP, *Plant Cell Rep*., **30**, 81–8.
36. Zhang, X., Yazaki, J., Sundaresan, A., et al. 2006, Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*, *Cell*, **126**, 1189–201.
37. Li, X., Zhu, J., Hu, F., et al. 2012, Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression, *BMC Genomics*, **13**, 300.
38. Yang, H., Chang, F., You, C., et al. 2015, Whole-genome DNA methylation patterns and complex associations with gene structure and expression during flower development in Arabidopsis, *Plant J*., **81**, 268–81.
39. Brenet, F., Moh, M., Funk, P., et al. 2011, DNA methylation of the first exon is tightly linked to transcriptional silencing, *PLoS One*, **6**, e14524.
40. Schmitz, R. J., He, Y., Valdés-López, O., et al. 2013, Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population, *Genome Res*., **23**, 1663–74.
41. Seymour, D. K., Koenig, D., Hagmann, J., Becker, C. and Weigel, D. 2014, Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization, *PLoS Genet*., **10**, e1004785.
42. Qiu, D., Gao, M., Li, G. and Quiros, C. 2009, Comparative sequence analysis for *Brassica oleracea* with similar sequences in *B. rapa* and *Arabidopsis thaliana*, *Plant Cell Rep*., **28**, 649–61.
43. Takuno, S. and Gaut, B. S. 2013, Gene body methylation is conserved between plant orthologs and is of evolutionary consequence, *Proc. Natl. Acad. Sci. U. S. A*., **110**, 1797–802.
44. Takuno, S., Ran, J. H. and Gaut, B. S. 2016, Evolutionary patterns of genic DNA methylation vary across land plants, *Nat. Plants*, **2**, 15222.
45. Bewick, A. J., Ji, L., Niederhuth, C. E., et al. 2016, On the origin and evolutionary consequences of gene body DNA methylation, *Proc. Natl. Acad. Sci. U. S. A*., **113**, 9111–6.
46. Ebbs, M. L. and Bender, J. 2006, Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase, *Plant Cell*, **18**, 1166–76.
47. Jackson, J. P., Lindroth, A. M., Cao, X. and Jacobsen, S. E. 2002, Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase, *Nature*, **416**, 556–60.
48. Gent, J. I., Madzima, T. F., Bader, R., et al. 2014, Accessible DNA and relative depletion of H3K9me2 at maize loci undergoing RNA-directed DNA methylation, *Plant Cell*., **26**, 4903–17.
49. Fujimoto, R., Sasaki, T. and Nishio, T. 2006, Characterization of DNA methyltransferase genes in *Brassica rapa*, *Genes Genet. Syst*., **81**, 235–42.
50. Dong, H., Liu, D., Han, T., et al. 2015, Diversification and evolution of the *SDG* gene family in *Brassica rapa* after the whole genome triplication, *Sci. Rep*., **5**, 16851.
51. Braszewska-Zalewska, A., Bernas, T. and Maluszynska, J. 2010, Epigenetic chromatin modifications in *Brassica* genomes, *Genome*, **53**, 203–10.
52. Braszewska-Zalewska, A., Dziurlikowska, A. and Maluszynska, J. 2012, Histone H3 methylation patterns in *Brassica nigra*, *Brassica juncea*, and *Brassica carinata* species, *Genome*, **55**, 68–74.