NKFIH-K 129171

FINAL REPORT

Engineering tomato fruit ripening variation by genome editing

György Szittya PhD

Hungarian University of Agriculture and Life Sciences
Institute of Genetics and Biotechnology
Department of Plant Biotechnology

16.12.2022

Abbreviations:

A. thaliana: *Arabidopsis thaliana*
CDS: coding sequence
CMT3: CHROMOMETHYLASE 3
CNR: COLORLESS NONRIPENING
CRE: cis-regulatory elements
Cas-9: Cas-9 endonuclease
CRISPR: clustered regularly interspaced palindromic repeat
DE: differential expression
DME: DEMETER
DML: DEMETER-like
DMR: differentially methylated region
DRM1: DOMAINS REARRANGED METHYLTRANSFERASE1
DRM2: DOMAINS REARRANGED METHYLTRANSFERASE2
gRNA: guide RNA
GWAS: genome-wide association study
IDM: INCREASED DNA METHYLATION
HDP2: HARBRINGER TRANSPOSON DERIVED DNA METHYL-BINDING-PROTEIN 2
kb: kilobase
KEGG: Kyoto Encyclopaedia of Genes and Genomes
MADS: MADS-box transcription factor
MBD7: METHYL-CpG-BINDING DOMAIN PROTEIN 7
MEMS: DNA methylation monitoring sequence
MET1: METHYLTRANSFERASE 1
mRNA: messenger RNA
NAC: NAC transcription factor
N. benthamiana: *Nicotiana benthamiana*
NOR: NON-RIPENING
nt: nucleotide
ORF: open reading frame
QTL: quantitative trait loci
RIN: RIPENING INHIBITOR
RNAi: RNA interference or RNA silencing
ROS1: REPRESSOR OF SILENCING 1
S. lycopersicum: *Solanum lycopersicum*
SPL: SQUAMOSA PROMOTER BINDING-LIKE PROTEIN
uORF: upper open reading frame
UTR: untranslated regions
TE: transposable elements
TES: transcription exit site
TSS: transcription start site
WT: wild type

Summary

Genomic DNA methylation is a major epigenetic mark that is fundamental to many aspects of chromatin function including the regulation of gene expression. In plants DNA methylation can occur at cytosine, both in symmetrical and non-symmetrical context and it is controlled by DNA methyltransferases. Genomic DNA methylation can be actively removed by DNA demethylases (DMLs). Genomic DNA methylation is an important mechanism that influences gene expression and methylation at promoters is known to inhibit gene transcription. Therefore, the active removal of methylation marks is an important mechanism during development and cell fate reprograming. In tomato SlDML2 controls the expression of genes encoding ripening transcription factors. During this project we created deletion series in the SlDML2 promoter with the help of CRISPR/Cas-9 based genome editing system. We designed eight, evenly spaced CRISPR guide RNAs within a 2.5 kb region upstream of the transcription start site (gRNA1-8) and another eight within a 5 kb region in a more distant part of the promoter (gRNA9-16) together with appropriate control constructs. We used these constructs to transform tomato, and we produced, regenerating lines (46 gRNA1-8, and 33 gRNA9-16) of the genome editing constructs. We observed, that the promoter mutagenesis was very efficient and precise, almost all of the gRNA sites were mutated. The genome edited mutant plants, carried large deletions either in the proximal (gRNA1-8) or the distal (gRNA9-16) part of the SlDML2 promoter. The fruit ripening times were similar to the WT in most of the obtained genome edited lines, however, there were few mutants that showed significantly shorter (A34 – 3 days earlier) while others longer ripening time (A12, A120 - 3 days later). Developmental defects were more obvious since many lines produced seedless fruits (eg. A120) more frequently than the WT. We concluded, that SlDML2 is also important for plant development. To examine the role of SlIDM3 in the target selection of SlDML2, we used CRISPR/Cas-9 to knock out SlIDM3. We also analysed the genome wide methylome of those plants.

The COVID lockdown mostly effected the dissemination of the results of the project. During this project we continuously published (3 peer reviewed article and 1 book chapter) and presented our results on international and national conferences (6 presentations). Furthermore, preparation of a manuscript and a doctoral dissertation based on the results on the targeted promoter mutagenesis of SlDML2 gene by CRISPR/Cas-9 is still expected. The results and figures shown in this report, outline the high quality of this study, which is anticipated to be published in an academic peer reviewed journal. The results of this work are expected to be a highly cited reference for future studies.

Content:

1. Brief Introduction

Heritable changes in gene expression that occur without modification of the underlying DNA sequence is the topic of epigenetic studies. This mainly involves histone post-translational modifications and DNA methylation which are transmitted through DNA replication and cell propagation, thereby determining and maintaining cell type specific gene expression patterns (Gallusci et al., 2016). Genomic DNA methylation is very important to many aspects of chromatin function, including DNA recombination, transposon silencing or regulation of gene expression (Law and Jacobsen, 2010; Matzke et al., 2016). DNA methylation in plants can occur at cytosine both in symmetrical (CG or CHG) and non-symmetrical (CHH) contexts (where H can be an A, C, or T) and is controlled by three classes of DNA methyltransferases. METHYLTRANSFERASE1 (MET1) maintains methylation at symmetrical 5'-CG-3'sites during DNA replication and its targets are mainly silenced transposons and heterochromatin, but MET1 also maintains marks on some imprinted genes. CHROMOMETHYLASE3 (CMT3) methylates at 5'-CHG-3' sequences and can initiate DNA methylation *de novo* at sites with certain histone modifications. DOMAINS REARRANGED METHYLTRANSFERASE1 (DRM1) and DRM2 are closely related proteins with redundant roles in methylating non-symmetrical CHH sequences (Zhang et al., 2018). Plants can also actively demethylate DNA via the action of DNA Glycosylase-Lyases, the so-called DEMETER-Like DNA demethylases (DMLs), which remove methylated cytosine, which is then replaced by a nonmethylated cytosine (Zhu, 2009). Genomic DNA methylation is an important mechanism that influences gene expression, and methylation at promoters is known to inhibit gene transcription (Gehring and Henikoff, 2007; Zhang et al., 2010). Therefore, it is likely that the active removal of methylation marks is an important mechanism during plant development and plant cell fate reprogramming, leading to the hypomethylation of sites important for DNA–protein interaction and gene expression (Liu et al., 2015). Early assessments of DNA methylation in tomato indicated an overall 20–30% decrease across the genome as fruits mature. These DNA-methylation changes were most pronounced in tomato fruit pericarp tissues (Giovannoni et al., 2017). Hadfield et al., (1993) reported that a decrease in DNA methylation in genes highly expressed in tomato fruits was coincident with the onset of ripening. A study by Zhong et al. (2013) provided new insights into the importance of DNA demethylation in tomato fruit ripening. They reported dynamic changes in methylation distribution during fruit development and revealed a loss of DNA methylation in the promoters of more than 200 ripening-related genes. These included genes encoding proteins involved in carotenoid accumulation, in ethylene synthesis, in fruit softening and several transcription factors, which controlling ripening induction such as RIN-MADS, CNR-SPL and a NAC transcription factor Non-Ripening (NOR-NAC) (Zhong et al., 2013). This data was consistent with the idea that active DNA demethylation is involved in the tomato fruit ripening process and it also indicated that demethylation does not occur in a random way.

The tomato genome contains four putative DML genes encoding proteins with characteristic domains of functional DNA Glycosylase-Lyases (Mok et al., 2010). SlDML1 and -2 are orthologous to the Arabidopsis AtROS1 (Repressor of Silencing 1) gene and SlDML3 to AtDME (DEMETER), whereas SlDML4 has no closely related Arabidopsis orthologue. All four SlDML genes are ubiquitously expressed in tomato plants, although SlDML4 is expressed at a very low level in all organs. SlDML1, SlDML3, and SlDML4 are barely expressed during fruit ripening. However, SlDML2 (Solyc10g083630) mRNA abundance increases dramatically

in ripening fruits. It was recently also reported that RNA silencing mediated repression of the SlDML2 gene expression in transgenic tomato, inhibits fruit DNA demethylation and strongly inhibits the onset and progression of fruit ripening (Liu et al., 2015). This demonstrated that SlDML2 is central in mediating the promoter DNA hypomethylation necessary for ripening progression of tomato (Liu et al., 2015).

During ripening, tomato fruits undergo a global DNA demethylation due to the highly induced expression level of SlDML2 (Liu et al., 2015). The increased expression level of SlDML2 leads to active DNA demethylation that mainly activates genes, involved in fruit ripening and also silences a set of genes involved in fruit growth which therefore become unnecessary during ripening (Lang et al., 2017). Therefore, SlDML2 mediated regulation of fruit ripening represents a unique model system of somatic cells to epigenetic reprogramming as it can switch the function of fruit pericarp cells from growth to ripening, without inducing cell differentiation. The molecular mechanism behind the target selection of SlDML2 mediated DNA demethylation and the regulation of SlDML2 expression has not been studied in any crop plants to date. Most of our knowledge of DNA demethylation comes from Arabidopsis, where the target selection of the DNA demethylases is defined by the IDM complex (Figure 1) which marks the target sites of ROS1.

A

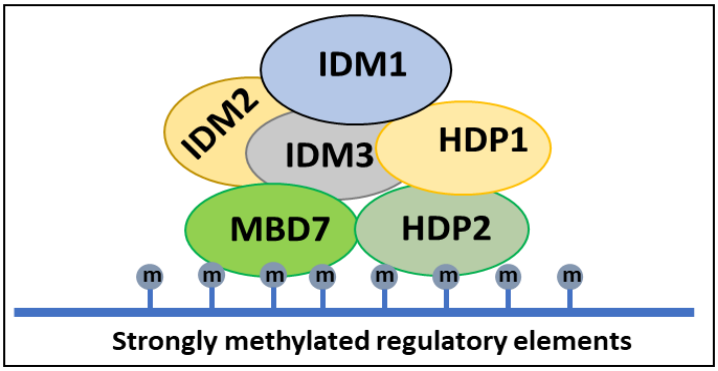


Figure 1. The protein composition of the Arabidopsis IDM protein complex.

The IDM complex contains two DNA binding proteins that recognize methylated cytosines, MBD7 (canonical Methyl-CpG-binding domain) and HDP2 (Harbringer transposon derived DNA methyl-binding-protein) which ensure that the IDM complex is targeted to highly methylated regulatory sequences, where active DNA demethylation is required for gene expression (Lang et al., 2015; Wen-Feng et al., 2019). Two α-crystallin domain containing proteins, IDM2 and IDM3 proteins hold the IDM complex together and connect the methylated DNA recognition components to the core enzyme, which is the IDM1 histone acetyltransferase. Furthermore, IDM2 and IDM3 also function as chaperone proteins to ensure IDM1 activity (Lang et al., 2017). The action of IDM1 create a permissive chromatin environment for ROS1 and related DNA demethylases recruitment and function by specifically binding to chromatin sites that lack histone H3K4 di- or trimethylation and acetylating H3K18 and H3K23 (Lang et al., 2017). Recently, 18 putative tomato MBD proteins (SlMBD) were identified using *in silico* experiments including orthologues of AtMBD7 protein (Parida et al., 2018). Furthermore, we

also identified the tomato orthologues of Arabidopsis IDM1, IDM3 and HDP2 (SlIDM1, SlIDM3 and SlHDP2). These data suggested to us, that target selection and DNA demethylation by SlDML2 in tomato might use similar molecular machineries as it was described in Arabidopsis.

Enhancing genetic and phenotypic variation in crops that cause subtle changes in quantitative traits are the most desired by breeders. It was shown by many quantitative trait loci (QTL) and genome-wide association study (GWAS) in both animals and plants that many of the genetic changes driving evolution, domestication and breeding occurred in cis-regulatory regions (Meyer et al., 2013; Olsen and Wendel, 2013; Wang et al., 2014). Since they do not alter protein structure, cis-regulatory variants are frequently less pleiotropic and often cause subtle phenotypic change by modifying the timing, pattern or the level of gene expression (Wittkopp and Kalay, 2011). Though widely favoured in evolution and domestication, mutations in cis-regulatory elements (CRE) are far from saturated and thus represent an untapped resource for expanding allelic diversity for breeding. Therefore, expanding cis regulatory variation holds promise for crop improvement and breeding. A recently introduced powerful new approach to create novel allelic variation is through genome editing (Doudna and Charpentier, 2014; Hsu et al., 2014).

As described above, the SlDML2 mediated active DNA demethylation is central to the control of fruit ripening in tomato. During this research project, we used CRISPR/Cas9 based genome editing to target cis-regulatory motifs of SlDML2 promoter to engineer fruit ripening variants of tomato. Genome editing driven mutagenesis of SlDML2 promoter created several mutant tomato lines that could shed light on our understanding which CREs influences fruit ripening in tomato. The findings written in this report are not published yet, therefore we give a detailed description of our results here.

2. Characterization of SlDML2 gene expression level variations in different tomato varieties.

SlDML2 is central in mediating DNA hypomethylation at several thousand loci which are necessary for ripening progression of tomato fruits (Liu et al., 2015). The regulation of ripening mediated by the DNA methylation/demethylation balance has evolved as double lock mechanism, along with changes in gene expression as a result of developmental cues, to prevent premature dispersal of seeds prior their full maturation (Gallusci et al., 2013). Cultivated tomato varieties and wild tomato species show considerable variation for ripening behaviour (Grumet et al., 1981), that could be used to survey natural variations in SlDML2 expression pattern, timing and level. Therefore, we have started to study the gene expression pattern of SlDML2 mRNA in Moneymaker and in other tomato varieties (Balkon, Mini, Manó, Microtom) with different growing times during fruit ripening. According to the literature, SlDML2 expression increases dramatically at the fruit breaker stage and this elevated expression are maintained at the orange and red stage (Liu et al., 2015). In our samples, we observed similar gene expression pattern (Figure 2) as it was published earlier in the literature, although we also found significant variations in SlDML2 expression during the different stages of the ripening process.

Based on the qPCR analysis of the mRNA expression patterns, our samples can be classified into two gene expression groups. The SlDML2 mRNA expression increased dramatically in *S. lycopersicum cv.* Microtom and Balkon to the breaker stage and then

decreased in the orange and red stage. In the other group (Manó and Mini) SlDML2 expression increased moderately to the breaker stage and slightly elevated in the orange and red stage (Figure 3A). Based on this result, it is plausible that timing and extent of demethylation may represent an important source of variation in the diversity of kinetics and intensity of ripening found among tomato varieties, thus presenting a frontier for further investigation for us.

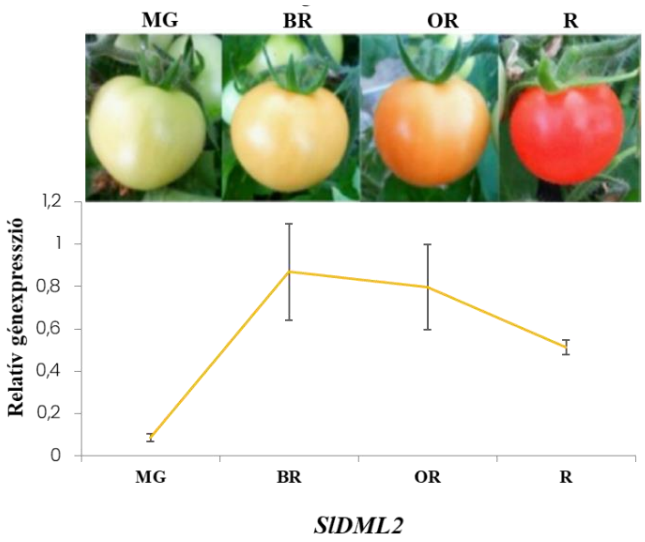


Figure 2. SlDML2 transcript profiling in Moneymaker tomato fruits at different stages of fruit ripening (MG - mature green; Br - breaker; OR - orange; R - red).

SlDML2 is highly induced at the onset of fruit ripening and it was demonstrated previously, that its repression resulted in DNA hypermethylation and substantially delaying ripening (Liu et al., 2015). It was also shown that transient increases in genome wide DNA methylation occur during chilling stress and the re-methylated regions includes many loci that were demethylated during ripening by SlDML2 (Zhang et al., 2016). We also analysed SlDML2 expression in post-harvest tomato fruits in response to chilling. We observed reduced transcript level of SlDML2 during cold storage (Figure 3B).

According to these observations SlDML2 may contribute to both developmental and chilling stress associated changes in fruit DNA methylation with corresponding effects on both fruit ripening and quality characteristics influenced by cold storage.

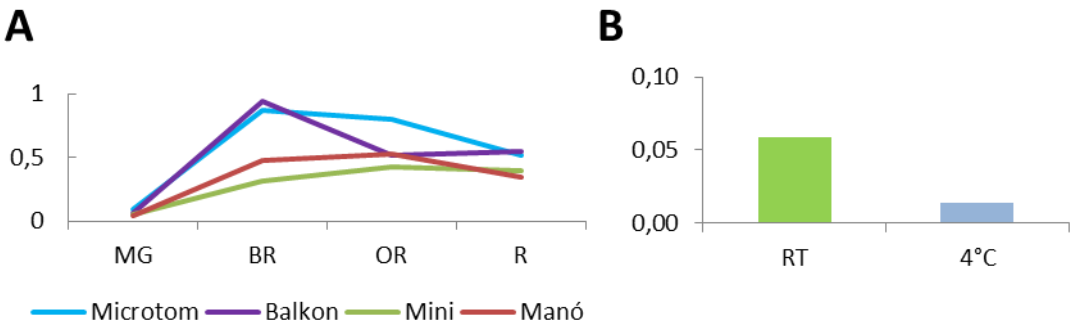
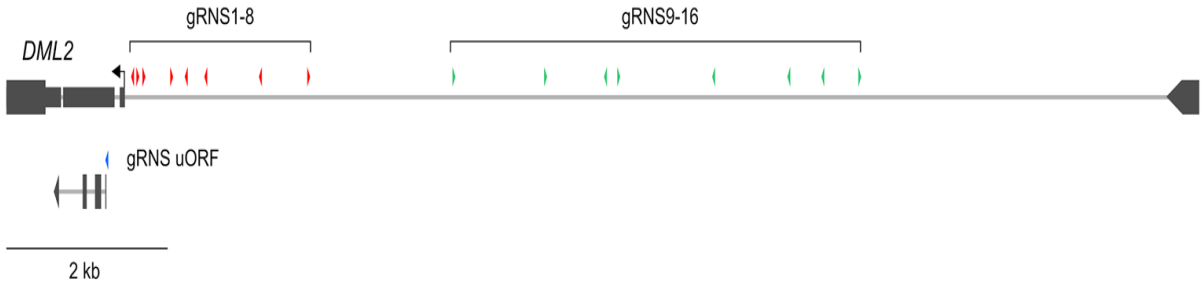


Figure 3. Transcript profiling of SlDML2 in developing and ripen tomato fruit. (A) qPCR analysis of SlDML2 gene expression in different tomato varieties in the developing fruit (MG - mature green; Br - breaker; OR - orange; R - red). **(B)** qPCR analysis of ripen tomato fruits after post-harvest storage for two days at room temperature (RT) and cold storage at 4 °C (4 °C).

3. Creating fruit ripening QTL by CRISPR/Cas9 mutagenesis of cis-regulatory elements of SlDML2 promoter.

The SlDML2 gene has a very large promoter region, since the closest predicted gene is more than 10 kb away from the predicted transcription site of SlDML2. The DNA sequence of the SlDML2 to promoter region is extremely AT-rich (77%), which made the design of specific guide RNAs difficult. According to the SL3.0 ITAG3.0 annotation, the SlDML2 gene also has a long UTR with two introns suggesting that the gene expression could be regulated by nonsense-mediated RNA decay (NMD). To create targeted deletion series in the SlDML2 promoter, we designed eight, evenly spaced CRISPR guide RNAs within a 2.5 kb region upstream of the transcription start site (gRNA1-8), another eight within a 5 kb region in a more distant part of the promoter (gRNA9-16), and another one in the UTR (Figure 4A). As a control, we created a construct carrying a gRNA targeting the first exon of the SlDML2 gene (CDS) (published by Lang et al., 2017) to create a DML2 null mutant in different tomato varieties (Microtom, Moneymaker, M82). We designed the gRNAs using the CRISPOR web server (http://crispor.tefor.net) selecting the PAM sequence of SpCas9 and the tomato genome version Solgenomics SL3.0 ITAG3.0.

A



B

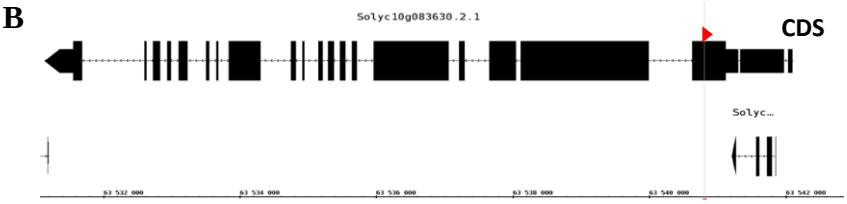


Figure 4. Position of CRISPR/Cas9 guide RNAs (gRNS) targeting the SlDML2 gene or the promoter region of the gene. (A). Red and green arrow heads represent the position of gRNAs in the promoter region of SlDML2. Blue arrow head shows the position of the gRNA targeting the upper ORF in the 5' UTR. **(B).** Red arrow head show the position of gRNA in the first exon of the gene.

To create the multiplex CRISPR constructs, we followed the protocol by Rodríguez-Leal et al., 2017. The components of the modular cloning were obtained from Addgene (MoClo Tool Kit #1000000044). The guide sequences were introduced into the gRNA expression unit (U6 promoter::unique guide sequence::gRNA scaffold::U6 terminator) by overlap extension PCR using pICH86966_AtU6-sgRNA-PDS as a template (Addgene #46966). First, the individual gRNA expression units were cloned into first-level vectors having the same internal

BsaI sites and distinct flanking BpiI sites (gRNA1 and 9: pICH47751, gRNA2 and 10: pICH47761, gRNA3 and 11: pICH47772, gRNA4 and 12: pICH47781, gRNA5 and 13: pICH47791, gRNA6 and 14: pICH47732, gRNA7 and 15: pICH47742, gRNA8 and 16: pICH47751) with BsaI digestion and ligation in one reaction (Golden Gate cloning method). From these first-level vectors, four gRNA expression units with compatible BpiI sites were assembled into one unit in an intermediate-level vector (pAGM8055) with the end-linker pICH50927, and the other four into another intermediate vector with compatible flanking BsaI sites (pAGM8093) with the end-linker pICH50892. From these two intermediate vectors, the 4-gRNA units (gRNA1-4 and gRNA5-8, gRNA9-12 and gRNA13-16) were assembled in the second-level vector pAGM4723 along with the plant kanamycin selection marker gene (pICH47732::NOSp-NPTII-OCST, Addgene #51144), the Cas9 expression unit (pICH47742::2x35S-5'UTR-hCas9(STOP)-NOST, Addgene #49771), and the end-linker pICH49277 resulting in the final gRNA1-8 and gRNA9-16 containing binary vectors. In the case of the CDS targeting and 5' UTR targeting constructs, the intermediate cloning step was skipped and pICH47751 containing a single gRNA was cloned directly in pAGM4723 along with the KanR and the Cas9 units and the end-linker pICH49277. The constructs were tested with restriction digestion and Sanger sequencing. The four different constructions (gRNA1-8, gRNA 9-16, UTR and CDS) were introduced to Agrobacterium tumefaciens (strain: LB4404) and first we transformed Microtom and M82 tomato varieties with them (Figure 5). Tomato transformation was done by a modified method of Fernandez et al., 2009.

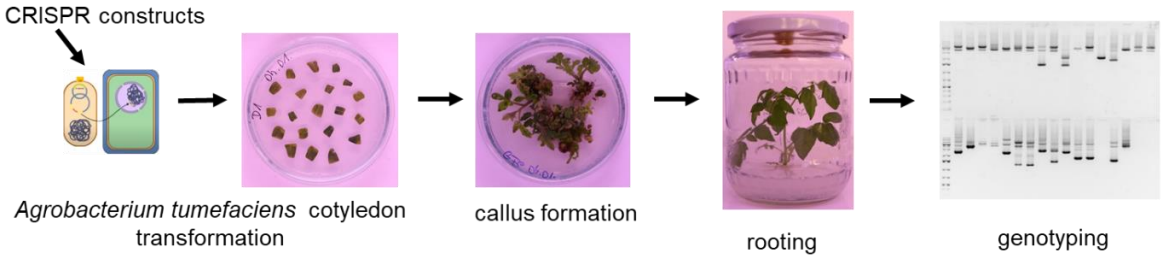


Figure 5. The workflow of tomato transformation.

To monitor the effect of the tissue culture process, we also regenerated wild type plants without transformation from a tissue culture that were used as controls in later experiments. After M82 transformation we got 4 rooted plants with CDS, gRNA1-8 and gRNA 9-16 constructions, after Microtom transformation we got five CDS, eight gRNA1-8 and two gRNA 9-16 rooted plants and we did not get any UTR plants. Since the transformation efficiency and the number of rooted plants was quite low, we decided to transform another tomato variety with the same CRISPR/Cas9 constructs (Moneymaker), which is known to have a good transformation efficiency. As expected in this case, we got many transformed regenerating lines (18 CDS, 21 UTR, 46 gRNA1-8, and 33 gRNA9-16) for all the four types of the genome editing constructs.

We genotyped the transformed tomato plants and it revealed that the lines carrying the promoter-targeting gRNA1-8 and gRNA9-16 constructs resulted in complex genome edited patterns. We have observed many different mutational events, including large deletion between

distant gRNAs, inversions, translocations, and small deletions at individual gRNA sites. We made the general observation that the mutagenesis was very efficient and precise, almost all of the gRNA sites were mutated, although some gRNA sites remained wild-type and mutations outside the gRNA sites were observed. We list the genotype of the most interesting genome edited lines in Table 1.

Table 1. Deletion series in the SlDML2 promoter region. Lines A, carries deletion in the proximal region of SlDML2 promoter targeted by gRNA1-8. Lines B, carries deletion in the dital region of SlDML2 promoter targeted by gRNA9-16. The distances between gRNAs are shown in the headers of the table. Yellow colour represents the deleted regions in the promoter regions between the gRNAs. The size of the deletion is also shown (eg. A12 line contain a 9 nt deletion at the target site of gRNA1 and 2104 nt deletion between gRNA2 and gRNA8). Pink represent insertion at the gRNA target sites. Blue represents inversions between the gRNA target sites. Grey represents deletions with mixed sequences where we were not able to define the size of the deleted region exactly. WT and white represents the wild type sequence without any modifications.

Vonal#	gRNS1	103 bp	gRNS2	84 bp	gRNS3	319 bp	gRNS4	181 bp	gRNS5	240 bp	gRNS6	681 bp	gRNS7	599 bp	gRNS8
A12	D9	WT	D2104												
A20	D106	WT	I1	WT	D9	WT	D30	WT	D2	WT	D1272				
	D11	WT	D835									WT	WT	WT	D1
A34	I1	WT	D1	WT	D751							WT	WT	WT	D1
A61	D10	WT	WT	WT	D3	WT	<< 130 bp D1744 501 bp >>							WT	D2
A120	I1	WT	D2104												
A170	I1	WT	D829									WT	WT	WT	D2
A201	D12	WT	D5	WT	D5	WT	<< 324 bp D505			WT	D5	WT	?	WT	D1
	D?	WT	D1512											WT	I1
A230	?	?	?	?	?	?	?	WT	D256			WT	WT	WT	D7
A250	D8	WT	D400					INV438				WT	WT	WT	D173
A280	I1	WT	D835									WT	WT	WT	D1
	D107	WT	I1	WT	D9	WT	D6	WT	D2	WT	D1272				
A291	D12	WT	D1517											WT	D1
	gRNS9	1142 bp	gRNS10	741 bp	gRNS11	174 bp	gRNS12	1175 bp	gRNS13	940 bp	gRNS14	422 bp	gRNS15	445 bp	gRNS16
B40	D14	WT	<< 300 bp D3809 72 bp >>											WT	I1
B60	D3	WT	D1	WT	D3143										
	D3	WT	D3903												
B90	D2	WT	INV751			D3147									
B91	D1	WT	D3462											WT	D1
B111	WT	WT	D3054									WT	D7	WT	I1
B120	WT	WT	?	?	?	?	?	WT	D1	D1739					I1
B140	D6	WT	?	?	?	?	?	?	?	?	?	WT	D5	WT	D3
B161	D4	WT	?	?	D2712									WT	D8
B170	WT	WT													
B171	D14	WT	<< 120 bp D3677										I38 117 bp >>	WT	D1
B190	I1	WT	<< 292 bp D3820 73 bp >>											WT	I1
B141 B150															

Regarding the phenotypes of the mutant tomato plants, we observed several interesting phenotypes. As a summary, we can say that the cds lines (they carry a mutation in the first exon of SlDML2 and do not produce functional SlDML2) in all the three varieties used for transformation (Microtom, M82, and Moneymaker) ripened much slower (wt: 6 days, cds: 14 days) but the fruits of the Microtom and M82 homozygous cds lines were mostly seedless (or only a few seeds were obtained). The fruits of the Moneymaker homozygous two independent cds lines did produce seeds (although contained many small, aborted seeds as well), had somewhat elongated shape, ripened much slower than WT tomatoes, and typically remained orange. The plants also showed an interesting physical appearance that were not previously reported: they have plenty of branches and lots of leaves growing from the axillary buds and

the leaves grew upward resulting in a broom-like overall look. This phenotype suggests that the plants have some problem with the apical dominance and/or a meristem-related function (Figure 6A). These plants were producing also less trichomes than the wild-type Moneymaker plants. It is worth to note that these plants carry the same mutations as the T0 Microtom plants reported by Lang et al., 2017. However, the published biallelic T0 Microtom plants (sldml2-1) did not show any developmental problem they had only late ripening phenotype (Figure 6B). The phenotype of later generations of sldml2-1 plants were not investigated by Lang et al., 2017.

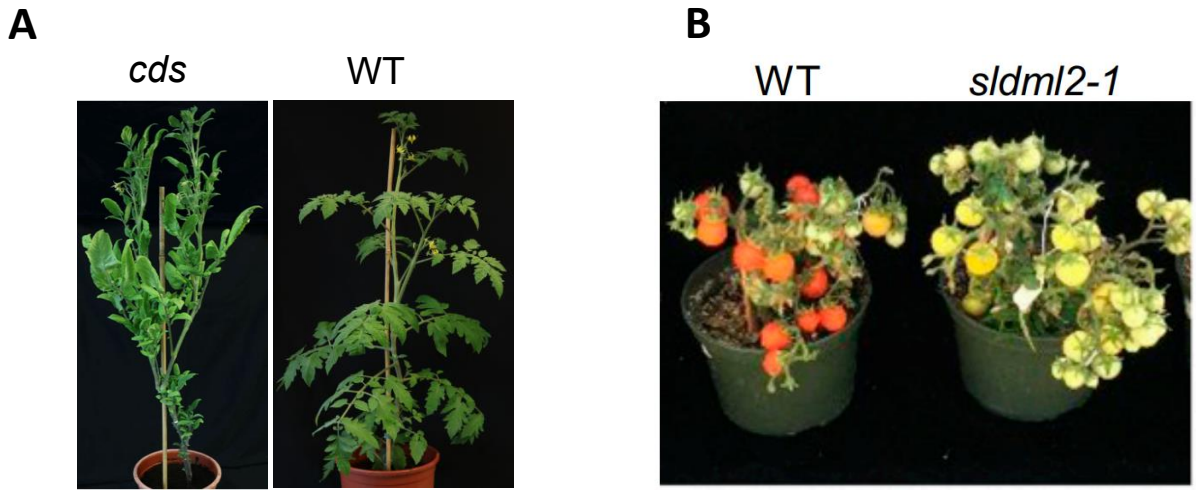


Figure 6. The phenotype of the SlDML2 cds mutants. (A). Comparison of the vegetative phenotype of T2 sldml2 mutant (cds) and WT Moneymaker plants prepared in our laboratory to assess the phenotype of plants not producing functional SlDML2. **(B)**. Ripening phenotype of the wild type and *sldml2-1* mutant Solanum lycopersicum 'Microtom' plants of the same age from the

Regarding the utr lines, we got several transgenic genome edited T2 homozygous lines that carried a small deletion or 1 nt insertion before the uORF (there is a small ORF in the 5'-UTR, that could have regulatory role) and we also have mutant T2 homozygous lines (4 utr lines) that contain larger deletions spanning the starting ATG of the uORF. Interestingly, the fruits of these lines ripened in a mosaic way (they showed red and green patches on their fruits for a prolonged time while wild-type plants quickly got through this phase) and sometimes had multiple locule numbers (WT Moneymaker plants have only two locules in their fruits). However, their total ripening time was similar to the wild-type. Both the small and large deletion-carrying utr plants contained more trichomes than the wild-type or the cds lines.

Regarding the promoter mutants, we have several interesting lines carrying large deletions either in the proximal (gRNA1-8) or the distal (gRNA9-16) part of the SlDML2 promoter. The fruit ripening time was similar to the wild-type in most of the obtained genome edited Moneymaker transgenic lines, however, there were few mutant lines that showed significantly shorter ripening time (A34 - 3, 4 days instead of 6) while others ripened longer (A12, A120, B60, B171, and B190 - 8, 9, 12 days). These measurements were performed in the greenhouse during the summers; therefore, they should be also confirmed under controlled environmental conditions in phytotrons, if they will be available. Developmental defects were

more obvious, however. For example, many lines produced seedless fruits (eg. A120) more frequently than the WT. There were mutant lines that produced only small, seedless fruits and many lines have fruits with more than two locules (Fruits of WT Moneymaker plants has two cules, only; Figure 7).

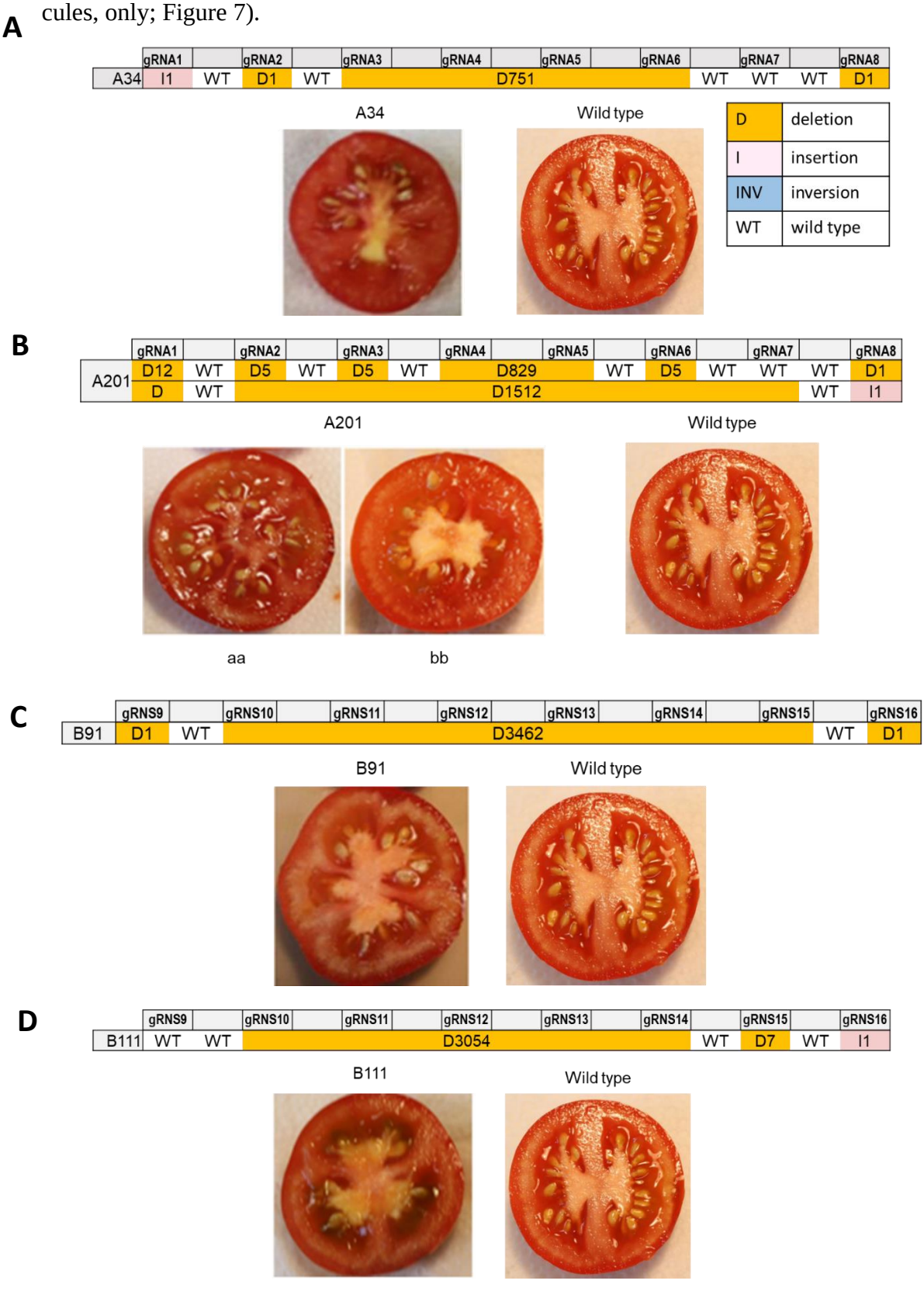


Figure 7. Large deletions either in the proximal (A and B) or in the distal part (C and D) of SlDML2 promoter region of Moneymaker plants results multi locule fruit phenotype.

Three lines (A120, B141, B150) showed serious developmental defects: they looked like a “plastic” plant with no trichomes (only the small, glandular type VI trichome) and very stiff leaves and cauliflower-like apical and axillary meristems. Most of them produced fruits, mostly seedless, but one line (B141-1) development was so seriously affected that it never produced any flowers (Figure 8). Its phenotype suggested that it has a meristem development defect (like in the case of cds mutants but much stronger).

Figure 8. The phenotype of B141 plants that carry a 4400 nt long deletions in the distal part of SlDML2 promoter region. The upper row shows the phenotype of the WT Moneymaker plants.

The fact that a very similar phenotype was observed in SlDML2 promoter mutant Moneymaker tomato lines carrying large deletions (several kb) in the proximal or the distal part of the promoter suggests that these two parts of the promoter may physically interact like a promoter and an enhancer/silencer region. However, this hypothesis needs to be rigorously tested in the future.

The other proximal promoter mutant, which carries a large deletion and has a very strong developmental phenotype was the A120 tomato line. This mutant contains a 2104 nt deletion between gRNA2 and gRNA8 in the proximal part of SlDML2 promoter and the expression level of SlDML2 mRNA decreases to the same level as of the C33 cds mutant line (although the C33 cds mutant do not produce functional SlDML2 proteins, Figure 9). The mutant plants have very strong and characteristic phenotypes throughout their life cycle (Figure 10). The juvenile plants have potato like leaves and these plants do not have apical dominance therefore they develop side branches that gives a bushy appearance to the plants although the height of the WT and A120 mutants are almost similar. The adult leaves of A120 mutant plants are very distorted, contains much less trichomes than WT plants and it mostly lacks type I glandular and type III non-glandular trichomes. The apical shoot of A120 plants are almost trichomless and

in many occasions they contain spontaneous necrotic regions. Most of the fruits of A120 plants is small, seedless and late ripening and only very occasionally produces fruits with a few seeds.

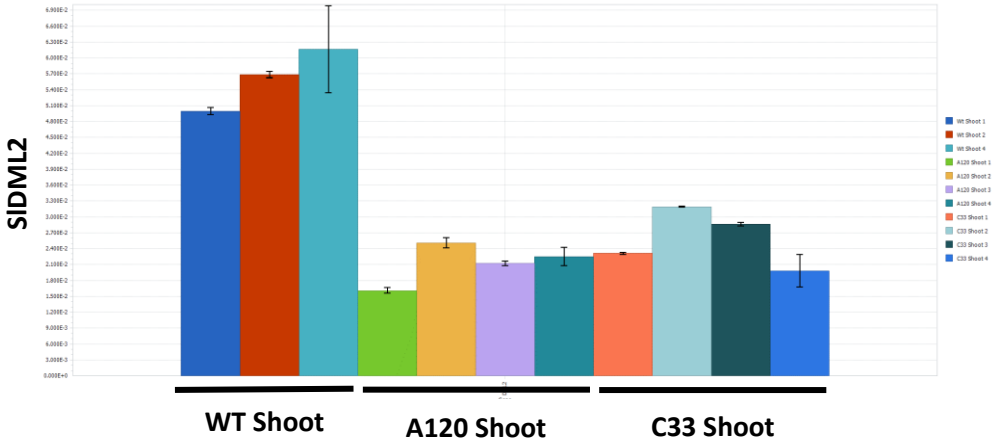


Figure 9. RT-qPCR analysis of SlDML2 expression level in wild type (WT), A120 proximal promoter mutant and cds mutant (C33) Moneymaker tomato shoots. The C33 plants do not produce functional SlDML2.

These strong developmental phenotypes clearly show that maintaining the proper spatial and temporal expression of SlDML2 is critical to normal development. Previous works only reported the role of SlDML2 in fruit ripening (Liu et al., 2015; Lang et al., 2017) and did not observed problems during other stages of plant development. The reason for the observed differences could be partly because: **(1)** different tomato varieties were used in each experiment (Ailsa Craig – Liu et al., 2015; Microtom – Lang et al., 2017; Moneymaker – in our laboratory).

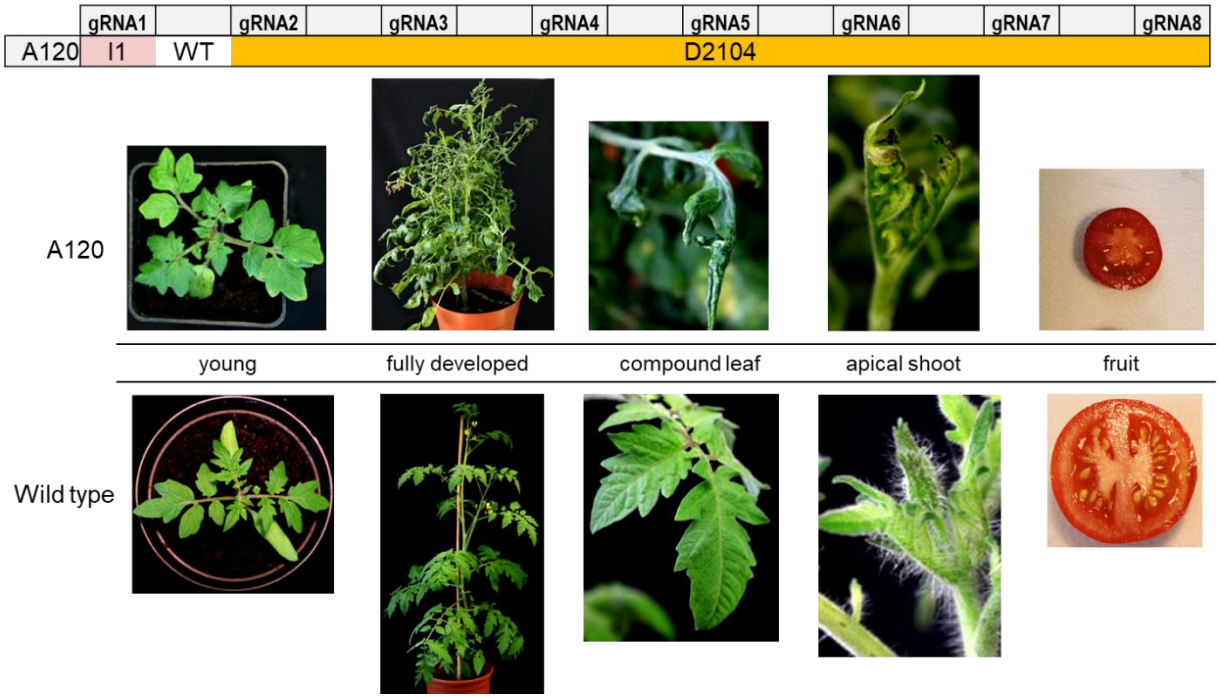


Figure 10. The phenotype of A120 mutant plants that carry a 2104 nt long deletions in the proximal part of SlDML2 promoter region between gRNA2 and gRNA8. The upper row shows the phenotype of the A120 and lower row of the WT Moneymaker plants.

(2) Liu et al., 2015, used knocked down mutant plants where the expression of SlDML2 was silenced by RNAi and Lang et al., 2017, used only T0 *sldml2* mutant plants in their experiments. Other very important difference that in our A120 mutant plants we deleted a large proximal region of the SlDML2 gene promoter, while the protein coding part remained unaltered. It was reported in Arabidopsis that the promoter region of the SlDML2 gene is a DNA methylation monitoring sequence (MEMS) or "methylstat" and it has a very important role to detect and regulate the methylation status of the genome (Lei et al., 2015). Therefore, deleting an important methylation level sensor region from the genome could results in developmental defects. In accordance with this we also observed highly methylated regions in the SlDML2 gene promoter region at both the proximal and distal part of the promoter where the guide RNAs were designed (Figure 11). To establish and investigate the exact role of the SlDML2 promoter as a "methylsat" in tomato and its effect in the regulation of other tomato demethylases requires further experimental work in the future.

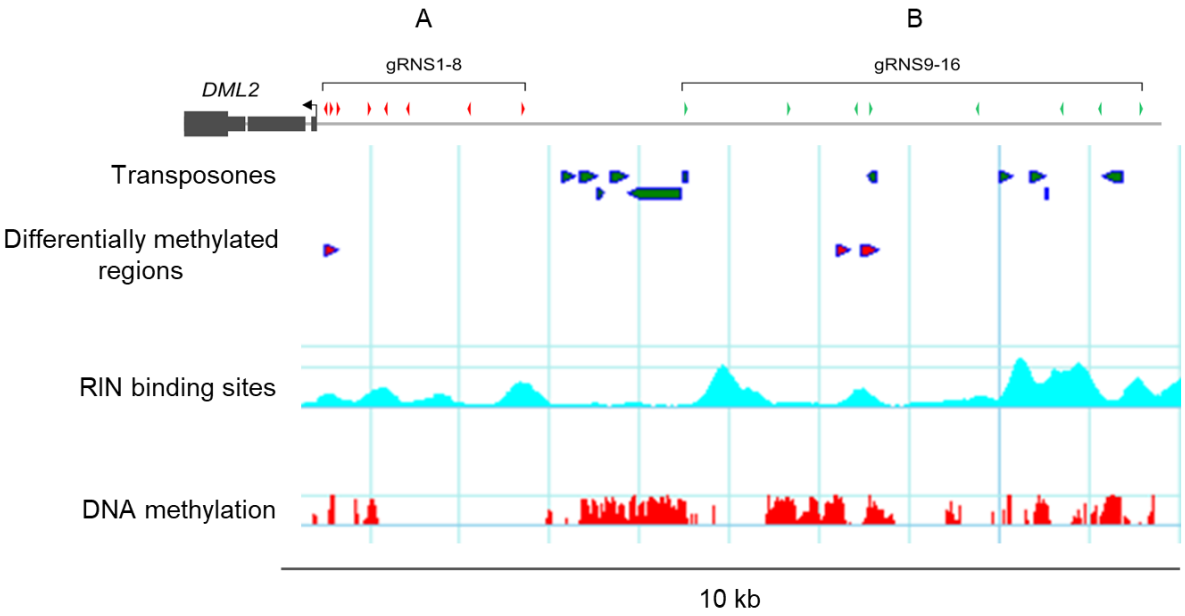


Figure 11. DNA methylation of the SlDML2 gene promoter region. Position of methylated cytosines is shown in red. Hypermethylated regions in the sldml2 mutant plants are represented by purple triangles. SlDML2 is regulated RIN (a MADS-box transcription factor) during fruit ripening. RIN binding to the SlDML2 promoter is shown by cyan.

4. Identifying putative SlDML2 target loci in *slidm3* mutant tomato by bisulfite sequencing

Most of our knowledge of DNA demethylation comes from Arabidopsis, where the target selection of the DNA demethylases is defined by the IDM complex (Figure 1) which marks the target sites of ROS1 (SlDML2 is the orthologue of ROS1). IDM3 is an α-crystallin domain containing protein, which hold the IDM complex together and connect the methylated DNA recognition components to the core enzyme, which is the IDM1 histone acetyltransferase. Arabidopsis *idm3* T-DNA mutants show DNA hypermethylation at transposable elements (TEs) and other repeats which partly overlap with hypermethylated regions of ROS1 (Lang et

al., 2017). First, we identified the tomato orthologue (SlIDM3) of Arabidopsis IDM3. Next, we characterised the expression level of SlIDM3 during plant development and fruit ripening (Figure 12).

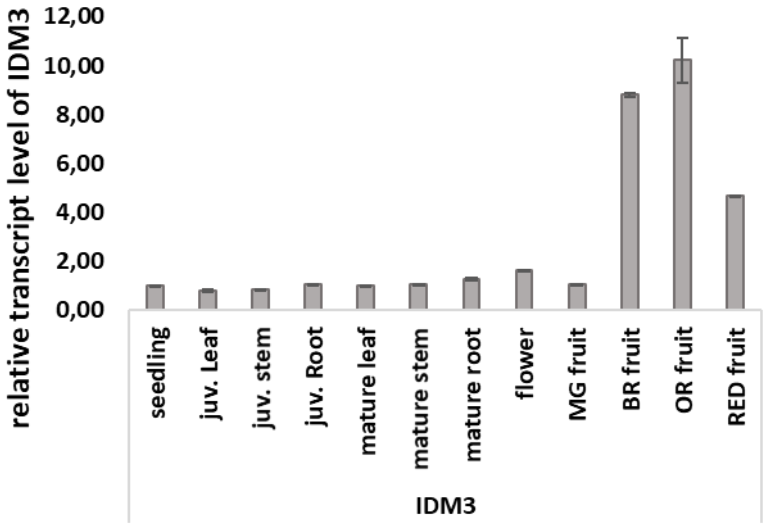


Figure 12. Tissue specific expression of SlIDM3 in Moneymaker tomato. Relative expression was determined by qRT-PCR using three biological replicates of different tissues from WT tomato plants.

SlIDM3 and SlDML2 showed the same expression pattern, both of them is constitutively expressed in all analysed tissues and reached the highest expression level during fruit ripening. To examine the role of SlIDM3 in the target selection of SlDML2 we used CRISPR/Cas-9 genome editing system with two independent gRNAs to knock out SlIDM3. We selected two independent T2 homozygous line, each harbouring a premature stop codon in the first exon. *slidm3-1* (gRNA11: 1 nt deletions (Figure 13), that results an early stop codon 153 nt after the

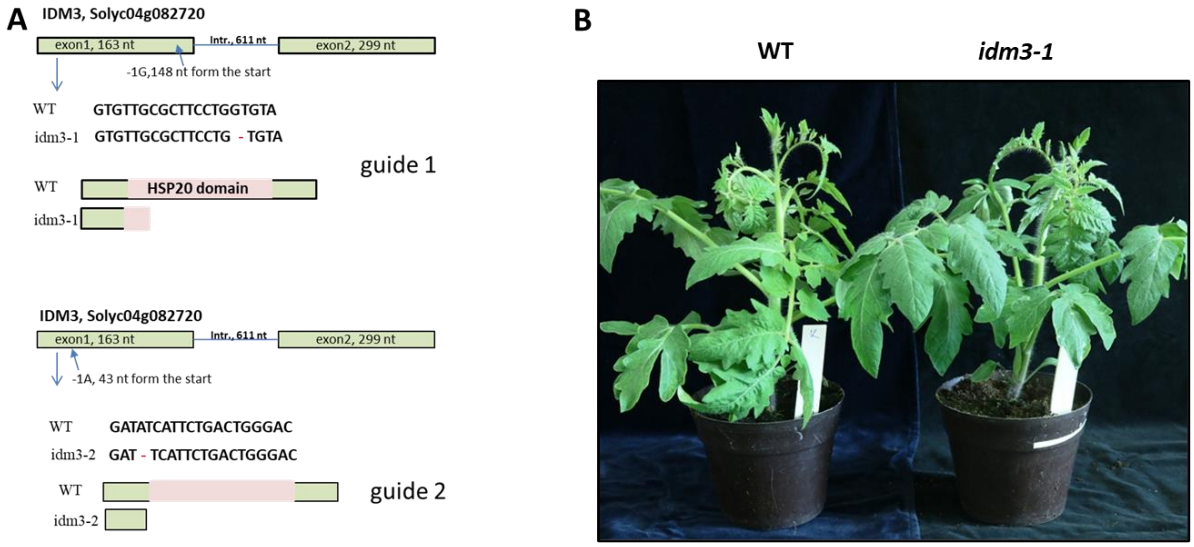


Figure 13. Preparation of *slidm3* mutant Moneymaker plants. (A) Schematic representation of SlIDM3 CRISPR mutations. Mutations (single nt deletion) in the first exon of SlIDM3 give rise to a premature stop codon. (**B)** Growth phenotype of WT and *idm3-1* plants.

start codon) and *slidm3-2* (gRNA2: 4 nt deletion, that also results an early stop codon 51 nt after the start codon). Homozygous *slidm3-1* and *slidm3-2* plants showed higher trichome density in leaves, vegetative stems and inflorescence stems when compared to wild-type (WT) plants, without any other morphological change in vegetative or reproductive organs. The fruit ripening of *slidm3* mutants were the same as WT plants. To elucidate its role in DNA demethylation by tomato demethylases (SlDML1, SlDML2, SlDML3 and SlDML4) *slidm3* mutant lines were analysed through the whole-genome bisulfite sequencing (WGBS) to generate single-base resolution maps of DNA methylation from the vegetative stem of *slidm3* and WT plants (3 replicates of WT were compared to the 3 replicates of *slidm3* mutant DNA). Over 99 % of the genomic cytozines were covered in each sample, and each methylome was sequenced with 20-fold coverage per DNA strand. The coverage and depth of these methylomes were higher than those of published methylomes of S. lycopersicum (Zhong et al., 2013; Lang et al., 2017). To identify IDM3 targets, we compared the methylome of WT with those of three biological replicates of *slidm3-1*. Our analysis identified 36143 hyper-DMR-s and 53177 hypo-DMRs. The distribution of hyper-DMRs and hypo-DMRs was context dependent (Figure 14). DNA hypomethylation in *slidm3-1* plants occurred in mCG sequence context (8632 hyper-DMRs vs. 30849 hypo-DMRs) while we observed hypermethylation in mCHH sequence context (19995 hyper-DMRs vs. 10822 hypo-DMRs).

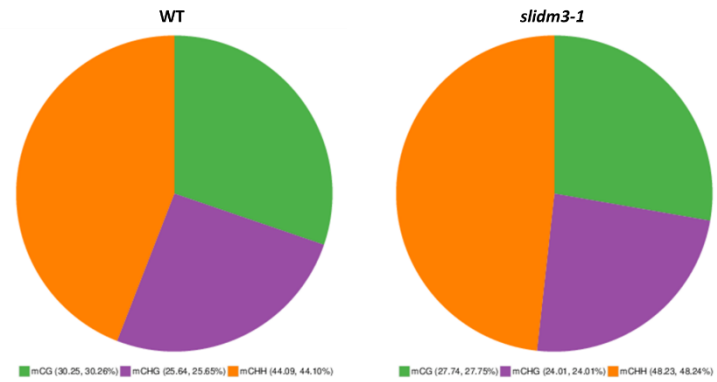


Figure 14. Classification of methylated cytosines. Different colours represent methylated cytosines (mCG – green; mCHG – purple; mCHH – orange) in different contexts, the area represents the percentage of methylated cytosines in the corresponding context.

The methylation status may vary among the genomes of various samples (tissue, cells, individuals, etc.). The differentially methylated regions involved could be participated in the transcriptional regulation of genes. Identification of differential methylation can be used to examine differences in epigenetics in different tissues. It is well known that DNA methylation is involved in cell differentiation and proliferation, and many DMRs have been found in embryo reprogramming (R-DMRs) and developmental stages (D-DMRs) (Reik et al., 2001; Meissner et al., 2008; Doi et al., 2009). Cancer also exhibits overall methylation loss and hypermethylation of localized areas such as CpG islands in the promoter region of the tumour suppressor gene compared to normal samples (Irizarry et. al, 2009). In order, to explore the target regions and also the role of DNA demethylation in tomato development, we focused on the differential methylation analysis between *slidm3-1* and WT samples. The methylation analysis for comparison groups includes: DMR distribution on chromosome (Figure 15), DMR distribution on functional genetic region (Figure 16), methylation distribution on up/downstream 2kb and gene body (Figure 17).

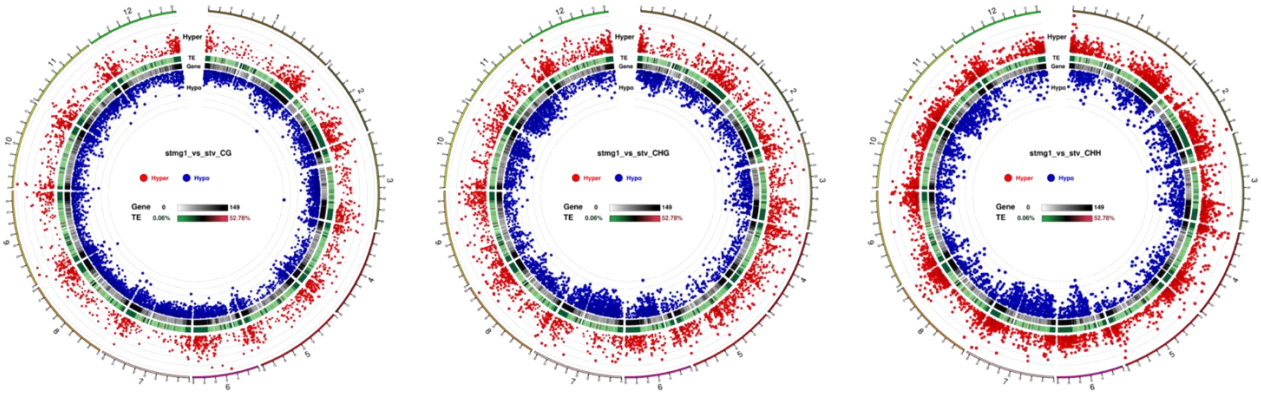


Figure 15. Circos plot for DMR condition in each sequence context (left-mCG/middle-mCHG/ right-mCHH). The 12 tomato chromosomes are shown as a circle. The circos plot represents (from outside to inside): 1. Hyper DMR statistical value. The higher and bigger the point, the larger differences between two groups. left-mCG/middle-mCHG/ right- mCHH is shown in red circle, 2. TE, the heatmap of percentage of repeat element. 3. heatmap of gene density 4. Hypo DMR statistical value. The higher and bigger the point, the larger differences between two groups. left-mCG/middle-mCHG/ right- mCHH is shown in blue circle,

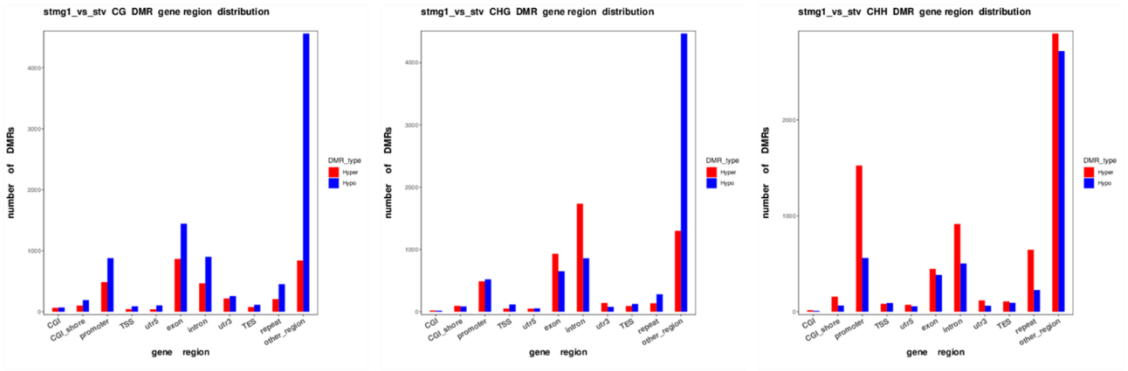


Figure 16. DMR distribution plot in mCG (left), mCHG (middle) and mCHH (right) context in different functional region. The x-axis is functional regions (from left to right: CGI, CGI-shore, promoter, TSS, 5'UTR, exon, intron, 3' UTR, TES, repeat, other region) the y-axis is number of hyper-DMRs (red) and hypo-DMRs (blue) in each region.

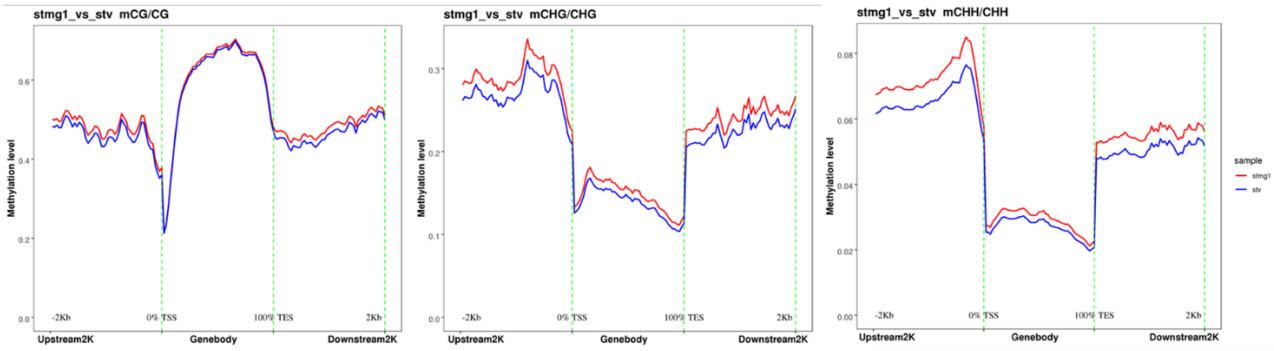


Figure 17. An overview of methylation level distribution at up/downstream 2kb and genebody in each sequence context (mCG, mCHG, mCHH). The x-axis is the functional genetic elements, the y-axis is the methylation level. After combining the samples with biological repeats, each region is divided into 50 bins and methylation level is calculate in each bin. WT is blue, slidm3-1 is red.

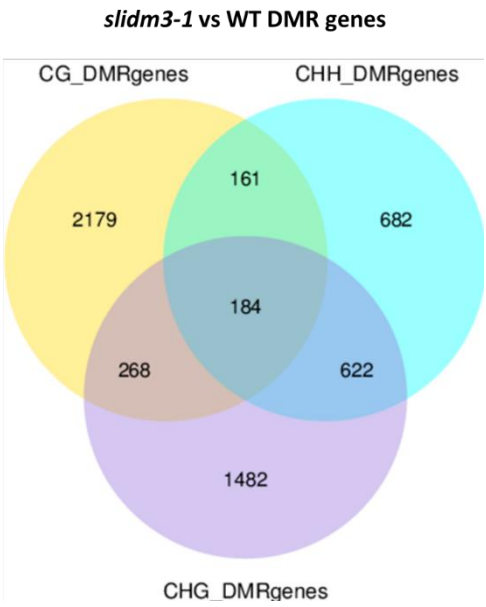


Figure 18. Venn plot for DMR target gene in three contexts (CG, CHG, CHH). The legend is the context type; the number means the number of genes in that context or in that overlapped context.

The distribution of the hypo-methylated mCG regions on the twelve tomato chromosomes were very similar (Figure 15) and the majority of them located in the intergenic regions (Figure 15 and Figure 16) that are depleted in genes and enriched in transposable elements. However, in the vicinity of protein coding genes we detected a slight increase in hyper-methylation also at mCG sites (Figure 17). The number of hypo-methylated and hyper-methylated DMRs (7519 hyper-DMRs vs. 11516 hypo-DMRs) of the mCHG context is very similar (Figure 15) and hypo-methylated DMRs are located mainly in intergenic regions while hyper-DMRs are mostly in exons and introns (Figure 16). Furthermore, mCHG sites are hyper-methylated in the vicinity of protein coding genes (Figure 17). The presence of hyper-methylated mCHH DMRs is the most obvious (Figure 15) and they mostly co-localize with gene rich regions of the genome. In the slidm3-1 mutant the promoter regions of protein coding genes are hyper-methylated in the mCHH sequence context (Figure 16 and Figure 17).

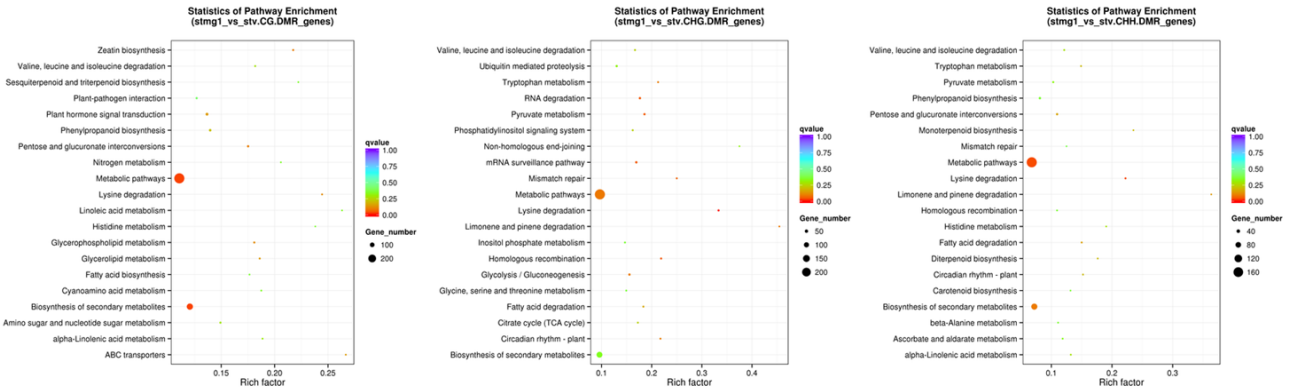


Figure 19. KEGG enrichment plot DMR related pathway. The x-axis represents Rich factor, and the y-axis represents pathway name. The size of points stand for DMRs related genes counts and the colours stand for different Qvalues range. Statistic of pathway enrichment: slidm3-1 vs wt mCG (left)/ mCHG (middle)/ mCHH (right) DMR genes

After filtering the DMRs we identified 184 protein coding genes that were differentially methylated in all sequence context (mCG/mCHG/mCHH) in *slidm3-1* plants (Figure 18). Next, KEGG enrichment analysis was performed on genes with overlapping body bodies (from TSS to TES) and DMR (sequence context). Different genes interact with each other to achieve biological functions. Pathway enrichment can help to understand the main signalling pathways involved in DMR-related genes. KEGG is an important public database of Pathway (Kanehisa et la., 2008).

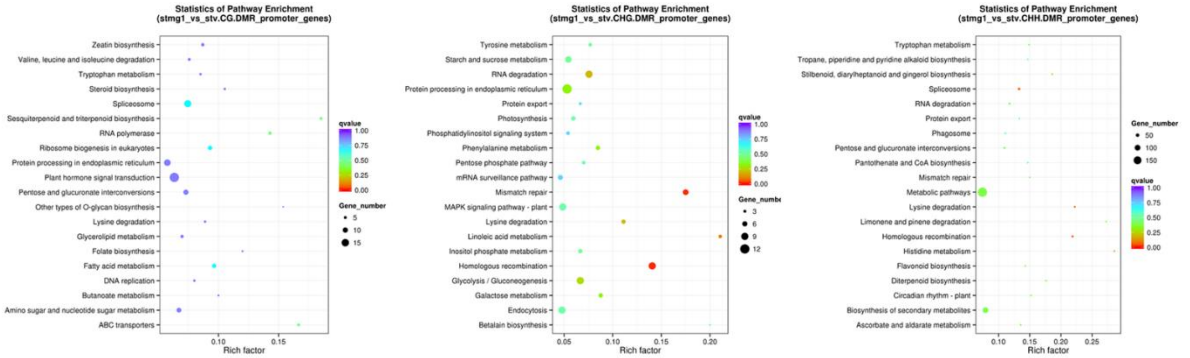


Figure 20. KEGG enrichment analysis of DMR anchoring promoter region related genes. The x-axis represents Rich factor, and the y-axis represents pathway name. The size of points stand for DMRs related genes counts and the colours stand for different Qvalues range. Statistic of pathway enrichment: slidm3-1 vs wt mCG (left)/ mCHG (middle)/ mCHH (right) DMR promoter genes

Pathway that was significantly enriched in DMR-related genes compared to the entire genomic background. The results of the KEGG metabolic pathway enrichment are shown on Figure 19. The scatter plots in Figure 19 are a graphical representation of the results of the KEGG enrichment analysis. In this figure, the degree of KEGG enrichment is measured by the number of genes, the Qvalue, and the number of genes enriched in this pathway. Where Rich factor refers to the ratio of the number of DMR-related genes enriched in the pathway to the number of annotated genes. The larger the Rich factor, the greater the degree of enrichment. The value of Qvalue is [0,1]. The enrichment will be more significant if the q-value is close to 0. The most significant enriched 20 pathways are presented in the scatter plot. Our results show that among the metabolic pathways the most prominent enrichment is within the genes of the biosynthesis of secondary metabolites pathway. We also performed the KEGG enrichment analysis of DMR anchoring promoter region related genes (Figure 20). Functional enrichment analysis based on the distribution of DMR on the genome and overlapping genes in the gene promoter region (2K upstream of the transcription start site). The most prominent hyper-DMRs in promoter regions are in mCHH sequence context and we also observed the enrichment of the genes of the biosynthesis of secondary metabolites pathway in their promoters. These finding correlates very well with the observed phenotypes of *slidm3* tomato plants, because they produce significantly more glandular trichomes than WT plants and glandular trichomes produce and store secondary metabolites.

In summary we can conclude that SlIDM3 in tomato participate in the targeting of DNA demethylases, since we observed mCHH context dependent hyper-methylation in the promoter regions of protein coding genes in the *slidm3* mutant plants. This observation is in line with previous report of Arabidopsis which showed, that active DNA demethylation by ROS1 is required for the maintenance of the expression of those genes that promoter regions are located to nearby transposable elements (Harris et al., 2018). The spatial and temporal gene expression pattern of SlDML2 (ROS1 orthologue) and SlIDM3 is identical, however SlIDM3 do not target SlDMl2 exclusively to its target loci during fruit ripening because the fruits of slidm3 mutants ripens in the same time as WT tomato fruits. Further research is required to identify the overlapping targets of the tomato IDM complex (containing SlIDM3) and SlDML2 in the future.

5. References

Doi A; Park IH, Wen B, Murakami P, et al. (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nature genetics, 41: 1350–3.

Doudna, J.A., and Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science 346:1258096.

Fernandez AI, Viron N, Alhagdow M, Karimi M, Jones M, Amsellem Z, Sicard A, Czerednik A, Angenent G, Grierson D, May S, Seymour G, Eshed Y, Lemaire-Chamley M, Rothan C, Hilson P. (2009). Flexible tools for gene expression and silencing in tomato. Plant Physiol. 151:1729-40. doi: 10.1104/pp.109.147546.

Gallusci P, Hodgman C, Teyssier E, Seymour GB. (2016). DNA methylation and chromatin regulation during fleshy fruit development and ripening. Front. Plant Sci. 7:807.

Gehring M, Henikoff S (2007). DNA methylation dynamics in plant genomes. Biochim Biophys Acta 1769:276–286.

Giovannoni, J., Nguyen, C., Ampofo, B., Zhong, S., Fei, Z. (2017). The Epigenome and Transcriptional Dynamics of Fruit Ripening. Annu. Rev. Plant Biol,. 68:61-84.

Hadfield, K.A., Dandekar, A.M., and Romani,R.J.(1993).Demethylation of ripening specificgenes intomato fruit. Plant Sci., 92:13–18.

Harris CJ, Scheibe M, Wongpalee SP, Liu W, Cornett EM, Vaughan RM, Li X, Chen W, Xue Y, Zhong Z, Yen L, Barshop WD, Rayatpisheh S, Gallego-Bartolome J, Groth M, Wang Z, Wohlschlegel JA, Du J, Rothbart SB, Butter F, Jacobsen SE. (2018). A DNA methylation reader complex that enhances gene transcription. Science. 362:1182-1186. doi: 10.1126/science.aar7854.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell, 157:1262–1278.

Irizarry RA, Ladd-Acosta C, Wen B, et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nature Genetics, 41: 178–86.

Kanehisa M, Araki M, Goto S, et al. (2008). KEGG for linking genomes to life and the environment. Nucleic Acids Res, 36 (Database issue): D480-4.(KEGG)

Lang Z, Lei M, Wang X, Tang K, Miki D, Zhang H, Mangrauthia SK, Liu W, Nie W, Ma G, Yan J, Duan CG, Hsu CC, Wang C, Tao WA, Gong Z, Zhu JK. (2015). The methyl-CpG-binding protein MBD7 facilitates active DNA demethylation to limit DNA hyper-methylation and transcriptional gene silencing. Mol Cell. 57:971-983. doi: 10.1016/j.molcel.2015.01.009.

Lang Z, Wang Y, Tang K, Tang D, Datsenka T, Cheng J, Zhang Y, Handa AK, Zhu JK. (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 U S A. 114:E4511-E4519:

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet, 11:204–220.

Lei M, Zhang H, Julian R, Tang K, Xie S, Zhu JK. Regulatory link between DNA methylation and active demethylation in Arabidopsis. (2015). Proc Natl Acad Sci U S A. 112:3553-7. doi: 10.1073/pnas.1502279112.

Liu R, How-Kit A, Stammitti L, Teyssier E, Rolin D, et al. (2015). A DEMETER-like DNA demethylase governs tomato fruit ripening. PNAS, 112:10804–9.

Matzke MA, Kanno T, Matzke AJM. (2015). RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. Annu. Rev. Plant Biol. 66:243–67.

Meissner A, Mikkelsen TS, Gu H, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature, 454: 766–70.

Meyer, R.S., and Purugganan, M.D. (2013). Evolution of crop species: Genetics of domestication and diversification. Nat. Rev. Genet., 14:840–852.

Mok YG, et al. (2010). Domain structure of the DEMETER 5-methylcytosine DNA glycosylase. PNAS, 107:19225–19230.

Nie WF, Lei M, Zhang M, Tang K, Huang H, Zhang C, Miki D, Liu P, Yang Y, Wang X, Zhang H, Lang Z, Liu N, Xu X, Yelagandula R, Zhang H, Wang Z, Chai X, Andreucci A, Yu JQ, Berger F, Lozano-Duran R, Zhu

JK. (2019). Histone acetylation recruits the SWR1 complex to regulate active DNA demethylation in Arabidopsis. Proc Natl Acad Sci U S A. 116:16641-16650. doi: 10.1073/pnas.1906023116.

Olsen, K.M., and Wendel, J.F. (2013). A bountiful harvest: Genomic insights into crop domestication phenotypes. Annu. Rev. Plant Biol, 64: 47–70.

Parida AP, Raghuvanshi U, Pareek A, Singh V, Kumar R, Sharma AK. (2018). Genome-wide analysis of genes encoding MBD domain-containing proteins from tomato suggest their role in fruit development and abiotic stress responses. Mol Biol Rep. 45:2653-2669. doi: 10.1007/s11033-018-4435-x.

Reik W, Dean W, Walter J. (2001) Epigenetic reprogramming in mammalian development. Science, 293:1089–93.

Rodriguez-Leal, D., Lemmon, H.Z., Man, J., Bartlett, M.E., Lippman, Z.B. (2017). Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. Cell, 171:1-11.

Wang, G.-D., Xie, H.-B., Peng, M.-S., Irwin, D., and Zhang, Y.-P. (2014). Domestication genomics: Evidence from animals. Annu. Rev. Anim. Biosci., 2:65–84.

Wittkopp, P.J., and Kalay, G. (2011). Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. Nat. Rev. Genet., 13:59–69.

Zhang M, Kimatu JN, Xu K, Liu B (2010). DNA cytosine methylation in plant development. J Genet Genomics, 37:1–12.

Zhang H, Lang Z, Zhu JK. (2018). Dynamics and function of DNA methylation in plants. Nat. Rev. Mol. Cell. Biol. 19:489-506. doi: 10.1038/s41580-018-0016-z

Zhang, B., Tieman, D. M., Jiao, C., Xu, Y., Chen, K., Fei, Z., Giovannoni J. J., Klee, H. J. (2016). Chilling-induced tomato flavour loss is associated with altered volatile synthesis and transient changes in DNA methylation. PNAS, 113:12580-12585.

Zhong S, Fei Z, Chen Y, Zheng Y, Huang M, et al. (2013). Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat. Biotechnol., 31:154–59.

Zhu J-K (2009) Active DNA demethylation mediated by DNA glycosylases. Annu Rev Genet, 43:143–166.