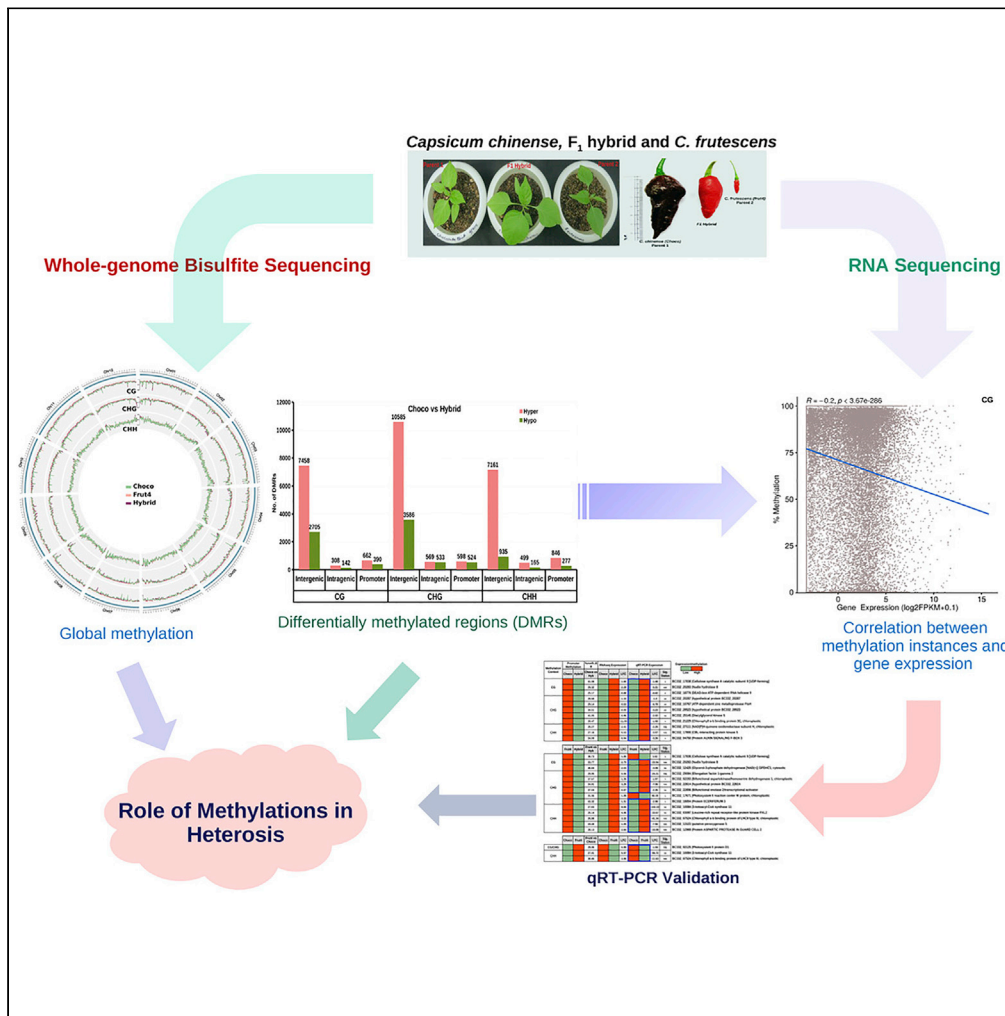


Article

Integrated analysis of DNA methylation, transcriptome, and global metabolites in interspecific heterotic *Capsicum* F<sub>1</sub> hybrid



Vandana Jaiswal, Abdul Rawoof, Vijay Gahlaut, Ilyas Ahmad, Sushil S. Chhapekar, Meenakshi Dubey, Nirala Ramchiary

vandana.jaiswal2009@gmail.com (V.J.)  
nrudsc@gmail.com (N.R.)

Highlights

Global methylation identified significantly different proportions of mCs in hybrid

Of common DMRs, 33.08% showed different methylation in hybrid from the mid-parental value

Negatively correlated DEG pDMR-genes were enriched in metabolic pathways

Significant higher expression of metabolites and DE-Gs were identified in the F<sub>1</sub> hybrid



## Article

Integrated analysis of DNA methylation, transcriptome, and global metabolites in interspecific heterotic *Capsicum* F<sub>1</sub> hybridVandana Jaiswal,<sup>1,2,5,7,\*</sup> Abdul Rawoof,<sup>1,5</sup> Vijay Gahlaut,<sup>2,6</sup> Ilyas Ahmad,<sup>1,6</sup> Sushil S. Chhapekar,<sup>1,3</sup> Meenakshi Dubey,<sup>1,4</sup> and Nirala Ramchiary<sup>1,\*</sup>

## SUMMARY

Hybrid breeding is one of the efficacious methods of crop improvement. Here, we report our work towards understanding the molecular basis of F<sub>1</sub> hybrid heterosis from *Capsicum chinense* and *C. frutescens* cross. Bisulfite sequencing identified a total of 70597 CG, 108797 CHG, and 38418 CHH differentially methylated regions (DMRs) across F<sub>1</sub> hybrid and parents, and of these, 4891 DMRs showed higher methylation in F<sub>1</sub> compared to the mid-parental methylation values (MPMV). Transcriptome analysis showed higher expression of 46–55% differentially expressed genes (DE-Gs) in the F<sub>1</sub> hybrid. The qRT-PCR analysis of 24 DE-Gs with negative promoter methylation revealed 91.66% expression similarity with the transcriptome data. A few metabolites and 65–72% enriched genes in metabolite biosynthetic pathways showed overall increased expression in the F<sub>1</sub> hybrid compared to parents. These findings, taken together, provided insights into the integrated role of DNA methylation, and genes and metabolites expression in the manifestation of heterosis in *Capsicum*.

## INTRODUCTION

The *Capsicum* ( $2n = 2x = 24$ ) crop, which belongs to the family Solanaceae, is one of the most important vegetable crops. Although more than 38 species are reported under the *Capsicum* genus, only *Capsicum annuum*, *C. frutescens*, *C. chinense*, *C. pubescens* and *C. baccatum* are cultivated (Bosland and Votava, 2000; Ibiza et al., 2012). Besides their use as a spice, *Capsicum*, commonly known as chilli peppers contains several important secondary metabolites and other phytochemicals compounds which makes them an economically important crop with high industrial and pharmacological potential (Ramchiary et al., 2014; Sarpras et al., 2020; Sricharoen et al., 2017; Sinisgalli et al., 2020; Olatunji and Afolayan, 2019; Baenas et al., 2019). For instance, the naturally found compounds such as phenolics, flavonoids, carotenoids, capsaicinoids, vitamins, and alkaloids in *Capsicum* can be used for multipurpose industry usage viz. in increasing food shelf-life, improving the sensory properties, seasoning, and coloring of foods, in cosmetic and pharma industry (Baenas et al., 2019). Several studies reported the presence of intra and interspecific diversity of fruit traits, metabolites content, and other agronomic traits in *Capsicum* species (Ahmad et al., 2021a, 2021b; Araceli et al., 2009; Chhapekar et al., 2020; Liu et al., 2020; Sarpras et al., 2016; Sarpras et al., 2019). Furthermore, many studies on marker development, QTL/gene(s) mapping of economically important traits, whole transcriptome, and genome sequencing have been reported on *Capsicum* species (Ahn et al., 2014; Chhapekar et al., 2021; Dubey et al., 2019; Jaiswal et al., 2020; Kim et al., 2020; Liu et al., 2019; Martínez et al., 2021; Razo-Mendivil et al., 2021). Few studies reported that inter- and intraspecific hybrids of cultivated *Capsicum* species i.e. *C. annuum*, *C. frutescens*, *C. chinense*, *C. baccatum*, and *C. pubescens*, showed vigor for different traits like fruit yield, plant growth, and earliness (Sood and Kumar, 2010; Shrestha et al., 2011; Singh et al., 2014; Gözen et al., 2018). Heterosis breeding has been extensively used in *Capsicum* crop to increase the production of economically important traits (Herath et al., 2021). A recent study reported the expression of higher capsaicinoids and sugars in heterotic hybrid (derived from crossing between *C. chinense* x *C. frutescens*) as compared to their parental lines (Zamljen et al., 2020). Also, the higher accumulation of several metabolites such as sugars (viz. glucose, fructose, and sucrose) and primary metabolites like citric acid was observed in leaves of *Capsicum* hybrids compared to the parents at lower temperatures (Shiragaki et al., 2021). These observations in different studies suggest that the higher expression of genes and metabolites in F<sub>1</sub> hybrids compared to the parental lines might

<sup>1</sup>Translational and Evolutionary Genomics Lab, School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India

<sup>2</sup>Biotechnology Division, CSIR-Institute of Himalayan Bioresource Technology, Palampur, Himachal Pradesh, India

<sup>3</sup>Department of Horticulture, Chungnam National University, Daejeon 34134, South Korea

<sup>4</sup>Department of Biotechnology, Delhi Technological University, Shahbad Daultpur, Bawana Road, Delhi 110042, India

<sup>5</sup>These authors contributed equally

<sup>6</sup>These authors contributed equally

<sup>7</sup>Lead contact

\*Correspondence: vandana.jaiswal2009@gmail.com (V.J.), nrudsc@gmail.com (N.R.)  
<https://doi.org/10.1016/j.isci.2022.105318>



have a direct role in the manifestation of heterosis in  $F_1$  hybrids. Furthermore, it was reported that DNA methylation plays an important role in the regulation of the expression of genes, which would eventually influence the biosynthesis of metabolites (Sinha et al., 2020; Lauss et al., 2018), thereby, suggesting further the need for integrated analysis of DNA methylation, and expression of genes and metabolites content for detail understanding of the manifestation of heterosis, as no such study is reported in *Capsicum* crops until date.

Although, primarily, DNA methylation causes the silencing of the gene(s) (He et al., 2011, 2021); however, a few studies also reported higher expression of genes despite methylation found in the gene body (Arechederra et al., 2018), and the exact mechanism is still not understood. In plants, DNA methylation occurs in three contexts: CG, CHG, and CHH (where H is any nucleotide other than G) and are maintained by *MET1*, *CM3*, and *DRM2*, respectively (He et al., 2021). Studies have also found that DNA methylation also inherits loyally across the generations (Becker et al., 2011; Schmitz et al., 2011) when progenies are derived by selfing. However, the inheritance trend of DNA methylations may be deferred in hybrid genotypes derived by crossing (He et al., 2010; Chodavarapu et al., 2012); and the degree of differential methylation pattern is directly proportional to the divergence of parental genotypes. These allelic and epi-allelic variations and their corresponding effect on the expression of genes in hybrids are considered one of the major causes of heterosis (Lauss et al., 2018).

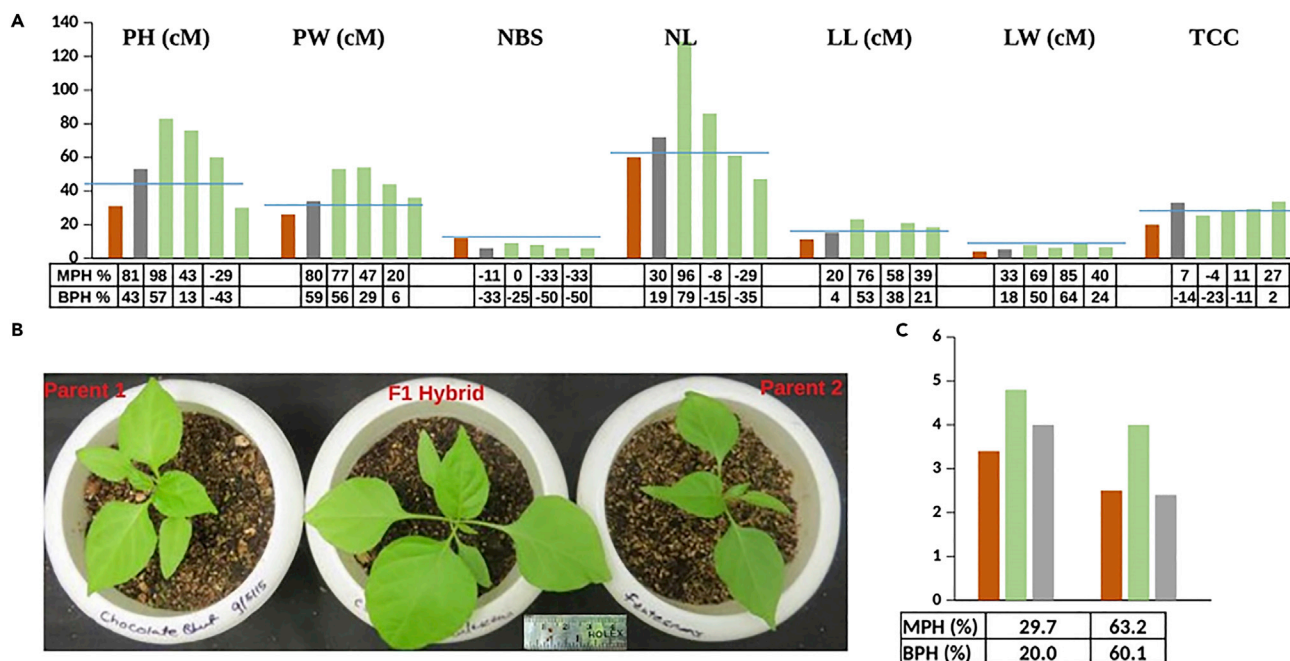
Recently, a number of studies were conducted to explore the DNA methylation pattern and their possible role in heterosis in model plants (Greaves et al., 2012; Shen et al., 2012; Chen, 2013; Groszmann et al., 2013; Dapp et al., 2015; Zhang et al., 2016; Lauss et al., 2018) as well as important crops (He et al., 2010, 2013; Shen et al., 2017; Li et al., 2018; Sinha et al., 2020; Zhou et al., 2021). In pigeon pea, the global DNA methylation and transcriptome study identified an overall increase in the methylation in  $F_1$  hybrids (Sinha et al., 2020). The same study reported the differential expression of genes involved in phytohormone signaling, defense response and growth. More than 65% of differentially expressed genes (DE-Gs) were also found to be differentially methylated in hybrids, thereby indicating the active role of DNA methylations in gene expression and ultimately hybrid vigor in  $F_1$  (Sinha et al., 2020). Similarly, in the case of *Brassica*, higher methylation was also observed in  $F_1$ s than their respective parents (Shen et al., 2017; Li et al., 2018). In *Brassica napus* (allotetraploid), the DNA methylation was also associated with higher expression of small RNA clusters of that region and was also involved in mechanisms of active growth and flower development (Shen et al., 2017). In *B. oleracea*, differential DNA methylations were found to contribute to curd yield heterosis (Li et al., 2018). Likewise, efforts were also made to understand the role of DNA methylations in hybrids in other major food crops like rice (He et al., 2010; Zhou et al., 2021), maize (He et al., 2013) and so forth.

However, despite the availability of whole genome sequences of *Capsicum* spp. (Kim et al., 2014, 2017; Qin et al., 2014), the integrated analysis of global DNA methylome and its correlation with gene(s) and metabolites expression are not studied in *Capsicum* hybrids, although few studies explored the role of epigenetics in fruit development and ripening using whole genome methylome (Rawoof et al., 2020; Xiao et al., 2020) and small RNA sequencing (Lopez-Ortiz et al., 2021; Chhapekar et al., 2021), and observed that the co-networking/correlation of methylome and small RNAs were potential players in fruit development. Keeping this knowledge gap in mind, in this study for the first time we report the integrated analysis wherein we observed significant correlations of DNA methylation with the expression of genes and metabolites in heterotic  $F_1$  *Capsicum* hybrid and their parental lines at the early plant growth stage. Furthermore, our study revealed the complex and dynamic relationships of the genomes of two *Capsicum* spp. and their inheritance in the interspecific  $F_1$  hybrid.

## RESULTS

### Phenotypic analysis of $F_1$ hybrid and parental lines

The interspecific *Capsicum*  $F_1$  hybrid derived from crossing between two contrasting *Capsicum* species, *C. chinense* (Choco, female parent) and *C. frutescens* (Fru4, pollen donor) was evaluated for heterotic traits along with their parental lines. The parent *C. chinense* bears large fruits and with extreme pungency, while the *C. frutescens* bears small and a large number of fruits, with robust plant type and medium pungency. Since the performance of the plants is determined by the early establishment and robust growth of plants from the seedling stage, we chose 60 days old plants for the present study. The developed  $F_1$  hybrid plants showed heterosis over the parental genotypes for phenotypic traits studied (Figure 1A). The  $F_1$  plants showed vigorous performance from early seedlings compared to the plants of parental



**Figure 1. Trait variation observed in F<sub>1</sub> hybrid and their parents**

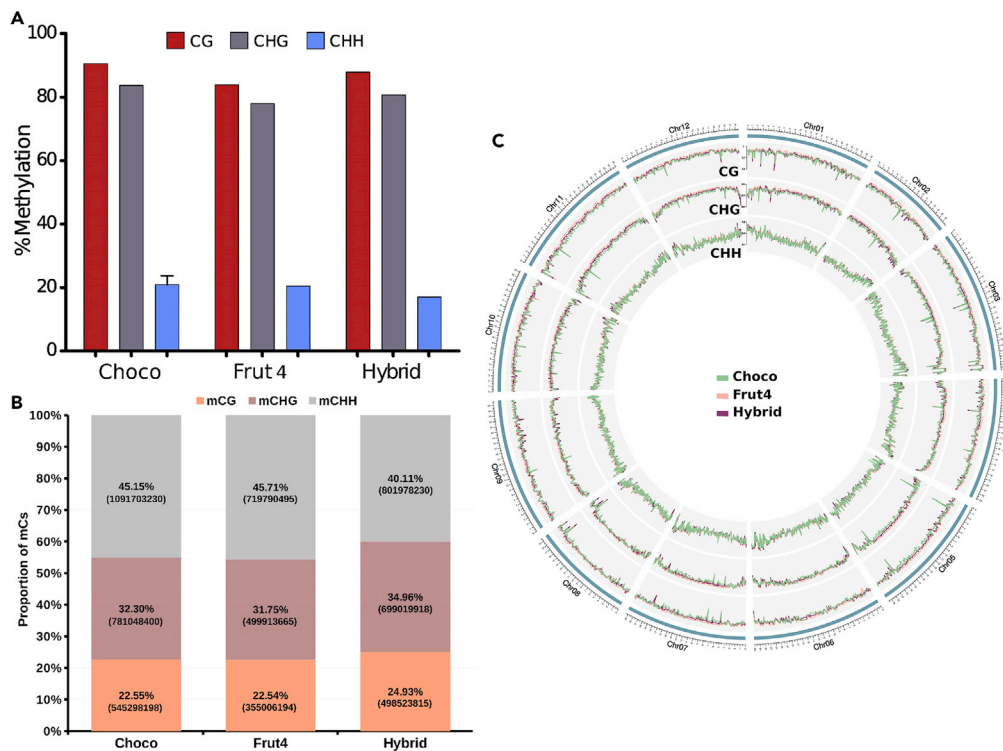
(A) Phenotypic variability observed in adult plants.

(B and C) Phenotypic variability in young plants (two months old). Blue line showing mid parent value. PH = plant height, PW = plant width, NBS = number of secondary branches, NL = number of leaves, LL = leaf length, LW = leaf width, TCC = total chlorophyll content. MPH = mid parent heterosis, BPH = better parent heterosis.

genotypes. For example, 29.7% mid-parent heterosis (MPH) and 20.0% better parent heterosis (BPH) were observed for leaf lengths (LL). In hybrids, leaves were broader than parental genotypes with high value of heterosis (MPH = 63.2%, BPH = 60.1%) (Figures 1B and 1C). Furthermore, at the time of the flowering height of parents was 31.5 (Choco) and 53.2 cm (Frut4), respectively, however, hybrids reached a height of 83.6 cm with 97.6% MPH and 56.6% BPH. Similarly, for plant width (PW), upto 80 and 58.8% MPH and BPH, respectively, were observed in hybrids. At the time of flowering, vigourity was also observed for leaf traits like number of leaves (NL; MPH = 95.5%, BPH = 79.2%), leaf length (LL; MPH = 75.5%, BPH = 52.6%), leaf width (LW; MPH = 84.9%, BPH = 63.7%) in hybrids. These results emphasized that the heterosis in plants was manifested from the early seedling stage. Thus we used DNA/RNA from two-month-old plant stage for methylome and transcriptome analysis.

### Whole-genome bisulfite and RNA sequencing of *Capsicum* hybrid and its parents

It is expected that the methylation of two distinct parental lines belonging to different *Capsicum* species might potentially contribute to the methylome of their F<sub>1</sub> hybrid. Therefore, to obtain an in-depth understanding of the global dynamics of cytosine methylation/epigenetic regulation owing to DNA methylation during hybrid formation at single-base resolution, whole-genome bisulfite sequencing of parental lines and their hybrid was performed. A total of 1660 million (~553 million/library including biological replicates) of length 150 bp clean bisulfite sequencing reads were obtained from parental lines and their hybrid. Around 1038 million reads from all samples were aligned to the reference genome with an average mapping rate of 62.54%. Overall, 384825557, 278031457, and 375544686 reads from Choco, Frut4, and hybrid were uniquely mapped to the *C. chinense* reference genome, respectively, thereby providing more than 10X coverage (Table S2A). The reproducibility of the results was also tested using Pearson correlation analysis and a high correlation (0.9–0.95) among the biological replicates of each sample was found (Figure S1). An average of 0.4–0.5% mapping efficiency of each sample against lambda DNA reference genome indicated >99% bisulfite conversion efficiency of sample replicates from parents and hybrids (Table S2B). In addition, to understand gene(s) expression changes along with methylome dynamics in the hybrid compared to its parents, RNA sequencing (RNAseq) was performed using the same leaf tissue used in WGBS. Approximately 476 million of 150 bp clean reads (~159 million/library including replicates) were



**Figure 2. Global DNA methylation profile observed in hybrid and parental lines Choco and Frut4**

(A) Average methylation, (B) methylated cytosine (mCs) proportion, and (C) average methylation distribution across 12 *Capsicum* chromosomes at CG, CHG, and CHH methylation context.

obtained from parents and their hybrid was aligned to the *C. chinense* reference genome. On an average, 84.5% reads of Choco, 93.8% of Frut4, and 96.8% reads of the hybrid aligned to the *C. chinense* reference genome. (Table S2C).

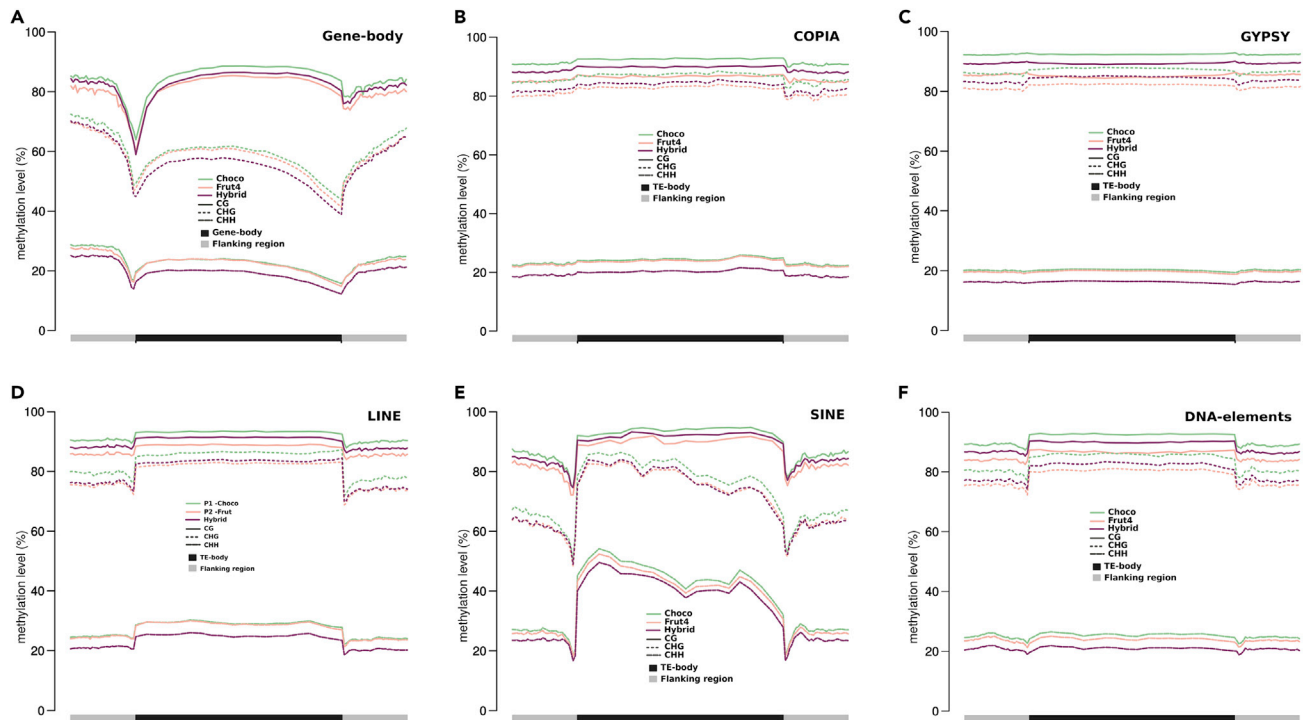
### Global methylation landscape in *Capsicum* hybrid and its parental lines

The genome-wide dynamics of cytosine methylation at CG, CHG, and CHH context varied in parental line and their hybrid. Choco showed the highest global methylation in all three contexts as compared to hybrid and Frut4. The hybrid showed higher global average methylation in CG and CHG contexts and lower methylations in the CHH context compared to parent Frut4 (Figure 2A). Also, the frequencies of mCs in *Capsicum* hybrid and its parental lines ranged from 32.8% to 36.02% (Table S3). In particular context, the fraction of mCs observed was approximately 22.5–24.9%, 31.7–34.9% and 40–45.7% at CG, CHG, and CHH, respectively, in three genotypes (Figure 2B). Notably, compared to both the parents, hybrid showed 1.2-fold higher mCs proportion (24.9 and 34.9%, respectively) at symmetric (CG, CHG) context, while contrasting to this, it showed 1.12-fold lower mCs proportion in asymmetric (CHH) context (Figure 2B). Segmentation of chromosome-wise average methylation showed higher methylation at symmetric CG and CHG context than asymmetric context throughout the 12 chromosomes in all three genotypes (Figure 2C). Also, at segments of chromosome (chr) chr2 to chr5, chr7, and chr11, both hybrid and Frut4 showed slightly higher methylation than Choco. However, the same chromosomes in hybrid showed slight decrease in methylation than Frut4 (Figure 2C).

### Methylation pattern at various genomic regions and transposable elements

The pattern of mCs at different genomic regions/elements was analyzed in hybrid and its parental line (Figures 3 and S2). Both parental line and hybrid showed the highest average gene-body methylation (GBM) at CG followed by the CHG context, while lowest at the CHH context (Figure 3A). At CG context, the level of GBM looked similar to that in the 2 kb flanking region and was highest in Choco, followed by hybrid, and lowest in the case of Frut4. The GBM at CHG and CHH context showed notably less methylation as compared to the 2 kb flanking region of genes. Unexpectedly, the average GBM at CHG and CHH context





**Figure 3. Metaplots representing methylation distribution at genomic elements and their flanking regions**

(A) gene-body, (B–F) transposable-elements (TEs) body including copia, gypsy, LINE, SINE, and DNA-elements, respectively, along with their 1.5 kb flanking regions. Methylation at gene/TE-body and flanking regions was analyzed across 20 and 30 bins, respectively.

was lowest in the hybrid; however, both parental lines showed similar methylation frequencies (Figure 3A). To avoid the potential noise (arising owing to transposable elements), the GBM was segmented into exonic and intronic regions. We observed that methylation at exon body was lower than its flanking regions, while it was higher at intron body than its flanking region at both symmetric and asymmetric methylation contexts in three genotypes (Figures S2A and S2B). On comparing methylations at three different contexts in hybrid and parents, we found that at CHG and CHH contexts, both exonic and intronic body methylations were lower in hybrid than its parents. However, methylation in CG context was at par with Frut4 in exon, lower and higher than Choco and Frut4, respectively in intronic regions (Figures S2A and S2B). Genome-wide proportion/distribution of cytosine methylation was analyzed at different genomic regions including exons, introns, intergenic, and promoter regions across 4 different methylation bins categorized based on percent methylation, i.e. <20%, 20–50%, 50–80%, and >80% DNA methylation (Figure S2C). At CG, these genomic regions showed a higher proportion of mCs with >80% methylation in Choco and hybrid than Frut4. However, with <20% methylation, the hybrid showed a slightly higher methylation proportion than its parental line across all genomic elements. Also, within 50–80% methylation bin at CG and CHG context, both hybrid and Frut4 showed a higher proportion of mCs than Choco (Figure S2C).

In addition, methylation pattern within transposable-elements body (TE-body) and their 2kb flanking regions was analyzed for long terminal repeats (LTRs; copia and gypsy), non-LTRs (LINEs and SINEs) and DNA-elements (Figures 3B–3F). Overall, higher TE-body methylation than its flanking region was observed at both symmetric (CG, CHG) and asymmetric (CHH) contexts in parental lines and their hybrid. The methylation in LTRs at both TE-body and flanking regions was consistently higher at all contexts (Figures 3B and 3C), while in non-LTRs and DNA-elements, it was preferentially higher at TE-body than their flanking regions (Figures 3D–3F and S2D). Across all TE methylation, at both CG and CHG contexts, hybrid showed lower methylations than Choco and higher methylation than Frut4, while in the CHH context, hybrid showed peculiarly lowest TE methylation (Figures 3B–3F). Likewise, in hybrid, at CHG context, SINE (non-LTRs) and at CHH context all TE (LTRs, non-LTRs, and DNA-elements) were hypo-methylated than its two parents (Figure S2E). Altogether, as far as the CG context is concerned, the hybrid had lower methylation than Choco and higher methylation than Frut4 in all the genomic regions (GBM, TEs, DNA elements); however, in the

CHH context, hybrid had lower methylations than both the parents. In the case of the CHG context, hybrid has lower methylations than both the parents in GBM and SINE, and in COPIA, GYPSY, LINE, DNA elements, and hybrid had lower and higher methylations than Choco and Frut4, respectively.

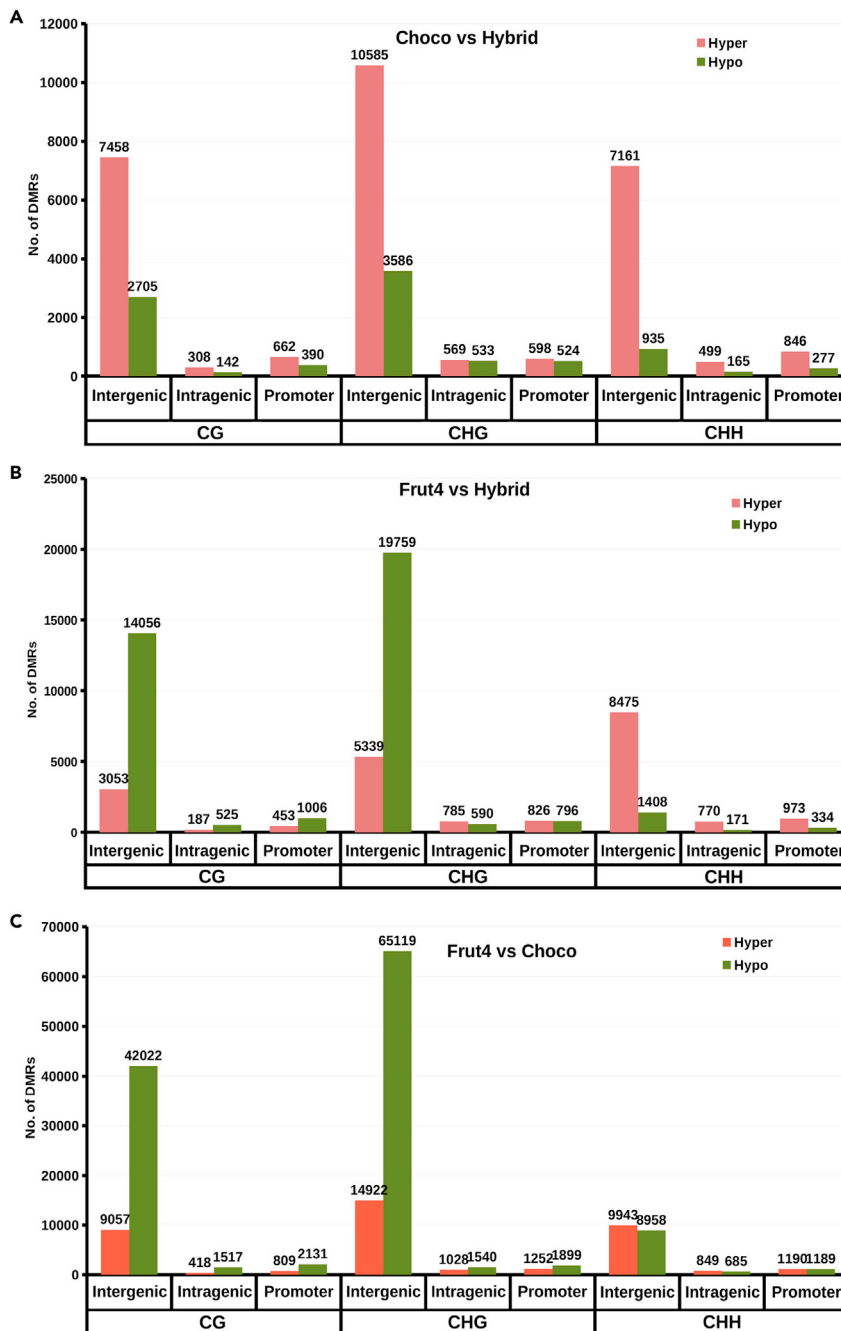
### Differentially methylated regions in *Capsicum* parental lines and their hybrid

Analysis of differentially methylated regions (DMRs) between parental lines (Choco, Frut4) and their hybrid led to the identification of a total of 70597, 108797, and 38418 unique DMRs at CG, CHG, and CHH methylation context, respectively (Tables S4, S5, and S6 and Figure S3A). In Choco vs hybrid and Frut4 vs hybrid, around 8.8–9.2% and 10.1–11.2% DMRs in symmetric CG and CHG context, respectively while, 19.9% DMRs in the CHH context, were uniquely methylated (Figures S3B–S3D; Table S7A). It was noted that, in both symmetric contexts, Frut4 had a high number of hypo-methylated DMRs, while in the CHH context; it had high hyper-methylated DMRs compared to hybrid and Choco (Figure S3A). Also, hybrid had a low contribution of DMRs, i.e. 27.7%, 28.3 and 13.9% from CG, CHG, and CHH context respectively when compared to Choco, indicating hybrid was hypo-methylated than parent Choco. However, when compared to the other parent Frut4, hybrid showed an increased contribution of DMRs, i.e. 80.8%, 75.2%, and 15.8% from CG, CHG, and CHH context, respectively, indicating that hybrid is hyper-methylated than parent Frut4 (Table S7B). Then the density of DMRs in hybrid and its parents was checked across 12 *Capsicum* chromosomes. It was observed that at CG and CHG context, the density of hyper DMR was more in Choco vs hybrid, while it was low in Frut4 vs hybrid and Frut4 vs Choco (Figures S4A–S4C). However, when analyzed at the individual chromosome level, low-density of hypo DMRs was observed in Choco vs hybrid across segments of chr2 to chr5, chr7, and chr11 at both symmetric contexts (Figure S4A). But in Frut4 vs hybrid at CG and CHG context and in Frut4 vs Choco at CG, CHG, and CHH context, the same aforementioned chromosome segments showed increased density of hypo-methylated DMRs (Figures S4B and S4C). After this, the genomic annotation/distribution of these DMRs was analyzed and observed that most of the DMRs in hybrid and its parents were from intergenic regions of the genome (Figures 4A–4C and Table S7C). Of total DMRs in Choco vs hybrid, around 81.9–87.12%, 3.8–6.7%, and 6.8–11.3% DMRs were from intergenic, intragenic, and promoter regions respectively (Figure 4A and Table S7C). Similarly, in Frut4 vs hybrid, of total DMRs, 81.4–89.3%, 3.6–7.7%, and 5.8–10.8% were from intergenic, intragenic, and promoter regions, respectively (Figure 4B and Table S7C). Also, in Choco vs hybrid, all intergenic, intragenic, and promoter DMRs were hyper-methylated at all three methylation contexts (Figures 4A and S5A). However, in Frut4 vs hybrid and Frut4 vs Choco at CG and CHG context, the intergenic DMRs were hypo-methylated (Figures 4B, 4C, S5B, and S5C).

Furthermore, we identified a total of 4891 (2206 CG, 2353 CHG, and 332 CHH) common DMRs between hybrids and both of its parents (Table S8). Based on common DMRs, the predictive level of methylation in hybrid was estimated against the mean average parental methylation level or mid-parental methylation values (MPMV). Of common DMRs, only 33.08% (1618) showed significant different methylation ( $p$  value  $< 0.01$ ; FDR $<0.05$ ) in hybrid from MPMV and DMRs in hybrid having higher methylation than MPMV were categorized as *trans*-chromosomal methylation (TCM), while those having methylation less than MPMV were categorized as *trans*-chromosomal demethylation (TCdM). Most of these DMRs (86.5%) were located in intergenic regions (Table S8). Furthermore, the percent methylation contribution of both parents in hybrid as well as newly originated mCs in hybrid were identified. A total of 77994412 mCs in hybrid were from both parents out of which 78.92% (61551482 mCs) were from Choco and 21.08% (16442930 mCs) were from Frut4 (Tables S9A and S9B). Also, around 76.5–76.7% mCHH sites in hybrid were contributed by both parents, while only 8.8–9.1% mCG and 14.4–14.5% mCHG sites arose from both parental lines. Additionally, a total of 15523997 new mCs originated in hybrid, which contributed to 76.99% mCHH, 13.94% mCHG, and 9.07% mCG sites (Tables S9A and S9B).

### Allele-specific methylation variations

The significantly methylated alleles (ASM) at single nucleotide positions were analyzed across the DMRs spanned over various genomic regions in parental lines and their  $F_1$  hybrid. Around 8803–11379 CG, 10964–13966 CHG, and 4901–7299 CHH SNVs associated with pDMRs negatively correlated with their corresponding gene expressions were identified from *Capsicum*  $F_1$  hybrid and its parental lines. Of these SNVs, 1449–2338 in Choco, 1310–1783 in Frut4, and 1468–2315 in hybrid were significantly methylated ASM (adjusted  $p$  value  $< 0.05$ ) and were associated with 523, 753, and 734 pDMRs genes at CG, CHG, and CHH context, respectively (Figure 5A and Tables S10A–S10C). In addition, significant methylated ASMs (adjusted  $p$  value  $< 0.05$ ) associated with TCM and TCdM DMRs were also identified in Choco (404

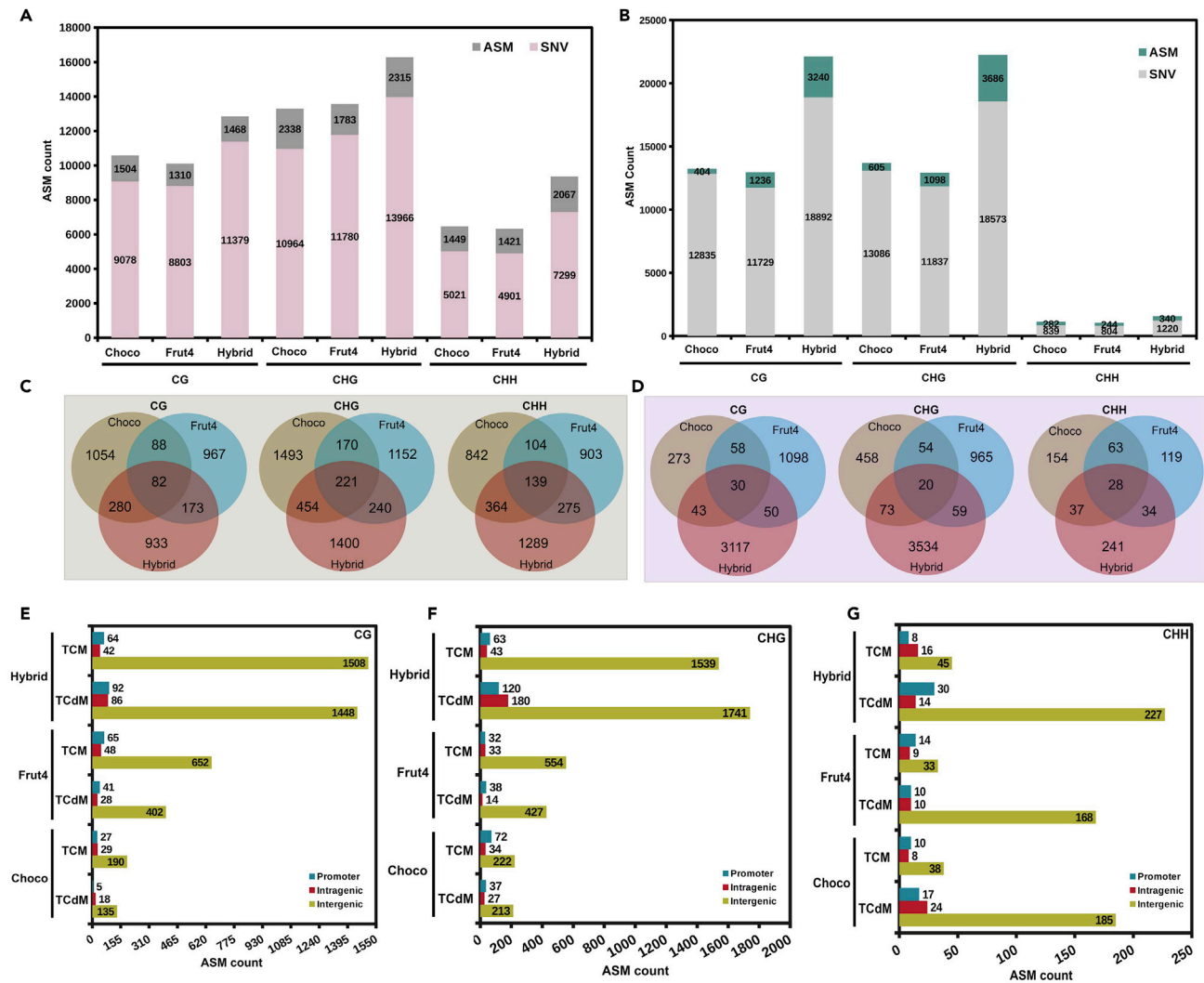


**Figure 4. Genomic distribution of differentially methylated regions (DMRs) at three methylation context**

Barplots representing DMRs distribution at intergenic, intragenic, and promoter regions in (A) Choco vs hybrid, (B) Frut4 vs hybrid, and (C) Frut 4 vs Choco.

CG, 605 CHG, 282 CHH), Frut4 (1236 CG, 1098 CHG, 244 CHH) and their  $F_1$  hybrid (3240 CG, 3686 CHG, 340 CHH, Figure 5B and Tables S10D–S10F). Furthermore, when comparing the ASM between the hybrid and its parents, a total of 3577 CG, 5130 CHG, and 3916 CHH ASM associated with promoter DMRs were identified collectively (Figure 5C), while, a total of 4669 CG, 5163 CHG and 680 CHH ASM associated with TCM and TCdM DMRs were collectively identified from the parental lines and hybrid (Figure 5D). Furthermore, at CG and CHG context, the majority of ASM associated with TCM and TCdM DMRs were from intergenic regions and both parents had comparatively lower number of ASM than hybrid (Figures 5E and 5F). Although,

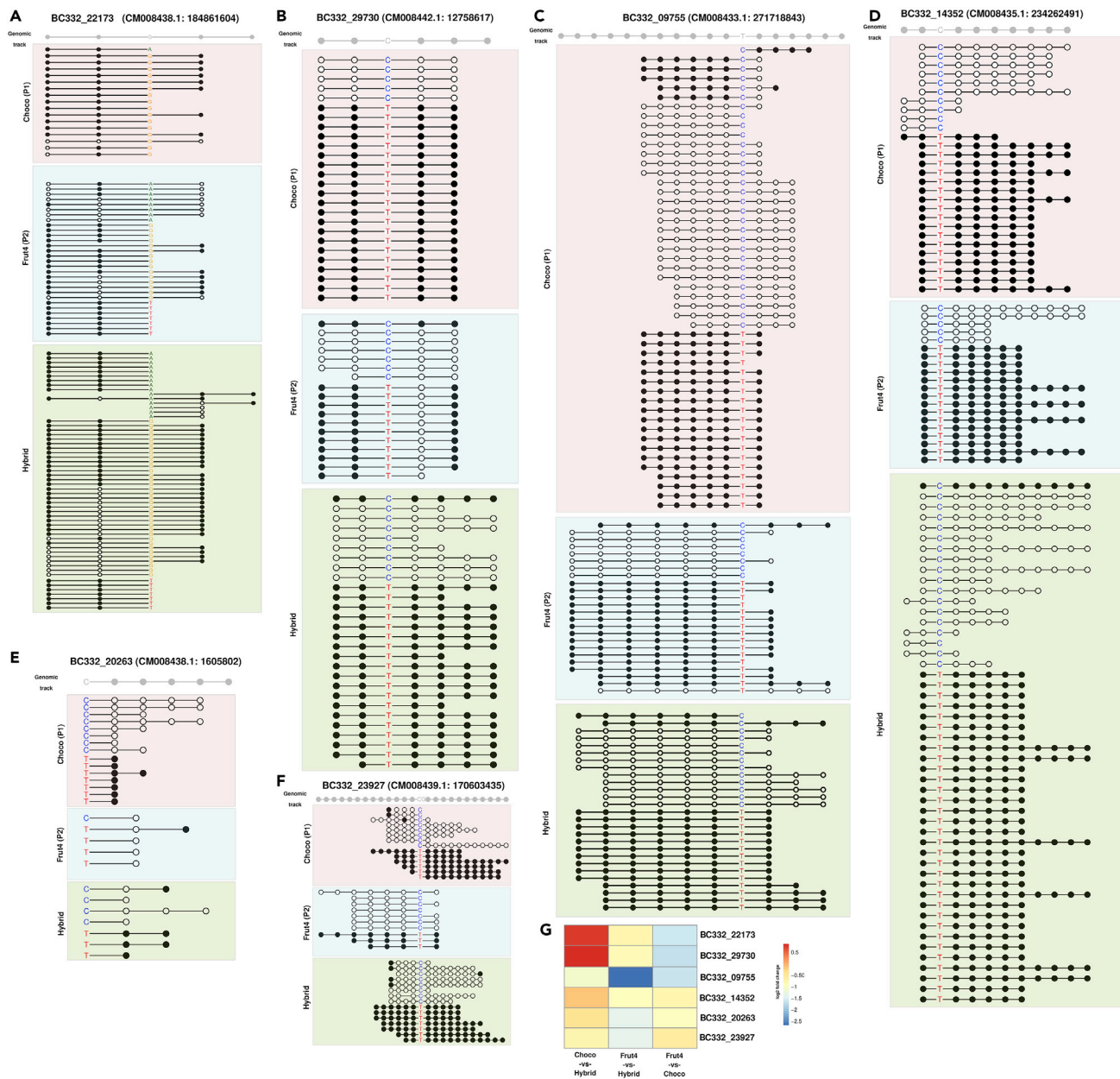




**Figure 5. Summary of allele-specific methylation (ASM) and single nucleotide variants (SNV) identified in *Capsicum* F<sub>1</sub> hybrid and its parental lines at CG, CHG, and CHH methylation context**

SNV and identified significant ASM associated with (A) Promoter DMRs (pDMRs), and (B) *trans* chromosomal methylation (TCM) and demethylation (TCdM) DMRs, common and unique ASM from (C) pDMRs and (D) TCM/TCdM DMRs, (E, F, and G) ASM distribution across the genomic region (promoter, intragenic and intergenic) at three different methylation context between the F<sub>1</sub> hybrid and its parental lines.

at CHH context, the number of intergenic ASMs associated with TCM DMRs was lower than TCdM DMRs in both parental lines and their F<sub>1</sub> hybrid (Figure 5G). Besides, the ASM showed a differential methylation pattern between both *Capsicum* parental lines and their F<sub>1</sub> hybrid. For instance, ASM with G>A/T alleles in the promoter region of gene *BC332\_22173* showed preferentially high methylation in the hybrid compared to both parents and it was observed that methylation in hybrid was preferentially affected by the A/T allele from parent Frut4 (Figure 6A). Also, the expression of the gene showed the overall negative correlation with the promoter methylation (Figure 6D). Another genes such as *BC332\_29730*, *BC332\_09755*, *BC332\_14352*, *BC332\_20263* and *BC332\_23927* with ASM having C>T allele showed preferential contribution of methylation in hybrid from parental lines (Figures 6B–6F). The ASM associated with the promoter of *BC332\_20263* and *BC332\_23927* genes from the glyoxylate and dicarboxylate metabolism pathway showed low or no methylation for allele T in parent Frut4. Also, in hybrid overall methylation for allele T was slightly lower than mid-parent methylation, while the expression of gene was also preferentially high in hybrid than in parent Frut4 (Figure 6G). Additionally, ASM associated with TCM and TCdM DMRs showed preferential methylation and demethylation (Figure 7). For example, in TCdM associated ASM with G>T allele at chr 9 (CM008439.1:4821650), the allele T was hypo methylated in hybrid compared to parent Frut4 (Figure 7A),



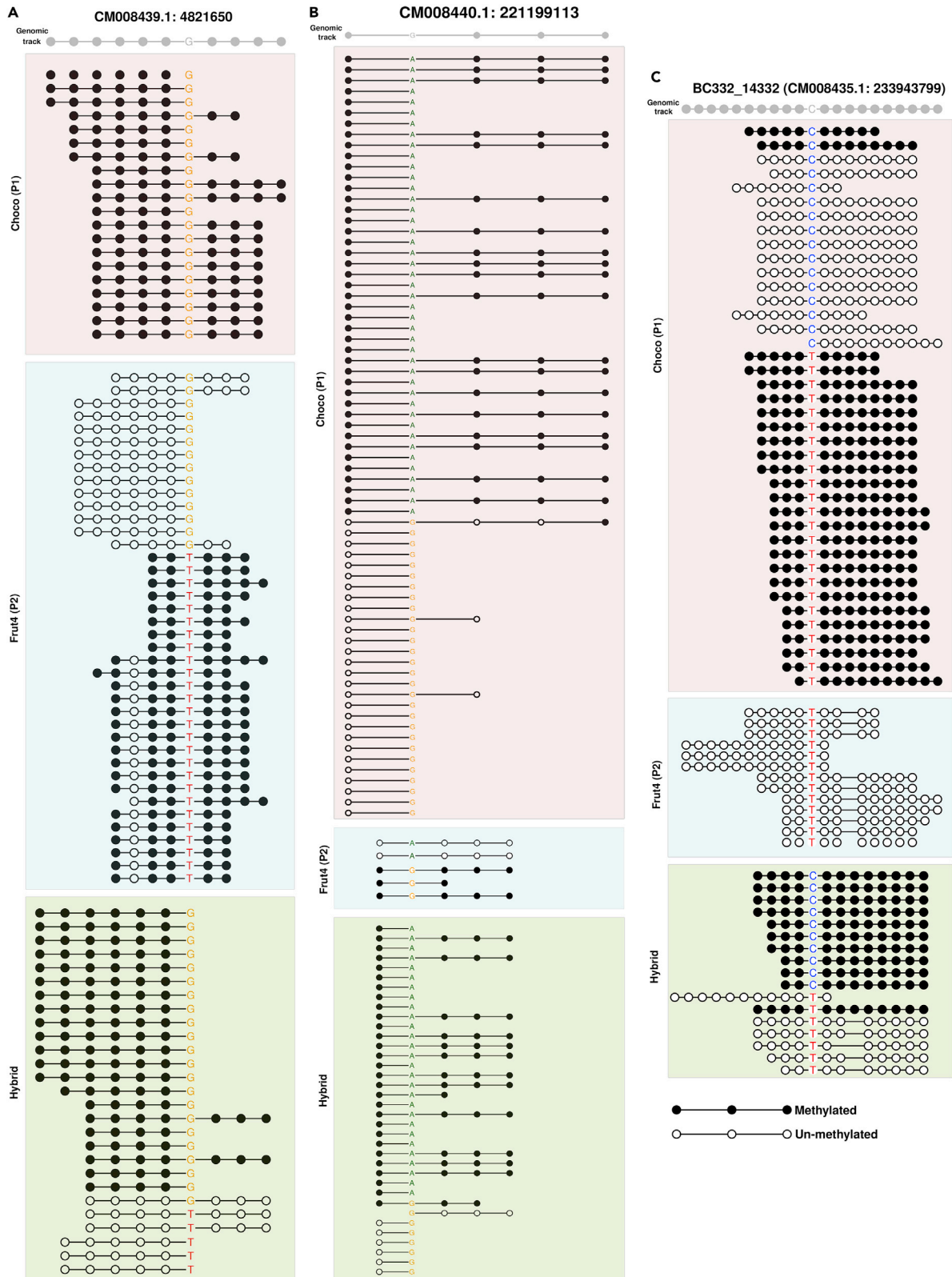
**Figure 6. Allele-specific methylation (ASM) associated with promoter differentially methylated regions (pDMR) genes in *Capsicum* F<sub>1</sub> hybrid and its parental lines.**

(A–F) Representative ASM locus at CG context associated with heterozygous SNV sites in DMRs located in promoters region of genes and (G) expression pattern of genes in both parental lines Choco and Frut4 with respect to F<sub>1</sub> hybrid.

while in ASM with G>A allele at chr 10 (CM008440.1:221199113), the allele G was hypo methylated in hybrid compared to parent Frut4, and the methylation at allele A was transferred to hybrid from parent Choco (Figure 7B). Similarly, in TCM-associated ASM with C>T allele, allele T showed hypo-methylation in both hybrid and parent Frut4 compared to other parent Choco, while allele C was preferentially hyper-methylated in hybrid compared to both parental lines (Figure 7C).

### Expression profiling in *Capsicum* hybrid and its parents

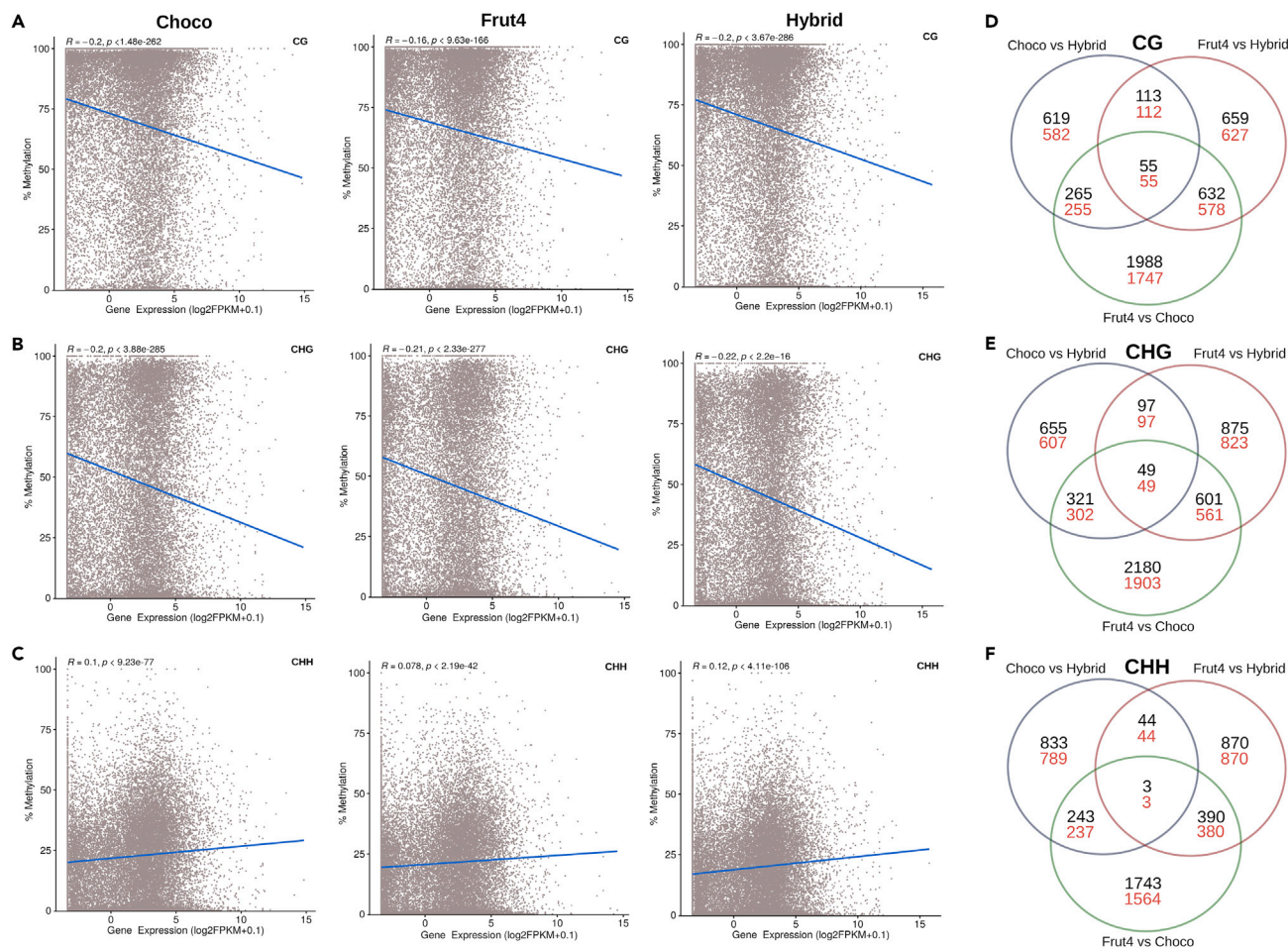
The expression profiling of transcripts and differential expression of genes between the *Capsicum* hybrid and its parents (Choco and Frut4) was analyzed using RNAseq data. A total of 6875 genes were significantly



**Figure 7. Allele-specific methylation (ASM) associated with non-additive trans chromosomal methylation and demethylation (TCM and TCdM) differentially methylated regions (DMRs).**

Representative heterozygous ASM locus with (A) G>T allele and (B) G>A allele associated with TCdM DMRs while (C) C>T allele associated with TCM DMRs in *Capsicum* F<sub>1</sub> hybrid and its parental lines.





**Figure 8. The correlation of promoter methylation with gene(s) expression**

Scatter plot showing a correlation between promoter methylation and gene expression in (A) Choco, (B) Frut4 and (C) hybrid at CG, CHG, and CHH context. Venn diagram representing common and unique promoter DMRs along with their gene count (in red text) among at (D) CG, (E) CHG and (F) CHH methylation context.

(p value < 0.001 and FDR < 0.01) differentially expressed (DE-Gs) between hybrid and its parents. A total of 2003 (1083 up-regulated; 920 down-regulated) genes were DE-Gs between Choco vs hybrid and 4815 DE-Gs (2187 up-regulated; 2628 down-regulated) in Frut4 vs Hybrid, while 4316 DE-Gs (1924 up-regulated; 2392 down-regulated) in Frut4 vs Choco, were significantly expressed (Table S11). Furthermore, it was observed that these DE-Gs were significantly enriched (adjusted p value < 0.01) in different pathways, including photosynthesis (ko00195), glyoxylate and dicarboxylate metabolism (ko00630), photosynthesis-antenna proteins (ko00196), carbon metabolism (ko01200), retinol metabolism (ko00830), and ubiquinone and other terpenoid-quinone biosynthesis (ko00130) (Table S12A). Also, gene ontology (GO) terms from biological processes (BP) such as photosynthesis and light harvesting (GO:0009765, GO:0015979), transport of inorganic anion (GO:0015698), precursor metabolites and energy generation (GO:0006091) and organic acid catabolic process (GO:0016054) were found to be significantly enriched. Although GO terms related to the photosystem I and II (GO:0009521, GO:0009523, GO:0009522) under cellular component (CC) and GO terms such as those related to binding of carbohydrate, chlorophyll, monosaccharide (GO:0030246, GO:0016168, GO:0048029) and related to protein serine/threonine kinase activity (GO:0004674) were significantly enriched (p value < 0.01; FDR < 0.05) under molecular function (MF) category (Table S12B).

### Correlation between methylation instances and expression of genes

Karl Pearson's correlation analysis was performed to understand the correlations between the overall expression of genes and their average methylation in *Capsicum* hybrid and its parents Choco and

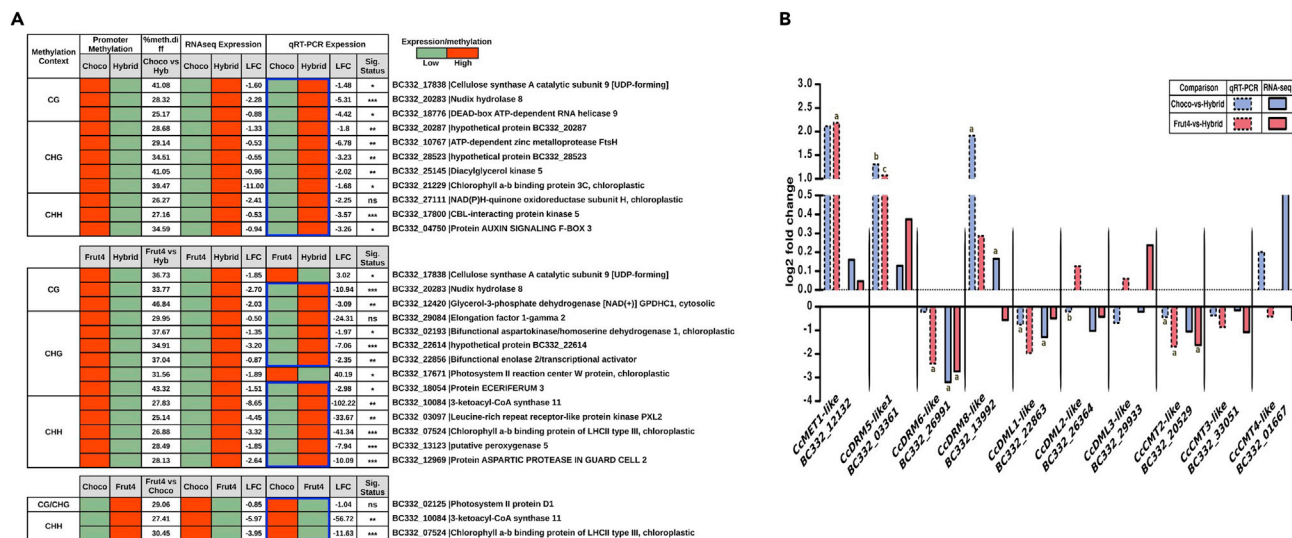
Frut4. Overall, a significant negative correlation ( $R = -0.16$  to  $-0.22$ ;  $p$  value  $< 0.001$ ) between the expression of genes and their CG and CHG methylation instances was observed in all the samples (Figures 8A and 8B). However, no negative correlation was observed at CHH methylation (Figure 8C). Furthermore, pDMRs were screened for their associated genes, and a total of 4332, 4782, and 4127 pDMRs underlying in 3154, 3476, and 3213 genes were identified at CG, CHG, and CHH context, respectively, between hybrid and its parental lines (Tables S13, S14, and S15). Furthermore, 15.3–24.25% pDMR-Genes at CG, 13.8–21.64% pDMR-Genes at CHG and 15.54–25.62% pDMR-Genes were significant DE-Gs between hybrid and its parental lines. Also, only 55 CG, 49 CHG, and 3 CHH pDMRs were common in hybrid and its parental lines (Figures 8D–8F). Out of total pDMRs, 1052 (662 hyper, 390 hypo) CG, 1122 (598 hyper, 524 hypo) CHG, and 1123 (846 hyper, 277 hypo) CHH pDMRs were identified in Choco vs Hybrid (Table S7C). In Frut4 vs hybrid, a total of 1459 (453 hyper, 1006 hypo) CG, 1622 (826 hyper, 796 hypo) CHG, and 1307 (973 hyper, 334 hypo) CHH pDMRs were identified, while, in parents (Frut4 vs Choco), a total of 2940 (809 hyper, 2131 hypo) CG, 3151 (1252 hyper, 1899 hypo) CHG and 2379 (1190 hyper, 1189 hypo) CHH pDMRs were identified (Tables S13, S14, and S15). Additionally, a total of 1854 (496 with DE-Gs) CG pDMRs, 1960 (532 with DE-Gs) CHG pDMRs, and 1800 (463 with DE-Gs) CHH pDMRs showed a negative correlation with the expression of genes between Choco vs hybrid, Frut4 vs hybrid and Frut4 vs Choco comparisons (Tables S13, S14, and S15 and Figures S6A–S6C), respectively. Moreover, these negatively correlated DEG pDMR-Genes were significantly enriched (adjusted  $p$  value  $< 0.01$ ) in pathways (Table S16A), including photosynthesis (ko00195), oxidative phosphorylation (ko00190) as well as were significantly enriched ( $p$  value  $< 0.01$ ;  $FDR < 0.05$ ) in several GO terms such as those related to hydrogen and proton transports (GO:0006818, GO:0015992, GO:1902600) under BP, transmembrane transporter activity of monovalent inorganic cation and hydrogen ions (GO:0015077, GO:0015078) under MF and mitochondrial inner membrane (GO:0005743), organelle envelope (GO:0031967) under CC (Figures S7 and S8A–S8C; Table S16B).

### qRT-PCR expression analysis

The expression of 24 DE-Gs which showed a negative correlation with their promoter methylation at CG, CHG, and CHH context was validated using qRT-PCR in hybrid and its parental lines. Of the 24 genes, 22 (91.66%) showed similar expression patterns with their respective RNAseq expression data (Figure 9A). The expression of these genes was higher while promoter methylation was lower in hybrid than both parents Choco and Frut4. Genes such as *Cellulose synthase A catalytic subunit 9* (BC332\_17838), *Nudix hydrolase 8* (BC332\_20283), *Dead-box ATP-dependent RNA helicase 9* (BC332\_18776), and *Glycerol-3-phosphate dehydrogenase* (BC332\_12420) at CG context, *Bifunctional aspartokinase/homoserine dehydrogenase 1* (BC332\_02193), *ATP-dependent zinc metalloprotease FtsH* (BC332\_10767), *Photosystem II reaction center W protein* (BC332\_17671), *Eceriferum 3* (BC332\_18054), *hypothetical protein BC332\_20287*, *Chlorophyll a-b binding protein 3C* (BC332\_21229), *hypothetical protein BC332\_22614*, *Bifunctional enolase 2/transcriptional activator* (BC332\_22856), *Diacylglycerol kinase 5* (BC332\_25145), *hypothetical protein BC332\_28523*, and *Elongation factor 1-gamma 2* (BC332\_29084) at CHG context, while *Leucine-rich repeat receptor-like protein kinase PXL2* (BC332\_03097), *protein AUXIN SIGNALING F BOX 3* (BC332\_04750), *Chlorophyll a-b binding protein of LHCII type III* (BC332\_07524), *3-ketoacyl-CoA synthase 11* (BC332\_10084), *protein ASPARTIC PROTEASE IN GUARD CELL 2* (BC332\_12969), *putative peroxygenase 5* (BC332\_13123), *CBL-interacting protein kinase 5* (BC332\_17800), and *NAD(P)H-quinone oxidoreductase subunit H* (BC332\_27111) at CHH context, had significantly ( $q$ -value  $< 0.001$ ) high promoter methylation in both Choco and Frut4 compared to hybrid. Also, genes such as *Photosystem II protein D1* (BC332\_02125) at CG and CHG context, *Chlorophyll a-b binding protein of LHCII type III* (BC332\_07524) at CHH context, had significantly ( $q$ -value  $< 0.001$ ) high promoter methylation in Frut4 compared to Choco. The qRT-PCR expression trend of these genes (except BC332\_17838, BC332\_17671 at CG and CHG in Frut4 vs Hybrid) between hybrid and both parents was similar to their RNAseq expression data. Also, the qRT-PCR expression of all genes was significantly different ( $p$  value  $< 0.05$ ,  $p$  value  $< 0.01$  and  $p$  value  $< 0.001$ ) among samples, except for genes BC332\_27111 in Choco vs hybrid, BC332\_29084 in Frut4 vs hybrid and BC332\_2125 in Frut4 vs Choco - which did not show a significant difference in their expression (Figure 9A).

Additionally, we have examined the expression patterns of 10 genes belonging to enzymes including DNA methyltransferases (METs), chromomethylases (CMTs), demeter-like (DML), and domain rearranged methyltransferases (DRMs) using qRT-PCR and RNA-seq analysis (Figure 9B). We observed similar expression patterns of 8 out of 10 genes in both the expression analysis. The expression pattern of genes such as





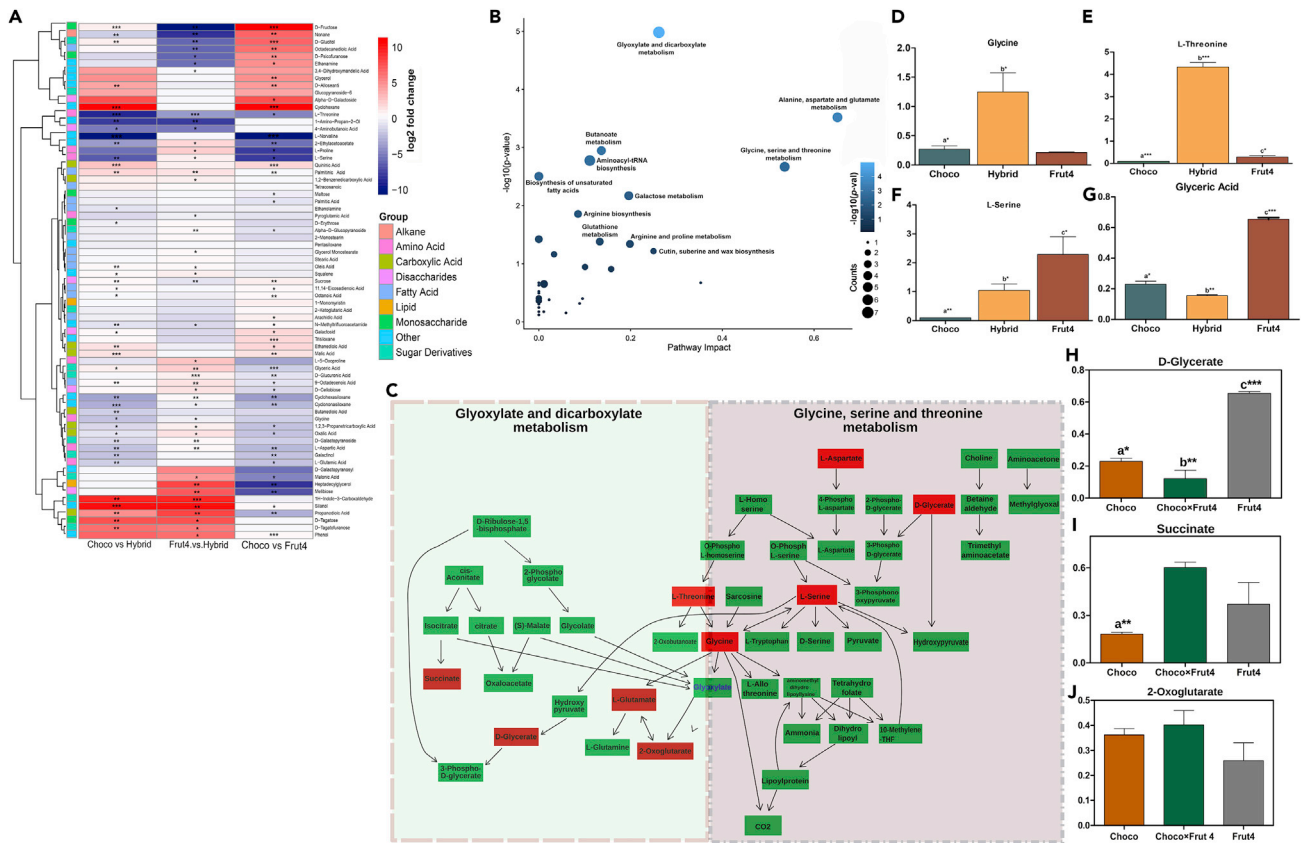
**Figure 9. Gene expression observed in quantitative real-time (qRT) PCR analysis**

(A) Twenty-four (24) significantly ( $p$  value  $< 0.01$ ) differentially expressed genes (DE-Gs) showing negative correlation with promoter methylation in parental line Choco, Fru4 and their hybrid. Genes showing qRT-PCR expression similar to their RNAseq data were represented in blue boxes. LFC represents expression in log<sub>2</sub> fold change and significant  $p$  values were denoted by \*  $< 0.05$ ; \*\*  $< 0.01$ ; \*\*\*  $< 0.001$ . (B) Differential expression pattern of DNA methyltransferase and demethylase genes in parental lines Choco and Fru4 compared to F<sub>1</sub> heterotic hybrids analyzed using RNA sequencing (RNA-seq) and qRT-PCR expression data. Letters a, b, and c shown on bar indicate the significant expression difference of a gene between each parent and hybrid with  $p$  value less than 0.001, 0.01, and 0.05, respectively.

*CcMET1* (BC332\_12132) and *CcDRM5-like 1* (BC332\_03361) was higher in parental lines compared to F<sub>1</sub> hybrid and was significantly different between Choco-vs hybrid and Fru4-vs-hybrid ( $p$  value  $< 0.001$ ;  $p$  value  $< 0.01$ ). Similarly, but opposite expression pattern (low in parental lines than hybrid) of genes including *CcDRM6-like* (BC332\_13992), *CcDML1-like* (BC332\_22864), *CcCMT2-like* (BC332\_20529) and *CcCMT3-like* (BC332\_33051) was observed.

### Significant metabolites in hybrid

Our GC-MS analysis identified more than 100 metabolites, of which only 71 were found in both parents and hybrid were used for further analysis. The overall changes in metabolic content between parents and their hybrid were analyzed in term of log<sub>2</sub> fold change or ratio of normalized area. Most commonly altered metabolites in hybrid compared to parental genotypes were amino acids and their derivatives, and carboxylic acid and sugar derivatives groups. For instance, L-proline, L-serine, L-threonine, L-5-oxoproline, 1-amino-propan-2-ol and 4-aminobutanoic acid from the amino group, 1,2,3-propanetricarboxylic acid from the carboxylic group and galactinol from sugar derivatives were significantly different in hybrid compared to parents (Figure 10). Also, metabolites such as octanoic acid, sucrose, palmitic acid, quininic acid, cyclohexane, D-fructose, and silanol were significantly higher ( $p > 0.001$ ) in the parent Choco than the hybrid, while most of the sugar and its derivatives including D-fructose, D-glucitol, sucrose, D-psicofuranose, and amino acid L-threonine were significantly higher in hybrid compared to the parent Fru4 (Figure 10A). Moreover, we have observed that two pathways namely glyoxylate and dicarboxylate metabolism and glycine, serine, and threonine metabolism were most significantly ( $p$  value  $< 0.01$ ) enriched (Figures 10B and 10C). In hybrid, metabolites such as glycine, L-threonine, succinate, and 2-oxoglutarate showed significantly ( $p$  value  $< 0.05$  &  $p$  value  $< 0.01$ ) high response ratio compared to both the parental genotypes, while metabolites L-serine, glyceric acid, and D-glycerate showed low response ratio than in both the parental genotypes (Figures 10D–10J). In addition, few metabolites showed slightly higher content in the F<sub>1</sub> hybrid compared to both parents but were not significantly different from either or both of the parental genotypes. For example, the content of 2-ketoglutaric acid (2-oxoglutarate) was higher in hybrid compared to both parents but was not significantly different (Figures 10A and 10J). Likewise, the content of metabolite butanedioic acid (succinate) was higher in the F<sub>1</sub> hybrid than both parental genotypes but was only significantly different ( $p$  value  $< 0.01$ ) compared to parent Choco (Figures 10A and 10I).



**Figure 10. Global metabolite profiling in hybrid and parental genotypes**

(A) Heatmap representing metabolite changes in term of log<sub>2</sub> fold change (ratio) of normalized area, (B) pathway enrichment of metabolites commonly identified in hybrid and parental genotypes, (C) pathway map of significantly enriched pathways, and (D–J) metabolites with significantly high and low response ratio compared to both parents Choco and Fru4. Metabolite response difference between Choco vs. hybrid, Fru4 vs. hybrid and Fru4 vs. Choco are represented as a, b, and c, respectively and the significant level are indicated as \* for p value <0.5, \*\* for p value <0.01 and \*\*\* for p value <0.001.

### Correlation analysis of promoter methylation with the expression of genes and metabolites

The integrated analysis between significant DE-Gs with their promoter methylation and metabolites biosynthesis pathway was performed to find the plausible impact of expression of genes and/or promoter methylation in *Capsicum* hybrid and parental genotypes. Our analysis observed the expression of 43 significant DE-Gs which were enriched in glyoxylate and dicarboxylate metabolism pathways. Of the total, 28 and 31 DE-Gs showed increased or higher expression in F<sub>1</sub> hybrid than parent Choco and Fru4, respectively (Figure S9). Also, the expression of 28 DE-Gs was higher in hybrid than the mid-parent expression value. Few genes including *Glycerate dehydrogenase*, *Glutamine synthetase*, *Glycine cleavage system H protein*, *Aminomethyltransferase*, *Aconitate hydratase*, and *Serine hydroxymethyltransferase* showed 1-3 fold higher expression than mid-parent expression values, while genes such as *ribulose bisphosphate carboxylase*, and *Malate dehydrogenase* showed 2-47 fold higher expression than mid-parental expression value. Furthermore, the promoter regions of these genes was scrutinized for methylated cytosine which revealed that the upstream promoter regions of 14 genes including *Glutamine synthetase*, *Glycine cleavage system H protein*, *Serine hydroxymethyltransferase*, *ribulose bisphosphate carboxylase*, and *Malate dehydrogenase* were less/low methylated than mid-parental methylation value (MPMV) and hence showed higher gene expression of those genes compared to the mid-parent expression values (Figure S9).

### DISCUSSION

Hybrid breeding is one of the most preferred methods in many crop plants owing to its heterotic performance and higher yield, among many advantages. However, the detailed mechanism of the heterotic performance of F<sub>1</sub> hybrids compared to parental lines is not still understood well especially with respect to the

DNA methylation and its correlation with the expression of genes and metabolites. In this study, we developed an interspecific hybrid by crossing genotypes (Choco and Frut4) belonging to two different *Capsicum* species and found heterosis in the developed hybrid (Figure 1). Through earlier, different genetic models like dominance and overdominance hypothesis have been proposed to explain hybrid vigour/heterosis in F<sub>1</sub> hybrids (Jones, 1917; Crow, 1948, 1999; Schnable and Springer, 2013); however, these models do not provide sufficient explanation or prediction of heterosis. Recently, the role of DNA methylation in hybrid vigour has also been reported in several important plants, including *Arabidopsis* (Shen et al., 2012; Kawana et al., 2016), tomato (Raza et al., 2017), rice (He et al., 2010; Chodavarapu et al., 2012), pigeon pea (Junaid et al., 2018; Sinha et al., 2020), and *Populus* (Gao et al., 2014). Using methylation-sensitive amplified polymorphism (MSAP) analysis, Xu et al. (2015) reported the adjustment of methylation patterns in heterotic reciprocal heterotic hybrids of two hot peppers. In the present study, we conducted an integrated high-resolution mapping of genome-wide DNA methylation at single-base resolution in interspecific *Capsicum* hybrid and its parental genotypes. This study also identified gene(s) expression signatures and showed expression correlation with their promoter methylation in both parental lines and their F<sub>1</sub> hybrid. Therefore, this study may prove to be a useful asset for research and investigations of molecular mechanisms influencing the heterotic phenomenon in *Capsicum*.

DNA methylations are highly variable in three different contexts i.e. CG, CHG, and CHH across the different plant species (Alonso et al., 2015; Niederhuth et al., 2016; Takuno et al., 2016). For instance, *B. vulgaris* showed maximum (~81%) methylations at CHG; however, *Eutrema salsugineum* showed the lowest (9.3%) methylation at CHG context (Niederhuth et al., 2016). In the case of tomato, CG methylations were prominent (71.6–72.8%) followed by CHG (52.5–53.0%) and CHH (10.7–12.5%) methylations (Zuo et al., 2018). Earlier 82.8–89.6% CG, 77.8–83.9% CHG, and 22.4–25% CHH methylation was reported in fruits of three different *Capsicum* species, i.e. *Capsicum annuum*, *C. chinense*, and *C. frutescens* (Rawoof et al., 2020). The present study also identified similar methylation i.e. 87.9–90.6% CG, 78–83.7% CHG and 17.2–21.1% CHH methylation context in leaves of hybrid and its parental lines (Figure 2A). In *Arabidopsis*, it was reported that CMT2 and RNA-directed DNA methylation (RdDM) facilitate and maintain methylation in the CHH context (Zemach et al., 2013). In hybrid, it was observed that the fractional proportion of CHH methylation was the lowest than both Choco and Frut4 parental lines, which suggested that CMT2 and RdDM mediated CHH methylation instances were less active in *Capsicum* hybrid (Figure 2B). Patterns of DNA methylation are ubiquitous in nature and occur throughout the genome, including genic regions, inter-genetic regions, transposable elements; DNA elements and so forth and their density/distribution across these regions may provide biological information regarding functional processes and methylation regulation (Takuno and Gaut, 2012, 2013). Furthermore, it was reported that promoter methylation affects the expression/transcription of genes. Also, high CHG and CHH methylation in flanking regions than gene-body was reported in fruits of *Capsicum* (Rawoof et al., 2020). In *Arabidopsis* and rice, high CHG methylation at GBM along with flanking regions was observed with low transcript/gene expression (Saze et al., 2008; Zemach et al., 2010; Klosinska et al., 2016). A similar GBM pattern was observed in *Capsicum* hybrid and its parental lines Choco and Frut4 (Figure 3A). However, in present study, the hybrid showed the lowest CHG and CHH GBM compared to both parental lines, suggesting methylation via RdDM pathway is less active in the *Capsicum* hybrid. One of the most important functions of DNA methylation is to stabilize transposable elements (TEs) (Chan et al., 2005; Law and Jacobsen, 2010). In this study, it was found that TE body along with flanking region of the majority of TEs including LTRs, non-LTRs, and DNA elements were densely methylated (Figures 3B–3F and S3), suggesting their possible role in genome stability. Methylation at both TE-body and TE flanking regions regulates transposon silencing (Inagaki and Kakutani, 2012). It was observed that TE-body methylation was higher than GBM (Figures 3 and S3), which was similar to earlier studies (Rawoof et al., 2020; Sinha et al., 2020); however, at CHH context, methylation in both GBM and TEs is lower in *Capsicum* hybrid than its parents Choco and Frut4, suggesting that *de novo* maintenance of CHH methylation instances in TEs and GBM via CMT2 pathway is less active in *Capsicum* hybrid.

The interaction between parental genomes and epigenomes within the nuclei of hybrid alters the cytosine methylation patterns (Chodavarapu et al., 2012; Greaves et al., 2012; Shen et al., 2012). These alterations of mCs occur at particular chromosomal segments/regions which are differentially methylated (DMRs) between hybrid and parents (Greaves et al., 2014). In this study, we observed that at CG and CHG context, hypo and hypermethylated DMRs were the most abundant in hybrid when compared with parent Choco and Frut4, respectively (Figures S3 and S4), while at the CHH context, hybrid had a higher abundance of hypo DMRs than both parents (Figures S3 and S4). It was reported that DNA methylations are associated

with regulatory function and are more prominent in intergenic regions than in genic regions (He et al., 2010). In this study, the genomic distribution of DMRs showed that the majority of DMRs in *Capsicum* hybrid and its parental lines belonged to the intergenic regions, followed by promoter regions and genic regions (Figures 4 and S5), which indicated their potential role in gene regulation and heterosis (Zhao et al., 2007; Jin et al., 2008; He et al., 2010; Greaves et al., 2012; Sinha et al., 2020).

There are two types of genetic regulators, first consisting of *cis*-acting regulators including promoter sequences, and the other being *trans*-acting regulators like transcription factors (Springer and Stupar, 2007). It has been suggested that *trans*-acting elements are mainly involved in non-additive variations; however, additive variations may be caused by both *cis*- and *trans*-acting elements (Swanson-Wagner et al., 2006; Springer and Stupar, 2007; Guo et al., 2008; He et al., 2010). The non-additive methylation changes in *Capsicum* hybrid were attributed to TCM and TCdM whereas compared to the parental segments; methylation in hybrid was higher in case of TCM and lower in case of TCdM (Greaves et al., 2014). Both TCM and TCdM are considered major epigenetic factors driving wide changes in hybrid methylome; however, their possible role in transcript expression regulations in heterotic hybrids remained unclear (Greaves et al., 2014). During the present study, 4891 TCM (2272) and TCdM (2619), majorly located in intergenic regions of the genome, were identified in *Capsicum* hybrid, and only 33.08% (1618) were significantly different from mid-parent-methylation (Table S8), suggesting significant *trans*-acting effects in *Capsicum* hybrid that may play an important role in heterosis.

In plants, the methylation of the DMRs in the promoter of genes not only involved in the alteration of gene expression but is also used to understand the maintenance of parental variations in hybrid (Zemach et al., 2010; Ma et al., 2021). This study identified several allelic methylation associated with pDMR gene which was preferentially methylated in heterotic *Capsicum* hybrid (Figures 5, 6, and 7). It was reported that the genetic diversity of parental genomes even at single nucleotide polymorphism might impact the expression of functional genes in hybrid owing to the expression of either of the parental allele and thus may contribute to heterosis (Botet and Keurentjes, 2020; Zebell, 2021). This study observed that the expression of *Malate dehydrogenase* (BC332\_23927) gene from glyoxylate and dicarboxylate metabolism pathway in F<sub>1</sub> hybrid was higher than mid-parent expression, while its promoter region was hypo-methylated than mid-parent methylation (Figures 6F and S9). Additionally, the allele T with ASM (C>T) was preferentially high methylated than one of the parental alleles from Frut4 (Figure 6F). Likewise, another gene *Glycine cleavage system H protein* (BC332\_20263) showed significantly higher expression in hybrid than both the parental lines (Figure 6G), while in its promoter region, methylation of allele T with ASM (C>T) was contributed from parental allele from Choco (Figure 6E), suggesting that parental allele associated with pDMRs with their corresponding gene expression might be involved in heterosis, however, its molecular mechanism remains elusive. Earlier in rice, it was reported that parental alleles associated with CG and CHG DMRs were stably transferred in hybrid (Ma et al., 2021). Therefore, to get insight into the conservation and transfer of the parental DMRs (TCM & TCdM) in the *Capsicum* hybrid, this study analyzed the methylation at the single base allele (ASM) associated with these TCM and TCdM DMRs in hybrid and its parental lines (Figures 7A–7G). This study observed that at both CG and CHG contexts, high number of ASM were associated with TCM and TCdM DMRs, suggesting that the variations at CG and CHG methylation contexts are substantially maintained for parental alleles in F<sub>1</sub> *Capsicum* hybrids, which is in concordance with the previous finding.

The stability of newly formed methylation instances in F<sub>1</sub> hybrid and the inheritance of methylated cytosines from parental lines to successive generations can be assessed using WGBS at single base resolution (Becker et al., 2011; Schmitz et al., 2011; Van Der Graaf et al., 2015). In *Arabidopsis*, the inheritance of ~45% of newly methylated sites was stable and maintained between generations (Hofmeister et al., 2017). In *Capsicum* hybrid, the majority of mCs (78.9%) were inherited from parent Choco, while only 21.1% were inherited from Frut4 (Table S9), suggesting major mCs in hybrid were inherited from parent Choco. The use of Choco as female parent in the development of hybrid could be one of the reasons for the higher proportion of mCs (78.9%) from Choco than Frut4. Furthermore, only a small fraction (9% of total inherited mCs) of newly formed methylated sites was identified in hybrid and of which ~76% of newly formed mCs belonged to the CHH context and only ~23% mCs originated at CG and CHG context (Table S9). Altogether, our data suggested that although the majority of methylations are stably inherited across the generations as suggested previously (Zhang et al., 2008; Becker et al., 2011; Van Der Graaf et al., 2015; Hofmeister et al., 2017), however, newly formed methylations and demethylation events in hybrid may play a possible role in providing heterosis to the plants which need to be further confirmed.

Hitherto, a number of studies have suggested that DNA methylation may differently regulate the transcript expression in hybrids and may contribute to the heterosis (Greaves et al., 2015; Zhang et al., 2016; Raza et al., 2017; Shen et al., 2017; Junaid et al., 2018; Sinha et al., 2020). Therefore, to understand the role of DNA methylations on gene expressions, we developed a complex of dynamic methylome and transcriptome and the associated gene functions; and thus conducted transcriptome analysis of the same samples. In *Capsicum* hybrid, ~46% significant DE-Gs were up-regulated compared to parent Choco, while ~55–56% significant DE-Gs were up-regulated in both hybrid and Choco compared to the other parent Frut4 (Table S11). Also, it was observed that several DE-Gs with higher expression in hybrid were having low promoter methylation than both parental genotypes, suggesting methylation may involve in expression regulation during the manifestation of heterosis in *Capsicum*. For instance, in hybrid DEG *DEAD-box ATP-dependent RNA helicase* which was reported to be involved in plant development and abiotic stress responses (Liu and Imai, 2018; Macovei et al., 2012) was observed with higher expression. Similarly, DEG *D. kinase (DGK)* which is a key enzyme involved in lipid signaling, signal transduction from hormones, growth factors, and in build-up metabolic network in responses to several stresses (Kue Foka et al., 2020), was also found to be highly expressed in *Capsicum* F<sub>1</sub> hybrid in both parental genotypes. Likewise, in hybrid around 2-fold increased expression of DEG *B. enolase 2/transcriptional activator* which is important for plant growth and development (Eremina et al., 2015) was observed than both parents (Figure 9A). In *Arabidopsis*, heterotic phenotypes showed an increment in the total amount of photosynthesis than parental lines (Fujimoto et al., 2012; Kawanabe et al., 2016). This study observed that photosynthesis-related gene *Chlorophyll a/b binding protein of Light-Harvesting Complex (LHC) II* was highly expressed in hybrid (4–11 LFC) than both parents and at the same time it showed low promoter methylation than MPMV of both parents suggesting increased photosynthesis in heterotic F<sub>1</sub> hybrid with promoter methylation may interplay during heterosis in *Capsicum* (Figure 9A). The enrichment analysis showed that DE-Gs in *Capsicum* hybrid and parental lines were significantly enriched in pathways and GO terms related to photosynthesis, glyoxylate and dicarboxylate metabolism, and metabolite biosynthesis (Table S12), which was similar to earlier reports in hybrids (Fujimoto et al., 2012; Kong et al., 2020; Li et al., 2020a, 2020b).

In this study, we also observed significant enrichment of several metabolites from sugar and sugar derivatives, amino acid and their derivatives, and carboxylic acid group between hybrid and parental genotypes (Figure 10). A similar observation was reported by recent studies on *Capsicum* hybrids wherein they observed the increased metabolite (mainly sugar, citric acid, and capsaicinoids) content in hybrids compared to parental genotypes (Zamljen et al., 2020; Shiragaki et al., 2021). Also, our metabolite profiling analysis showed that metabolites related to glyoxylate and dicarboxylate metabolism pathway were most enriched ( $p$  value < 0.01) between *Capsicum* hybrid and parents (Figure S8; Table S12), and it was observed that the expression of genes (~65–72% of enriched genes) related to this pathway were also higher in F<sub>1</sub> hybrid than mid-parent average expression values of parental genotypes (Figure S9). The promoter methylation analysis showed that around 32.6% of total significant DE-Gs enriched in glyoxylate and dicarboxylate pathway were having low promoter methylation than MPMV and increased expression than both parental genotypes (Figure S9), which suggested that, methylation in hybrid might potentially involve in the regulation of gene expression thereby increasing certain metabolite content in hybrid leading to early heterosis in *Capsicum* F<sub>1</sub> hybrid.

Additionally, the overall significant negative correlation of promoter methylations with gene expression (Figure 8) suggested the overall repressive nature of DNA methylations which is also widely reported in different plant species (Zilberman et al., 2007; Zemach et al., 2010; Lang et al., 2017; Raza et al., 2017). Earlier, it was reported that promoter CHH methylation positively correlated with the gene expression (Bhatia et al., 2018; Rawoof et al., 2020) and similar results were observed in *Capsicum* hybrid and parental lines; however, the molecular mechanism behind this correlation is still unknown. To get insight into the regulation of expression in hybrid, we investigated the promoter DMR (pDMR) of genes and observed that the majority of genes showing increased expression had low/decreased promoter methylation in the hybrid compared to their parents (Figure S6). Also, these pDMR DE-Gs were significantly enriched in photosynthesis and oxidative phosphorylation-related pathways (Table S16; Figures S7 and S8), suggesting DNA methylation in *Capsicum* hybrid may be involved in enhanced photosynthesis and may contribute to heterosis, which further needs to be investigated. Furthermore, we have validated the expression of randomly selected 24 DE-Gs which notably have low promoter methylation and high expression in hybrid compared to both parental lines (Figure 9A). The qRT-PCR expression analysis also showed that most of these genes had significant differential expression in hybrid, which was much more than that observed



in both parental lines and suggested that promoter methylation in *Capsicum* hybrid is correlated with altered expression of genes.

In plants, the establishment of DNA methylation at all context is carried out by DNA methyltransferase Domain Rearranged Methyltransferase 2 (DRM2) via siRNA-guided RdDM pathways, while, methylation maintenance at CG, CHG, and CHH context is carried out by DNA methyltransferase 1 (MET1), chromomethylase 3 (CMT3), and CMT2, respectively. Additionally, the demethylation of methylated cytosine is actively arbitrated by several DNA demethylases including demeter (DME), demeter-like (DML), and repressor of silencing 1 (ROS1) (Liang et al., 2019). Therefore, to understand the expression dynamics of the DNA methylase and demethylase genes, we have analyzed the expression of 10 different genes (1 MET1, 3 DRM, 3 DML, and 3 CMT) using RNA-seq and qRT-PCR analysis (Figure 9B). Of these, 8 genes including *CcMET1-like* (BC332\_12132), *CcDRM5-like 1* (BC332\_03361), *CcDRM6-like* (BC332\_26991), *CcDML1-like* (BC332\_22863), *CcDML3-like* (BC332\_29933), *CcCMT2-like* (BC332\_20529), *CcCMT3-like* (BC332\_33051), and *CcCMT4-like* (BC332\_01667) showed similar expression trends among the samples in both RNA-seq and qRT-PCR analysis. We observed that compared to F<sub>1</sub> hybrid, both the parents (Choco and Frut4) showed the upregulation of *CcMET1-like*, while *CcCMT4-like* was upregulated in parent Choco and downregulated in Frut4 compared F<sub>1</sub> hybrid. Also, in Arabidopsis hybrids generated from met1-RNAi knockdown, it was observed that MET1 is not involved in heterosis (Kawanabe et al., 2016). In this study, we observed that the high expression of MET1 and CMT4 with relatively low CG and CHG methylation in F<sub>1</sub> hybrid than parental lines and their role in heterosis needs further investigation. Furthermore, the increased expression of *CcCMT2-like* than the mid-parent expression and additional newly formed mCs (9% of total inherited mCs), chiefly from CHH context was observed in the F<sub>1</sub> hybrid (Table S9). Previously, it was reported that in addition to CHH methylation maintenance via RdDM, CMT2 along with a decrease in DNA methylase 1 (DDM1) can alternatively also be involved in the generation of new methylation marks (Greaves et al., 2015; Zemach et al., 2013) and in corroboration with, our observation suggests that CMT2 plausibly might be involved in the establishment of new mCs partially in F<sub>1</sub> *Capsicum* hybrid, which in turn might lead to heterosis.

In conclusion, the present study identified the inheritance pattern of DNA methylation marks from parent to hybrids. Some of the regions were hyper/hypo methylated with significant MPMV in hybrids as compared to parental genotypes. Expression dynamics and global metabolite profiling suggested the potential role of cross-talk between the three components (methylation, expression of genes, and metabolites) in the development of early heterosis in the F<sub>1</sub> hybrid of *Capsicum*. The expression analysis of genes identified a significant number of DE-Gs in hybrids which overall showed negative correlations with the promoter methylation. Furthermore, our metabolite profiling showed significantly higher expression of metabolites such as glycine, carboxylic acid, D-fructose, D-glucitol, sucrose, D-psicofuranose, and L-threonine in hybrid. Glyoxylate and dicarboxylate metabolism pathway was significantly enriched for metabolites and DE-Gs in F<sub>1</sub> hybrid and parental lines. Of the total, 65–72% enriched genes in this pathway showed overall increased expression (of these 32.6% with low promoter methylation than MPMV) in F<sub>1</sub> hybrid than mid-parent value. Altogether, the present study provided insightful dimensions into the plausible integrated role of DNA methylation, and expression of genes and metabolites toward the early heterosis in *Capsicum* hybrid.

### Limitations of the study

Out of two *Capsicum* spp used to develop hybrid, the reference genome was available only for *C. chinense*, and thus present study might escape the detection of some essential epigenomic regions associated with heterosis that varies among both parental genomes (such as SNP/InDels) can be explored in future on the availability of reference genome of *C. frutescens*. Secondly, the present study focused on early-stage heterosis, however, hybrid vigor for fruit characters is also a potential area to be explored.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability

**METHOD DETAILS**

- Plant material and phenotyping
- Whole-genome bisulfite sequencing and data processing
- RNA sequencing and data processing
- Analysis of differential methylation levels and their correlation with gene expression
- Analysis of allele-specific methylation (ASM) associated with DMRs
- Gene ontology and pathway enrichment analysis
- qRT-PCR expression validation of DE-Gs negatively correlated with promoter methylation
- Untargeted GC-MS based metabolomics analysis

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.105318>.

**ACKNOWLEDGMENTS**

This work was supported by the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Govt. of India in the form ECR grant (ECR/2017/001932, GAP-260) awarded to VJ. CSIR grant (MLP-201) also acknowledged for funding. Partial funding was received as UGC DRS-I to the School of Life Sciences, Jawaharlal Nehru University from the University Grant Commission, India. AR and IA acknowledge the fellowships received from the Council of Scientific and Industrial Research (CSIR), Govt. of India, New Delhi in the form of Senior Research fellowships. VG acknowledges the DST-INSPIRE faculty fellowship. CSIR-IHBT communication number is 5052.

**AUTHOR CONTRIBUTIONS**

VJ and NR conceived and designed research. VJ, AR, VG, IA, SSC, and MD conducted field and lab experiments. AR performed analysis. VJ and AR analyzed and interpreted results data. VJ and AR drafted the manuscript. VJ and NR corrected and finalized the manuscript. All authors read and approved the final manuscript.

**DECLARATION OF INTERESTS**

The authors declare no conflict of interest.

Received: February 21, 2022

Revised: June 4, 2022

Accepted: October 6, 2022

Published: November 18, 2022

**REFERENCES**

- Ahmad, I., Rawoof, A., Dubey, M., and Ramchiary, N. (2021a). ICP-MS based analysis of mineral elements composition during fruit development in *Capsicum* germplasm. *J. Food Compos. Anal.* 101, 103977.
- Ahmad, I., Rawoof, A., Islam, K., Momo, J., and Ramchiary, N. (2021b). Identification and expression analysis of phosphate transporter genes and metabolites in response to phosphate stress in *Capsicum annuum*. *Environ. Exp. Bot.* 190, 104597.
- Ahn, Y.-K., Tripathi, S., Kim, J.-H., Cho, Y.-I., Lee, H.-E., Kim, D.-S., Woo, J.-G., and Cho, M.-C. (2014). Transcriptome analysis of *Capsicum annuum* varieties Mandarin and Blackcluster: assembly, annotation and molecular marker discovery. *Gene* 533, 494–499.
- Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., and Mason, C.E. (2012). methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* 13, R87.
- Alonso, C., Pérez, R., Bazaga, P., and Herrera, C.M. (2015). Global DNA cytosine methylation as an evolving trait: phylogenetic signal and correlated evolution with genome size in angiosperms. *Front. Genet.* 6, 4.
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Aguilar-Meléndez, A., Morrell, P.L., Roose, M.L., and Kim, S.C. (2009). Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum annuum*; Solanaceae) from Mexico. *Am. J. Bot.* 96, 1190–1202.
- Arechederra, M., Daian, F., Yim, A., Bazai, S.K., Richelme, S., Dono, R., Saurin, A.J., Habermann, B.H., and Maina, F. (2018). Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer. *Nat. Commun.* 9, 3164.
- Arnon, D.I. (1949). Copper enzymes in isolated chloroplasts. polyphenoloxidase in *Beta vulgaris*. *Plant Physiol.* 24, 1–15.
- Baenas, N., Belović, M., Ilic, N., Moreno, D.A., and García-Viguera, C. (2019). Industrial use of pepper (*Capsicum annuum* L.) derived products: technological benefits and biological advantages. *Food Chem.* 274, 872–885.
- Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., and Weigel, D. (2011). Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* 480, 245–249.
- Bhatia, H., Khemka, N., Jain, M., and Garg, R. (2018). Genome-wide bisulphite-sequencing reveals organ-specific methylation patterns in chickpea. *Sci. Rep.* 8, 9704.
- Bosland, P.W., and Votava, E.J. (2000). Peppers: Vegetable and Spice Capsicums (CABI Publishing).

- Botet, R., and Keurentjes, J.J.B. (2020). The role of transcriptional regulation in hybrid vigor. *Front. Plant Sci.* **11**, 410.
- Chan, S.W.-L., Henderson, I.R., and Jacobsen, S.E. (2005). Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat. Rev. Genet.* **6**, 351–360.
- Chen, Z.J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* **14**, 471–482.
- Chhakekar, S.S., Brahma, V., Rawoof, A., Kumar, N., Gaur, R., Jaiswal, V., Kumar, A., Yadava, S.K., Kumar, R., Sharma, V., et al. (2020). Transcriptome profiling, simple sequence repeat markers development and genetic diversity analysis of potential industrial crops *Capsicum chinense* and *C. frutescens* of Northeast India. *Ind. Crops Prod.* **154**, 112687.
- Chhakekar, S.S., Kumar, N., M, S., Brahma, V., Rawoof, A., Jaiswal, V., Kumar, A., Sarmah, B.K., Ramchiary, N., and Ramchiary, N. (2021). Profiling of miRNAs in bhut jolokia (*Capsicum chinense*) and kon jolokia (*C. frutescens*) of northeast India. *Sci. Hortic.* **281**, 109952.
- Chodavarapu, R.K., Feng, S., Ding, B., Simon, S.A., Lopez, D., Jia, Y., Wang, G.L., Meyers, B.C., Jacobsen, S.E., and Pellegrini, M. (2012). Transcriptome and methylome interactions in rice hybrids. *Proc. Natl. Acad. Sci. USA* **109**, 12040–12045.
- Crow, J.F. (1948). Alternative hypotheses of hybrid vigor. *Genetics* **33**, 477–487.
- Crow, J.F. (1999). Dominance and overdominance. In *Genetics and Exploitation Heterosis in Crops*, J.G. Coors and S. Pandey, eds., pp. 49–58. <https://doi.org/10.2134/1999.geneticsandexploitation.c5>.
- Dapp, M., Reinders, J., Bédie, A., Balsera, C., Bucher, E., Theiler, G., Granier, C., and Paszkowski, J. (2015). Heterosis and inbreeding depression of epigenetic *Arabidopsis* hybrids. *Nat. Plants* **1**, 15092.
- Dubey, M., Jaiswal, V., Rawoof, A., Kumar, A., Nitin, M., Chhakekar, S.S., Kumar, N., Ahmad, I., Islam, K., Brahma, V., and Ramchiary, N. (2019). Identification of genes involved in fruit development/ripening in *Capsicum* and development of functional markers. *Genomics* **111**, 1913–1922. <https://doi.org/10.1016/j.ygeno.2019.01.002>.
- Eremina, M., Rozhon, W., Yang, S., and Poppenberger, B. (2015). ENO 2 activity is required for the development and reproductive success of plants, and is feedback-repressed by A t MBP-1. *Plant J.* **81**, 895–906.
- Fujimoto, R., Taylor, J.M., Shirasawa, S., Peacock, W.J., and Dennis, E.S. (2012). Heterosis of *Arabidopsis* hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proc. Natl. Acad. Sci. USA* **109**, 7109–7114.
- Gao, M., Huang, Q., Chu, Y., Ding, C., Zhang, B., and Su, X. (2014). Analysis of the leaf methylomes of parents and their hybrids provides new insight into hybrid vigor in *Populus deltoides*. *BMC Genet* **15**, 1–17. <https://doi.org/10.1186/1471-2156-15-S1-58>.
- Gözen, V., Özalp, R., and Çelik, I. (2018). Evaluation of heterosis on yield and fruit number of hybrid bell pepper (*Capsicum annuum* L.) in unheated greenhouse condition. In XXX International Horticultural Congress IHC2018: II International Symposium on Plant Breeding in Horticulture, 1282XXX International Horticultural Congress IHC2018: II International Symposium on Plant Breeding in Horticulture, pp. 75–82.
- Greaves, I.K., Gonzalez-Bayon, R., Wang, L., Zhu, A., Liu, P.-C., Groszmann, M., Peacock, W.J., and Dennis, E.S. (2015). Epigenetic changes in hybrids. *Plant Physiol.* **168**, 1197–1205.
- Greaves, I.K., Groszmann, M., Wang, A., Peacock, W.J., and Dennis, E.S. (2014). Inheritance of trans chromosomal methylation patterns from *Arabidopsis* F1 hybrids. *Proc. Natl. Acad. Sci. USA* **111**, 2017–2022.
- Greaves, I.K., Groszmann, M., Ying, H., Taylor, J.M., Peacock, W.J., and Dennis, E.S. (2012). Trans chromosomal methylation in *Arabidopsis* hybrids. *Proc. Natl. Acad. Sci. USA* **109**, 3570–3575.
- Groszmann, M., Greaves, I.K., Fujimoto, R., Peacock, W.J., and Dennis, E.S. (2013). The role of epigenetics in hybrid vigour. *Trends Genet.* **29**, 684–690.
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). Circize implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812. <https://doi.org/10.1093/bioinformatics/btu393>.
- Guo, M., Yang, S., Rupe, M., Hu, B., Bickel, D.R., Arthur, L., and Smith, O. (2008). Genome-wide allele-specific expression analysis using massively parallel signature sequencing (MPSS™) reveals cis- and trans-effects on gene expression in maize hybrid meristem tissue. *Plant Mol. Biol.* **66**, 551–563.
- Guo, W., Zhu, P., Pellegrini, M., Zhang, M.Q., Wang, X., and Ni, Z. (2018). CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* **34**, 381–387.
- He, G., Chen, B., Wang, X., Li, X., Li, J., He, H., Yang, M., Lu, L., Qi, Y., Wang, X., and Deng, X.W. (2013). Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. *Genome Biol.* **14**, R57.
- He, G., Zhu, X., Elling, A.A., Chen, L., Wang, X., Guo, L., Liang, M., He, H., Zhang, H., Chen, F., et al. (2010). Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* **22**, 17–33.
- He, L., Zhao, C., Zhang, Q., Zinta, G., Wang, D., Lozano-Durán, R., and Zhu, J.K. (2021). Pathway conversion enables a double-lock mechanism to maintain DNA methylation and genome stability. *Proc. Natl. Acad. Sci. USA* **118**. e2107320118.
- He, X.-J., Chen, T., and Zhu, J.-K. (2011). Regulation and function of DNA methylation in plants and animals. *Cell Res.* **21**, 442–465.
- Herath, H.M.S.N., Rafii, M.Y., Ismail, S.I., JJ, N., and Ramlee, S.I. (2021). Improvement of important economic traits in chili through heterosis breeding: a review. *J. Hortic. Sci. Biotechnol.* **96**, 14–23.
- Hofmeister, B.T., Lee, K., Rohr, N.A., Hall, D.W., and Schmitz, R.J. (2017). Stable inheritance of DNA methylation allows creation of epigenotype maps and the study of epiallele inheritance patterns in the absence of genetic variation. *Genome Biol.* **18**, 155.
- Ibiza, V.P., Blanca, J., Cañizares, J., and Nuez, F. (2012). Taxonomy and genetic diversity of domesticated *Capsicum* species in the Andean region. *Genet. Resour. Crop Evol.* **59**, 1077–1088.
- Inagaki, S., and Kakutani, T. (2012). What triggers differential DNA methylation of genes and TEs: contribution of body methylation? *Cold Spring Harb. Symp. Quant. Biol.* **77**, 155–160.
- Islam, K., Rawoof, A., Ahmad, I., Dubey, M., Momo, J., and Ramchiary, N. (2021). *Capsicum chinense* MYB transcription factor genes: identification, expression analysis and their conservation and diversification with other Solanaceae genomes. *Front. Plant Sci.* **12**, 721265.
- Jaiswal, V., Rawoof, A., Dubey, M., Chhakekar, S.S., Sharma, V., and Ramchiary, N. (2020). Development and characterization of non-coding RNA based simple sequence repeat markers in *Capsicum* species. *Genomics* **112**, 1554–1564. <https://doi.org/10.1016/j.ygeno.2019.09.005>.
- Jin, H., Hu, W., Wei, Z., Wan, L., Li, G., Tan, G., Zhu, L., and He, G. (2008). Alterations in cytosine methylation and species-specific transcription induced by interspecific hybridization between *Oryza sativa* and *O. officinalis*. *Theor. Appl. Genet.* **117**, 1271–1279.
- Jones, D.F. (1917). Dominance of linked factors as a means of accounting for heterosis. *Genetics* **2**, 466–479.
- Junaid, A., Kumar, H., Rao, A.R., Patil, A.N., Singh, N.K., and Gaikwad, K. (2018). Unravelling the epigenomic interactions between parental inbreds resulting in an altered hybrid methylome in pigeonpea. *DNA Res.* **25**, 361–373.
- Kawanabe, T., Ishikura, S., Miyaji, N., Sasaki, T., Wu, L.M., Itabashi, E., Takada, S., Shimizu, M., Takasaki-Yasuda, T., Osabe, K., et al. (2016). Role of DNA methylation in hybrid vigor in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **113**, E6704–E6711.
- Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915. <https://doi.org/10.1038/s41587-019-0201-4>.
- Kim, S., Park, M., Yeom, S.-I., Kim, Y.-M., Lee, J.M., Lee, H.-A., Seo, E., Choi, J., Cheong, K., Kim, K.T., et al. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* **46**, 270–278. <https://doi.org/10.1038/ng.2877>.
- Kim, S., Park, J., Yeom, S.I., Kim, Y.M., Seo, E., Kim, K.T., Kim, M.S., Lee, J.M., Cheong, K., Shin, H.S., et al. (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. *Genome Biol.* **18**, 210. <https://doi.org/10.1186/s13059-017-1341-9>.
- Kim, T.J., Hyeon, H., Park, N.I., Yi, T.G., Lim, S.-H., Park, S.-Y., Ha, S.-H., and Kim, J.K.

- (2020). A high-throughput platform for interpretation of metabolite profile data from pepper (*Capsicum*) fruits of 13 phenotypes associated with different fruit maturity states. *Food Chem.* 331, 127286.
- Klosinska, M., Picard, C.L., and Gehring, M. (2016). Conserved imprinting associated with unique epigenetic signatures in the *Arabidopsis* genus. *Nat. Plants* 2, 16145. <https://doi.org/10.1038/nplants.2016.145>.
- Kolde, R. (2019). Pheatmap: pretty heatmaps. R package version 1.0.12.
- Kong, X., Chen, L., Wei, T., Zhou, H., Bai, C., Yan, X., Miao, Z., Xie, J., and Zhang, L. (2020). Transcriptome analysis of biological pathways associated with heterosis in Chinese cabbage. *Genomics* 112, 4732–4741.
- Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27, 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>.
- Kue Foka, I.C., Keteouli, T., Zhou, Y., Li, X.W., Wang, F.W., and Li, H. (2020). The emerging roles of Diacylglycerol kinase (DGK) in plant stress tolerance, growth, and development. *Agronomy* 10, 1375.
- Lang, Z., Wang, Y., Tang, K., Tang, D., Datsenka, T., Cheng, J., Zhang, Y., Handa, A.K., and Zhu, J.K. (2017). Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. *Proc. Natl. Acad. Sci. USA* 114, E4511–E4519.
- Lauss, K., Wardenaar, R., Oka, R., van Hulst, M.H.A., Guryev, V., Keurentjes, J.J.B., Stam, M., and Johannes, F. (2018). Parental DNA methylation states are associated with heterosis in epigenetic hybrids. *Plant Physiol.* 176, 1627–1645.
- Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220. <https://doi.org/10.1038/nrg2719>.
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9, e1003118.
- Li, H., Yuan, J., Wu, M., Han, Z., Li, L., Jiang, H., Jia, Y., Han, X., Liu, M., Sun, D., et al. (2018). Transcriptome and DNA methylome reveal insights into yield heterosis in the curds of broccoli (*Brassica oleracea* L var. *italica*). *BMC Plant Biol.* 18, 168.
- Li, R., Hu, F., Li, B., Zhang, Y., Chen, M., Fan, T., and Wang, T. (2020a). Whole genome bisulfite sequencing methylome analysis of mulberry (*Morus alba*) reveals epigenome modifications in response to drought stress. *Sci. Rep.* 10, 8013.
- Li, Z., Zhu, A., Song, Q., Chen, H.Y., Harmon, F.G., and Chen, Z.J. (2020b). Temporal regulation of the metabolome and proteome in photosynthetic and photorespiratory pathways contributes to maize heterosis. *Plant Cell* 32, 3706–3722.
- Liang, X., Hou, X., Li, J., Han, Y., Zhang, Y., Feng, N., Du, J., Zhang, W., Zheng, D., and Fang, S. (2019). High-resolution DNA methylome reveals that demethylation enhances adaptability to continuous cropping comprehensive stress in soybean. *BMC Plant Biol.* 19, 1–17.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.
- Lisec, J., Schauer, N., Kopka, J., Willmitzer, L., and Fernie, A.R. (2006). Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* 1, 387–396. <https://doi.org/10.1038/nprot.2006.59>.
- Liu, Y., and Imai, R. (2018). Function of plant DExD/H-box RNA helicases associated with ribosomal RNA biogenesis. *Front. Plant Sci.* 9, 125.
- Liu, Y., Lv, J., Liu, Z., Wang, J., Yang, B., Chen, W., Ou, L., Dai, X., Zhang, Z., and Zou, X. (2020). Integrative analysis of metabolome and transcriptome reveals the mechanism of color formation in pepper fruit (*Capsicum annuum* L.). *Food Chem.* 306, 125629.
- Liu, Z., Lv, J., Zhang, Z., Li, H., Yang, B., Chen, W., Dai, X., Li, X., Yang, S., Liu, L., et al. (2019). Integrative transcriptome and proteome analysis identifies major metabolic pathways involved in pepper fruit development. *J. Proteome Res.* 18, 982–994.
- Lopez-Ortiz, C., Peña-García, Y., Bhandari, M., Abburi, V.L., Natarajan, P., Stommel, J., Nimmakayala, P., and Reddy, U.K. (2021). Identification of miRNAs and their targets involved in flower and fruit development across domesticated and wild *Capsicum* species. *Int. J. Mol. Sci.* 22, 4866.
- Ma, X., Xing, F., Jia, Q., Zhang, Q., Hu, T., Wu, B., Shao, L., Zhao, Y., Zhang, Q., and Zhou, D.X. (2021). Parental variation in CHG methylation is associated with allelic-specific expression in elite hybrid rice. *Plant Physiol.* 186, 1025–1041.
- Macovei, A., Vaid, N., Tula, S., and Tuteja, N. (2012). A new DEAD-box helicase ATP-binding protein (OsABP) from rice is responsive to abiotic stress. *Plant Signal. Behav.* 7, 1138–1143.
- Martínez, O., Arce-Rodríguez, M.L., Hernández-Godínez, F., Escoto-Sandoval, C., Cervantes-Hernández, F., Hayano-Kanashiro, C., Ordaz-Ortiz, J.J., Reyes-Valdés, M.H., Razo-Mendivil, F.G., Garcés-Claver, A., and Ochoa-Alejo, N. (2021). Transcriptome analyses throughout chili pepper fruit development reveal novel insights into the domestication process. *Plants* 10, 585.
- Niederhuth, C.E., Bewick, A.J., Ji, L., Alabady, M.S., Kim, K.D., Li, Q., Rohr, N.A., Rambani, A., Burke, J.M., Udall, J.A., et al. (2016). Widespread natural variation of DNA methylation within angiosperms. *Genome Biol.* 17, 194. <https://doi.org/10.1186/s13059-016-1059-0>.
- Olatunji, T.L., and Afolayan, A.J. (2019). Comparative quantitative study on phytochemical contents and antioxidant activities of *Capsicum annuum* L. and *Capsicum frutescens* L. *Sci. World J.* 2019.
- Pang, Z., Chong, J., Zhou, G., de Lima Morais, D.A., Chang, L., Barrette, M., Gauthier, C., Jacques, P.E., Li, S., and Xia, J. (2021). MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res.* 49, W388–W396. <https://doi.org/10.1093/nar/gkab382>.
- Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., Cheng, J., Zhao, S., Xu, M., Luo, Y., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proc. Natl. Acad. Sci. USA* 111, 5135–5140. <https://doi.org/10.1073/pnas.12168-9881.s1.013>.
- Ramchiary, N., Kehie, M., Brahma, V., Kumaria, S., and Tandon, P. (2014). Application of genetics and genomics towards *Capsicum* translational research. *Plant Biotechnol. Rep.* 8, 101–123. <https://doi.org/10.1007/s11816-013-0306-z>.
- Rawoof, A., Chhakekar, S.S., Jaiswal, V., Brahma, V., Kumar, N., and Ramchiary, N. (2020). Single-base cytosine methylation analysis in fruits of three *Capsicum* species. *Genomics* 112, 3342–3353.
- Raza, M.A., Yu, N., Wang, D., Cao, L., Gan, S., and Chen, L. (2017). Differential DNA methylation and gene expression in reciprocal hybrids between *Solanum lycopersicum* and *S. pimpinellifolium*. *DNA Res.* 24, 597–607.
- Razo-Mendivil, F.G., Hernández-Godínez, F., Hayano-Kanashiro, C., and Martínez, O. (2021). Transcriptomic analysis of a wild and a cultivated varieties of *Capsicum annuum* over fruit development and ripening. *PLoS One* 16, e0256319.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. <https://doi.org/10.1093/bioinformatics/btp616>.
- Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11, R25. <https://doi.org/10.1186/gb-2010-11-3-r25>.
- Sarpras, M., Ahmad, I., Rawoof, A., and Ramchiary, N. (2019). Comparative analysis of developmental changes of fruit metabolites, antioxidant activities and mineral elements content in Bhut jolokia and other *Capsicum* species. *LWT* 105, 363–370.
- Sarpras, M., Chhakekar, S.S., Ahmad, I., Abraham, S.K., and Ramchiary, N. (2020). Analysis of bioactive components in Ghost chili (*Capsicum chinense*) for antioxidant, genotoxic, and apoptotic effects in mice. *Drug Chem. Toxicol.* 43, 182–191. <https://doi.org/10.1080/01480545.2018.1483945>.
- Sarpras, M., Gaur, R., Sharma, V., Chhakekar, S.S., Das, J., Kumar, A., Yadava, S.K., Nitin, M., Brahma, V., Abraham, S.K., and Ramchiary, N. (2016). Comparative analysis of fruit metabolites and pungency candidate genes expression between Bhut Jolokia and other *Capsicum* species. *PLoS One* 11, e0167791. <https://doi.org/10.1371/journal.pone.0167791>.
- Saze, H., Shiraishi, A., Miura, A., and Kakutani, T. (2008). Control of genetic DNA methylation by a jmjC domain-containing protein in *Arabidopsis*

- thaliana*. *Science* 319, 462–465. <https://doi.org/10.1126/science.1150987>.
- Schmitz, R.J., Schultz, M.D., Lewsey, M.G., O'Malley, R.C., Ulrich, M.A., Libiger, O., Schork, N.J., and Ecker, J.R. (2011). Transgenerational epigenetic instability is a source of novel methylation variants. *Science* 334, 369–373.
- Schnable, P.S., and Springer, N.M. (2013). Progress toward understanding heterosis in crop plants. *Annu. Rev. Plant Biol.* 64, 71–88.
- Shen, H., He, H., Li, J., Chen, W., Wang, X., Guo, L., Peng, Z., He, G., Zhong, S., Qi, Y., et al. (2012). Genome-wide analysis of DNA methylation and gene expression changes in two Arabidopsis ecotypes and their reciprocal hybrids. *Plant Cell* 24, 875–892.
- Shen, Y., Sun, S., Hua, S., Shen, E., Ye, C.Y., Cai, D., Timko, M.P., Zhu, Q.H., and Fan, L. (2017). Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid *Brassica napus* uncovers key roles for small RNA s. *Plant J.* 91, 874–893.
- Shiragaki, K., Furukawa, H., Yokoi, S., and Tezuka, T. (2021). Temperature-dependent sugar accumulation in interspecific *Capsicum* F1 plants showing hybrid weakness. *J. Plant Res.* 134, 1199–1211.
- Shrestha, S.L., Luitel, B.P., and Kang, W.H. (2011). Heterosis and heterobeltiosis studies in sweet pepper (*Capsicum annuum* L.). *Hortic. Environ. Biotechnol.* 52, 278–283. <https://doi.org/10.1007/s13580-011-0106-8>.
- Singh, P., Cheema, D.S., Dhaliwal, M.S., and Garg, N. (2014). Heterosis and combining ability for earliness, plant growth, yield and fruit attributes in hot pepper (*Capsicum annuum* L.) involving genetic and cytoplasmic-genetic male sterile lines. *Sci. Hortic.* 168, 175–188.
- Sinha, P., Singh, V.K., Saxena, R.K., Kale, S.M., Li, Y., Garg, V., Meifang, T., Khan, A.W., Kim, K.D., Chitkineni, A., et al. (2020). Genome-wide analysis of epigenetic and transcriptional changes associated with heterosis in pigeonpea. *Plant Biotechnol. J.* 18, 1697–1710.
- Sinigalli, C., Faraone, I., Vassallo, A., Caddeo, C., Bisaccia, F., Armentano, M.F., Milella, L., and Ostuni, A. (2020). Phytochemical profile of *Capsicum annuum* L. cv Senise, incorporation into liposomes, and evaluation of cellular antioxidant activity. *Antioxidants* 9, 428.
- Sood, S., and Kumar, N. (2010). Heterosis for fruit yield and related horticultural traits in bell pepper. *Int. J. Veg. Sci.* 16, 361–373.
- Springer, N.M., and Stupar, R.M. (2007). Allelic variation and heterosis in maize: how do two halves make more than a whole? *Genome Res.* 17, 264–275.
- Srichaeroen, P., Lamaiphan, N., Patthawaro, P., Limchoowong, N., Techawongstien, S., and Chanthai, S. (2017). Phytochemicals in *Capsicum* oleoresin from different varieties of hot chili peppers with their antidiabetic and antioxidant activities due to some phenolic compounds. *Ultrason. Sonochem.* 38, 629–639.
- Swanson-Wagner, R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D., and Schnable, P.S. (2006). All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. *Proc. Natl. Acad. Sci. USA* 103, 6805–6810.
- Takuno, S., and Gaut, B.S. (2012). Body-methylated genes in Arabidopsis thaliana are functionally important and evolve slowly. *Mol. Biol. Evol.* 29, 219–227.
- Takuno, S., and Gaut, B.S. (2013). Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proc. Natl. Acad. Sci. USA* 110, 1797–1802.
- Takuno, S., Ran, J.-H., and Gaut, B.S. (2016). Evolutionary patterns of genic DNA methylation vary across land plants. *Nat. Plants* 2, 15222. <https://doi.org/10.1038/nplants.2015.222>.
- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. <https://doi.org/10.1093/nar/gkx382>.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40, e115. <https://doi.org/10.1093/nar/gks596>.
- Van Der Graaf, A., Wardenaar, R., Neumann, D.A., Taudt, A., Shaw, R.G., Jansen, R.C., Schmitz, R.J., Colomé-Tatché, M., and Johannes, F. (2015). Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proc. Natl. Acad. Sci. USA* 112, 6676–6681.
- Xiao, K., Chen, J., He, Q., Wang, Y., Shen, H., and Sun, L. (2020). DNA methylation is involved in regulation of pepper fruit ripening and interacts with phytohormones. *J. Exp. Bot.* 71, 1928–1942. <https://doi.org/10.1093/jxb/era003>.
- Xu, X.W., Li, T., Li, Y., and Wang, H.M. (2015). Variation of DNA cytosine methylation patterns among parent lines and reciprocal hybrids in hot pepper. *Chem. Eng. Trans.* 46, 1345–1350.
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omi. a J. Integr. Biol.* 16, 284–287.
- Zamljen, T., Zupanc, V., and Slatnar, A. (2020). Influence of irrigation on yield and primary and secondary metabolites in two chilies species, *Capsicum annuum* L. and *Capsicum chinense* Jacq. *Agric. Water Manag.* 234, 106104.
- Zebell, S. (2021). Gene body silencing parental alleles for vigorous hybrid rice. *Plant Physiol.* 186, 822–823.
- Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153, 193–205.
- Zemach, A., Kim, M.Y., Silva, P., Rodrigues, J.A., Dotson, B., Brooks, M.D., and Zilberman, D. (2010). Local DNA hypomethylation activates genes in rice endosperm. *Proc. Natl. Acad. Sci. USA* 107, 18729–18734.
- Zhang, Q., Wang, D., Lang, Z., He, L., Yang, L., Zeng, L., Li, Y., Zhao, C., Huang, H., Zhang, H., et al. (2016). Methylation interactions in Arabidopsis hybrids require RNA-directed DNA methylation and are influenced by genetic variation. *Proc. Natl. Acad. Sci. USA* 113, E4248–E4256.
- Zhang, X., Shiu, S.H., Shiu, S., Cal, A., and Borevitz, J.O. (2008). Global analysis of genetic, epigenetic and transcriptional polymorphisms in Arabidopsis thaliana using whole genome tiling arrays. *PLoS Genet.* 4, e1000032.
- Zhao, X., Chai, Y., and Liu, B. (2007). Epigenetic inheritance and variation of DNA methylation level and pattern in maize intra-specific hybrids. *Plant Sci.* 172, 930–938.
- Zhou, S., Xing, M., Zhao, Z., Gu, Y., Xiao, Y., Liu, Q., and Xue, H. (2021). DNA methylation modification in heterosis initiation through analyzing rice hybrid contemporary seeds. *Crop J.* 9, 1179–1190. <https://doi.org/10.1016/j.cj.2020.12.003>.
- 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, 61–69.
- Zuo, J., Wang, Y., Zhu, B., Luo, Y., Wang, Q., and Gao, L. (2018). Comparative analysis of DNA methylation reveals specific regulations on ethylene pathway in tomato fruit. *Genes* 9, 266.



## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological samples</b>		
<i>Capsicum chinense</i> (Choco)	This study	N/A
<i>Capsicum frutescens</i> (Frut4)	This study	N/A
<i>Capsicum</i> F1 Hybrid (Choco X Frut4)	This study	N/A
<b>Chemicals, peptides, and recombinant proteins</b>		
Acetone	Sigma-Aldrich	34850
CTAB	Sigma-Aldrich	H5882
Zymo EZ-DNA Methylation-Gold Kit	Zymo Research	D5005
NucleoSpin RNA plant Kit	Macherey-Nagel	740949.50
Illumina TruSeq RNA sample preparation kit	Illumina	RS-930-2001
PrimeScript IV 1 <sup>st</sup> strand cDNA Synthesis Mix	TaKaRa Bio	6215A
TB Green Premix Ex Taq II Master Mix	TaKaRa Bio	RR820L
Methanol	Sigma-Aldrich	34860
Adonitol (Adonite, Ribitol)	Sigma-Aldrich	A5502
Methoxyamine hydrochloride	MP Biomedicals	02155405-CF
Pyridine	Sigma-Aldrich	360570
N-Methyl-N-(trimethylsilyl) trifluoroacetamide	Sigma-Aldrich	69479
<b>Deposited data</b>		
Raw Data for WGBS and RNA sequencing	This study	PRJNA770309
<i>Capsicum chinense</i> reference genome ASM227189v2	NCBI (Kim et al., 2017)	<a href="https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002/271/895/GCA_002271895.2_ASM227189v2/">https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002/271/895/GCA_002271895.2_ASM227189v2/</a>
<b>Oligonucleotides</b>		
Gene specific Primers for qRT-PCR analysis, see Table S1	This study	N/A
<b>Software and algorithms</b>		
FastQC tool	Andrews (2010)	<a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
TrimGalore	N/A	<a href="https://github.com/FelixKrueger/TrimGalore">https://github.com/FelixKrueger/TrimGalore</a>
Bismark	Krueger and Andrews (2011)	<a href="https://github.com/FelixKrueger/Bismark">https://github.com/FelixKrueger/Bismark</a>
Hisat2	Kim et al. (2019)	<a href="http://daehwankimlab.github.io/hisat2/">http://daehwankimlab.github.io/hisat2/</a>
featureCounts(v1.5.1)	Liao et al. (2014)	<a href="http://subread.sourceforge.net/">http://subread.sourceforge.net/</a>
methylKit	Akalin et al. (2012)	<a href="https://bioconductor.org/packages/methylKit/">https://bioconductor.org/packages/methylKit/</a>
edgeR	Robinson et al. (2010)	<a href="https://bioconductor.org/packages/edgeR/">https://bioconductor.org/packages/edgeR/</a>
CircIize	Gu et al. (2014)	<a href="https://github.com/jokergoo/circIize">https://github.com/jokergoo/circIize</a>
GenomicFeatures	Lawrence et al. (2013)	<a href="https://bioconductor.org/packages/GenomicFeatures/">https://bioconductor.org/packages/GenomicFeatures/</a>
CGmapTools	Guo et al. (2018)	<a href="https://cgmaptools.github.io/">https://cgmaptools.github.io/</a>
Pheatmap	Kolde (2019)	<a href="https://CRAN.R-project.org/package=pheatmap">https://CRAN.R-project.org/package=pheatmap</a>
agriGO tool	Tian et al. (2017)	<a href="http://systemsbiology.cau.edu.cn/agriGOv2/">http://systemsbiology.cau.edu.cn/agriGOv2/</a>
clusterProfiler	Yu et al. (2012)	<a href="https://github.com/YuLab-SMU/clusterProfiler">https://github.com/YuLab-SMU/clusterProfiler</a>
Primer3	Untergasser et al. (2012)	<a href="https://primer3.ut.ee/">https://primer3.ut.ee/</a>
GC-MS	Shimadzu	GCMS-QP2010 Plus
Reference mass spectral library databases	NIST14	<a href="https://www.nist.gov/">https://www.nist.gov/</a>

(Continued on next page)

<i>Continued</i>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Reference mass spectral library databases	Wiley 9	<a href="https://www.sisweb.com/software/ms/wiley.htm">https://www.sisweb.com/software/ms/wiley.htm</a>
MetaboAnalystR	Pang et al. (2021)	<a href="https://www.metaboanalyst.ca/docs/RTutorial.xhtml">https://www.metaboanalyst.ca/docs/RTutorial.xhtml</a>
Other		
CFX96 Touch Real-Time PCR Detection System	Bio-Rad	CFX96
HiSeq X Ten Sequencing System	Illumina	N/A
HiSeq 4000 Sequencing System	Illumina	N/A

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contacts, Vandana Jaiswal ([vandana.jaiswal2009@gmail.com](mailto:vandana.jaiswal2009@gmail.com)).

### Materials availability

This study did not generate new unique reagents or material.

### Data and code availability

- The whole genome-bisulfite sequencing and RNA sequencing data generated in this study has been deposited in NCBI Sequence Read Archive (SRA) under Bioproject PRJNA770309.
- This study did not generate original code.
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## METHOD DETAILS

### Plant material and phenotyping

To generate *Capsicum* hybrid, two *Capsicum* accessions Choco and Frut4 (pollen donor) belonging to *Capsicum chinense* and *C. frutescens*, respectively, were crossed by hand pollination. The resultant F<sub>1</sub> seeds were collected, sown and allowed to grow at standard growth conditions (day/night temperature: 27°C/19°C and 16 h/day light) in the greenhouse of the School of Life Sciences, Jawaharlal Nehru University, New Delhi, India. Then leaf samples were collected from two months old plants from two parents and hybrid, each with two biological replicates. Leaves from each sample were immediately frozen and preserved in liquid nitrogen and used for DNA/RNA extraction and metabolome profiling.

Hybrids and their parents were phenotyped for seven traits at the time of flowering each with five replicates. These traits include plant height (cm), plant width (cm), number of secondary branches, number of leaves, leaf length (cm), leaf width (cm), and total chlorophyll content (µg/mg). The total chlorophyll content was estimated according to the method of Arnon (1949). Fresh leaves were taken, washed, blotted in tissue paper to dry the external moisture and the leaf weight of the material was recorded. Approximately, 100 mg leaf samples were homogenized in a pre-chilled mortar using cold 80% acetone and tissue was pulverized completely. The homogenate was centrifuged at 3200 rpm for 30 min and the supernatant was collected. 80% acetone was taken as blank and the optical density of the acetone extract was measured at 645 and 663 nm in a spectrophotometer (Thermo Scientific). Total chlorophyll content was calculated from three biological replicates using the following formula:

$$\text{Total Chlorophyll} = 20.1 \times \text{O.D. (at 645 nm)} + 8.02 \times \text{O.D. (at 663 nm)}$$

Mid parent heterosis (MPH) and better parent heterosis (BPH) were also calculated for these seven traits using following formula-

$$\text{MPH} = \left[ \frac{(F_1 - MP)}{MP} \right] \times 100$$

$$\text{BPH} = \left[ \frac{(F_1 - BP)}{BP} \right] \times 100$$

where, MP and BP are mid parent and better parent trait values respectively.

### Whole-genome bisulfite sequencing and data processing

From each sample, high molecular weight genomic DNA was extracted using the cetyltrimethylammonium bromide (CTAB) method. Around 50–100 ng of DNA was used for bisulfite conversion using Zymo EZ-DNA Methylation-Gold kit (Zymo Research, Irvine, USA), following manufacturer's protocol. Briefly, fragmented DNA of 100–300 bp size was subjected to anneal to random primers having tagged sequences, providing ssDNA copy. Then, using ssDNA as a template, dsDNA was synthesized using terminal tagging oligo (TTO) with 3' end blocked. Further, polymerase chain reaction (PCR) was performed with Illumina TruSeq indices followed by assessment of fragment size, concentration and library size using Agilent 2200 TapeStation. At last, six 2 × 150 bp paired-end libraries from three samples (Choco, Frut4 and hybrid; each with two biological replicates) were constructed on Illumina's HiSeq X Ten platform (Illumina, San Diego, USA).

Each sample's raw sequencing reads were converted to FASTQ format and quality was checked using FastQC tool (Andrews, 2010). Adaptor sequences and reads with low quality scores were filtered out using TrimGalore (<https://github.com/FelixKrueger/TrimGalore>). The clean reads were then aligned to bismark-transformed *C. chinense* (Kim et al., 2017) reference genome using Bismark (Krueger and Andrews, 2011) and genome-wide CG, CHG and CHH methylation at single-base resolution were identified using criteria as described earlier by (Rawoof et al., 2020). Bisulfite conversion efficiency was determined using spiked-in un-methylated 1% lambda DNA (GenBank: J02459.1). Average methylation at 12 *Capsicum* chromosomes was represented across bins of equal 1MB length using the R circlize package (Gu et al., 2014).

### RNA sequencing and data processing

Total RNA from each sample was extracted using NucleoSpin RNA plant Kit (Macherey-Nagel, Germany) following standard protocol. The quality and integrity of RNA were checked using an Agilent bioanalyzer and then RNA with RNA integrity number (RIN) > 8 was sequenced as described earlier (Rawoof et al., 2020; Islam et al., 2021). Briefly, an equal amount of total RNA (5–10 µg) from each sample was used for library preparation using Illumina TruSeq RNA sample preparation kit (Illumina, San Diego, CA) following manufacturer's protocols and paired-end sequencing was performed on Illumina HiSeq 4000 to develop 150 bp paired-end reads from three samples (two biological replicate of each sample). The raw RNAseq were subjected for quality check and then adapter sequences, as well as reads with quality score <30, were filtered out. Further, clean reads were aligned to the *C. chinense* reference genome (Kim et al., 2017) using Hisat2 (Kim et al., 2019) program with default parameters. Read counts of transcripts were quantified using the featureCounts (v1.5.1) (Liao et al., 2014). Raw read counts were normalized using TMM methods (Robinson and Oshlack, 2010), and differentially expressed genes (DE-Gs) between two sample comparisons were identified using the glmQLFit and glmQLFTest functions in the edgeR package (Robinson et al., 2010). Genes with p value < 0.01 and false discovery rate (FDR) < 0.01 were considered significantly differentially expressed between the two samples.

### Analysis of differential methylation levels and their correlation with gene expression

For the identification of differentially methylated regions (DMRs) at each methylation context and their genomic annotation, parameters earlier described by Rawoof et al. (2020) were used. In brief, differential methylation analysis was performed using a sliding window of 100bp with 50 bp step size and overlapping DMRs were merged, and methylation information of each DMR was assessed using the R methylKit package (Akalin et al., 2012) with parameters including Fisher's exact test, sliding linear model for q-values and p values. Also, each window having bases >10X coverage, with at-least 3 mCs for CG, CHG and CHH context along with methylation difference (meth.diff) > 25% with q-values <0.01 were considered as significant DMRs. For common DMRs between hybrid and its parents, all samples were analyzed together. Then for additive methylation level in hybrid, mid-parental methylation values (MPMV) were calculated and common DMRs with higher methylation than MPMV were categorized as *trans*-chromosomal methylation (TCM), while lower than MPMV were categorized as *trans*-chromosomal demethylation (TCdM). A significance test was performed using pairwise t-test with both FDR and p value <0.01. DMRs were annotated using an in-house built txdb SQLite database of *C. chinense* using R GenomicFeatures (Lawrence et al., 2013) and were categorised as intergenic, intragenic and promoter DMRs. Repeat-elements information as reported by (Rawoof et al., 2020) was used for methylation analysis at transposable elements (TEs). Gene-body methylation (GBM) and TE methylation were analyzed using CGmapTools (Guo et al., 2018). Moreover, to avoid noise from TEs at GBM, methylation at exonic and intronic levels along with 2 kb flanking regions were also analyzed. Methylation and gene expression correlation was performed using Pearson correlation analysis between the RNAseq expression of total *C. chinense* genes and their promoter

methylation at each methylation context, and correlations with p value <0.001 were considered as significant. Additionally, promoter-DMR genes (PDG) having differential expression between parents and hybrids were investigated for methylation and their correlation with expression was represented as a heatmap using R pheatmap package (Kolde, 2019).

### Analysis of allele-specific methylation (ASM) associated with DMRs

To analyze allele specific methylation, single nucleotide variants (SNVs) were identified using CGmapTools snv function (Guo et al., 2018). The SNVs detection was performed using bismark mapped methylated reads (>10x coverage) from hybrid and parental lines against *C. chinense* reference genome with BayesWC mode and dynamic p value cut-off value of <0.001. Prior to ASM calling, the SNVs associated with TCM, TCdM and promoter DMRs were filtered out using vcftools –recode function. These DMRs specific SNVs were further used for ASM calling using asm mode of CGmapTools with filter criteria of minimum 5 reads for each allele linked sites, adjusted p value < 0.05 and high methylation threshold  $\geq 0.8$  to identify significant ASM associated with DMRs at CG, CHG and CHH methylation context. The tanghulu plots representing ASM were visualized using CGmapTools.

### Gene ontology and pathway enrichment analysis

Gene ontology (GO) analysis of DE-Gs between hybrid and parental lines and DE-Gs having a negative correlation between their expression and promoter methylation, was performed using agriGO tool (Tian et al., 2017). For GO information, a hypergeometric test with p value <0.01 and FDR <0.01 was used. For pathway enrichment analysis, SQLite-based custom annotation package for *C. chinense* was prepared in R, and then enrichment analysis was performed using R clusterProfiler (Yu et al., 2012). GO terms and pathways with adjusted p values <0.01 were considered as significant.

### qRT-PCR expression validation of DE-Gs negatively correlated with promoter methylation

At least 24 DE-Gs genes that negatively correlated with promoter methylation were selected for qRT-PCR validation, and gene-specific primers (Table S1) were designed using Primer3 (Untergasser et al., 2012). RNA was extracted from leaves of hybrid and parental lines (same samples used in WGBS and RNAseq) using NucleoSpin RNA plant kit (Macherey-Nagel, Germany) following manufacturer's protocol. The quality and integrity of extracted RNA was checked using NanoDrop 1000 (Thermo Scientific) and 1% Agarose gel, respectively. Equally,  $\sim 1 \mu\text{g}$  of total RNA was used to synthesize cDNA using PrimeScript IV 1st strand cDNA Synthesis Mix (Takara). The real-time PCR reaction was set up using TB Green Premix Ex Taq II Master Mix (Clontech, USA) and run on the Bio-Rad CFX96 Real-Time System (Bio-Rad, USA) using the thermal profile of initial denaturation at 95°C for 30 s, followed by amplification step with 40 cycles of 95°C for 5 s and 60°C for 30 s. Then final amplification was completed at 65°C for 5 s, as described earlier (Dubey et al., 2019; Ahmad et al., 2021b). Further, qRT-PCR expression was compared with *in silico* RNAseq expression data. EF1-alpha was used as internal control, and significant expression difference was analyzed using two-way ANOVA test.

### Untargeted GC-MS based metabolomics analysis

The global metabolome of hybrid and parental genotypes, three biological replicates each, was analyzed using gas chromatography coupled with mass spectrometry (GC-MS) following the method described by Lisec et al. (2006) and Ahmad et al. (2021b) with slight modification. Briefly, 100 mg of each leaf sample was homogenized with mortar and pestle using liquid nitrogen and then 1400  $\mu\text{L}$  of pure methanol (chilled) was added. For internal quantitative standard, 60  $\mu\text{L}$  of ribitol ( $0.2 \text{ mg mL}^{-1}$ ) was added and kept for shaking for 10 min at 70°C, followed by vortexing at 11000g for 10 min. Further, 750  $\mu\text{L}$  of chloroform and 1500  $\mu\text{L}$  of  $\text{dH}_2\text{O}$  were added and centrifuged for 15 min at 2,200g. Finally, the supernatant of each sample was dried in a vacuum concentrator without heating. For derivatization, 40  $\mu\text{L}$  of methoxyamine hydrochloride ( $20 \text{ mg mL}^{-1}$ ) in pyridine was added in all samples, vortexed and kept in a shaker for 2 h at 37°C. For trimethylsilylation, 70  $\mu\text{L}$  of N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) as a silylating agent was added and then vortexed and incubated with shaking for 30 min at 37°C. The GC-MS running conditions includes: oven temperature – 100°C for 2 min followed by increment in oven temperature upto 250°C at the rate of 5°C per min, then finally 280°C at the rate of 2°C per min. Further, 1  $\mu\text{L}$  of derivatised sample was injected at 10:1 split ratio into a GC-MS (Shimadzu QP2010 Plus), equipped with a Rtx- 5 MS capillary column (0.25 mm film thickness, 0.25 mm internal diameter, and 30 m in length) with helium as the carrier gas with flow-rate of 1.2 mL/min. The quantification of detected metabolites was analyzed using GCMS post run Analysis

Program (Shimadzu Scientific Instruments) following criteria described by [Ahmad et al. \(2021b\)](#). In short, metabolites were identified by MS detector via full scan mode using their distinct peak fragmentation pattern and peaks of identified metabolites were compared with retention indices and mass spectra from reference mass spectral library databases (NIST14 and Wiley 9). At last, relative quantification was based on peak area ratio of analyte to internal standard ribitol. The metabolites in hybrid were assessed by relative changes in abundance as response ratios compared to its parents. The response ratio was calculated by dividing the average metabolite concentration from both parents (Choco and Frut4) by the average metabolite concentration from the hybrid. Metabolite differences between parents and hybrid with fold change  $>2$  and p value  $< 0.05$  were considered significant. Pathway enrichment of metabolites were analyzed using MetaboAnalystR ([Pang et al., 2021](#)) and pathways number of metabolites  $>3$  and pathway impact  $>0.2$  along with p value  $<0.01$  were considered as significantly enriched.