OXFORD

Full Paper

Differential DNA methylation and gene expression in reciprocal hybrids between *Solanum lycopersicum* and *S. pimpinellifolium*

Muhammad Ammar Raza[1,†], Ningning Yu[1,†], Dan Wang[1], Liwen Cao[1], Susheng Gan[1,2], and Liping Chen[1,*]

[1]Department of Horticulture, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, P. R. China, and [2]Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY, USA

*To whom correspondence should be addressed. Tel. +86-571-88982006. Fax. +86-571-88982006.
Email: chenliping@zju.edu.cn
[†]These authors contributed equally to this work.

Edited by Prof. Kazuhiro Sato

Abstract

Wide hybridization is a common and efficient breeding strategy for enhancing crop yield and quality. An interesting phenomenon is that the reciprocal hybrids usually show different phenotypes, and its underlying mechanism is not well understood. Here, we reported our comparative analysis of the DNA methylation patterns in *Solanum lycopersicum*, *Solanum pimpinellifolium* and their reciprocal hybrids by methylated DNA immunoprecipitation sequencing. The reciprocal hybrids had lower levels of DNA methylation in CpG islands and LTR retroelements when compared with those of their parents. Importantly, remarkable differences in DNA methylation patterns, mainly in introns and CDS regions, were revealed between the reciprocal hybrids. These different methylated regions were mapped to 79 genes, 14 of which were selected for analysis of gene expression levels. While there was an inverse correlation between DNA methylation and gene expression in promoter regions, the relationship was complicated in gene body regions. Further association analysis revealed that there were 15 differentially methylated genes associated with siRNAs, and that the methylation levels of these genes were inversely correlated with respective siRNAs. All these data raised the possibility that the direction of hybridization induced the divergent epigenomes leading to changes in the transcription levels of reciprocal hybrids.

Key words: tomato, reciprocal hybridization, MeDIP-seq, siRNAs

1. Introduction

Wide hybridization or distant crossing refers to the inter-specific or above-specific hybridization.[1] It is a useful breeding strategy that has been used for crop improvement by transferring many desired traits from wild species to crops.[2–4] The hybridization involves the selection of maternal and paternal parents, and often one parent can be maternal or paternal, and the other paternal or maternal, respectively, which is referred to as reciprocal crossings. It is well known that the reciprocal hybrids share identical nuclei but have different maternal cytoplasms. Therefore, the major difference between the reciprocal hybrids is the source of their cytoplasms.

Reciprocal hybrids often show different phenotypes. A good example is crosses between *Arabidopsis thaliana* and *A. arenosa*. When *A. thaliana* was used as the maternal parent and *A. arenosa* as the

paternal parent, many viable seeds were produced. In contrast, the reciprocal cross did not yield any viable seeds.[5] Several genetic mechanisms have been proposed to explain the differences between the reciprocal hybrids. For example, maternal effects model is defined as the causal influence of the maternal genotype or phenotype on the offspring phenotype,[6] while the parent-origin effects model hypothesizes that female and male gametes contribute different sets of active alleles or else different complements of gene products to the seed.[7] Cytoplasmic inheritance.[8,9] and cytoplasmic-nuclear interaction[10] models mainly elucidate the interactions between chloroplast and mitochondrial genomes and the exogenous nuclear genome on the influence of cytoplasmic inheritance, respectively. Gene imprinting,[11] the dominance model, over-dominant effects and epistasis hypothesis[12–14] have also been proposed to explain the phenomenon. However, the underlying biological mechanisms accounting for the differences between reciprocal hybrids remain to be deciphered. Because the main differences between the reciprocal hybrids are in their cytoplasms, it is reasonable to assume that the most essential question is how the cytoplasm influences gene expression. Recent studies have pointed out that gene expression can be regulated by epigenetic modification, especially DNA methylation.[15] Therefore, whether different cytoplasms can lead to the changes in methylation profile, and then affect gene expression levels is the key point to explain the phenotypic differences in reciprocal hybrids.

Cytoplasm is where many biochemical reactions take place, including the formation of mature small RNAs (sRNAs) such as miRNAs and siRNAs that are often involved in regulation of gene expression.[16,17] miRNAs are a class of small non-coding RNA molecules that regulate eukaryotic gene expression by cleaving their respective target genes at the posttranscriptional level. They specifically bind to mRNAs based on sequence pairing, leading to the degradation of their target mRNAs.[18,19] In reciprocal crossings between *A. thaliana* ecotypes Ler and C24, researchers observed increased levels of 21-nt miRNAs in both hybrids compared with those in their parents.[20] siRNAs are a class of double stranded RNA molecules, 20–25 bp in length, and may direct chromatin modification through RNA-directed DNA methylation (RdDM) and transposon silencing.[21–23] Changes in 24-nt siRNA levels in two *A. thaliana* ecotypes showed correlated with changes in DNA methylation, creating allelic variants that might contribute to the phenotypic heterosis.[24] An increase in genome-wide DNA methylation in *A. thaliana* reciprocal hybrids, possibly due to RdDM pathway, might alter circadian rhythms and gene expression in F1 hybrids.[20] The maternal siRNAs as regulators of parental genome imbalance and gene expression could be transmitted to hybrids and regulate the seed size in reciprocal crosses.[25] It was also reported that there were different profiles of 24-nt sRNAs in reciprocal hybrids, which might lead to the diversity of phenotypes.[26] However, such questions as what the differences of DNA methylation profiles between two reciprocal hybrids are, whether the differences are induced by cytoplasmic siRNAs, and whether or how they regulate gene expression and subsequent phenotypes remain elusive.

In our previous studies, reciprocal crossings between cultivated species *Solanum lycopersicum* cv. Micro-Tom (for simplicity hereafter M was used for Micro-Tom) and wild type *S. Pimpinellifolium* line WVa700 (W was similarly used for WVa700) were performed, and the resulting reciprocal hybrids (M × W and W × M were similarly used for Micro-Tom_WVa700 and WVa700_Micro-Tom, respectively) displayed different phenotypes in terms of plant height (43.79 cm for M × W and 41.89 cm for W × M), fruit shape index (0.95 for M × W and 0.92 for W × M) and single fruit weight (3.54 g for M × W and

4.40 g for W × M).[26] To unveil the underlying possible epigenetic mechanisms, we profiled, using methylated DNA immunoprecipitation sequencing (MeDIP-seq) method, and compared the DNA methylation patterns of the entire genomes of M, W and their reciprocal hybrids (M × W and W × M). We found that both reciprocal hybrids had lower methylation levels in CpG islands (CGIs) and major types of retroelements than their parents, suggesting that the crossing process per se may reduce the methylation levels of reciprocal hybrids in these regions. Importantly, the two reciprocal hybrids showed different DNA methylation patterns across their entire genomes, with major differences in gene body regions. We found 79 different methylated genes (DMGs) between the reciprocal hybrids, and 14 of which were randomly selected for analyses of their expression levels using qRT-PCR. We found that DNA methylation correlated negatively with gene expression in the promoter regions, while in gene body regions the correlation was complicated. Further association analysis of DNA methylation with siRNAs expression revealed that the methylation levels mainly inversely correlated with the siRNAs expression levels in gene body regions of the related genes. Our research suggested that the direction of hybridization induced the differences of epigenomes in the reciprocal hybrids. And the differences of DNA methylation profiles might be modified by siRNAs existed in maternal cytoplasms, leading to the distinct gene expression levels between the reciprocal hybrids.

2. Materials and methods

2.1. Plant materials

Inbred lines *of S. lycopersicum* cv. Micro-Tom (for simplicity, we used "M" for Micro-Tom in this study) (2n = 24) and *S. pimpinellifolium* line WVa700 (similarly "W" for WVa700) (2n = 24) were used as parents, and reciprocal crosses between the parental lines were made to produce two different types of hybrids M × W (i.e., Micro-Tom × WVa700) and W × M (WVa700 × Micro-Tom), respectively. A total of 80 tomato plants, with 20 plants for each of the four genotypes (W, M, W × M and M × W), were grown in a greenhouse at 23 °C with a photoperiod of 16-h light and 8-h dark and 60% relative humidity.

2.2. Extraction, 5mC antibody immunoprecipitation and sequencing of DNA

Total genomic DNA was extracted from 45-day-old leaves of the reciprocal hybrids and their parent plants using CTAB method as described by Murray and Thompson.[27] At this stage, the differences in leaf phenotype and plant morphology among the four genotypes of tomato plants became evident. Three seedlings were pooled in each genotype sample for genomic DNA extraction. The quality of DNA was analysed using agarose gel electrophoresis. Five μg DNA was then randomly fragmented into small pieces with most of them 100–500 bp in length and purified by QIA quick PCR Purification Kit (Qiagen, Hilden, Germany). Four DNA libraries were constructed, namely the M, W, M × W and W × M DNA libraries. After DNA-end repair and 3'-dA overhang, DNA was ligated to an Illumina sequencing primer adaptor. The double-stranded DNA was subsequently denatured and immunoprecipitated using 5mC antibody. Real time quantitative PCR analysis was performed to validate the quality of immunoprecipitated DNA fragments. DNA with high quality was used for PCR amplification and bands between 200-300 bp were selected, and subsequently sequenced using the HiSeq 2000 system (SBS KIT-HS V3, Illumina).

2.3. Bioinformatics analysis

Sequence data were obtained through filtering out the adapter, reads containing more than 10% N and low quality reads. MeDIP-Seq data was mapped to reference genome (ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG2.3_release/ITAG2.3_genomic.fasta (12 June 2015, data last accessed)) by soapaligner, only unique alignments with no more than 2 mismatches were chosen for further analyses. The uniquely mapped reads were used to analyse the distribution of reads in genome chromosomes and different genomic regions. Peak scanning was performed using MACS (version 1.4.0) to identify genomic regions that were enriched with methylated reads. The scan graphs of two samples were merged to identify putative different methylation regions. The read numbers of individual methylated region from any two of the four genotypes were used to calculate the normalized log2 values and P-value, with chi-square statistics and a false discovery rate (FDR) statistics. The filtering criteria for P-value and the fold change were ≤ 0.05 and ≥ 2, respectively. DMRs were divided into two groups: the hypo-methylated (down) DMRs and the hyper-methylated (up) DMRs. The genes overlapped with DMRs were deemed as DMGs. DMGs were functionally clustered via Gene Ontology analysis to explore their biological functions, using the DAVID functional annotation tool with $P \leq 0.05$. KEGG pathway enrichment analysis was carried out to identify significant enriched metabolic pathway or signal transduction pathway. The calculation formula of analysis is the same as that of GO function analysis. Q-value ≤ 0.05 was used as the threshold to determine significant enrichment pathway in DEGs.

2.4. Bisulfite sequencing

To validate the results of MeDIP-seq, the bisulfite sequencing polymerase chain reaction was performed using the same genomic DNA materials. Briefly, 2 µg DNA was treated with sodium bisulfate using the EZ DNA Methylation Gold Kit according to the manufacturer's instruction (Zymo Research, Orange Country, US) and then incubated at 98 °C for10 minutes and 64 °C for 2.5 hours. 10 pairs of primers for selected genes were designed by the MethPrimer online software (http://www.urogene.org/methprimer) and were shown as Supplementary Table S1. The bisulfite-treated DNA was PCR amplified by with these primers. PCR products were purified with the Wizard® SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA), and then cloned into the pMD18-T vector (Takara, Dalian, China) and sequenced (Sunny, Shanghai, China). At least 15 clones were sequenced for each sample. The methylation level was calculated by the ratio of unconverted cytosines over the total cytosines, including three types: CG, CHG and CHH. The final bisulfate sequencing results were analysed according to the Kismeth website (http://Katahdin.mssm.edu/Kismeth (8 January 2016, data last accessed)).[28]

2.5. Quantitative real-time PCR

Total RNA from the 45-day-old leaves of the two reciprocal hybrids was extracted using the Trizol method per the manufacturer's instruction (Invitrogen, Carlsbad, CA) and used to construct small RNA libraries.[26] Three biological replicates were included for all the samples. The first strand cDNA was synthesized with oligo (dT) primer from 2 µg RNA using the Reverse Transcriptase M-MLV kit (Takara). Primers used for real-time PCR were listed in Supplementary Table S1. The real-time PCR was performed using a SYBR® Premix Ex Taq™ (Takara) on the ABI STEPONE Real-Time PCR System. Relative quantification was performed using the

CT ($2^{-\triangle\triangle Ct}$) method. Each sample was performed in triplicate for real-time PCR reaction.

2.6. Small RNA sequencing and analysis of siRNAs

Small RNA high-throughput sequencing of four libraries (representing M, W, W × M and M × W genotypes, respectively) was as previously described.[26] The small RNA-seq clean data have been deposited in the SRA database of NCBI with accession number SRX722032, SRX722033, SRX722034, and SRX722035. The leaf materials used for the small RNA libraries were same as those for DNA extraction and MeDIP-seq. After eliminating the contaminant reads and low quality reads, tags were used to predict siRNA using tag2siRNA software.[29] The clean reads were analysed by length distribution and common sequences. Then, the siRNAs sequences were matched to the DMGs via blast searching of NCBI databases (http://blast.ncbi.nlm.nih.gov/Blast.cgi (18 August 2016, data last accessed)).

3. Results

3.1. Genome-wide DNA methylation profiles of two parental lines and their reciprocal hybrids

Whole genome-wide DNA methylation patterns of two parental lines (M and W) and their reciprocal hybrids (M × W and W × M) were profiled using MeDIP-seq. The MeDIP-seq data have been deposited in the SRA database of NCBI with accession number SRR5397373, SRR5397378, SRR5397384 and SRR5397386. We obtained up to 4.5 Gb of MeDIP-seq data for each genotype (Table 1). After removal of adapter sequences, contaminations, and low-quality reads, more than 90 million reads, regarded as clean data, from each genotype were analysed. Total reads were mapped to the reference genome (ftp://ftp.solgenomics.net/tomato_genome/annotation/ITAG2.3_release/ITAG2.3_genomic.fasta (12 June 2015, data last accessed)), with mapping rates higher than 96% for individual genotypes. Reads were distributed on each chromosome region, with good genome coverage (see Supplementary Fig. S1). 57.3–58.79% of the mapped reads were uniquely mapped to reference genome for each genotype (Table 1). The distribution of MeDIP-seq reads in different CpG density regions was shown, and the majority of the reads in all samples were located in regions with 5 to 25 CpGs (see Supplementary Fig. S2).

We further analysed the methylation enrichment in different components of genome. CGIs are mainly located in promoter regions and play important roles in the epigenetic regulation of gene expression. Therefore, we compared the methylation status of CGIs and CGI shores (spanning 2kb fragment up- and down-stream of each CGI). We found that in both parents the CGIs had more methylation reads than the CGI shores had. In contrast, CGIs did not enrich much more methylated reads than the CGI shores enriched in the reciprocal hybrids (Fig. 1a). In particular, CGI shores had more methylated reads than CGIs had in M × W. More interestingly, the methylation levels of CGIs were lower in reciprocal hybrids than in their parents, suggesting that the process of hybridization led to the decrease in DNA methylation levels in these regions. Furthermore, W × M had higher methylation levels than M × W in CGIs (Fig. 1a), which was one of the major differences between the reciprocal hybrids in DNA methylation profiles.

We also analysed the distribution of the reads in intragenic region and nearby 2kb sequences, and found that 2kb upstream region showed higher methylation levels than other regions in all four genotypes (parents and reciprocal hybrids). The DNA methylation levels decreased sharply before intragenic region, indicating that promoter regions were hypermethylated and might play a key role in the modification of DNA

Table 1. General information of read alignment and peak scanning in four samples

	Parents		Reciprocal hybrids	
Sample	Micro-Tom	WVa700	Micro-Tom_WVa700	WVa700_Micro-Tom
Data Size (Gb)	4.5	4.6	4.6	4.6
Reads Number (No.)	91836734	93207132	93085356	92982838
Mapped Reads (No.)	88334928	90089549	89689126	89518553
Mapping Rate (%)	96.19	96.66	96.35	96.27
Effective Chain Depth[a]	0.53	0.54	0.53	0.53
Unique Mapped Bases (No.)	53987715	54363987	54273327	53301136
Unique Mapped Rate (%)[b]	58.79	58.33	58.30	57.32
Peak Number (No.)	55100	50316	48935	48179
Average Peak Length (bp)	1654	1725	1725	1765
Peak Median Length (bp)	1442	1500	1506	1543
Peak Total Length (bp)	91129487	86803207	84405012	85016601
Peak Coverage (%)	11.66	11.10	10.80	10.88

[a]Effective Chain Depth = Mapped bases/the Size of the Reference Genome.
[b]Unique Mapping Rate = Unique Mapped Reads Count/Total Reads.
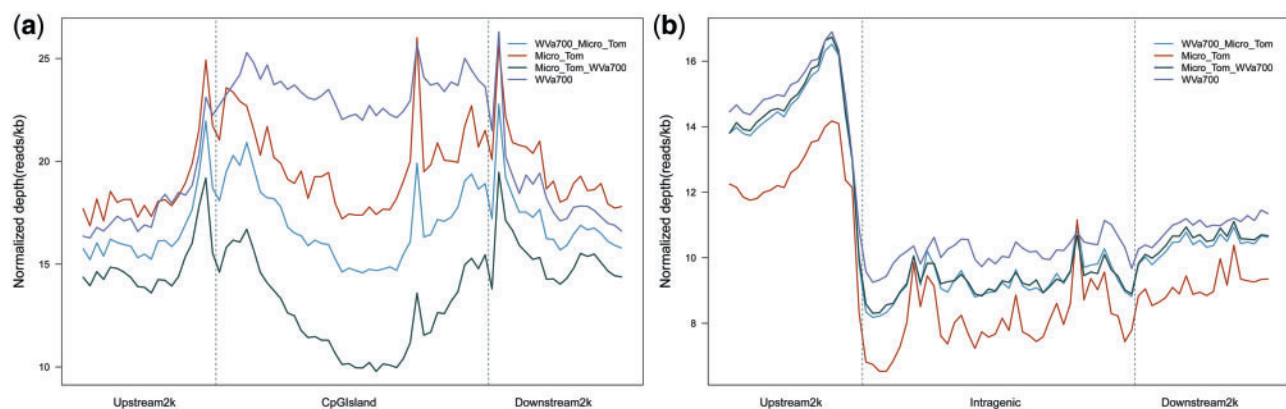

Figure 1. Distribution of reads around (a) CpG islands and (b) intragenic regions. The X axis indicates the position around CpG islands (intragenic regions) and the Y axis indicates the normalized read number. These panels can reflect the methylation levels around CpG islands and gene body regions. (See online article for colour version of this figure).

methylation (Fig. 1b). Interestingly, W (wild species) and the two reciprocal hybrids showed similar methylation levels in promoter regions, while M (cultivated species) had the lowest methylation levels.

Considering that methylation of repeat elements was usually associated with genomic instability through structural changes such as transposition, translocation, and recombination,[30] we compared the distribution of reads in such different repeat elements as LTR (long terminal repeat)/Gypsy, LTR/Copia, DNA/CMC-Enspm and SINE (short interactive nuclear element)/SINE. These four main repeat elements had the vast majority of the reads in all four samples (Table 2), and the LTR retroelements were especially rich with more than 90% reads, suggesting that such retroelements may play an important part in DNA methylation modification. It is noteworthy that the methylation levels in the major type of LTR retroelements of the two reciprocal hybrids were different, and lower than those of their parents (Table 2).

3.2. Divergent DNA methylation patterns in the reciprocal hybrids

Our previous research showed that the two reciprocal hybrids displayed significantly different phenotypes in terms of fruit shape

index, single fruit weight and plant height.[26] We here investigated DNA methylation patterns of the reciprocal hybrids to explore the possibility that these phenotypic changes were associated with or induced by changes in DNA methylation. Peak scanning software (MACS 1.4.0) was used to identify reads-enriched regions. Methylated peak, an important parameter in analysing DNA methylation profile,[31,32] was used to detect the highly methylated regions (HMRs) and only uniquely mapped reads were utilized in the whole genome peak scanning. As shown in Table 1, 48935 and 48179 peaks were identified in the reciprocal hybrids W × M and M × W samples, respectively. The methylation levels in the 2kb region upstream of 5'UTR, in the gene body (including 5'UTR, coding sequence (CDS) region, intron, and 3'UTR) and in the 2kb downstream region of 3'UTR in the genomes of the two hybrids were further analysed. The ratio of peaks was used to represent the methylation density of individual specific genome elements. As shown in Fig. 2, the 2kb upstream regions had the most abundant peaks, followed by the CDS regions and the 2kb downstream regions. Notably, the frequencies of the methylated peaks in both CDS region and intron were higher in W × M (with 13.32% in the CDS region and 9.02% in intron) than in M × W (12.85%, 8.72%) (Fig. 2).

Table 2. Proportion of peaks of the major repeat elements in four samples

Sample	Parents		Reciprocal hybrids	
	Micro-Tom	WVa700	Micro-Tom_WVa700	WVa700_Micro-Tom
LTR[a]/Gypsy (%)	69.95	67.22	66.42	67.01
LTR/Copia (%)	22.21	23.51	23.67	23.53
DNA/CMC-EnSpm (%)	1.61	1.99	2.10	2.04
SINE[b]-SINE (%)	1.33	1.73	1.85	1.81

[a]LTR: long terminal repeat
[b]SINE: short interactive nuclear element
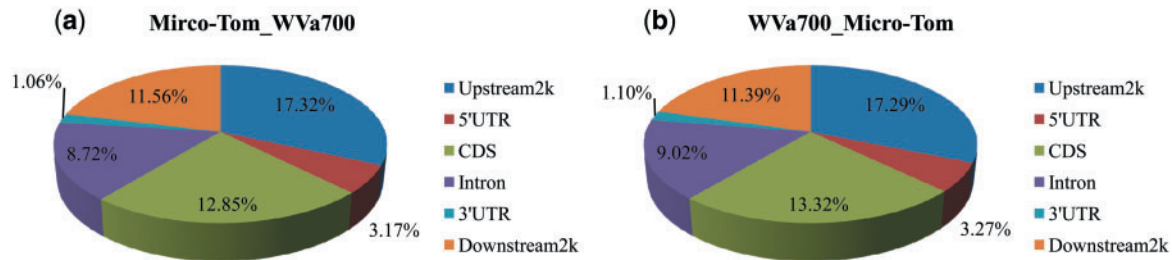
Figure 2. Distribution of peaks in different gene elements in the reciprocal hybrids, including 2kb upstream region, 5'UTR, CDS region, Intron, 3'UTR and 2kb downstream region. The proportion of peak was shown in each specific gene element. (a) Micro-Tom_WVa700 (i.e. M × W); (b) WVa700_ Micro-Tom (W × M). (See online article for colour version of this figure).

The differences between the two reciprocal hybrids as well as the two parents were further investigated by searching differentially methylated regions (DMRs), and the DMRs data-based whole-genome methylomes of M × W and W × M were shown in Fig. 3 There were 109 DMRs (between the two reciprocal hybrids) and 3208 DMRs (between the two parents), obviously many more differences between the parents than those between the two reciprocal hybrids. Comparison of gene methylation status between the two reciprocal hybrids revealed that there were 60 hypermethylated and 49 hypomethylated DMRs in six gene elements of W × M. Most of the DMRs were observed in introns (31.19%) and CDS regions (23.85%) (Table 3). Interestingly, more than one DMR were found in some of the DMGs. For examples, Solyc03g025770.2 gene was hypermethylated in the 2kb upstream region and CDS region, and Solyc00g090430.2 gene was hypomethylated in the 2kb upstream region, 3'UTR, intron and CDS region in W × M. Excluding repeat regions, there were 79 genes with different methylation patterns between the two reciprocal hybrids.

3.3. Gene ontology (GO) and KEGG pathway analysis of DMGs

In this study, we referred any gene with methylation peaks in its promoter or gene body regions as a methylated gene. As mentioned above, a total of 79 DMGs were found upon comparison of the two reciprocal hybrids, including 41 hypermethylated genes and 38 hypomethylated genes in W × M. These differentially methylated genes were annotated, including a cell division protease ftsH, a TPR repeat protein, a histone-lysine n-methyltransferase, a b-zip transcription factor-like, among others (Supplementary Table S2). GO analysis revealed that the DMGs in M × W and W × M were functionally belonging to one or more following categories: biological process (BP), cellular component (CC), and molecular function (MF). These DMGs were further divided into 17 functional clusters, with 10 in BP, 5 in CC and 2 in MF. Interestingly, hypermethylated gene Solyc08g013860.2 in M × W

was related to mitochondrial part, participating in dicarboxylic acid metabolic process, and also hypermethylated in the parental line M than the other parental line W (Supplementary Table S3). Per the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, we observed that gene Solyc01g099150.2 and gene Solyc02g092780.1 were highly enriched in pathways: linoleic acid metabolism (Ko00591) and other types of O-glycan biosynthesis (Ko00514), with Q-value ≤ 0.05. Furthermore, two DMGs (Solyc05g056450.2 and Solyc08g 013860.2) were involved in fructose and mannose metabolism and in carbon fixation in photosynthetic pathways (Supplementary Table S4).

3.4. Validation of MeDIP-seq data using bisulfite sequencing

To validate the MeDIP-seq data, we performed bisulfite sequencing of ten randomly selected DMRs of associated genes which showed the same methylation levels between the two parents. While, compared with W × M, four DMRs of associated genes [Solyc00g 020840.2 (Intron), Solyc00g282510.1 (CDS), Solyc08g065650.1 (CDS) and Solyc00g020940.1 (Upstream 2k)] had higher methylation levels and six DMRs of associated genes [Solyc05g008030.1 (Intron), Solyc03g115180.1 (Intron), Solyc06g065730.2 (Intron) Solyc03g115180.1 (CDS), Solyc07g009320.2 (Intron), Solyc03g 025770.2.1 (CDS)] had lower methylation level in M × W, essentially the same as those results revealed by MeDIP-seq. The bisulfite sequencing results proved that methylation levels of both Solyc05g008030.1 (Intron) and Solyc03g115180.1 (Intron) were up-regulated in W × M, while methylation levels of Solyc00g282510.1 (CDS) and Solyc00g020840.2 (Intron) were down-regulated (Fig. 4). Solyc08g065650.1 (CDS) and Solyc00g020940.1 (Upstream 2k) had higher levels of methylation in M × W, while Solyc03g115180.1 (CDS), Solyc07g009320.2 (Intron), Solyc03g025770.2 (CDS) and Solyc06g065730.2 (Intron) had higher methylation levels in W × M (see Supplementary Fig. S3). Hence, the bisulfite sequencing results were
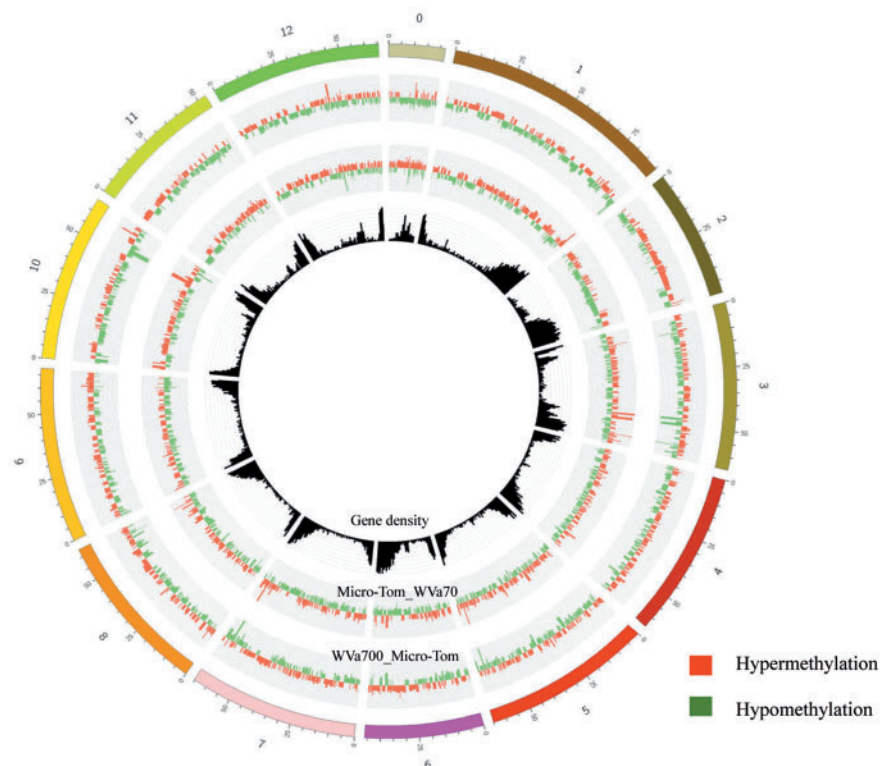
Figure 3. Comprehensive maps of the entire genome methylome of the reciprocal hybrids. Inner black circle represents gene density corresponding to each chromosome. Outer two circles represent the hyper (up-regulated) - and hypo (down-regulated) - methylated regions of the two reciprocal hybrids. (See online article for colour version of this figure).

Table 3. Numbers of differentially methylated genes between the two reciprocal hybrids in six gene elements

Contrast	Micro-Tom_WVa700VSWVa700_Micro-Tom (Hypermethylation)	Micro-Tom_WVa700VSWVa700_Micro-Tom (Hypomethylation)
CDS	17	9
Downstream 2k	12	12
5'UTR	0	1
Intron	19	15
3'UTR	2	1
Upstream 2k	10	11
Total	60	49

consistent with the MeDIP-seq data, demonstrating that the genome-wide DNA methylation data from MeDIP-seq was reasonably reliable.

3.5. Association analysis of gene expression and DNA methylation

To investigate the correlation between DNA methylation and gene expression, we randomly selected 14 DMGs, including seven genes with up-regulated methylation levels and the other seven genes with down-regulated methylation levels in M × W, for qRT-PCR analysis of their expression levels in the reciprocal hybrids (Fig. 5). Correlated with the different methylation patterns, the selected genes showed different expression levels between the two reciprocal hybrids. Specifically, in the M × W hybrid, three genes, namely Solyc04g 071320.2, Solyc02g094350.2 and Solyc07g062890.2, were hypermethylated in the 2kb region upstream of the gene body, correlated with lower levels of transcripts. In contrast, the gene named

Solyc05g018390.2 was hypomethylated in the 2kb region upstream of its gene body and had higher transcript levels. Notably, DNA methylation in the gene body regions is also mostly (but not always) inversely correlated with gene expression. Eight genes, namely Solyc05g 008030.1, Solyc06g065730.2, Solyc00g282510.1, Solyc01g111020.2, Solyc02g092780.1, Solyc03g115180.1, Solyc07g009320.2 and Solyc03g025770.2, exhibited inverse correlation between the DNA methylation levels and their gene expression levels in the gene body regions. For examples, Solyc01g111020.2 had higher methylation level in its intron and CDS region, but dramatically down-regulated expression level in W × M. Solyc02g092780.1 showed a lower methylation level and a higher expression level in W × M (Fig. 5 and Supplementary Table S2). However, the intron of Solyc06g051760.2 and both the intron and CDS of Solyc03g097460.2 were hypermethylated in M × W, but both genes had higher levels of transcripts (Fig. 5 and Supplementary Table S2). These results suggested that there may be a negative correlation between the DNA methylation levels in the
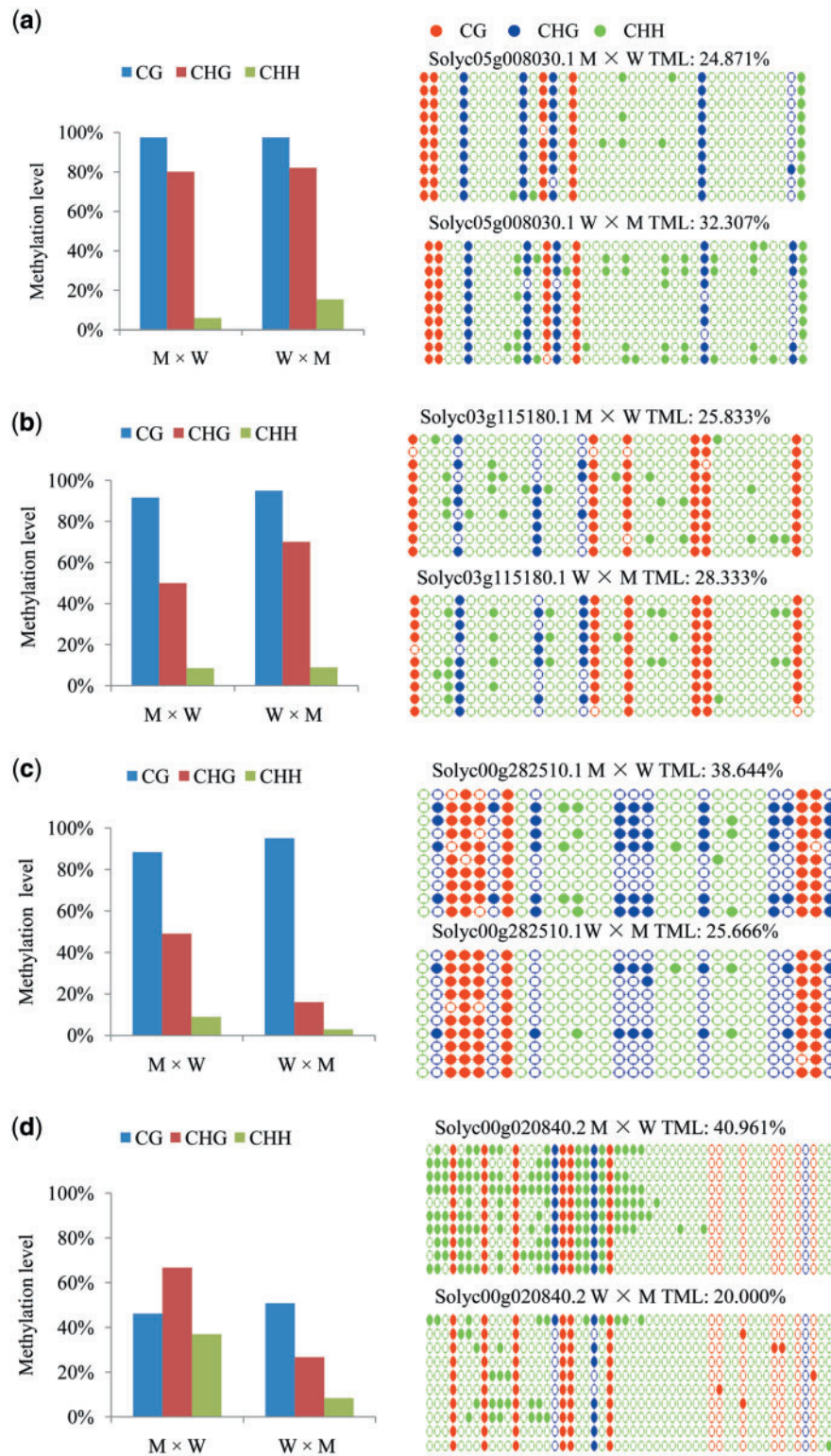
Figure 4. Validation of DNA methylation levels of four randomly selected DMR-associated genes as examples. (a) and (b) Bisulfite sequencing of hypomethy-lated genes (Solyc05g008030.1 and Solyc03g115180.1) in M × W; (c) and (d) Bisulfite sequencing of hypermethylated genes (Solyc00g282510.1 and Solyc00g020840.2) in M × W. M × W represents Micro-Tom_WVa700, and W × M represents WVa700_Micro-Tom. The Y axis indicates the methylation of each genotype. The dot plot shows only the cytosines as circles, colored according to the type of cytosine. The filled circles represent methylated cytosines and empty circles represent un-methylated cytosines. There were 10 positive clones for every primer pairs of each genotype. TML stands for total methylation level. (See online article for colour version of this figure).
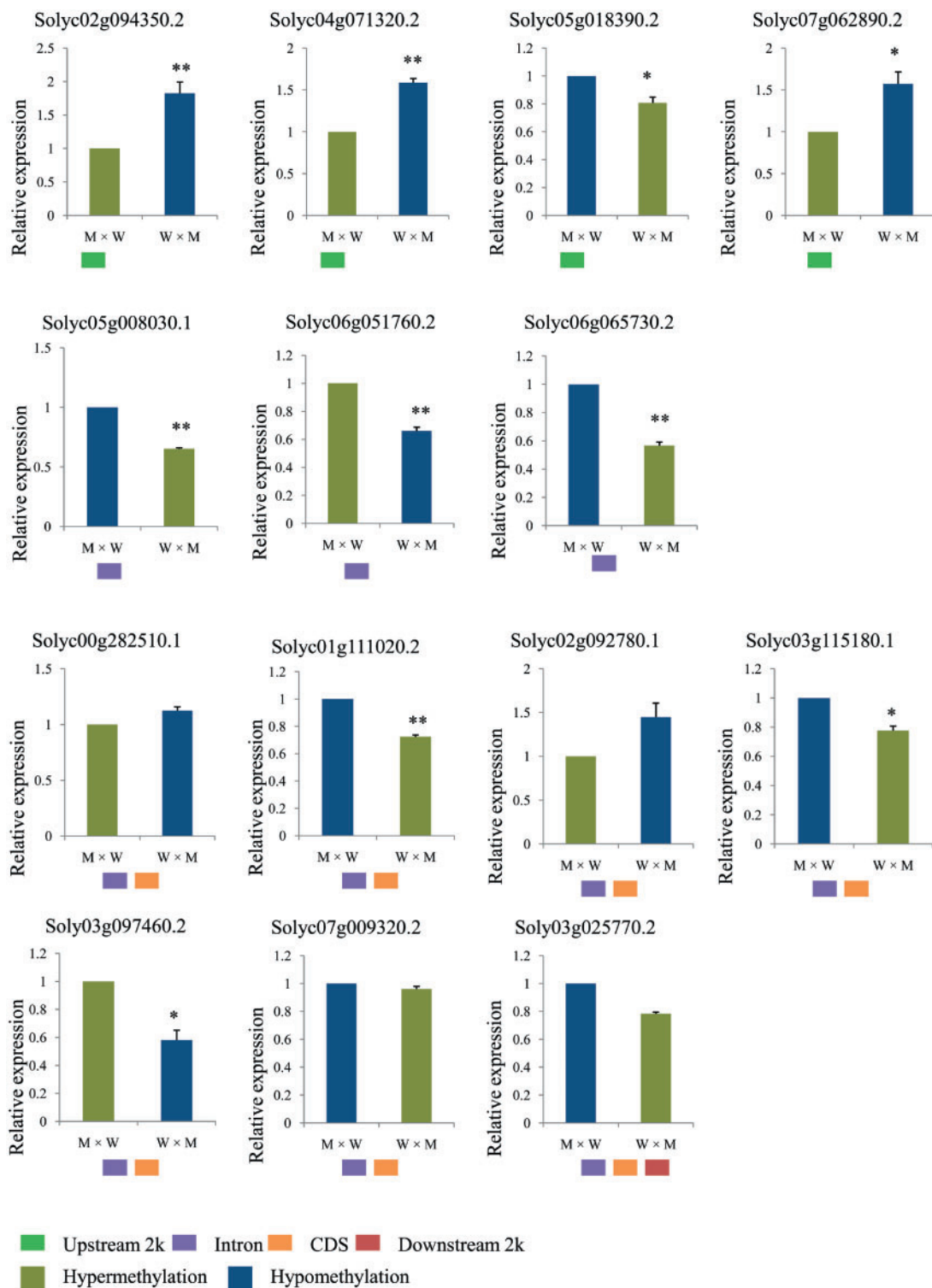
Figure 5. Relative gene expression levels of 14 randomly selected DMR-associated genes between the reciprocal hybrids. Asterisks indicate significant differences [Student's *t* test ($n = 3$)] between the reciprocal hybrids. *$P < 0.05$, **$P < 0.01$. The blocks with different colors indicate different gene elements. Error bars represent SD of triplicate experiments. (See online article for colour version of this figure).
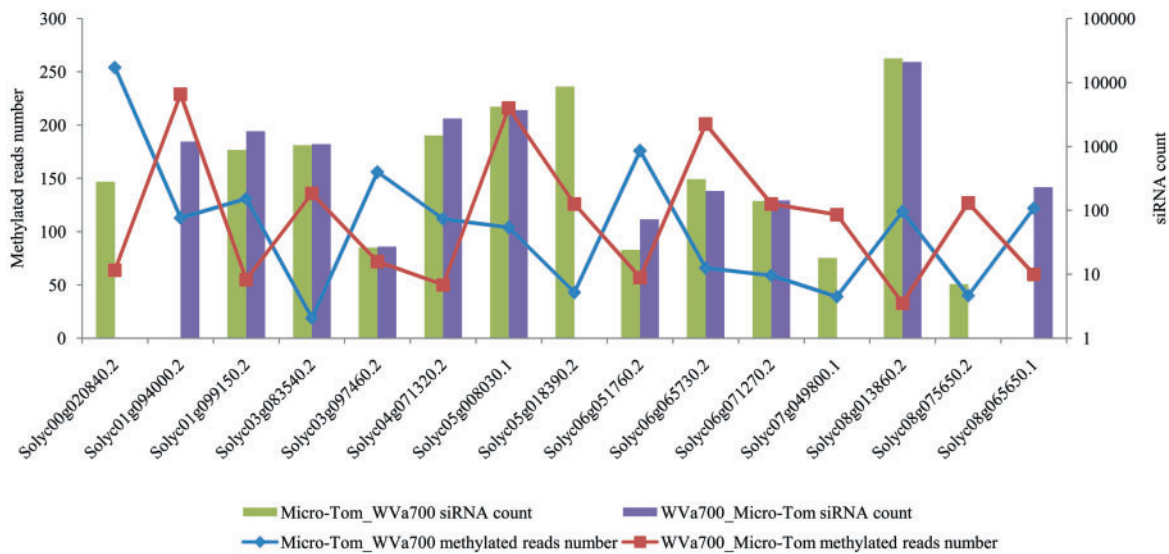
Figure 6. Methylation levels and siRNAs profiles of 15 DMGs in the reciprocal hybrids. The X axis shows 15 DMGs. The Y axis indicates the number of methylated reads or the count of siRNAs. The siRNAs sequences were matched to the DMGs via NCBI blast search. The counts of methylated reads and mapped siRNAs in each gene were displayed in Supplementary Table S6. (See online article for colour version of this figure).

promoter regions and the gene expression levels, but in the gene body regions the correlation was complicated.

3.6. Association analysis of DNA methylation and small RNAs

It has been recently reported that siRNAs, generated via RNA interfering (RNAi) pathway, can target homologous genomic DNA sequences for cytosine methylation through the RdDM pathway.[33,34] To investigate the possible relationship between siRNAs of different cytoplasms and differentially methylated patterns of the reciprocal hybrids, we analysed the accumulation of 24-nt siRNAs from small RNA high-throughput sequencing (small RNA-seq) data. Firstly, we compared the distribution of clean reads in chromosome from MeDIP-seq and small RNA-seq in the two reciprocal hybrids (see Supplementary Fig. S4). After removing reads mapped to rRNAs, tRNAs, and small nuclear and nucleolar RNAs, we newly data-mined 15,021,181 and 12,515,123 reads of 24-nt siRNAs from M × W and W × M, respectively (Supplementary Table S5). In both reciprocal hybrids, siRNAs were mapped to DMGs. We found that the expression levels of siRNAs, which were mapped to DMGs, were different between the two reciprocal hybrids. As shown in Fig. 6, those 15 genes with different methylation levels between the reciprocal hybrids were covered by different numbers of siRNAs. For example, Solyc00g0208400.2 was hypermethylated in M × W and mapped with 282 siRNAs, whereas in W × M there was no associated siRNAs at all. However, there was also opposite circumstance. Solyc06g065730.2 gene was hypermethylated in W × M and mapped with 201 siRNAs, while in M × W the number of associated siRNAs was 307 (Supplementary Table S6). As a result, the methylation levels of ten associated DMGs showed a negative correlation with their siRNAs' expression levels, while methylation levels of the other five DMGs showed a positive correlation with siRNAs' expression levels between the two reciprocal hybrids (Fig. 6).

4. Discussion

Wide hybridization, as a useful strategy for hybrid offspring to change genetic composition and phenotype, has been used in basic research

and practice in plants.[1,35] It's well observed that reciprocal hybrids usually display different phenotypes. Several genetic models such as maternal effects, parent-origin effects and cytoplasmic inheritance have been proposed to determine if reciprocal differences are interacting with different genetic backgrounds.[36,37] It has also been found that epigenetic changes influence gene expression and subsequently govern relative characteristics.[38,39] However the underlying molecular, genetic and epigenetic mechanisms remain elusive. In our current study, we established genome-wide methylation and siRNAs profiles between the two reciprocal hybrids and performed a comparative analysis of these profiles and their association with gene expression. This research revealed an epigenetic mechanism possibly accounting for the different phenotypes between the reciprocal hybrids.

In the present study, we established the global methylation profiles in S. lycopersicum, S. pimpinellifolium and their reciprocal hybrids by high-throughput sequencing, and identified the differentially methylated genes between the reciprocal hybrids. By bisulfite sequencing, we confirmed the reliability of MeDIP-seq results. At the genome-wide level, we compared the differences of methylation patterns between the reciprocal hybrids and their parents. Firstly, the W × M hybrid had higher methylation levels of methylation CGIs than its reciprocal hybrid M × W had (Fig. 1a). CpG islands normally remain unmethylated, whereas the methylation of CpG islands, especially in promoter regions, is often associated with gene silencing.[40,41] Therefore, we hypothesized that the differences in CGIs methylation between the reciprocal hybrids might be correlated with the different gene expression levels between them. Secondly, sharply-decreased methylation levels after 2kb upstream region were observed, and the parents showed remarkable differences of methylation levels in these promoter regions (Fig. 1b). Interestingly, the parental line M (cultivated species) showed lower methylation levels than the other parental line W (wild species) in 2kb upstream region. It has been reported that 2kb upstream region (the promoter region) as a repressive epigenetic mark had a role in down-regulating gene expression.[42,43] The lower methylation levels of promoter region might be inversely correlated with higher transcription levels of certain genes in M than in W. Thirdly, we analysed the distribution of reads in different repeat elements and found that LTR

retroelements had vast majority of the reads (Table 2). It is noteworthy that the reciprocal hybrids had different methylation levels in these major types of retroelements. Generally, when there was a genomic hybridization, parental genomes suffered from gene shock caused by genome recombination.[36] Methylation of these repeat elements was an important modification in the maintenance of genomic stability.[30] The differences of methylation levels between the reciprocal hybrids in major types of retroelements might be correlated with the different cytoplasmic genome backgrounds during the process of hybridization.

Specifically, we analysed the methylation patterns in different gene elements, including 2kb region upstream of 5'UTR, 5'UTR, CDS region, intron, 3'UTR and 2kb downstream of 3'UTR. The data showed that 2kb upstream region had the highest methylation level, whereas the main differences between the two reciprocal hybrids concentrated in introns and CDS regions (Fig. 2). It meant that the differences in DNA methylation between the reciprocal hybrids were predominantly located in gene body regions. Totally, we searched 109 DMRs and mapped them to 79 genes (Table 3 and Supplementary Table S2). Hence, an interesting phenomenon was that some methylated genes had more than one DMR, indicating that there might be the genomic region-specific DNA methylation patterns. After annotating the differentially methylated genes, the functions of these genes were revealed. Per GO analysis, we found that one DMG Solyc08g013860.2 was related to mitochondrial part and participated in dicarboxylic acid metabolic process, also showed differential methylation levels between the two parents (Supplementary Table S3). KEGG pathway analysis revealed that gene Solyc01g099150.2 and gene Solyc02g092780.1 were highly enriched in pathways: linoleic acid metabolish and other types of O-glycan biosynthesis. Furthermore, two DMGs (Solyc05g056450.2 and Solyc08g013860.2) were involved in carbon fixation in photosynthetic organisms and fructose and mannose metabolism pathways (Supplementary Table S4). It has been reported that carbohydrate content and sugar metabolism can regulate tomato fruit development.[44,45] And the reciprocal hybrids showed significantly different phenotypes in fruit shape index and single fruit weight, as mentioned above. These results suggested that there existed differences of epigenetic regulation in the reciprocal hybrids that might be responsible for phenotypic diversities.

As a common epigenetic process, DNA methylation can inactivate mobile elements and silence the repetitive elements to regulate gene expression or influence other epigenetic courses.[46,47] To explore how the differences in methylation levels are responsible for the distinctive phenotypes in the reciprocal hybrids, we carried out the association analysis of DNA methylation and gene expression. In our study, randomly selected 14 DMGs were used to analyse the relationship between DNA methylation and gene expression, and most of them showed remarkably different transcription levels between the two reciprocal hybrids, suggesting that DNA methylation changes in the reciprocal hybrids are correlated with altered gene expression. Specifically, the negative relationship between DNA methylation and transcription has been reported in several papers.[20,48] There are also some reports putting forward that gene body methylation, especially intronic DNA methylation, is positively correlated with gene expression.[49,50] Our results showed that gene expression was negatively correlated with DNA methylation in the promoter region, while in the gene body region the relationship between DNA methylation and gene expression was complicated (Fig. 5). Here, we assumed that, DNA methylation in the promoter regions as well as the gene body regions modulated the expression of related genes, and the differences in transcription levels may subsequently contribute to the different

phenotypes between the reciprocal hybrids. To fully address the point, the transcriptomes of all the DMGs revealed should be established by RNA sequencing and comparatively analysed in further research.

In the process of reciprocal hybridization, the source of cytoplasm (mainly comes from maternal parent) is a primary difference between the reciprocal hybrids. Therefore, small RNAs, which are matured in cytoplasms are different between the reciprocal hybrids. Moreover, 24-nt siRNAs have been reported to direct DNA methylation and regulate gene expression.[51,21] In the current study, we analysed the siRNAs which were generated in M × W and W × M hybrids, and associated these siRNAs with DMGs in the reciprocal hybrids. And we found 15 genes (among 79 DMGs) were associated with siRNAs in both reciprocal hybrids. Interestingly, siRNAs' expression levels were mostly negatively correlated with DNA methylation levels (Fig. 6). It is notable that most DMRs of associated genes that were mapped with siRNAs were located in the gene body regions (Supplementary Table S6). Thus, these results are consistent with a recent study in *Arabidopsis* that methylation levels were inversely correlated with their siRNAs profiles in genomic transcribed region of protein coding genes.[52] So it is reasonable to hypothesize that different sources of siRNAs induced the different DNA methylation patterns between the two reciprocal hybrids, and the correlation might be also associated with specific genomic regions.

In conclusion, this study explored the epigenomic differences of tomato reciprocal hybrids. The reciprocal hybrids showed divergent methylation levels in different genomic regions, especially in the gene body regions. Gene expression analysis suggested that methylation in the promoter region inversely regulates transcription, while in gene body regions the correlation between DNA methylation and gene expression is complicated. Moreover, association analysis of DNA methylation and siRNAs implied the differences of DNA methylation might be induced by siRNAs existed in different cytoplasms. And these correlations were relative to specific genomic regions. Taken together, this research proposed that the direction of hybridization caused variance in DNA methylation profiles between the reciprocal hybrids. Furthermore, the distinct methylation levels might be regulated by siRNAs and then lead to the changes in gene expression levels and ultimately the different phenotypes in the reciprocal hybrids.

Supplementary data

Supplementary data are available at *DNARES* online.

Funding

Conflict of interest

None declared.

References

1. Liu, S., Qin, Q., Xiao, J., et al. 2007, The formation of the polyploid hybrids from different subfamily fish crossings and its evolutionary significance, *Genetics*, **176**, 1023–34.

2. Martin-Sanchez, J. A., Gómez-Colmenarejo, M., Moral, J. D., et al. 2003, A new Hessian fly resistance gene (H30) transferred from the wild grass *Aegilops triuncialis* to hexaploid wheat, *Theor. Appl. Genet.*, **106**, 1248–55.

3. Ronicke, S., Hahn, V., Horn, R. F., et al. 2004, Interspecific hybrids of sunflower as a source of Sclerotinia resistance, *Plant Breeding*, **123**, 152–7.

4. Amantebordeos, A., Sitch, L. A., Nelson, R. and Friedt, W. 1992, Transfer of bacterial blight and blast resistance from the tetraploid wild rice *Oryza minuta* to cultivated rice, *Oryza sativa*, *Theor. Appl. Genet.*, **84**, 345–54.

5. Josefsson, C., Dilkes, B. and Comai, L. 2006, Parent-dependent loss of gene silencing during interspecies hybridization, *Curr. Biol.*, **16**, 1322–8.

6. Wolf, J. B. and Wade, M. J. 2009, What are maternal effects (and what are they not)? *Philos. Trans. R. Soc. Lond, B, Biol. Sci.*, **364**, 1107–15.

7. Adams, S., Vinkenoog, R., Spielman, M., Dickinson, H. G. and Scott, R. J. 2000, Parent-of-origin effects on seed development in *Arabidopsis thaliana* require DNA methylation, *Development*, **127**, 2493–502.

8. Galloway, L. F. and Fenster, C. B. 2001, Nuclear and cytoplasmic contributions to intraspecific divergence in an annual legume, *Evolution*, **55**, 488–97.

9. Levin, D. A. 2003, The cytoplasmic factor in plant speciation, *Syst. Bot.*, **28**, 5–11.

10. Song, K., Lu, P., Tang, K. L. and Osborn, T. C. 1995, Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution, *Proc. Natl. Acad. Sci. USA*, **92**, 7719–23.

11. Kinoshita, T., Yadegari, R., Harada, J. J., Goldberg, R. B. and Fischer, R. L. 1999, Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm, *Plant Cell*, **11**, 1945–52.

12. Crow J. F. 1948, Alternative hypotheses of hybrid vigor, *Genetics*, **33**, 477–87.

13. Yu, S. B., Li, J. X., Xu, C. G., et al. 1997, Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid, *Proc. Natl. Acad. Sci. USA*, **94**, 9226–31.

14. Birchler, J. A., Auger, D. L. and Riddle, N. C. 2003, In search of the molecular basis of heterosis, *Plant Cell*, **15**, 2236–39.

15. Huang, Y., Sun, J., Zhang, L., et al. 2014, Genome-wide DNA methylation profiles and their relationships with mRNA and the microRNA transcriptome in bovine muscle tissue (*Bos taurine*), *Sci. Rep.*, **4**, 1–17.

16. Chen, X. 2009, Small RNAs and their roles in plant development, *Annu. Rev. Cell Dev. Biol.*, **25**, 21–44.

17. Chapman, E. J. and Carrington, J. C. 2007, Specialization and evolution of endogenous small RNA pathways, *Nat. Rev. Genet.*, **8**, 884–96.

18. Llave, C., Xie, Z. X., Kasschau, K. D. and Carrington, J. C. 2002, Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA, *Science*, **297**, 2053–6.

19. Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., et al. 2008, Widespread translational inhibition by plant miRNAs and siRNAs, *Science*, **320**, 1185–90.

20. Shen, H., He, H., Li, J., et al. 2012, Genome-wide analysis of DNA methylation and gene expression changes in two Arabidopsis ecotypes and their reciprocal hybrids, *Plant Cell*, **24**, 875–92.

21. Baulcombe, D. 2004, RNA silencing in plants, *Nature*, **431**, 356–63.

22. Lippman, Z. and Martienssen, R. 2004, The role of RNA interference in heterochromatic silencing, *Nature*, **431**, 364–70.

23. Mcclintock, B. 1984, The significance of responses of the genome to challenge, *Science*, **226**, 792–801.

24. Groszmann, M., Greaves, I. K., Albertyn, Z. I., Scofield, G. N., Peacock, W. J. and Dennis, E. S. 2011, Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor, *Proc. Natl. Acad. Sci. USA*, **108**, 2617–22.

25. Lu, J., Zhang, C., Baulcombe, D. C. and Chen, Z. J. 2012, Maternal siRNAs as regulators of parental genome imbalance and gene expression in endosperm of Arabidopsis seeds. *Proc, Natl. Acad. Sci. USA*, **109**, 5529–34.

26. Li, J., Sun, Q., Yu, N., et al. 2014, The role of small RNAs on phenotypes in reciprocal hybrids between *Solanum lycopersicum* and *S. pimpinellifolium*, *BMC Plant Biol.*, **14**, 296.

27. Murray, M. G. and Thompson, W. F. 1980, Rapid isolation of high molecular weight plant DNA, *Nucleic Acids Res.*, **8**, 4321–5.

28. Gruntman, E., Qi, Y., Slotkin, R. K., Roeder, T., Martienssen, R. A. and Sachidanandam, R. 2008, Kismeth: analyzer of plant methylation states through bisulfite sequencing, *BMC Bioinformatics*, **9**, 371.

29. Cao, L., Yu, N., Li, J., Qi, Z., Wang D., Chen, L. 2016, Heritability and reversibility of DNA methylation induced by *in vitro* grafting between *Brassica juncea and B. oleracea*, *Sci. Rep.*, **6**, 27233.

30. Gordenin, D. A., Lobachev, K. S., Degtyareva, N. P., Malkova, A. L., Perkins, E and Resnick, M. A. 1993, Inverted DNA repeats -a source of eukaryotic genomic instability, *Mol. Cell, Biol.*, **13**, 5315–22.

31. Yan, H., Kikuchi S., Neumann P., et al. 2010, Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice, *Plant J.*, **63**, 353–65.

32. Hu, Y., Xu, H., Li, Z., et al. 2013, Comparison of the genome-wide DNA methylation profiles between fast-growing and slow-growing broiler, *Plos One*, **8**, e56411.

33. Henderson, I. R. and Jacobsen, S. E. 2007, Epigenetic inheritance in plant, *Nature*, **447**, 418–24.

34. Law, J. A. and Jacobsen, S. E. 2010, Establishing, maintaining and modifying DNA methylation patterns in plants and animals, *Nat. Rev. Genet.*, **11**, 204–20.

35. Mallet, J. 2005, Hybridization as an invasion of the genome, *Trends Ecol. Evol.*, **20**, 229–37.

36. Chen, Z. J. 2010, Molecular mechanisms of polyploidy and hybrid vigor, *Trends Plant Sci.*, **15**, 57–71.

37. Ordas, B., Malvar, R. A., Ordas, A. and Revilla, P. 2008, Reciprocal differences in sugary x sugary enhancer sweet corn hybrids, *J. Am. Soc. Hortic. Sci.*, **133**, 777–82.

38. Liu, R., How, K. A., Stammitti, L., et al. 2015, A DEMETER-like DNA demethylase governs tomato fruit ripening, *Proc. Natl. Acad. Sci. USA*, **112**, 10804–9.

39. Li, J., Wang, Y., Zhang, L., et al. 2013, Heritable variation and small RNAs in the progeny of chimeras of *Brassica juncea* and *Brassica oleracea*, *J. Exp. Bot.*, **64**, 4851–62.

40. Bird, A. 2002, DNA methylation patterns and epigenetic memory, *Genes Dev.*, **16**, 6–21.

41. Chuang, J. C. and Jones, P. A. 2007, Epigenetics and microRNAs, *Pediatr. Res.*, **61**, 24R–29R.

42. Curradi, M., Izzo, A., Badaracco, G. and Landsberger, N. 2002, Molecular mechanisms of gene silencing mediated by DNA methylation, *Mol. Cell. Biol.*, **22**, 3157–73.

43. Saze, H., Sasaki, T. and Kakutani, T. 2008, Negative regulation of DNA methylation in plants, *Epigenetics*, **3**, 122–124.

44. Schaffer, A. A., Petreikov, M., Miron, D., et al. 1999, Modification of carbohydrate content in developing tomato fruit, *HortScience*, **34**, 1024–7.

45. Miron, D. and Schaffer, A. A. 1991, Sucrose phosphate synthase, sucrose synthase, and invertase activities in developing fruit of *Lycopersicon esculentum* Mill. and the sucrose accumulating *Lycopersicon hirsutum* Humb. and Bonp, *Plant Physiol.*, **95**, 623–7.

46. Xiao, J., Song, C., Liu, S., et al. 2013, DNA methylation analysis of allotetraploid hybrids of red crucian carp (*Carassius auratus* red var.) and common carp (*Cyprinus carpio* L.), *Plos One*, **8**, e56409.

47. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. and Henikoff, S. 2007, Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription, *Nat. Genet.*, **39**, 61–9.

48. Hu, J., Chen, X., Zhang, H. and Ding, Y. 2015, Genome-wide analysis of DNA methylation in photoperiod- and thermo-sensitive male sterile rice Peiai 64S, *BMC Genomics*, **16**, 102.

49. Ball, M. P., Li, J. B., Gao, Y., et al. 2009, Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells, *Nat Biotechnol.*, **27**, 361–8.

50. Tran, R. K., Zilberman, D., Bustos, C. D., et al. 2005, Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in Arabidopsis, *Genome Biol.*, **6**, R90.

51. Kasschau, K. D., Fahlgren, N., Chapma, E. J., et al. 2007, Genome-wide profiling and analysis of Arabidopsis siRNAs, *Plos Biol.*, **5**, 479–93.

52. Cokus, S. J., Feng, S., Zhang, X., et al. 2008, Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning, *Nature*, **452**, 215–19.