Generation and Characterisation of a Tomato Epigenetic Recombinant Inbred Line (epiRIL) Population

Paramita C. Kuswandi

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

Division of Plant and Crop Sciences
School of Biosciences
The University of Nottingham

2019

Abstract

Heritable changes in gene expression, which are not due to alterations in the DNA sequence, are termed epigenetic modifications and they play an important role in plant development (Pickaard and Scheid, 2014). These changes include those in DNA methylation, chromatin remodelling and levels of small RNAs. This project focuses on the effects of variations in DNA methylation on the cultivated tomato (*Solanum lycopersicum*).

In Arabidopsis, the effects of DNA methylation on plant development have been investigated using mutants in DNA methyltransferase genes. These mutants have also been used to develop epigenetic recombinant inbred line (epiRIL) populations where the genetic background of individuals in the population is identical, but they show variation in genome wide DNA methylation. The aim of this project was to generate and characterize a novel population of tomato epiRILs and identify the effects of the epigenetic variation on the development of tomato.

We have established an epiRIL population in a cherry tomato line (WVA 106) where *methyltransferase 1* (*MET1*) gene has been silenced. F_2 seeds were generated by colleagues at INRA, (France). Then the epiRIL population, generations F_3-F_8, were propagated in Nottingham. The population was screened for altered phenotypes. Morphological changes were observed including twisted and fasciated stems, abnormal leaves, exerted style, twisted stamens and low fruit weight. Targeted analysis of the expression of candidate genes underlying the potential phenotypes focused on genes including *FW2.2*, *SlEZ1* and *TM8.* RNASeq analysis indicated significant changes in gene expression across the genome of selected epiRILs including altered expression of transposons. Analysis using the enzyme McrBC showed that the level of methylation for target genes was lower in the epiRILs than wild type. Changes in DNA methylation in epiRIL lines was also demonstrated using bisulfite conversion and targeted Sanger sequencing and whole genome bisulfite sequencing. The epiRIL population provides a unique resource for investigating the epigenetic basis of trait variation in tomato which could also be applied in other crop species.

Acknowledgements

I would like to express my sincere gratitude to Professor Graham Seymour for his guidance, motivation and patience during my study. I would also like to thank Professor Philippe Galusci from Bordeaux University and colleagues at INRA who have provided the initial generation of the tomato epiRILs. To Dr Michael Wilson, I would like to thank for his analysis on the RNA sequencing data and to Dr. Nicolas Bouché for the WGBS analysis.

I would also like to thank Professor Andrew Thompson and Dr. John Foulkes as examiners. I would like to thank Dr. Fereshteh Malekpoor Mansoorkhani, Dr. Hassan Moeiniyan Bagheri, Dr. Fabio Fedi and Guiping Sun for their kind help and technical guidance. To former Phd students in the Seymour group, Dr. Nurul Hidayah Samsulrizal and Dr. Duoduo Wang, I thank you both for your help and motivation. AIso I would like to say a big thank you to Dr. Fahmi Yunus and Parisatcha Sangsuwan, without you both this study would almost certainly take longer to finish.

I would like to thank all the staff from Plant and Crop Sciences Division who have helped me in many ways. Special thank you for Mark Meacham, Sue Flint, Catherine Tomlinson, Tim, Jools Marquez, Zoe Phillips, Marlene Oldman and Nicola Leftley for your invaluable work.

I would like to thank the Indonesia Endowment Fund for Education (LPDP) for the scholarship and research funding used in this study. I would also like to thank the Indonesian Ministry of Research, Technology and Higher education and Yogyakarta State University for giving me the time for further study.

Special thank you for my parents and brothers for their prayers and support during my study. Last but not least, thank you to my husband Khusnul Mubarok and to my children Jasmine, Lilly and Kynaan for their encouragement, prayers, patience and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANOVA	Analysis of Variance
ATP	Adenosine triphosphate
B + 7	Breaker + 7 days
BC	Backcross
BC1-S1	Backcross 1-Selfing 1
β Gal	β-galactosidase
BGS	Bisulfite Genome Sequencing
cDNA	Complementary Deoxyribonucleic Acid
CMT3	Chromomethylase 3
Cnr	Colourless non ripening
COBRA	Combined Bisulfite Restriction Analysis
DDM1	Deficient in DNA Methylation 1
DH	Double Haploid
DMR	Differentially Methylated Region
DNA	Deoxyribose Nucleic Acid
DPA	Days Post Anthesis
DRM2	Domains Rerarranged Methylase 2
EB	Elution Buffer
EZ	Enhancer of Zeste
EMS	Ethylmethanesulfonate
FW	Fruit Weight
epiQTL	Epigenetic Quantitative Trait Loci
epiRIL	Epigenetic Recombinant Inbred Line
FAOSTAT	Food and Agriculture Organization Statistics
H3K9me2	Histone H3 Lysine 9 Dimethylation
INRA	Institut National de la Recherche Agronomique
ITAG	International Tomato Annotation Group
MET1	Methyltransferase 1
miRNA	Micro Ribonucleic Acid
mRNA	Messenger Ribonucleic Acid
MSP	Methylated Specific Polymerase Chain reaction
NGS	Next Generation Sequencing
NIL	Near Isogenic Line
nor	Non ripening
Nr	Never-ripe
PCR	Polymerase Chain Reaction
PTGS	Post Transcriptional Gene Silencing
QPCR	Quantitative Polymerase Chain Reaction
QTL	Quantitative Trait Loci

RdDM	RNA Directed DNA Methylation
RIL	Recombinant Inbred Line
RNA	Ribonucleic Acid
RNAi	RNA interference
RNAseq	RNA Sequencing
RISC	RNA-Induced Silencing Complex
SUVH4/KYP	Su(var)3-9 homolog 4/Kryptonite
SMRT	Single Molecule Real Time Sequencing
sRNA	Small Ribonucleic Acid
TSS	Transcription Start Site
TEs	Transposable Elements
UV	Ultraviolet
WGBS	Whole Genome Bisulfite Sequencing
WT	Wild Type

CHAPTER 1 INTRODUCTION

1.1 Tomato

1.1.1 Tomato Classification

The tomato (*Solanum lycopersicum*) is a member of the Solanaceae family. This family comprises 3000–4000 species in approximately 90 genera and includes perennial trees as well as herbaceous annual species. Plants of this family occupy a wide range of terrestrial habitats from deserts to rainforests. The Solanaceae contains a number of species that are important to agriculture. These include tomato, potato (*Solanum tuberosum*), pepper (*Capsicum* spp.) and aubergine (*Solanum melongena*), which are major crops worldwide. Other uses of plants from this family include members of the genus *Petunia* which are used as ornamentals and tobacco (*Nicotiana tabacum*) from which nicotine is extracted (Costa and Heuvelink, 2004; Knapp *et al.* 2004; Gebhardt, 2016).

Members of the Solanaceae can contain a range of alkaloids and these may have medicinal properties. For example, the active agents in *Atropa belladonna* (deadly nightshade) are atropine, hyoscine (scopolamine) and hyosymine that have anticholinergic properties and are used for treating headaches, menstrual pain and peptic ulcers. *Solanum nigrum* (black nightshade) is another example of a Solanaceae plant that has medicinal properties. It is used as an antiseptic, anti-inflammatory and antidysentry medicine (Shah *et al*., 2013).

The classification of tomato was changed recently with the integration of the *Lycopersicon* into the *Solanum* genus. From the genus *Solanum*, the domesticated tomato species was named *Solanum lycopersicum* (Bai and Lindhout, 2007) and this is the name that is now recognised and used in the 21st century. The change in the classification of tomato started in the 18th century where Phillip Miller, an English botanist, used the term *Lycopersicon* (meaning wolf peach) for all variants of the cultivated tomato, along with other multilocular fruits described as 'roundish, soft fleshy fruit divided into several cells and contain many flat seeds'. When the system of giving plants a genus and species name began in the first edition of Species Plantarum in 1753, Carolus Linnaeus classified tomatoes in the genus *Solanum*. However, Miller continued to use the generic name *Lycopersicon* and named the tomato *Lycopersicon esculentum* Mill. (*esculentum* means "edible"). At the time, many people thought that the tomato fruit was poisonous so Miller may have wanted to emphasize that it was edible. He included potato in this genus and stated that '*Lycopersicon* has been established as a distinct Genus, on account of the fruit being divided into several cells, by intermediate partitions'.

The most recent classification of tomato, which uses molecular data recognizes the genus *Solanum* sect. Lycopersicon. It consists of the cultivated tomato and its 12 wild relatives *Solanum arcanum*, *S. cheesmaniae*, *S. chilense*, *S. chmielewskii*, *S. corneliomulleri*, *S. galapagense*, *S. habrochaites*, *S. huaylasense*, *S. neorickii*, *S. pennellii*, *S. peruvianum* and *S. pimpinellifolium* (Figure 1.1). Tomato-like

morphological characters which differentiate them from most of other *Solanum* species include yellow corollas, pedicels articulated above the base, pinnately segmented non-prickly leaves and lack of tubers. There are also morphological differences between the domesticated tomato and its wild relatives (Figure 1.2) such as leaf shape and fruit size (Bergougnoux, 2014; Knapp and Peralta, 2016).

Genus: Solanum	Lycopersicon synonyms
S. lycopersicum L.	*L. esculentum* Miller
S. cheesmaniae (Riley) Fosberg	*L. cheesmaniae* Riley
S. galapagense, S. Darwin and Peralta	*L. cheesmaniae, L. Riley* (forma or var. minor)
S. pimpinellifolium B. Juss.	*L. pimpinellifolium* (B. Juss.) Miller
S. chmielewski[a]	*L. chmielewski* C. M. Rick, Kesicki, Foboes and M. Holle
S. habrochaites S. Knapp and D. M. Spooner	*L. hirsutum*
S. neorickii[a]	*L. parviflorum* C. M. Rick, Kesicki, Foboes and M. Holle
S. pennellii Corell	*L. pennellii* (Correll) D'Arcy
S. chilense (Dunal) Reiche	*L. chilense* Dunal
S. corneliomuelleri J. F. Machr.	*L. peruvianum* (L.) Miller*L. glandulosum* C. F. Mull.
S. huaylasense Peralta	*L. peruvianum* (L.) Miller
S. peruvianum L.	*L. peruvianum* (L.) Miller
S. arcanum Peralta	*L. peruvianum* (L.) Miller

Figure 1. 1 Classification of the tomato species.The new classification in the genus Solanum and its Lycopersicon synonyms (Schwarz *et al*., 2014)

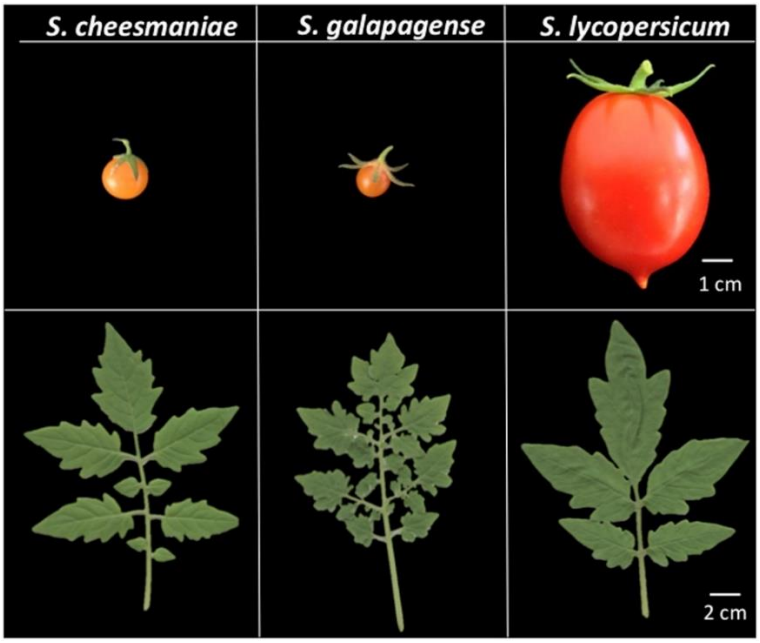


Figure 1. 2 Morphological differences between three species of Solanum. Fruits of *Solanum cheesmaniae*,*Solanum galapagense* and *Solanum lycopersicum* fruit (upper panel) and leaves (lower panel) (Pailles *et al*., 2017)

1.1.2 Tomato Domestication and Economic Importance

The tomato originated from South America but was introduced in to Europe in the 16th century. It was distributed from Ecuador then to Peru, Chile and the Galapagos Islands. The wide range of distribution showed that early tomato species grew in a variety of habitats ranging from sea level to the highlands and from arid to rainy climates. However, the exact origin of the cultivated tomato and the events related to its early domestication are still debated. Two hypotheses which have arisen involve a Peruvian or Mexican origin. The cultivated tomato was thought to have been the result of domestication from the cherry tomato, which originated in the coastal part of Peru and that the domestication occurred before the Europeans discovered America. The Mexican origin is based on the name 'tomatl' which is the Mexican word used to refer to tomato. There was also more evidence of tomato cultivation in Mexico compared to South America. Tomato is widely adapted to a variety of habitats and domestication likely occurred independently in several regions (Peralta *et al*., 2005; Bergougnoux, 2014).

The wild cherry tomato *Solanum lycopersicum* var *cerasiforme* was thought to be the most likely candidate as the ancestor of present domesticated tomato due to its wide distribution in Central America. However, recent molecular studies have shown 'cerasiforme' is not the direct ancestor of the cultivated tomato, but is a mixture of wild species *S.pimpinellifolium* and *S.lycopersicum* (Bergougnoux, 2014).

Tomato domestication had already been carried out by people in its native region and it was accepted as a cultivated crop when it arrived

in Europe (Costa and Heuvelink, 2004). The earliest record of a cultivated tomato appeared in Italy in 1544 describing a poisonous distant relative of the tomato, the mandrake (*Mandragora officinarum*); also from the Solanaceae family. The fruit was described as segmented and yellow. In most of Europe, tomato was considered poisonous and was used as an ornamental. It was not until the late 17th or early 18th century that tomato was used as vegetable. However, the real domestication of tomato in Europe was not until the 19th century when it was finally accepted as a nutritious food crop (Paran and Knaap, 2007; Bergougnoux, 2014; Knapp and Peralta, 2016).

At the end of the 19th century many cultivars of tomato were already available. These cultivars can be considered as landraces and products of early breeding. They were open pollinated by farmers and growers and because tomatoes did not out-cross naturally, seeds of tomatoes produced plants resembling the parents. However, private companies carried out commercial breeding by making hybrids. Crosses from selected parents with good characters were used to produce hybrid seeds (Bai and Lindhout, 2007).

Important traits selected in breeding programmes today are still similar to early domestication of tomatoes. Fruit appearance and quality, plant architecture, disease resistance and ease of mechanical harvest, have all contributed to the development of tomato breeding. As a result, many varieties of tomato can be found with a range of colours, shapes, sizes and texture. Plant stature also changed from shrub to a more erect and thick-stemmed plant (Paran and Knaap, 2007).

In the domestication of tomato there has been a focus on yield traits and this has had a detrimental effect on other characteristics. An example is the loss of flavour in modern commercial varieties. Whole-genome sequencing and genome-wide association studies have identified loci which have the potential to affect flavour chemicals. These discoveries are likely to provide ways to recover favourable traits through breeding that have been lost during domestication (Tieman *et al*., 2017).

In breeding for desirable traits, wild tomato species are being used as a source of traits that are beneficial for human consumption or agronomic practice. *Solanum pimpinellifolium* has been used as a source to improve fruit colour and quality. It is autogamous and is closely related to the domesticated tomato *S. lycopersicum*. Other wild relatives, *Solanum habrochaites* (syn*. L. hirsutum*) and *Solanum pennellii* have also been used in tomato breeding due to their tolerance to biotic and abiotic stress. Each of these wild relatives have a different fruit morphology to the domesticated tomato, including fruits that remain green even when ripe (Figure 1.3) (Top *et al*., 2014; Bolger *et al*., 2014).

Figure 1. 3 Fruit morphology in domesticated tomato and its three wild relatives. From left to right: *Solanum lycopersicum*, *S. pimpinellifolium*, *S.habrochaites* and *S.pennellii* (University of California, 2014)

Today, tomato breeding objectives are to produce and distribute new tomato cultivars with improved agronomical traits, depending on the market (fresh or processed market). Even though breeding objectives have changed over time, three main objectives remain the same: (1) adaptability to the environment, (2) resistance to pests and diseases, and (3) high fruit yield and good fruit quality (Bergougnoux, 2014).

Tomatoes are now grown world-wide, with high levels of production and consumption in China, North America and Europe. In 2012, global production of tomatoes reached more than 161 million metric ton with a value of over $59 billion (FAOSTAT, 2012). The tomato provides high levels of minerals, vitamins and phytochemicals to the human diet due to the amount eaten and is probably the most important fruit in the human diet in this respect.

1.1.3 Tomato Biology

The morphological characteristics that differentiate tomato from other *Solanum sp* include sympodial shoots, compound leaves, bright yellow flowers and fleshy fruits with many flat white seeds (Figure 1.4). The sympodial shoots grow from a primary shoot that after the production of several leaves, is terminated by the initiation of the first inflorescence. The growth of the bud at the axil of the last leaf will take the place of the previous primary shoot. This bud continues to grow until it is above the inflorescence when it then produces some leaves. A second inflorescence will also grow and will again be replaced by the growth of the next axillary bud. Depending on the growth of the main

shoot, the tomato growth habit can be classified as indeterminate or determinate. Indeterminate growth involves continuous growth following the main shoot growth. In the varieties with a determinate structure, growth of the leaves stops, but an inflorescence can form after the last leaf on the primary shoot (Elkind *et al*., 1991; Quinet and Kinet 2007).

As described above, in tomato the vegetative and reproductive phases alternate along the compound (sympodial) shoots. The number of leaves before the first inflorescence depends on genetic background and environmental cues. The transition time from initial vegetative stage to the first inflorescence can be measured by counting the number of leaves or nodes formed on the initial apical meristem before the first inflorescence (Samach and Lotan, 2007).

Environmental factors such as daily light energy strongly affects flowering time in tomato, along with the expression of several genes such as *SINGLE FLOWER TRUSS* (*SFT*) and *SELF PRUNING* (*SP*). The *SFT* gene encodes a protein that promotes flowering and is active in initial sympodial segments while *SP* regulates the vegetative-reproductive switch of the sympodial segments (Quinet and Kinet 2007).

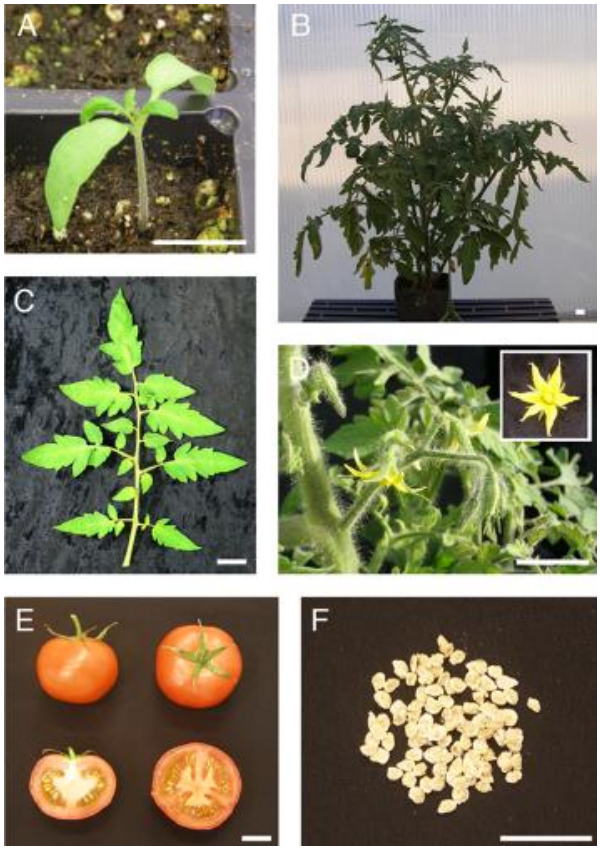


Figure 1. 4 Tomato plants and parts.(A) seedling; (B) 40-day-old plant; (C) leaf; (D) flowers; (E) fruit; (F) seeds. Scale Bar = 2 cm (Kimura and Sinha, 2008)

The tomato flower consists of four whorls as described in the ABC model of flower development (Bowman *et al*., 2012; Irish, 2017). In the outer whorl is the calyx, then followed by the corolla, androecium and in the centre is the gynoecium (Figure 1.5). The calyx is formed from individual sepals. The corolla consists of petals which in tomatoes are bright yellow and function to attract pollinators. The androecium or the male reproductive organs are the stamens which produce pollen. In tomato, the stamens are normally fused to form the stamen cone. The female organ found inside the stamen cone is the carpel. The carpel is made up of the ovary, stigma and ovule (Pesaresi *et al*., 2014; http://www-plb.ucdavis.edu, 2018; Dreni and Zhang, 2016).

The tomato fruit is composed of a thick pericarp and placental tissue surrounding the seeds. There are four stages in the fruit development of tomato (Figure 1.5).The first stage is the flower development which lasts 2-3 weeks. The fertilization process then initiates fruit development. The second stage is approximately 2 weeks of intensive mitotic division leading to an increase in pericarp cell number. The third stage is the period of rapid cell expansion and there is a significant increase in fruit weight. In the last phase, the fruit attains its final size at the mature green stage. This is then followed by a ripening or maturation phase (Bergougnoux, 2014; Pesaresi *et al.,* 2014).

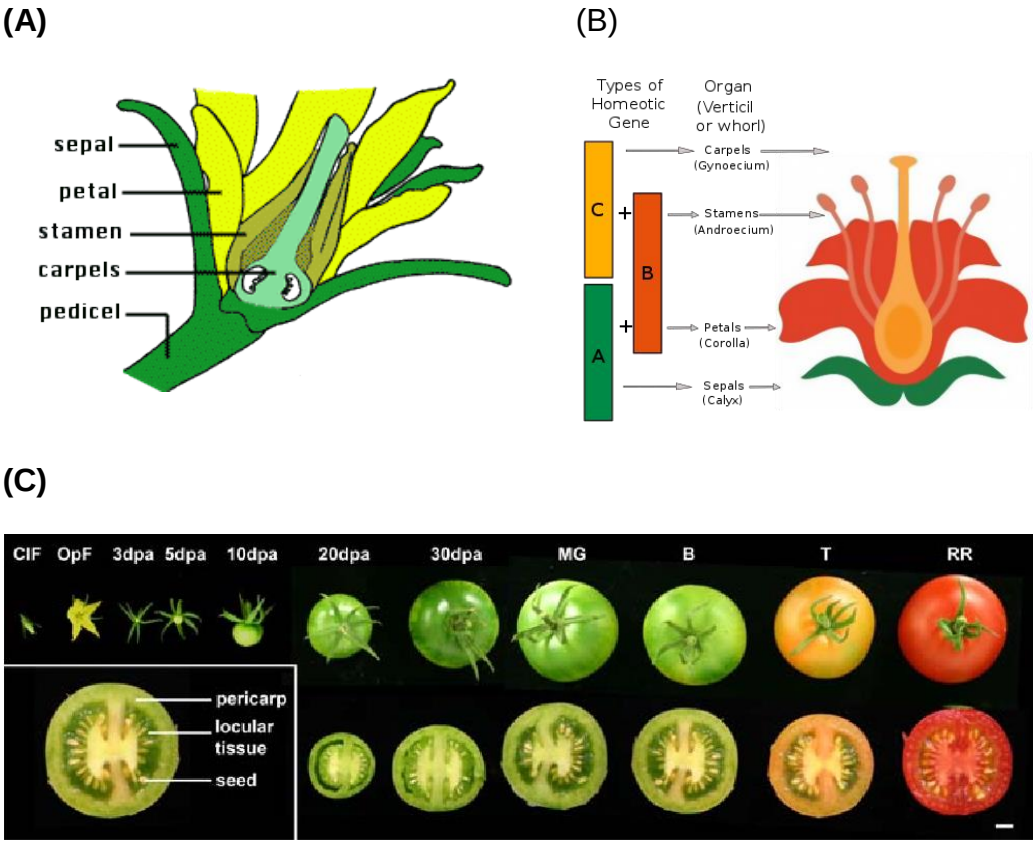


Figure 1. 5 Diagram of tomato flower and fruit development.Longitudinal section of a tomato flower is shown in (A) and the ABC model of flower development (B). In (C) the development of tomato fruit is shown from closed flower (CLF) stage until red ripe (RR). Early stages are mature green (MG), breaker (B), turning (T). dpa = days post anthesis and scale bar = 1cm (Pesaresi *et al*., 2014; http://www-plb.ucdavis.edu, 2018; Teyssier *et al*., 2008)

Domestication of tomato resulted in larger fruit size which is more beneficial for human diets. In addition, harvesting a given mass of fruit would require less time if the size of the fruits were larger. Smaller fruit size is more beneficial in nature because the plant relies on small animals for seed dispersal so smaller size is easier for the fruits to be carried and dispersed. Fruit size in tomato is also correlated to fruit shape, where larger fruits show more extreme shapes than the smaller size fruits (van der Knapp and Tanksley, 2003; Tanksley, 2004).

The increase tomato fruit size due to breeding has caused a change in other traits that could be beneficial to human consumption, such as fruit sugar content. Increase in tomato fruit size has been associated with low sugar content (Prudent *et al*., 2009; Kanayama, 2017). Fruit weight and fruit composition such as sugar and acid contents depend on the inward and outward fluxes into and from fruit. These fluxes which comprise mainly of carbon and water involve many processes such as transpiration, cell division and cell expansion. Leaf size and shape can affect sugar production via photosynthesis and changes in the environment and agricultural practices could also affect carbon supply. The number of cells determine the fruit final size by affecting the amount of structural dry matter and cell size affects the capacity to store soluble dry matter (Prudent *et al*., 2009).

Other than environmental effects that can affect the fluxes of carbon and water, genetic mapping studies showed that there are more than 10 loci affecting tomato size and shape. Four loci (*fw1.1*, *fw2.2*, *fw3.1* and *fw4.1*) were identified as QTLs in crosses between small and

large fruit varieties. Variation in these loci has a larger impact on fruit mass than fruit shape, except for *fw3.1* (Tanksley, 2004). The *fw2.2* has been cloned and studied at the molecular level and results showed that it does not affect fertility and sink source relationship but acts is a negative regulator of cell division in early fruit development (Cong *et al*., 2002).

1.1.4 Tomato as a model plant

The model species for plant science research has been *Arabidopsis thaliana* due to its small genome and wide range of genetic resources (Koornneef and Meinke, 2009). However, Arabidopsis produces fruits which are dry and inedible. It is therefore not a useful model for understanding fleshy fruit development. Tomato has become an especially tractable system for molecular genetic analysis of fleshy fruit development and ripening due to decades of work on the biochemistry and molecular biology of the fruit, simple diploid genetics and an excellent genome sequence (The Tomato Genome Consortium, 2012). Tomato can also be used as a model for studying other aspects of plant development and for research on the ripening of other fleshy fruit species.

As a model plant, tomato was used to investigate the control of flower architecture involving the *tomato terminating flower* (*tmf*) mutant. This *tmf* mutant flowers early and converts the normal multi-flowered inflorescence of tomato into a solitary flower. The *TMF* gene encodes a member of the ALOG (Arabidopsis LIGHT-SENSITIVE HYPOCOTYL 1, Oryza G1) family of proteins. The absence of this protein in the *tmf*

mutant results in the expression of transcription factors that induce early flowering altering the normal pattern of vegetative to reproductive transition and inflorescence formation (MacAlister *et al*., 2012).

Tomato has also been used to investigate adaptation to abiotic stress by investigating the response of wild tomato species germplasm to low temperature stress (Nosenko *et al*., 2016) and in studying plant leaf shape (Shani *et al*., 2009). However, tomato has been most useful in understanding the biological basis of ripening. Indeed studies on tomato fruit ripening have led to the discovery that this developmental process is under epigenetic control.

Ripening in fleshy fruits involves substantial metabolic changes that can include alterations in colour, carbohydrate metabolism and texture (Seymour *et al.*, 2013). Much work using tomato as a model system has demonstrated that ripening is under tight genetic and epigenetic control (Grierson, 2016).

In tomato, the initiation of the ripening process is characterized by increased respiration and a burst of ethylene production. The change in respiration is known as the climacteric rise. Fruits that show increased respiration and ethylene production are known as climacteric and these include tomato, apple, avocado and banana. In these climacteric fruits ripening is initiated and coordinated by the phytohormone ethylene and a range of developmental cues (Grierson, 2013).

Mutants have been used to study the regulation of ripening in tomato. These include the *Neverripe* (*nr*) which abolishes normal ripening and the mutation is a lesion in a gene encoding an ethylene

receptor (Wilkinson *et al*., 1995). Other mutants that result in altered ripening phenotypes include *ripening inhibitor* (*rin*), and *non-ripening* (*nor*). The *rin* mutation is a lesion in a gene encoding a MADS-box transcription factor and in *nor* the function of a gene encoding a transcription-factor related to NAC-domain proteins is compromised (Vrebalov *et al*., 2002; Giovannoni, 2004). These genes act up-stream of the ethylene signalling and perception pathway and their expression is dependent on developmental cues.

Another mutant showing altered ripening is the *Colorless non-ripening* (*Cnr*) (Figure 1.6). *Cnr* fruits fail to ripen normally. Unlike the other non-ripening tomato mutants, *Cnr* is the result, not of a DNA sequence change, but of an epigenetic change resulting in hypomethylation of the *CNR* gene promoter. The *CNR* gene encodes an SBP-box transcription factor that is likely to influence the expression of *RIN* and other ripening regulatory genes (Manning *et al*., 2006). The discovery of an epigenetic mutation governing ripening indicated that epigenetic processes may play a more important role in tomato development than previously thought.

In this project we are focusing on tomato as a model crop to understand epigenetic variation and its impact on phenotypes of agronomic interest.

(A)

(B)

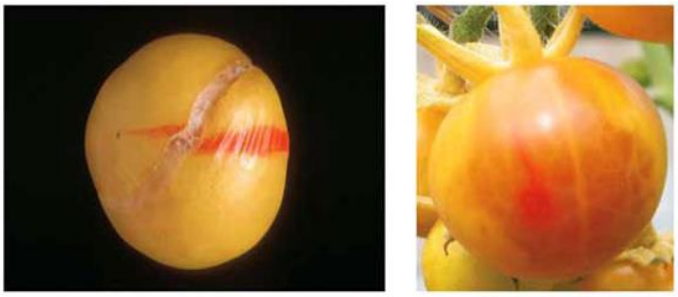


Figure 1. 6 Fruits of the Colourless non-ripening tomato.(A) *Cnr* fruit and its cross section and (B) revertant 'ripening' sectors on *Cnr* fruits showing wild-type ripening phenotype (Manning *et al*., 2006)

1.2 Epigenetics

1.2.1 Genetics and epigenetics

Genetics is the study of genes, how they are inherited and their impact on phenotypes. DNA sequence variation between organisms of the same species and between species can lead to differences in gene function or expression and hence phenotypic variation. The term epigenetics, was originally conceived by C.H.Waddington in 1956, to describe the existence of mechanisms of inheritance in addition to (over and above) standard genetics (Nobel, 2015). Today, epigenetics is widely used to describe modifications that cause variation between organisms of the same species or different species that are stably inherited without changes in the DNA sequence (Weigel and Colot, 2012) (Figure 1.7). For instance, histone modifications, change in chromatin structure, and the methylation of cytosine (Lauria and Rossi, 2011) can change the gene expression, producing differences in phenotype without changes in the nucleotide sequence (Figure 1.8).

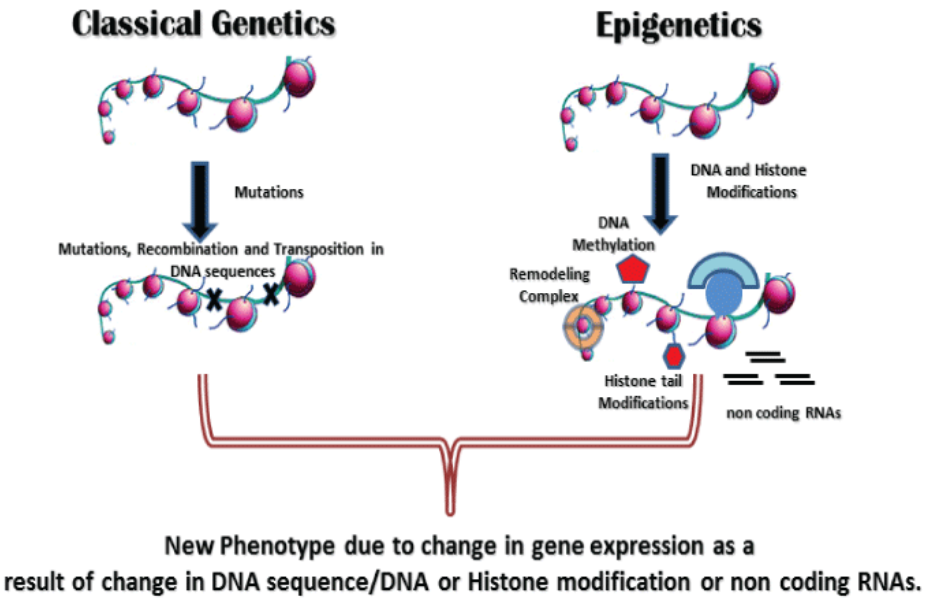


Figure 1. 7 Diagram representation of the difference between genetics and epigenetics (Yadav *et al.* 2015).

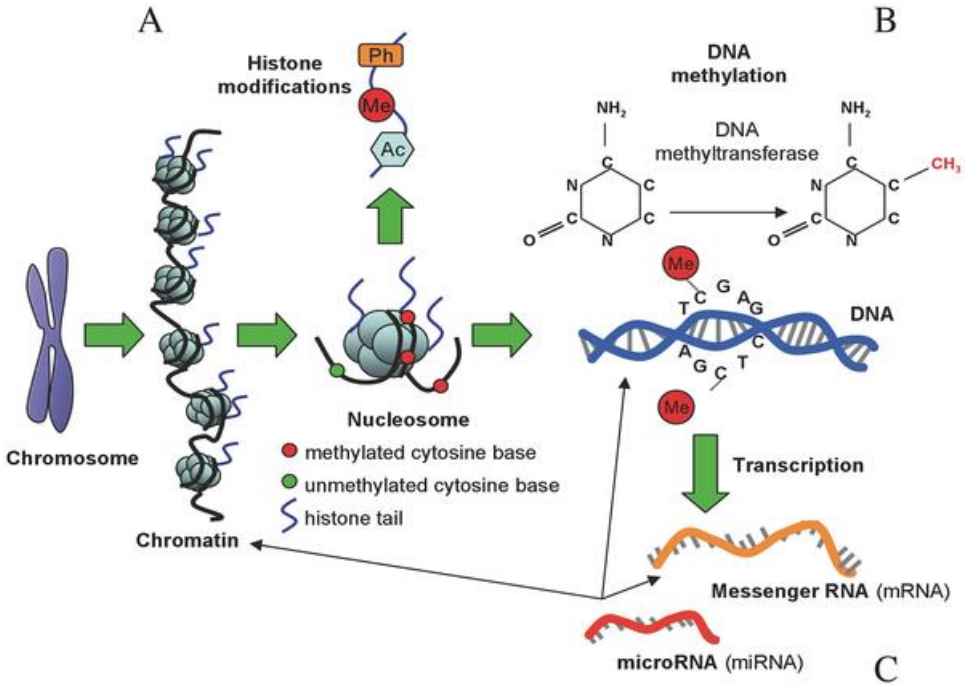


Figure 1. 8 Epigenetic modifications. (A) Histones can undergo phosphorylation (Ph), methylation (Me), and acetylation (Ac). (B) DNA molecules are methylated by the addition of a methyl group to carbon position 5 on cytosine bases, a reaction catalyzed by DNA methyltransferase enzymes. (C) mRNA is translated into a protein product, but this process can be repressed by microRNAs (miRNA) (Relton and Smith 2010, Gomez-Diaz *et al*., 2012).

1.2.2 Types of Epigenetic Modifications

Histone Modification

Chromatin is formed from nucleosomes in which DNA is wrapped around an octamer of histone proteins. This structure acts as a barrier to transcription, replication and repair of the DNA. It also acts as a regulator for the production of proteins and enzymes necessary for the appropriate cellular pathways. Histones can be modified by several processes such as acetylation, methylation, phosphorylation and ubiquitinylation (Margueron and Reinberg, 2013).

Histone acetylation results in transcriptionally active DNA sequences. Acetylation is carried out by the histone acetyltransferase enzyme (HAT), while deacetylation is caused by histone deacetylase (HDAC) enzyme (Figure 1.9). This allows changes in epigenetic marks as acetylation can be removed by HDAC. Methylation can occur in histones due to the histone lysine methyltransferases (HKMT). Depending on the histone that is methylated this mechanism can also change gene expression by promoting or inhibiting transcription. Other changes to histones such as phosphorylation are involved in DNA repair, chromosome segregation and cell division. Epigenetic marks in histones can influence the phosphorylation of adjacent histones. While histone ubiquitination is involved in the regulatory functions and can influence cell cycle, development and pathogen resistance (Pikaard and and Scheid, 2014).

The roles of histone modification and DNA methylation are interlinked in epigenetic regulation. Histone H3 lysine 9 dimethylation

(H3K9me2) plays a role in histone modification of transcriptional gene silencing. Investigations in Arabidopsis revealed that H3K9me2 is catalysed by the histone methyltransferase SUVH4/KYP. Further results revealed that histone methylation caused the recruitment of the methyltransferase CMT3 to methylate DNA loci (He *et al*., 2011).

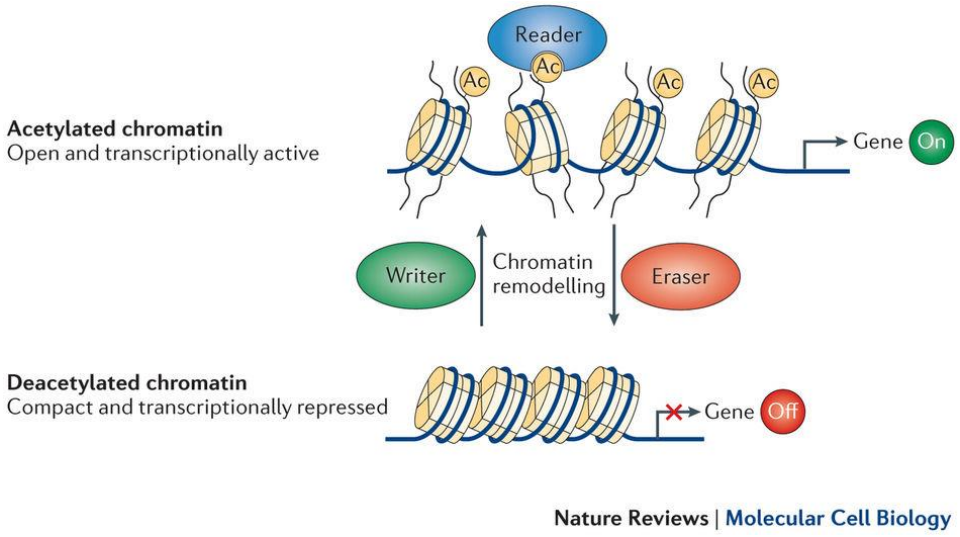


Figure 1. 9 Diagram of chromatin remodelling.The process involves histone acetylation, chromatin condensation and altered gene expression. Acetylation targets Lys residues in the amino-terminal tails of core histone proteins. A string of nucleosomes is shown with the tails protruding when acetylated. Acetylation of the tail domains inhibits the folding of nucleosome arrays into secondary and tertiary chromatin structures. This results in chromatin decondensation, allowing access to transcription factors and other transcription co-activators. Reader, writer and eraser refers to the enzymes involved in the process and Ac is acetylation (Verdin and Ott, 2015)

RNA interference

Another type of epigenetic modification is carried out by small RNAs. In general, the regulation of gene expression by small RNA is also called RNA interference (RNAi) or Post-Transcriptional Gene Silencing (PTGS). The small RNAs target complementary mRNAs for degradation or translational repression in the cytoplasm (Matzke and Mosher, 2014).

The small RNAs are non-coding RNA, approximately 22 nucleotides long and act as a guide for post-transcriptional control of protein coding genes (Mocellin and Provenzo, 2004; Mallory *et al*., 2004). MicroRNAs (miRNAs) along with small interfering RNA (siRNA) are classes of non-coding RNA. They are RNAs that do not encode a protein but play an important role in the development of an organism. The siRNAs and miRNAs are both short duplex RNA the difference is that siRNAs are transcribed from cellular genes or infecting pathogens, or artificially introduced into the cells and miRNAs are transcribed from micro-RNA (MIR) genes (Lam *et al.,* 2015).

The process of gene silencing by PTGS occurs in the cytoplasm and is caused by a short duplex RNA that can be double stranded or hairpin shaped. In the case of miRNAs, they are then processed by the RNase II enzymes Dicer or Dicer-like (DCL) protein into small RNA molecules with 20-24 nucleotides. The Dicer protein along with Argonaute (AGO) are the two core proteins involved in the process. The resulting small RNA (sRNA) molecules are formed into an RNA-induced silencing complex (RISC). The sRNA molecule guides the RISC to a

complementary single stranded RNA and the Argonaute cleaves the RNA. If there is a mismatch between the small RNA molecule with the mRNA, the mRNA is not cleaved but gene silencing still occurs (Figure 1.10) (Chuang and Jones, 2007; Matzke and Mosher, 2014; Guo *et al*., 2016).

In the nucleus, small RNAs regulate gene expression by directing epigenetic modifications such as methylation of cytosines and histone methylation (Figure 1.11). The mechanism where RNA is involved in the methylation of DNA is known as RdDM (RNA-directed DNA Methylation). The proteins DICER-like3 (DCL3), Argonaute4 (AGO4), the DNA dependent RNA polymerases (Pol IV and Pol V) and RNA-dependent polymerase RDR2 are involved in RdDM. In the process, the Domain Rearranged Methyltransferase 2 (DRM2) protein is directed to add methyl groups to cytosines within the target region (Lopez and Wilkinson, 2015; Baulcombe, 2015)

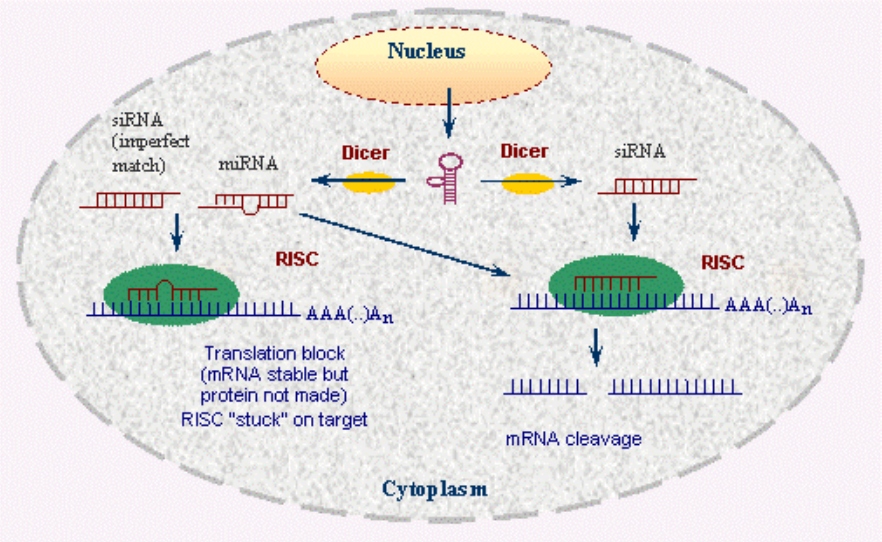


Figure 1. 10 A simplified model of the RNAi pathway

(https://www.ncbi.nlm.nih.gov/probe/docs/techrnai/, 2018)

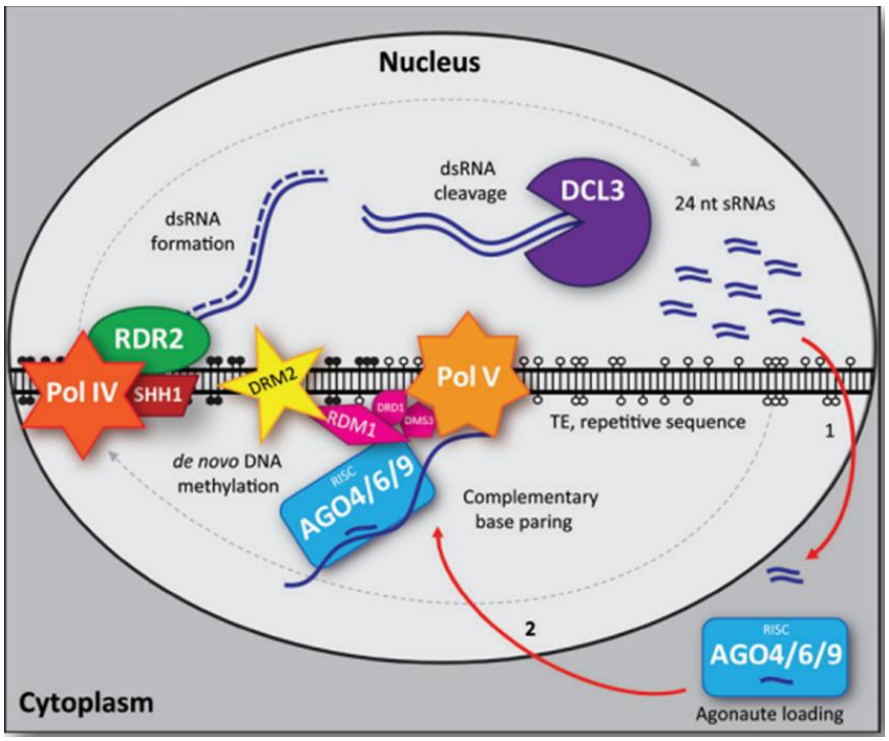


Figure 1. 11 Nuclear RNA silencing.Polymerase IV (Pol IV) transcribes a single-stranded RNA that is converted into a double-stranded form (dsRNA) by RNA-dependent RNA polymerase (RDR2). Dicer (DCL3) generates 24-nt siRNAs from this double-stranded RNA. The siRNAs are then bound by Argonaute (AGO) proteins and they base-pair to a transcript produced by polymerase V (Pol V). The AGO complex then recruits DNA methyltransferase that methylates the adjacent DNA. SHH1, RDM1, DRD1 and DMS3 are accessory factors causing the process to operate as a positive-feedback system. RISC (RNA-induced silencing complex); sRNA (short RNA); TE (transposable element) (Baulcombe, 2015).

DNA Methylation

Methylation at the 5' position of cytosine forms a part of epigenetic regulation along with histone modification and non-coding RNA. In DNA methylation, cytosine is changed to 5-methylcytosine (5mC). The methylation of DNA occurs after DNA synthesis and is catalysed by the methyltransferase enzymes. These enzymes transfer a methyl group from S-adenosyl-L-methionine (S-AdoMet) to carbon 5 of cytosine (Figure 1.12). There is strong evidence that this modified cytosine inhibits gene expression and cell differentiation (Lauria and Rossi, 2011; Moore *et al*., 2013; Pikaard and Scheid, 2015). Differences in DNA methylation are transmitted across mitosis and meiosis in mammals and plants and their pattern of inheritance can be similar to 'conventional' mutations caused by change in DNA sequence (Stokes *et al*., 2002; Saze *et al*., 2003). In Arabidopsis, depletion of the MET1 enzyme (the enzyme responsible for CG methylation in plants) during gametogenesis caused a loss of DNA methylation and resulted in a variation of DNA methylation patterns in the gametes.This was likely caused by random chromosome segregation during postmeiotic cell divisions. The new variation of methylation was then propagated through successive rounds of DNA replication (Saze *et al*., 2003).

In mammals, most of DNA methylation occurs on cytosines preceding a guanine nucleotide or CpG sites. DNA methylation plays an important role in silencing retroviral elements, regulating gene expression, genomic imprinting and X chromosome activation. In

different genomic regions DNA methylation have different influence in gene activities (Moore *et al*., 2012).

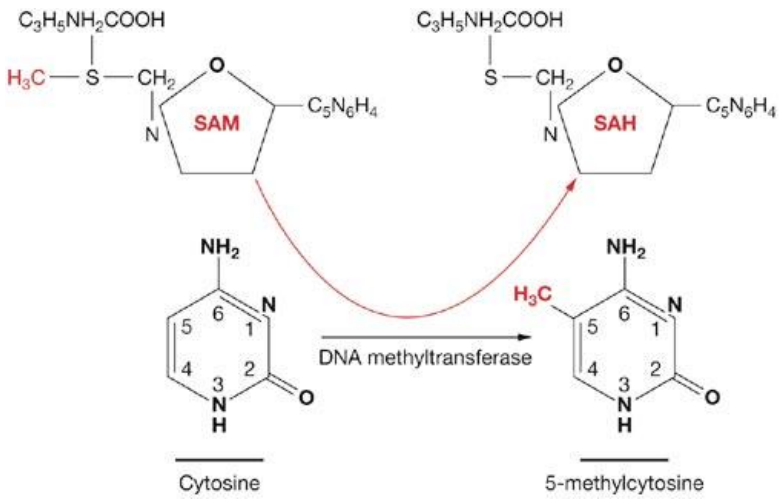


Figure 1. 12 Diagram of cytosine methylation.

(Richardson, 2007)

DNA methylation in plants occurs not only in CG sites, but in various other sequence contexts, CG, CHG and CHH (H = A, C, or T). Using a methylcytosine immunoprecipitation method, it was found that in Arabidopsis, methylation was found in all of the various C contexts in the genome. A comprehensive DNA methylation map of the Arabidopsis genome showed that repetitive sequences are heavily methylated and that over 1/3 of expressed genes are methylated in its transcribed region and 5% of genes are methylated in their promoter regions (Zhang *et al*., 2006). Methylation in promoter regions represses the binding of transacting factors and gene expression (Laird, 2010). The methyl groups on methylated cytosines lie in the major groove of DNA and interfere directly with protein binding required for transcription (Alberts *et al*., 2015).

The role of DNA methylation in the regulation of gene expression is also found to be correlated with the regulation of transposable elements. Repetitive elements and centromeric and pericentromeric regions are usually methylated (He *et al*., 2011). Studies in *Arabidopsis thaliana* using a *DECREASE IN DNA METHYLATION (DDM1)* mutant showed that hypomethylation not only reactivates silent repeat sequences but also causes the activation of transposons. The CAC1, an endogenous CACTA family transposon, transposes and increases in copy number at high frequency in the Arabidopsis mutant *ddm1* hypomethylated background (Miura *et al*., 2001). The mobility of transposons has been examined in Arabidopsis using other hypomethylated mutants. A higher frequency of transposon movement

(the CACTA elements) was found in the *cmt3-met1* double mutants. This indicates that both CG and non-CG methylation are needed for the immobilization of transposons (Kato *et al*., 2003).

The Regulation of DNA Methylation in Plants

Cytosine methylation in the plant genome is modulated by the coordination of several gene products and these include DNA methyltransferases, DNA demethylases, histone-modifying or remodelling enzymes and RNA interference (RNAi) components (Teixeira *et al*., 2009). In plants, there are three DNA methyltransferases that have been identified. These are METHYLTRANSFERASE 1 (MET1), CHROMOMETHYLASE 3 (CMT 3) and DOMAINS REARRANGED METHYLASE 2 (DRM2) (Figure 1.13) (He *et al*., 2011; Zhang *et al*., 2018). MET1 is a homolog of the mammalian DNMT1 DNA METHYLTRANSFERASE1 (DNMT1) which is responsible for the maintenance of symmetric CG methylation. CMT3 is needed for the maintenance of DNA methylation at CHG sites, while DRM2 is responsible for *de novo* DNA methylation at all CHH sites (Chen *et al.*, 2015). There are other methyltransferases which still need to be characterized and these include those encoded by the genes *MET2a*, *MET2b* and *MET3*, other *CMT* genes (*CMT1* and *CMT2*) and genes from the *DRM* class (*DRM1* and *DRM2*) (Martienssen and Colot, 2001).

The first plant gene encoding cytosine methyltransferase was isolated from Arabidopsis. The Arabidopsis *MET1* gene is a member of a small multigene family and *MET1* is the predominant

methyltransferase, being expressed in vegetative and floral tissues. *MET1* homologues have been identified in a wide variety of plant species and the main role of *MET1* is suggested to be in restoring parental pattern of cytosine methylation to the newly replicated DNA daughter strands (Finnegan and Kovac, 2000).

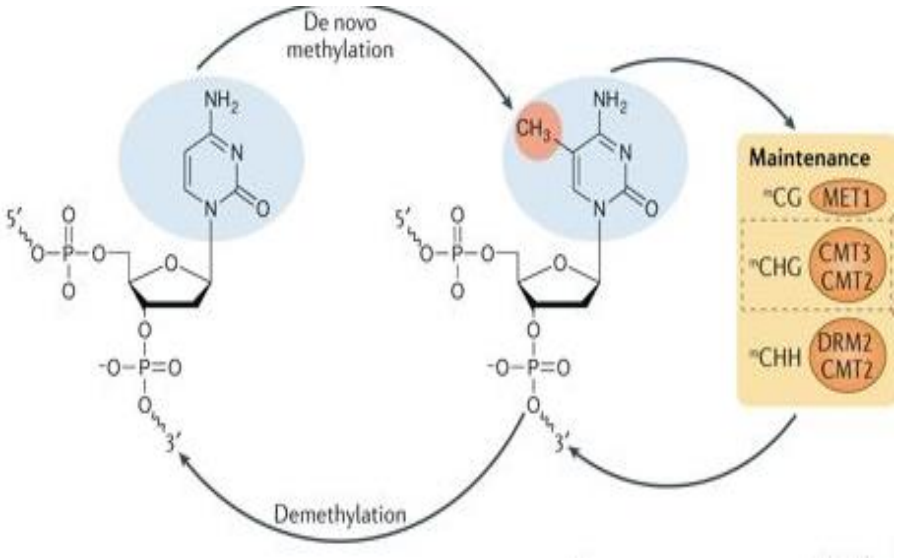


Figure 1. 13 The regulation of DNA methylation in plants. *De novo* DNA methylation can occur on all cytosine context: CG, CHG or CHH sequence (where H represents A, T or C). De novo methylation is initiated by the enzyme DRM2 through RNA-directed DNA Methylation (RdDM) and demethylation by DNA demethylases or due to failure in methylation maintenance following DNA replication (Zhang *et al*., 2018).

DNA Methylation in Tomato

The discovery of the naturally occurring tomato epimutant *Cnr* suggested that there was an important role of epigenetic regulation in fruit ripening (Manning *et al*, 2006). More recent work has shown that demethylation of ripening-related genes is important for tomato fruit developmental processes (Zhong *et al*., 2013; Liu *et al*., 2015). The role of methyltransferase in the methylation of the *CNR* promoter was investigated using 5-azacytidine which is an inhibitor of methyltransferase (Zhong *et al*., 2013). The inhibition caused premature ripening of tomato fruit (Figure 1.16).

Additionally Zhong *et al*. (2013) also demonstrated that the promoters of ripening-related genes were demethylated at the onset of ripening in tomato. This study was supported by data from Liu *et al*. (2015) who demonstrated that a *DEMETER-like* gene encoding a DNA demethylase was necessary for normal ripening in this fruit. Recently it has been shown that silencing of *CHROMOMETHYLASE3* (*SlCMF3*), can reduce methylation in the *LeSPL-CNR* promoter region and trigger normal fruit ripening (Chen *et al*., 2015).

The role of miRNAs in the regulation of tomato fruit ripening has also been investigated. The 3'-UTR of *LeSPL-CNR* mRNA possesses a potential target site for micro RNAs: miRNA156 and miRNA157. MiRNA-based virus induced gene silencing (Mr VIGS) was used to express pre-SylmiR157. SlymiR157 was found to regulate the *LeSPL-CNR* gene in a dose-dependent manner through miRNA-induced mRNA degradation and translation repression. This caused a delayed ripening effect in

tomato fruit. A related miRNA, SlymiR156, affects fruit softening after the red ripe stage (Chen *et al*., 2015).

(A)

(B)

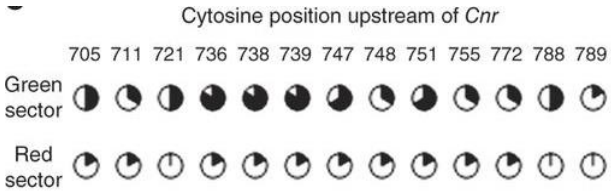


Figure 1. 14 Methylation in tomato fruit (cultivar Ailsa Craig). Different fruits showing (A) normal ripening (left) at 42d.p.a, in fruit treated with 5-azacytidine at 17 d.p.a. showing premature ripening at 30 d.p.a. (centre), and the control fruit (right) was treated with water (d.p.a = day post anthesis). (B) Sanger bisulfite sequencing showing demethylation in *Cnr* promoter region. The numbers indicate position of cytosine relative to the start of *Cnr* open reading frame (black parts showing frequency of methylated cytosine) (Zhong *et al*., 2013).

Detection of DNA Methylation

Methods for the detection of DNA methylation can be divided into two groups: bisulfite dependent methods and bisulfite independent methods (Figure 1.14). Bisulfite treatment converts the unmethylated cytosines to uracil, but leaves 5-methylcytosine residues unaffected. In DNA, uracil is a demethylated form of thymine turning an epigenetic mark into a genetic difference. Analysis such as methylated specific polymerase chain reaction (MSP), combined bisulfite restriction analysis (COBRA) and bisulfite genome sequencing (BGS) are used to determine the change in methylation state of a specific sequence. Sequencing can then be carried out after bisulfite treatment, either by the Sanger method, Next Generation Sequencing (NGS) or pyrosequencing (Saze *et al*., 2003; Jin *et al*., 2012; How-Kit *et al*., 2015; Zhong *et al*., 2013).

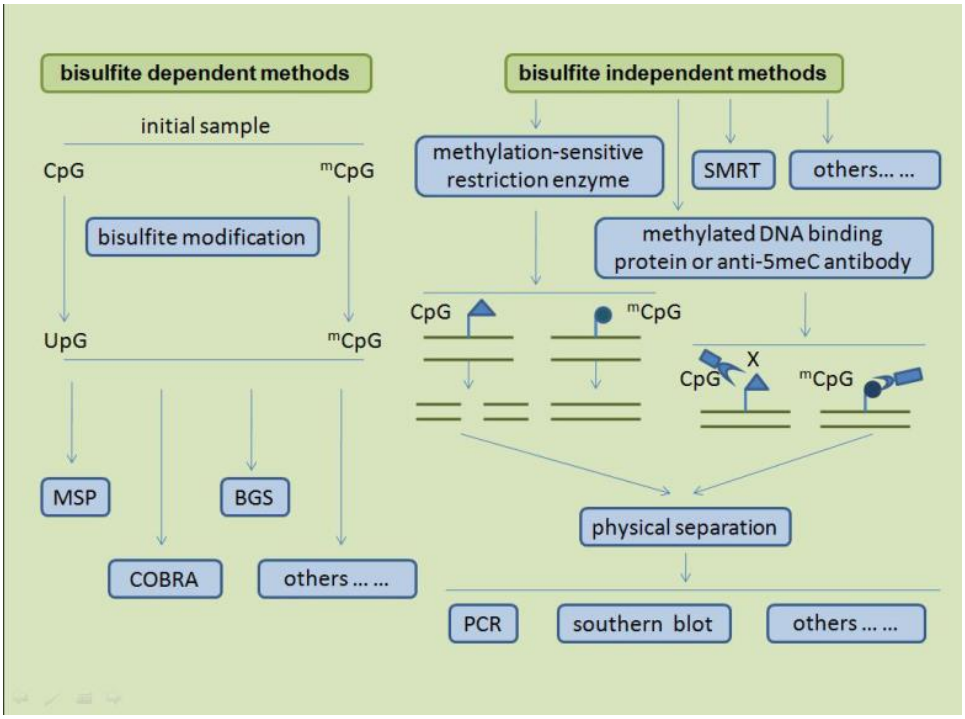


Figure 1. 15 Various methods for methylation analyses. Bisulfite dependent methods such as MSP, BGS and COBRA are based on bisulfite-mediated conversion of unmethylated cytosines into uracils. Bisulfite independent methods such as DNA digestion by some restriction enzymes, making it possible to determine methylation status. DNA fragments containing methylated CpG sites could be enriched by anti-methylcytosine antibody or methylation binding proteins. Advances in next generation genome sequencing technology have led to the development of novel techniques such as SMRT (single molecule real time sequencing) which can analyze 5-methylcytosines with genome wide coverage (Jin *et al*., 2012).

In non-bisulfite methods, restriction enzymes and antibodies specific for methylated regions are used (Figure 1.15). Some restriction enzymes are inhibited by methylated cytosines so the patterns of cutting can provide a tool in detecting methylated sites. The most widely used restriction enzymes for detection of methylation are *HpaII* and *MspI* (Vongs *et al*., 1993; Kankel *et al*., 2002; Saze *et al*., 2003; Reinders *et al*., 2009). The enzyme *HpaII* cleaves only CCGG sites that are unmethylated while *MspI* cleaves DNA at CCGG sites irrespective of 5-methyl group at the internal C residue (Waalwijk and Flavell, 1978).

Another restriction enzyme which can be used to analyse the presence of methylation sites is McrBC. McrBC is a well characterized methylation dependent restriction enzyme from *Eschericia coli* K12 (Panne *et al*., 2001). McrBC requires nucleoside triphosphates but does not have an associated methyltransferase activity and its cofactor is GTP. McrBC recognizes two methylated or hemimethylated cytosines at a distance of approximately 30 to more than 2000 base pairs and cleaves the DNA close to one of the methylated sites (Pieper *et al*., 2002).

Affinity enrichment of methylated regions using antibodies specific for methylated cytosines or protocols using methyl-binding proteins have also become useful methods for profiling of DNA methylation in complex genomes (Yan *et al*., 2004; Mathieu *et al.,* 2007; Zhong *et al*., 2013; Laird, 2010).

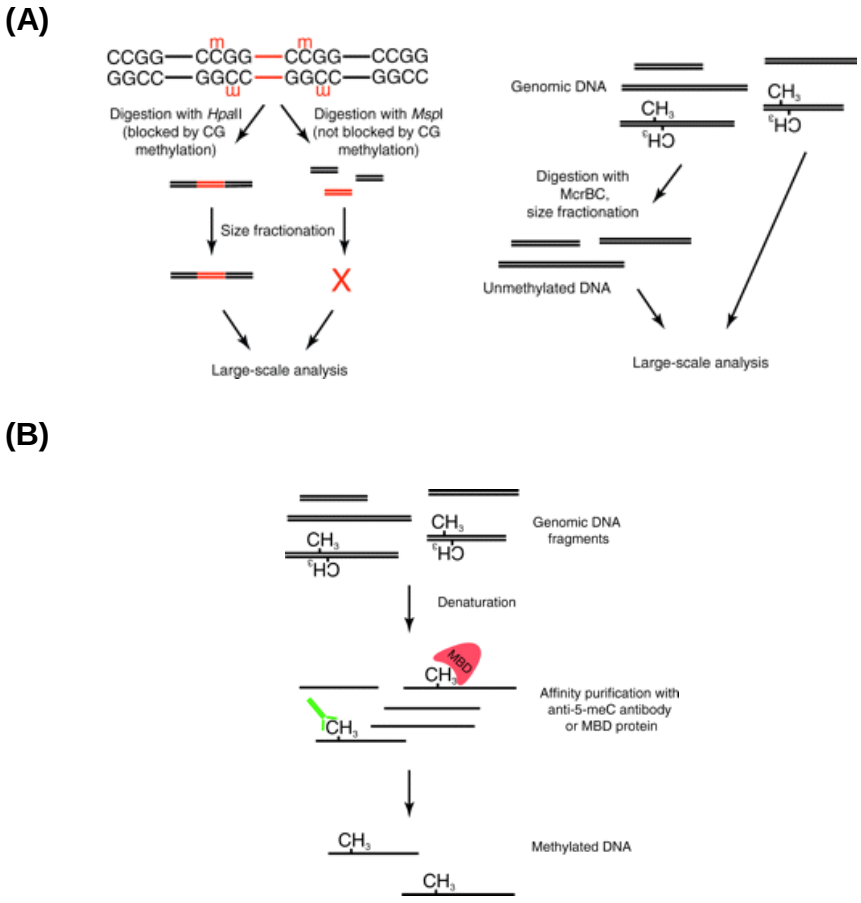


Figure 1. 16 Methylation detection using restriction enzymes. (A) Restriction enzymes *Hpa*II and *Msp*I (left) and McrBC (right) can be used to digest DNA before further analysis to detect differentially methylated region. Detection can also be done using (B) affinity purification where DNA is denatured and affinity purified using antibody (green) or methyl-binding domain (MBD, red) protein which is then attached to a column (Zilberman and Henikoff, 2007)

1.3 Mapping Populations and their potential for studying epigenetic variation

1.3.1 Types of Mapping Population

Studies on epigenetic variation and how it can be inherited have been undertaken using similar methods to those used in studies of genetic variation. In genetic studies, doubled haploids (DH), recombinant inbred lines (RIL), and near isogenic lines (NIL) populations are used for gene discovery and mapping (Figure 1.17). The differences between each lines are from the different parental lines and the crosses involved in generating the population. Doubled haploid populations are generated from induced doubling of pollen microspore while NILs and RILs are produced from parents that are both highly homozygous (inbred) but show significant genetic variation. NILs are generated from backcrossing to a recurrent parent and several generations of selfings, while RILs are derived from selfing of F_2 plants which are then propagated using the single seed descent method (Collard *et al*., 2005; Xu, 2010; Sehgal *et al*., 2016).

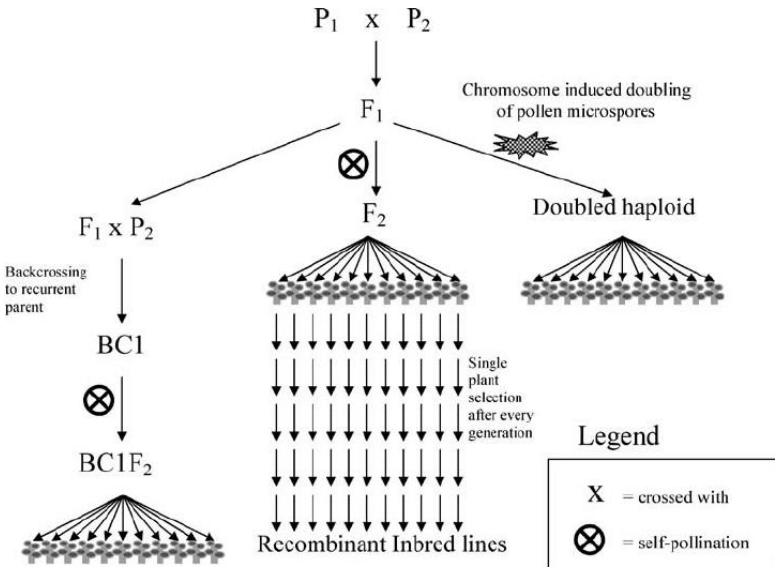


Figure 1. 17 Three types of mapping populations for self-pollinating species. Backcross lines, recombinant inbred lines and double haploid population (Collard *et al*., 2005).

.

Each mapping population has advantages and disadvantages. The type of plant, length of time to produce the mapping population and the type of marker that will be observed all must be considered when choosing a mapping population, DH and RILs are two populations that are highly homozygous and can be used easily by other researches once they have been established. A DH population takes less time than RIL but is limited to only plants that can be easily propagated by tissue culture. A RIL population takes a longer time to produce, but represents a homozygous population after 8-9 generations, while DH lines represents the segregation of individual F_1 progenies and a homozygous population in 2 generations (Sehgal *et al*., 2016).

RILs consist of a series of homozygous lines, each containing a unique combination of chromosomal segments from the original parent (Collard *et al*., 2005). This population of individuals with a mosaic of parental genomes can be used to fine map quantitative trait loci. With recent knowledge that epigenetic modifications can also be stable and inherited, epigenetic RIL (epiRIL) populations have been generated and used to observe transgenerational inheritance of those modifications in the model plant Arabidopsis (Johannes *et al*., 2009; Reinders *et al*., 2009) but none have been generated from crop species.

1.3.2 Epigenetic Recombinant Inbred Lines (epiRILs)

Transgenerational epigenetic variation in plants has been studied in *Arabidopsis thaliana*. Parental lines differing in DNA methylation profiles, experimental populations such as backcrosses, F_2 population and RILs have been used in those studies (Kankel *et al*., 2003; Mathieu *et al*., 2007; Reinders *et al*., 2009; Johannes *et al*., 2009).

In a recent investigation, an Arabidopsis epiRIL population was produced from a cross between parents with different methylation levels in two different investigations using two different hypomethylated mutants (Figure 1.18 and Figure 1.19). The population was generated using parents with different epigenetic marks and from the cross, progenies were self-fertilized for seven generations before bulking for the final generation. The parental lines used were a wild type and a mutant for a gene involved in the maintenance of DNA methylation.

Johannes *et al*. (2009) used the Colombia (Col) accession with a homozygous wild type with *DDM1* allele (Col-wt) as one parent and the other parent is homozygous for the *ddm1*-2 mutant allele (Col-*ddm1*, 4^{th} generation). The *ddm1* mutant was isolated from an ethyl methanesulfonate mutagenized populations. The level of 5-methylcytosine was reduced over 70% in the *ddm1* mutants (Vongs *et al*., 1993). Successive selfings are shown in Figure 1.18 causing segregation of the epialleles and becoming more homozygous in later generations. The resulting epiRIL population showed variation and high heritability for flowering time and plant height.

A different epiRIL population was generated by Reinders *et al*. (2009) using a wild type and a homozygous *met1*-3 mutant (both Col-0 ecotype) as parents of the epiRIL population. The *met1* mutant was identified in a screen for insertion mutants impaired in transcriptional gene silencing. The *met1*-3 mutant contained T-DNA integrated in the *MET1* gene. The insert was 7.1 kb and disrupted the conserved motif region of the gene (Saze *et al*., 2003). The mosaic of epialleles and increasing homozygosity of the population is presented in Figure 1.19. Self-fertilisation for several generations caused segregation of epialleles and the percentage of homozygosity from each segregation locus (assuming no epistasis) is $Fn = 1-(1/2)^{n-1}$, where n is the number of generation (Snape and Riggs, 1975). In the resulting epiRIL populations there were variation in flowering time and different resistance to some abiotic and biotic stresses.

Investigation on the methylome of the Arabidopsis epiRILs has identified hundreds of parental differentially methylated regions (DMRs) showing Mendelian segregation patterns. Analysis showed that induced DMRs were stably inherited independent of the DNA sequence and functioned as epigenetic quantitative trait loci (epiQTL) (Cortijo *et al*., 2014).

1.3.3 Mapping epiQTL

The final generation of a RIL population consists of more than 99% loci in a homozygous state (Snape and Riggs, 1978). In Arabidopsis, this hypothesis was tested using generation F_9 of the epiRILs from parents with different methylation state of the DDM1 alelle (Johannes *et al*., 2009). The methylation state of eleven target sequences were analysed in 22 Arabidopsis epiRILs. The results showed that five of the differentially methylated sequence segregated in a Mendelian or near-Mendelian manner. The percentage of homozygous state of methylation for the flowering time distribution showed that it was as expected from a segregating locus in a single seed descent population, with 99.6 % in homozygous state.

A recombinant map was constructed using transgenerationally differentially methylated region (DMR) of the Arabidopsis epiRILs (Colomé-Tatché *et al*, 2012). Methylated DNA immunoprecipitation followed by hybridization to a whole genome DNA tiling array (MeDIP-chip) was undertaken using 123 epiRILs and the parental lines. Further analysis was undertaken by utilizing the recombinant map with classical linkage analysis to reveal epigenetic quantitative trait loci (epiQTL) of complex traits in the Arabidopsis epiRIL population (Cortijo *et al.*, 2014). Interval mapping of two highly heritable and weakly correlated traits (flowering time and primary root length) detected highly significant QTLs in several chromosomes. The linkage mapping could also explain the causal variants of the broad-sense heritability in the epiRILs were from the original parents and not from later generations. The DMRs in the

Arabidopsis epiRILs were shown to be stably inherited independently of DNA sequence changes and can function as an epiQTL.

In Arabidopsis, differentially methylated regions (DMRs) were found in the epiRILs showing randomization of methylation marks and became fixed unique marks in different lines. The utilization of DMRs to identify epiQTL for specific traits would be useful in an economically important crop such as tomato.

In this project, a tomato epiRIL population was generated from parental lines with the same/similar genetic variations and differences in their epigenetic background. To obtain the epigenetic difference between the parent lines, RNA interference (RNAi) was used to reduce the expression of the *MET1* gene (MET1 is responsible for the maintenance of CG methylation in plants) in one of the parental line. For the RNAi, a pMET1 construct was designed and used for the transformation of cherry tomato variety WVA 106. The transgene was removed by successive selfings after the initial cross of the parents. Only plants without the transgene (azygous) was used to generate the tomato epiRIL population.Changes in phenotype, gene expression, cytosine methylation and transposons movements were investigated in different generations of the tomato epiRILs.

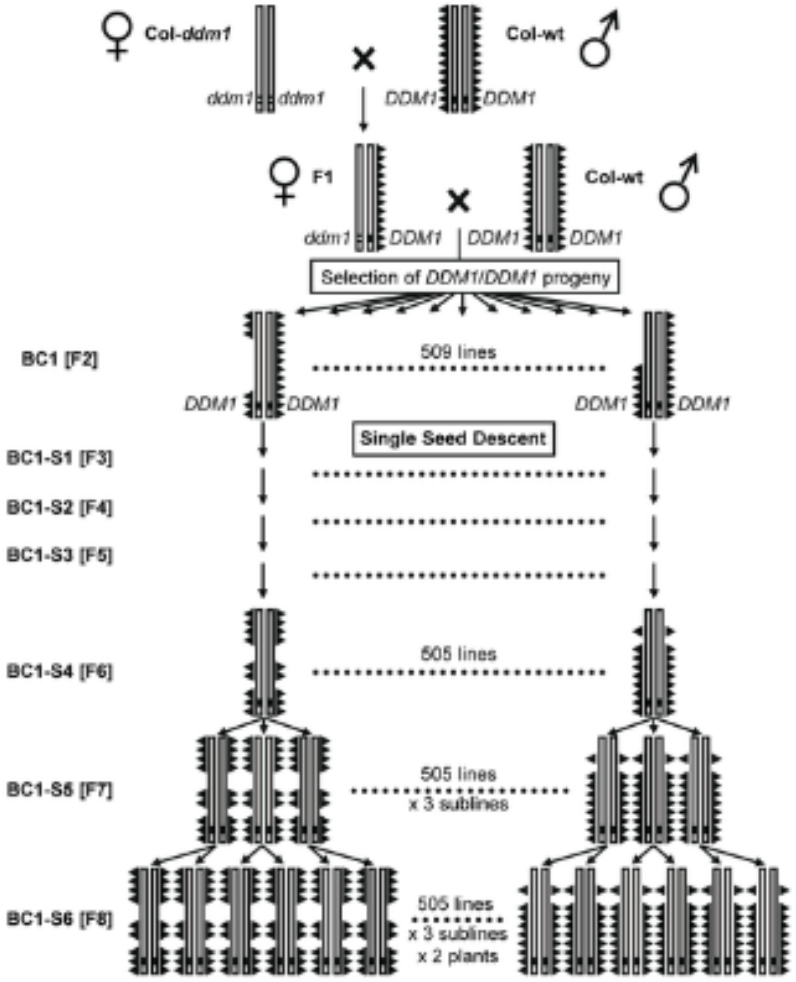


Figure 1. 18 Construction of an epiRIL population illustrated by Johannes *et al.* (2009). The grey bars represent the genome of Arabidopsis and triangles represent DNA methylation. Black and white squares represent the *DDM1* locus on chromosome 5. The parents were isogenic but differ in DNA methylation levels. An F_1 progeny was backcrossed to the wild type parent and its progenies were selfed until F_6. In F_7 three independent sublines were established and the seeds were bulked for F_8.

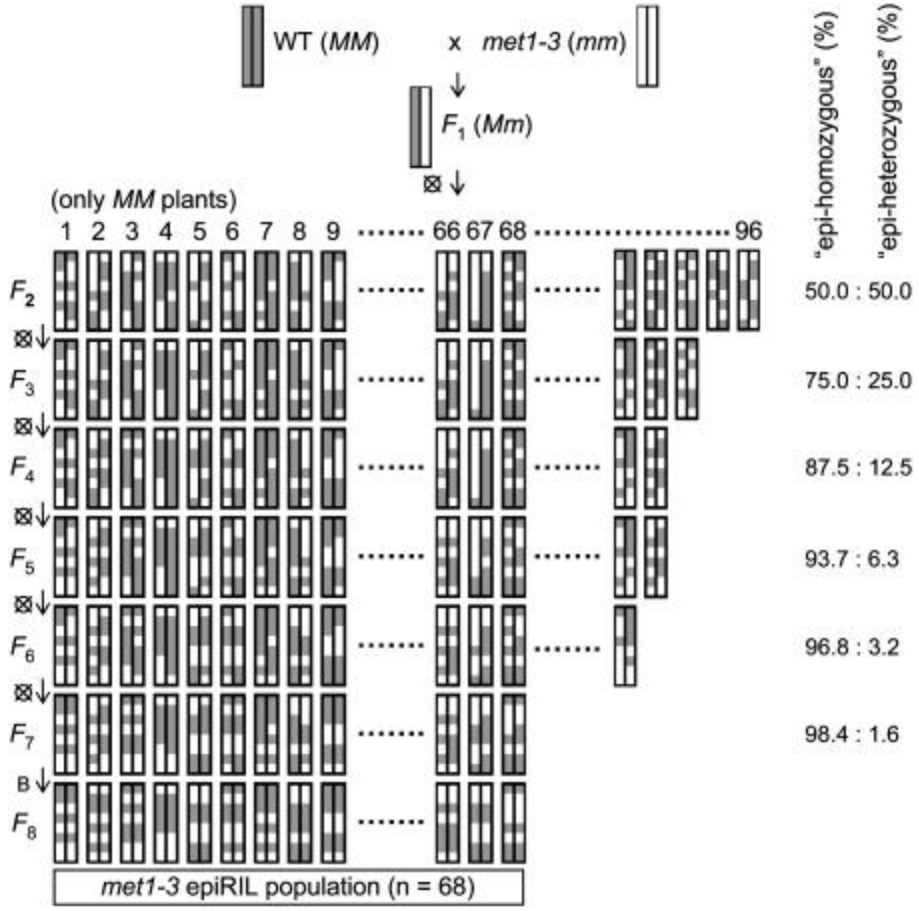


Figure 1. 19 Construction of the epiRIL population, illustrated by Reinders *et al*., (2009). Parental chromosomes and their segments are marked wild type (gray, WT, MM) and met1-3 (white, met1-3,mm). Eight generations of inbreeding (vertical) are marked as F_2 to F_8 (crossed circles at the left mark single-seed descent). A bulk harvest of individuals at F_7 is marked by B. Predicted levels of “epi-homozygosity” and “epi-heterozygosity” at each generation are indicated on the right

AIMS AND OBJECTIVES

The aim of this project was to generate and characterize a novel population of tomato epiRILs and identify the effects of the epigenetic variation on the development of tomato. This work was part of a collaboration with Professor Philippe Gallusci at Bordeaux University, France and INRA (The French National Institute for Agricultural Research).

Specific Objectives:

1. To use RNAi *MET1* tomato lines to generate a tomato epiRIL population by single seed descent to F_8.
2. To determine stable phenotype generated in the epiRILs.
3. To determine the methylome of the epiRIL parents and follow selected loci through epiRIL development.
4. To determine the effect of *MET1* knockdown on gene expression, level of methylation and transposon movement in the epiRIL population

CHAPTER 2 THE GENERATION AND PHENOTYPING OF THE epiRILs

2.1 Introduction

2.1.1 Epigenetics and natural variation

Phenotypic variation is difficult to explain adequately by taking into account genetic variation alone. This is especially true when considering diversity and the adaptability of plants in changing environments. Evidence from a several species proved that there is a 'missing' cause of variations. Some phenotypes such as adaptive responses are too rapid to be explained by genetic mutations which are rare events and require long periods of time to accumulate (Manolio *et al*., 2009; Thorson *et al*., 2017). Epigenetic modifications such as DNA methylation, chromatin remodelling, histone modification and non-coding RNAs can all result in changes of gene expression without changing the sequence of the DNA (Gallusci *et al*., 2017).

To explore the role of epigenetic variation in plant development, Arabidopsis was used to build populations of epiRILs (epigenetic Recombinant Inbred Lines) (Reinders *et al*., 2009; Johannes *et al*., 2009). In these experimental populations, the genetic background of all the lines is similar, but there is variation in the levels of cytosine methylation across the genome. In the Arabidopsis epiRILs, the effects on phenotype ranged from differences in plant height and flowering type to changes in disease resistance (Kakutani *et al*., 1995, Reinders *et al.*, 2009, Johannes *et al.,* 2009).

In this investigation, RNAi was used to silence the tomato *MET1* gene to investigate the effects of reducing cytosine methylation across the genome. The *MET1* silenced lines were then used to generate the tomato epiRIL population. This was a joint project between Bordeaux University and INRA (The French National Institute for Agricultural Research) with the University of Nottingham. The generation of the *MET1* RNAi lines and the initial crosses were undertaken by colleagues at INRA.

Tomato has been used as a model plant to understand the control of fruit ripening. It also has a large and complex genome in comparison to Arabidopsis with many repetitive sequences. A major objective of the current project was to generate a tomato epiRIL population that could be used to study the effect of altered DNA methylation on plant phenotypes including those in fleshy fruits. The hypothesis for this part of the project is that changes in methylation from the reduced *MET1* expression can cause changes in the phenotype of the tomato epiRILs due to the role of *MET1* in the maintenance of CG methylation in the plant genome.

2.2 Materials and Methods

2.2.1 Establishment of the epiRILs

The epiRIL population was generated from tomato (*Solanum lycopersicum*) cv. Cherry WVA 106 (wild type). The parental lines were genetically identical but one of the parents was a *MET1* RNAi line in the wild type background. The transgenic line was constructed by Philippe Gallusci and colleagues at Bordeaux University, France.

In the initial *MET1* transgenic experiments, three T_0 transgenic plants were obtained (Appendix 1) andthese T_0 lines were then backcrossed to wild type WVA106. T_1 individuals were then selected that had lost the transgene (azygous). These lines were then selfed to produce the F_2 generation. Unfortunately, a very low germination rate was observed in the F_2 seeds (Gallusci *et al*., unpublished data) and only a small number of individuals could be recovered.

To produce further azygous lines, T_1 lines containing the *MET1* transgene (hemizygous) were backcrossed with wild type to obtain BC1 lines. Azygous F_2 progenies from these BC1lines were selected and selfed to produce the BC1-S1 lines. There were 8 azygous lines from the original F_2 population and 109 lines from the BC1 used in the generation of the tomato epiRILs. The lines planted in each generations of the epiRILs consisted of lines from the two different parental cross (either from F2 or BC1). Because of the space requirements, in each generation there was a difference in the number of plants per line (table 2.1). In early generations up to five plants per genotype (line)were grown but in later generations, with more lines obtained from the backcross, a maximum

of two plants were grown to fruiting from each line. All epiRILs were then fixed by continuous rounds of selfing and single seed decent following the scheme shown in Figure 2.1. The seeds from each individual line were bulked at F_7.

To simplify the description of the generation of the epiRIL population, lines will be indicated as filial generations or F_2, F_3, etc. The origin of the lines is presented in Table 2.1. The seasons varied between each generations as seeds were collected from one generation and planted in the next season for the next generation (Table 2.1).

From the 118 epiRILs, only several lines were chosen for further detailed analysis. There were 10 lines selected consisting of eight lines (line 2.2, 3.3, 3.4, 4.1, 4.15, 4.18, 4.23 and 4.5) from the first cross between the parental lines (from F_2) and two lines (2.1 and 3.2) from the backcross (from BC_1). These lines were selected because the data were available from early generations and so they could be used to investigate the changes in transgenerational inheritance until F_7. The wild type was also planted in each generation as a control.

The experimental design of the early generations (F_2 and F_3) used a randomized complete block design and in later generations (F_4 - F_8) a complete randomized design was used due to the limited space in the glasshouse (Appendix 4-15). There were more plants per line in the first two generations and only two plants per line from generation F_4 until F_8 (Table 2.1). However, for the plant height measurements were taken from three plants in the sowing trays so the number of plants was the same for all of the epiRIs generations.

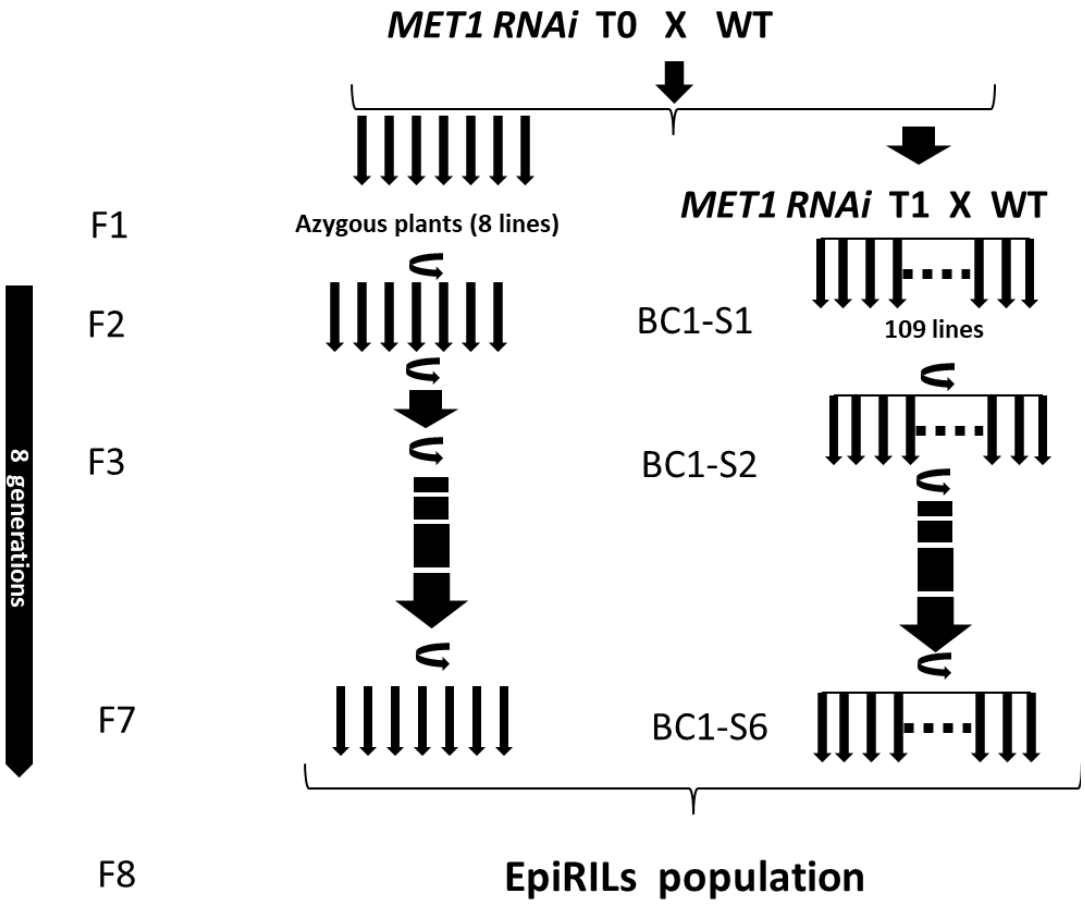


Figure 2. 1 Diagram for the generation of the tomato epiRILs. On the left side of the diagram are the lines (F_2, F_3, etc) from the first cross between transgenic line and wild type. On the right side of the diagram are backcross lines (BC1-S1, BC1-S2, etc) from a cross between T1 hemizygous line and wild type. Progenies from both crossings were used and the generations were labelled as F1, F2, etc. ↻ represents selfings in each generation.

Table 2. 1 The number of lines and planting dates of each generation of epiRILs

epiRIL Generation	Generation of the lines according to parental cross	Number and generation of lines planted	Number of plants per line	Planting date
F_2	F2	1 WT, 8 F2	3	February – June 2016
F_3	F3 and BC1-S2	1WT, 8 F3, 2 BC1-S2	5	October 2016 – February 2017
F_4	F3 and BC1-S3	1WT, 8F4, 109 BC1-S3	2	July – November 2017
F_5	F4 and BC1-S4	1WT, 8F5, 106 BC1-S4	2	December 2017 – April 2018
F_6	F5 and BC1-S5	1WT, 7F6, 107 BC1-S5	2	May – September 2018
F_7	F6 and BC1-S6	1WT, 7F7, 107 BC1-S6	2	October 2018 – February 2019
F_8	F7 and BC1-S7	1WT, 7F8, 107 BC1-S7	2	February – June 2019

2.2.2 Seed Sterilization

The seeds were sterilized to reduce infection from seed borne virus. The tomato seeds processing protocol was adapted from the protocol used by Cornell University (https://cuaes.cals.cornell.edu, 2019). The seeds were extracted from the pericarp and placed in 2 mL microfuge tubes and soaked in 50% (v/v) HCl for 10 minutes. They were then rinsed three times with sterile water. The seeds were then soaked in 10% (w/v) TSP (tri sodium phosphate) for 15 minutes and again washed with sterile water. After removing all the remaining water, the seeds were soaked in 70% ethanol for 10 minutes and rinsed once with sterile water. In the last step the seeds were soaked in 50% (v/v) sodium hypochlorite for 20 minutes and rinsed three times with sterile water. The seeds were then left in the opened tube to dry in a laminar flow hood.

2.2.3 Growing conditions and experimental design

Seeds were sown in nursery stock seed and modular compost (Levington® F2+S) and transferred after 4-5 weeks to large pots (7.5 L) in a potting compost (Levington CNSC). Watering and fertilizer were provided by an automated watering system. The glasshouse temperature was maintained at 24-26 ^{0}C during the day and 16 ^{0}C at night. The biocontrols used were Amblyline Flo (*Amblyscius cucumeris*) thrips predator and Oriline L (*Orius laevigatus*), both from Bioline Agriscience (Essex, U.K.).

2.2.4 Plant genotyping

To check the presence of the *MET1* transgene in the lines, PCR was undertaken using the *MET1* primers. The forward primer (SlMet1iF) was CGACGACAAGACCCTCTGCTAA and the reverse primer (SlMet1iR) was GAGGAGAAGAGCCCTATTTGCC. The primer sequences were designed by Philippe Gallusci at Bordeaux University, France. The primers were designed based on the sequence of tomato mRNA for DNA (cytosine-5)-methyltransferase 1 (*LeMET1*). The target length of the amplicon was 250bp. The PCR reactions were undertaken using Promega GoTaq® HotSTart Colorless MasterMix (Wisconsin, USA) using the conditions shown in Table 2.1. The PCR products were separated by electrophoresis on 1.5 % (w/v) agarose gels in 0.5X TAE (Tris-acetate-EDTA) at 100 Volts for 30 minutes.

Table 2. 2 PCR Reaction mixture and conditions for plant genotyping

PCR Reaction

Component	Final Volume (µL)	Final Concentration
GoTaq® HotStart Colorless Master Mix 2X	12.5	1 x
Upstream primer	2.5	1.0µM
Downstream primer	2.5	1.0µM
template DNA	X	50ng
Nuclease-Free Water	Up to 25 µL	

PCR Conditions

Step	Temperature	Time	Cycles
Initial denaturation	95°C	5 min	1
Denaturation	95°C	30 sec	30
Annealing	58°C	30 sec	30
Extension	72°C	30 sec	30
Final extension	72°C	5 min	1
Hold	4°C		

2.2.5 Plant Height Measurement

Plant height was measured using a 30 cm ruler and determined from the base of the plant on soil to the shoot tip four weeks after germination.

2.2.6 Observation of floral abnormality using a light microscope

Flowers were harvested at around 10 am in the morning and were placed on a black background and observed under a Zeiss light microscope, using 0.8x and 1x zoom, 10x objective, 100.00ms exposure and 100% intensity.

2.2.7 Pollen count

Pollen was taken from flowers at anthesis and the ‘iodine pollen starch test’ was used to identify viable pollen. The method used was adapted from Chang *et al*., (2014) and Abejide *et al*., (2014). Three flowers were taken from each plant. The anther cones from the three flowers were placed on a microscope slide and the pollen was extracted from the anthers using a needle. The pollen was then stained using two to three drops of iodine potassium iodide (IKI) solution. The IKI solution was 0.2 % (w/v) potassium iodide and 1 % iodine crystals (Sigma, St.Louis, USA) in distilled water. 0.2 g of potassium iodide was dissolved in a small amount of distilled water. Then 1g of iodine crystals were added with continuous stirring. Once dissolved, the solution was made up to a final volume of 100mL with distilled water.

The pollen grains were observed under a Leica DMRB fluorescence microscope. Pictures were taken from three randomly

selected areas of the slide and used to calculate the percentage of viable pollen grains.

2.2.8 Measurement of fruit weight, colour and total soluble solids (TSS).

Samples of fruits were taken from stage B+7 (breaker+7 days). From each plant, three fruits were analysed. The fruits were weighed using a digital balance CS series (OHAUS ®).

Fruit colour was measured using a Konica Minolta CR-400 handheld Chromameter. When using the Chromameter, it was first calibrated using a white background. It was calibrated for the measurements of a^*, b^* and L^*. The letters a^*, b^* and L^* are three perpendicular axes in the colour sphere used by the CIELAB colour space system (Lopez Camelo and Gomez, 2004). Each of the axes corresponds to different colours: L* (from white to black), a* (green to red) and b*(blue to yellow). The colour of the pericarp was measured in three regions of the fruit and the measurements were undertaken on three fruits per line. The total colour index (TCI) was calculated using the formula:

$$TCI = 2000a^*/L^* (a^{*2} + b^{*2})^{0.5}$$

An Atago digital pocket refractometer was used to measure total soluble solids of the tomato fruits at ripe stage (Breaker+7). The instrument was set to zero with a distilled water blank. To measure the TSS of the samples, fruit of each line were cut in half and a drop of juice squeezed onto the prism surface in the refractometer and the reading was taken. The prism surface was then cleaned using a tissue.

2.2.9 Statistical Analysis

Statistical analysis was undertaken using Genstat® software (19th edition). The variation among plants was analysed using Analysis of Variance (ANOVA). A post-hoc multiple comparison was undertaken to test all possible pairs of means using the Duncan Multiple Range Test (DMRT). Results of DMRT is presented in Appendix 3-6. Due to the difference in experimental design and the number of plants per line, for some traits only the standard error of difference (s.e.d) is available (Appendix 17-20).

2.2.10 Environment data

Daily maximum, minimum and mean temperature and photoperiod data were downloaded from metoffice.com for the year 2015 - 2019. Tinytag Data Loggers were used to record temperature and humidity inside the glasshouse in generation F_3 to observe flunctuations in the two abiotic factors (Appendix 2 and 3).

2.3 Results and Discussion

2.3.1 Genotyping for *MET1* transgene

The RNAi knockdown of *MET1* is likely to affect many aspects of plant development due the changes in the genome methylation. In this chapter, we focus on the phenotypic variation that was seen in several generations of the epiRILs. Some phenotypes that were stable across generations and these were further investigated using qPCR and enzymatic analysis.

An important first step in the generation of the epiRIL population was genotyping to determine whether the transgene was still present in any of the individual lines. Primers were designed from the sequence of the plasmid used for the transformation and also from a small part of the *MET1* sequence.

Genomic DNA from T_0 and T_1 plants were analysed to confirm the loss of g the transgene (Figure 2.2). A band of size 250 bp was present in the lane from the T_0 lines as would be expected since these harboured the *MET1* transgene.

To confirm the amplification of the transgene, T_1 lines that were known to carry the transgene were also used in the genotyping. The 250 bp band was also present in a range of T_1 lines. In contrast, there were no amplicons visible in the wild type control line and azygous T_1 lines. A similar analysis of T_2 azygous plants confirmed the absence of the transgene in these lines (Figure 2.3).

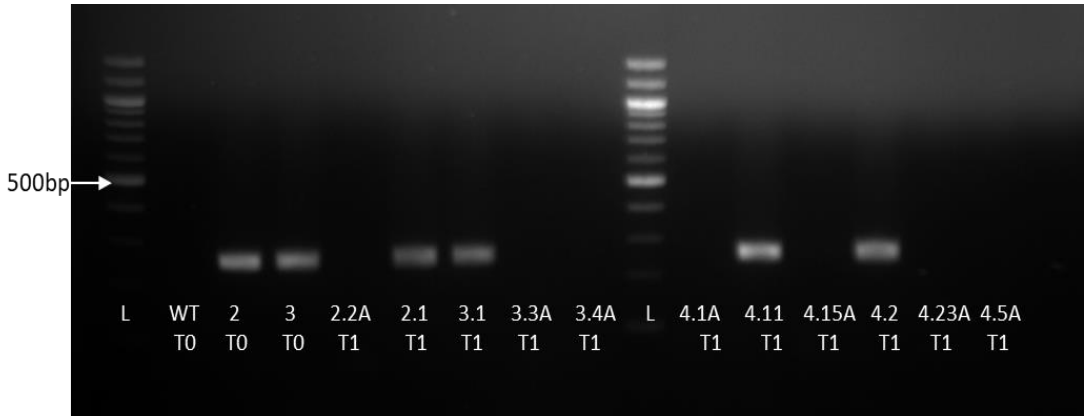


Figure 2. 2 Testing for the presence of *MET 1* transgene in the T_0 and T_1 generation. The *MET1* transgene was detected by PCR and generated an amplicon of 250 bp. Lanes 1-15 were (1) DNA Ladder, (2) Wild Type, (3) T_0 line 2, (4) T_0 line 3, (5) T_1 line 2.2A, (6) T_1 3.1 line, (7) T_1 line 3.3A, (8) T_1 line 3.4A, (9) DNA ladder, (10) T_1 line 4.1A, (11) T_1 line 4.11, (12) T_1 line 4.15A, (13) T_1 line 4.2, (14) T_1 line 4.2, (15) T_1 line 4.23A and (16) T_1 line 4.5 A. The letters L = DNA ladder and A = azygous.

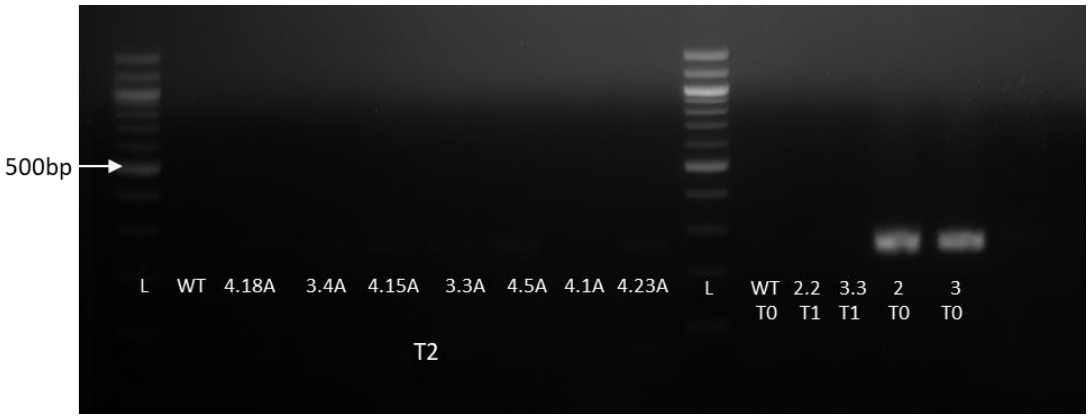


Figure 2. 3 Testing for the presence of *MET 1* transgene in the F_2 compared to the T_0 and T_1 generation. The *MET1* transgene was detected by PCR and generated an amplicon of 250 bp. Lanes 1-15 were (1) DNA Ladder, (2) Wild Type, (3) T_2 line 4.18, (4) T_2 line 3.4, (5) T_2 line 4.15, (6) T_2 3.3 line, (7) T_2 line 4.5, (8) T_2 line 4.1, (9) T_2 line 4.23, (10) DNA ladder, (11) T_0 wild type, (12) T_1 line 2.2A, (13) T_1 line 3.3A, (14) T_0 line 2, and (15) T_0 line 3. The letters L = DNA ladder and A = azygous.

The removal of the *MET1* transgene by segregation is an important step in the generation of the epiRILs. Methylation marks present after transgene removal will be maintained during DNA replication by the presence of working copies of the *MET1* gene. The methylation levels in the epiRILs will then be 'fixed' in a homozygous condition going forward by several rounds of selfing.

The selection for plants without the transgene (azygous plants) in both the initial parental cross and the backcross means that the initial lines were selected against genomes containing the transgene. Lines resulting from segregation carrying the transgene were lost and are not represented in the final tomato epiRIL population.

2.3.2 Phenotyping the vegetative development in the epiRILs

The phenotypes of the lines were assessed between the F_2 to F_7 generations. In many cases phenotypes became more pronounced in later generations presumably as epi-loci became fixed. In some cases abnormal phenotypes were observed in a single generation and one explanation is that these phenotypes were the result of transient changes caused by environmental conditions. Only persistent and stable phenotypic changes were characterised further.

Twisted and flat stems

Phenotypic changes were seen in stems and leaves in certain lines. These included flattening and broadening of stems (Figure 2.4). Fasciated stems have long been reported in tomato (Zielinski, 1948) and

this phenotype was seen in several lines. For example, line 4.15 showed twisted and / or fasciated stems in generations F_2 until F_4 (Table 2.2). In other lines, such as 3.3, the altered stem phenotype became prominent in the F_5 generation and was maintained from that point through F_7.

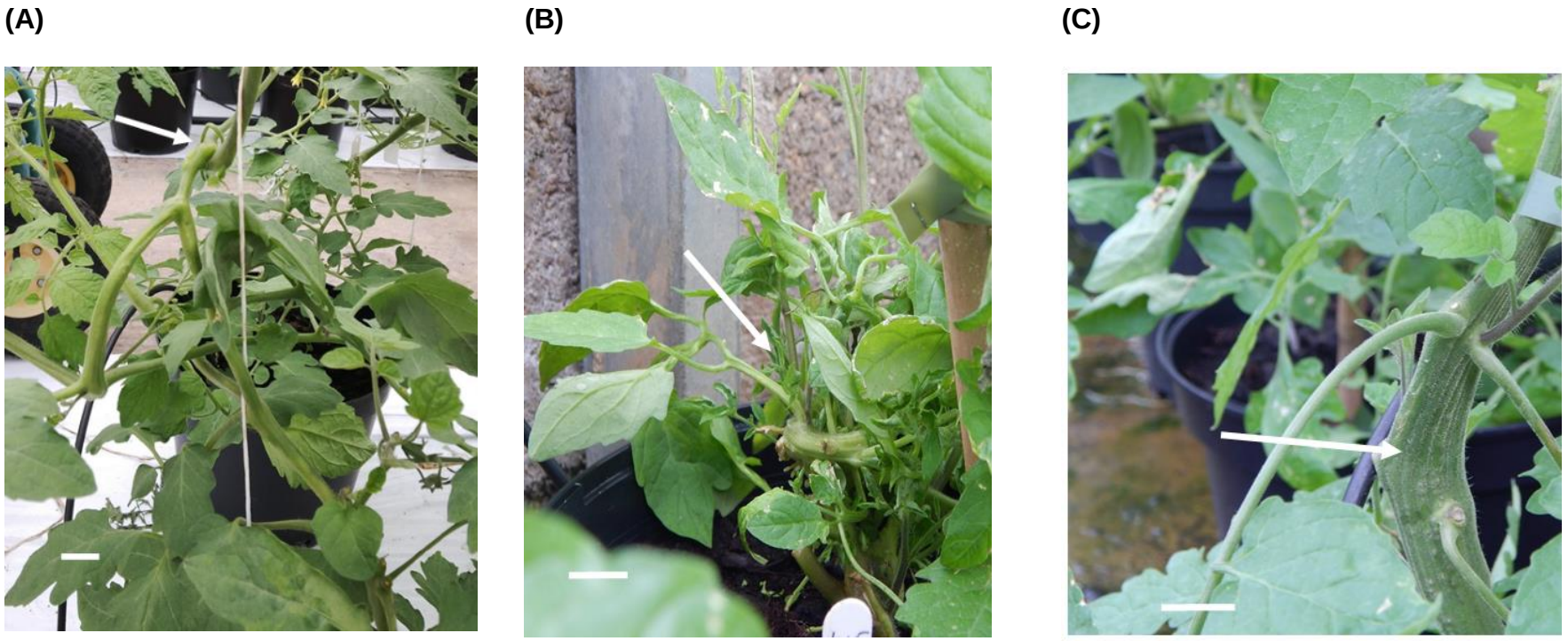


Figure 2. 4 Changes in stem phenotypes in certain epiRILs. Abnormal twisted stems in lines 4.15 (A and B) and flat stem in lines (C) 2.1 in the F_4 generation are indicated by arrows. Scale bar = 1 cm

Table 2. 3 Lines showing abnormal stems in several generations of epiRILs

Generation	Lines showing twisted and / or fasciated stems
F_2	4.1, 4.15
F_3	4.15, 4.5
F_4	4.5, 4.15, 4.35, 4.37, 2.123, 4.310,4.312, 4.1132, 2.1,329
F_5	33, 2.1, 2.18
F_6	33, 2.1, 2.18
F_7	3.3,4.69,4.1012,2.18

Fasciation is a phenomenon where a plant organ becomes flat or ribbon-like, in contrast to the normal cylindrical shape. Fasciation in plants has been found in several plant species and has been associated with pathogen infection, mutations, wounding or hormone application (Fambrini *et al*., 2006). In this study the phenotype in the epiRILs was usually accompanied by twisted stems and the plants were often shorter than wild type.

The change in the epiRILs stem shape suggests possible changes in the expression of genes controlling shoot apical meristem (SAM) formation. This meristem has several functions in plant development, which includes initiation of tissues and organs. The SAM has several zones, each with different roles in producing new organs with cell layers undergoing growth different directions. Any changes in the development and maintenance of zones and layers in SAM can cause changes in the phenotype of a plant. For example, an increase in SAM size can cause a loss of the normal arrangement of organ primordia and flattening of organs (fasciation) can occur (Fambrini *et al*., 2006).

Abnormal leaflets and leaf chimera

Changes in the SAM can affect the development of organs such as the leaves, which develop from the flank of the SAM (Busch *et al.*, 2011). In the epiRILs, some leaves showed reduced numbers of leaflets and curled margins. There were also changes in leaf colour and variegation (Figure 2.5).

Leaves with abnormal leaflets were first observed in generation F_4. In later generations, more lines showed this phenotype. The lines showing abnormal leaves were often not consistent in each generation (Table 2.3). However, line 3.3 showed abnormal leaves at F_5 and in the later generations.

Changes in the leaflet formation indicates alteration in the expression of genes regulating leaf development. The *KNOTTED1-LIKE HOMEOBOX (KNOXI)* gene encodes a transcription factor that is a master regulator of leaf development. In tomato, the *KNOXI* genes are called Tomato *KNOTTED 1* (*Tkn1*) and *Tkn2*. These genes are expressed in the SAM and in young leaves. Overexpression of *Tkn1* and *Tkn2* in young leaves can result in small, narrow and simple leaves (Jasinski *et al*., 2007; Shani *et al*., 2009; Busch *et al.*, 2011).

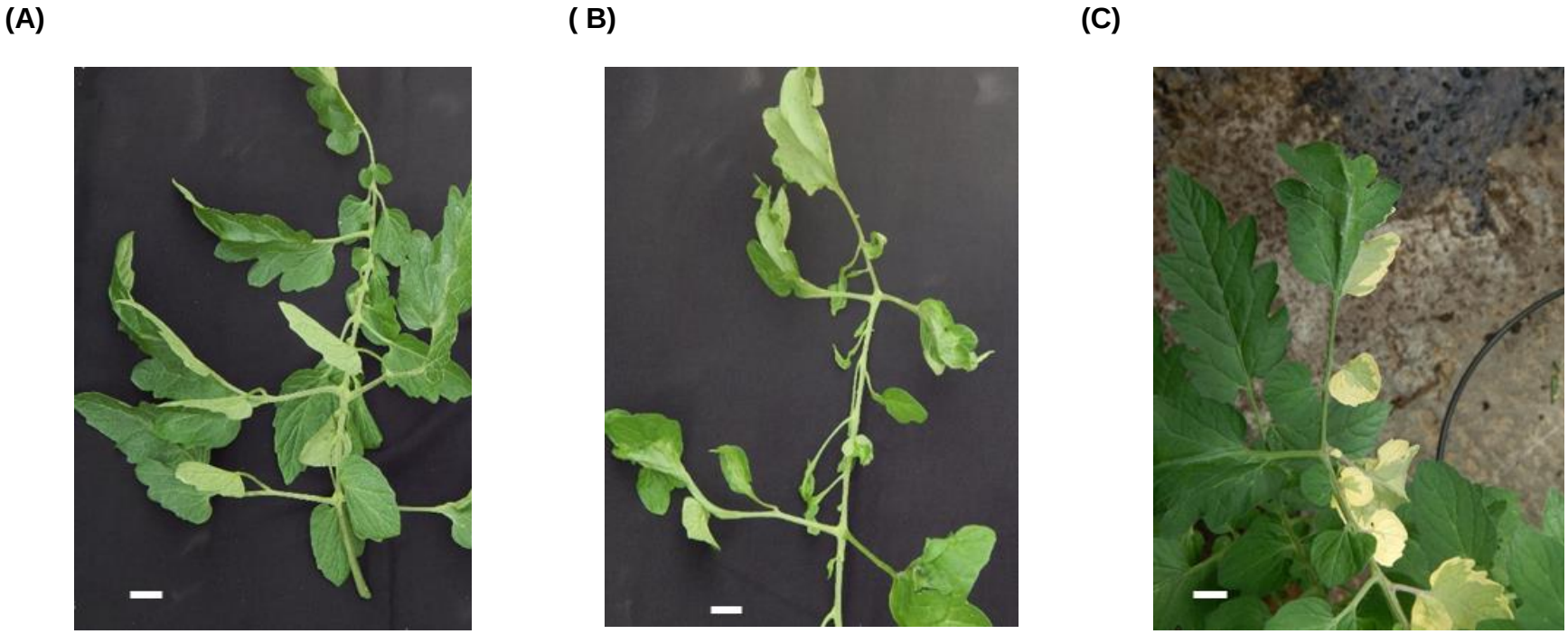


Figure 2. 5 Abnormal leaflet shapes in the epiRILs. (A) wild type leaf, (B) leaf from line 4.1017 and (C) variation in leaf colour of line 2.124. Pictures are from F_4 plants. Scale bars = 1 cm

Table 2.4 Lines showing abnormal leaves

in several generations of epiRIls

epiRIL Generation	Lines showing abnormal leaflets	Lines showing variation in leaf colour (chimera)
F_4	410.17, 4.5	2.124, 4.314
F_5	3.3, 4.5	2.11, 2.123
F_6	3.3,2.1,4.15, 4.35	2.123
F_7	41.11, 4.312, 4.1019, 3.3,4.15,4.2,4.94,	3.2, 4.33, 4.612,4.1019

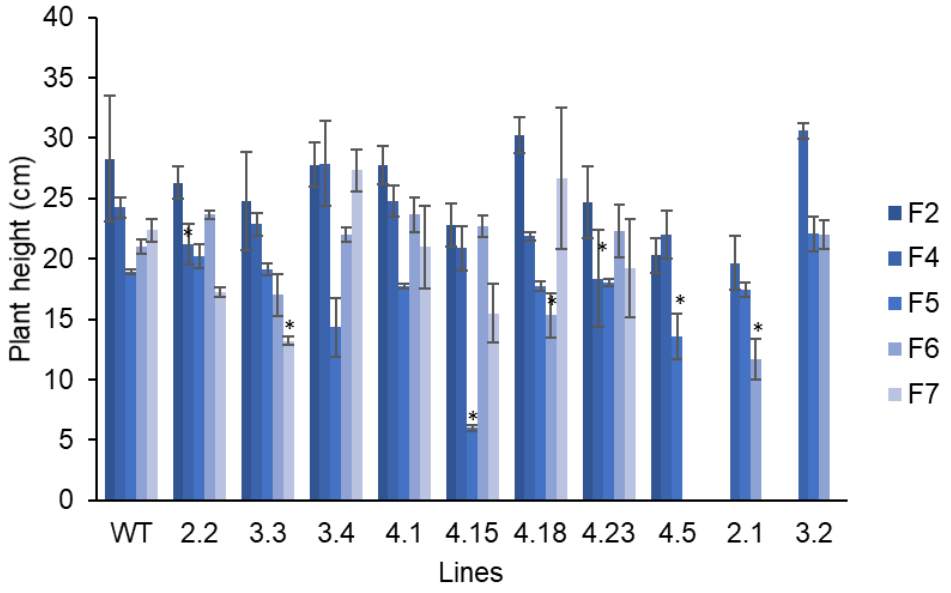
Plant Height

Plant height was lower in some epiRILs at 4-5 weeks after sowing. This trait showed fluctuation between each generation (Figure 2.6a). However, significant ($P<0.05$) differences were seen for some lines compared to wild type. Combining data from five generations of epiRILs indicated that mean plant height was significantly ($P<0.05$) reduced in several epiRILs in comparison to the wild type by F_7 (Figure 2.6b).

The shorter epiRILs often had abnormal twisted stems and altered leaf shapes (Figure 2.7) including line 4.5 in generation F_5. This line did not produce viable seeds in F_5 so was not included in the later generations. In generation F_7, line 3.3 showed a significantly ($P <0.05$) lower plant height compared to wild type and had abnormal stem, leaf and flower features.

Differences in plant height can be caused by both genetic and environmental factors. Environmental factors such as light, water and nutrition can affect plant height. The role of genetics (and epigenetics) in the growth of a plant could result from alterations in various genes and pathways. The tomato *procera* (*pro*) mutant shows suppression of axillary bud development and increased plant height compared to wild type (Bassel *et al.*, 2008). This phenotype is similar to a plant with constitutive response to gibberellic acid (GA), a plant hormone that stimulates cell elongation.

(A)



(B)

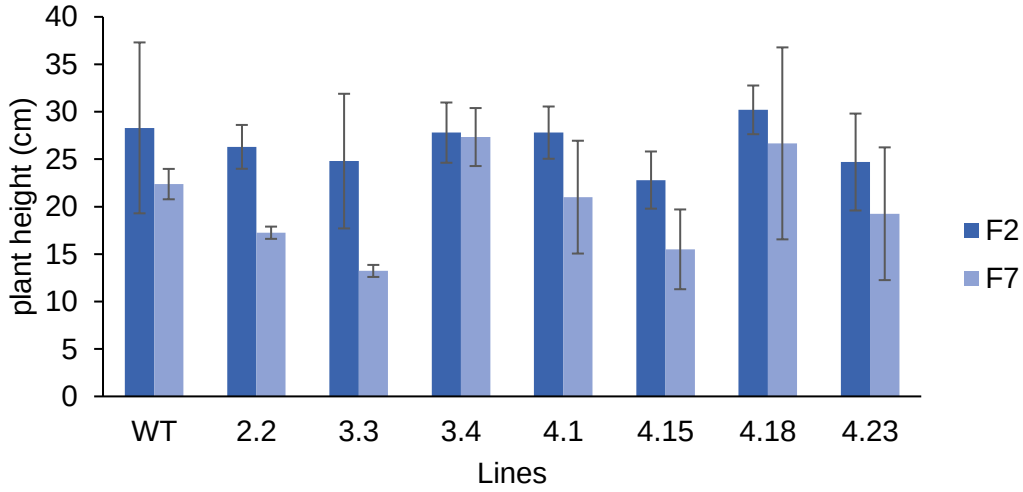


Figure 2. 6 Plant height of selected epiRILs at 4 weeks after sowing. (A) Average plant height for 10 epiRILs and wild type in generations F_2 until F_7. (B) Difference in plant height between generations F_2 and F_7. Error bars = standard error. Asterisks (*) show significant difference ($P<0.05$) to the wild type. Plant height value is the mean of three plants/biological replicates.

(A) **(B)**

Figure 2. 7 Difference in plant height in epiRIL 4.5. (A) Comparison of wild type (left) and line 4.5 (right) at 5 weeks after sowing and (B) plant 4.5 showing twisted stem and abnormal leaves. Pictures of plants taken at generation F_5. Scale bar = 5cm

2.3.3 Phenotyping reproductive development of the epiRILs

Abnormal flowers

Abnormal flowers were observed in the epiRILs in generations F_2 until F_7. Flowers with twisted stamens showed an open anther cone and exerted style (Figure 2.8). The lines showing this abnormal trait mostly originated from the T_1 lines 4.5, 4.1, 4.23, 2.1, 2.2, 4.6 and 3.3 (Table 2.4).

Changes in the stamen and style of tomato flowers can result from environmental factors such as temperature and heat stress (Giorno *et al*., 2013; Muller *et al*., 2016). However, as this trait was seen in several generations of the epiRILs, and in some cases maintained in certain lines, it suggests that abnormal stamen phenotypes could be caused as a result of reduced methylation across the genome due to silencing of *MET1*. This trait was found consistently in lines 4.5 and 2.1 in generations F_3 until F_6. These two lines eventually produced no fruits or seedless fruits. In line 4.69, abnormal stamens were seen from generations F_4 until F_7, but this line still produced some fruits with viable seeds.

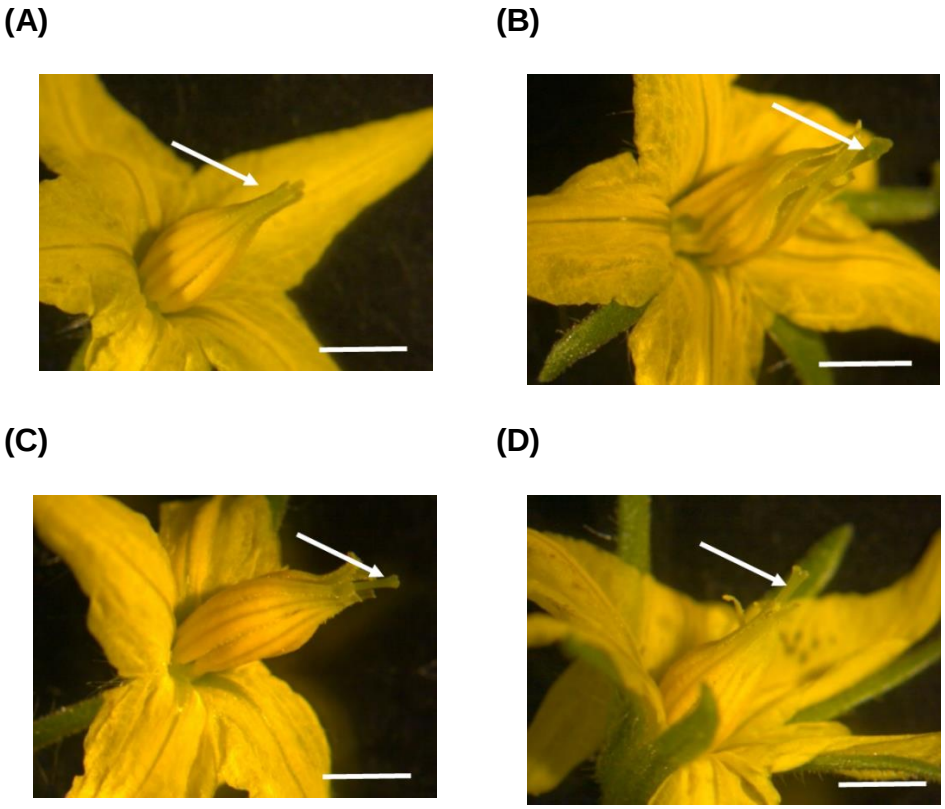


Figure 2. 8 Altered flower phenotypes. Flowers were from (A) wild type, (B) 4.5, (C) 2.1 and (D) 4.69. Arrows point to a normal stamen cone in wild type and exerted style in the other lines. Scale bar = 5mm

Table 2. 5 Lines showing abnormal flowers

in several generations of epiRIIs

Generation	Lines showing twisted anther or exerted style
F_2	4.5,4.15,4.23,2.2
F_3	4.5,4.15,4.23,2.2,2.1
F_4	2.18,4.5,4.15,4.35,4.69,2.17,4.37,2.117,4.611,2.1
F_5	2.2,3.3,3.4,4.1,4.23,4.5,2.17,2.18,2.117,2.123,2.127,4.69,2.1
F_6	2.2,3.3,4.1,2.17,2.18,2.117,2.1,4.69,4.33
F_7	3.3,2.18, 4.69

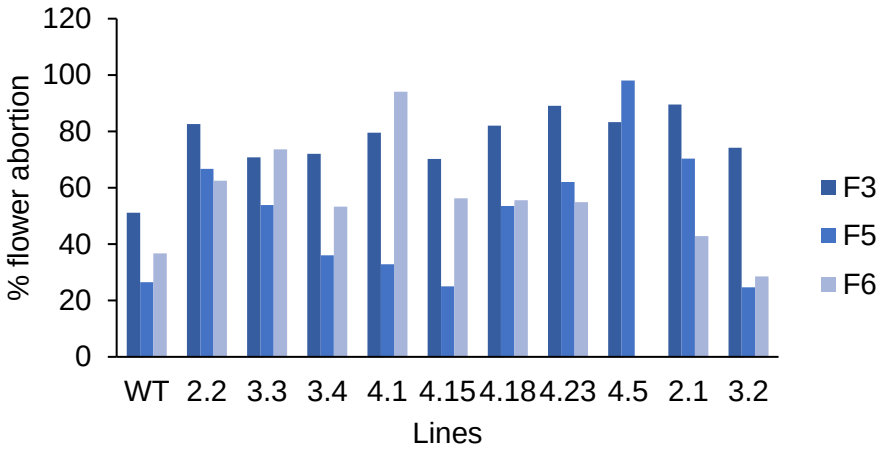
The abnormal stamen cone resembles the phenotype of the *SlEZ2* and *SlEZ1* knockdown plants (How Kit *et al*., 2010). The *SLEZ* genes encode the Enhancer of Zeste (E(z)) Polycomb group (PcG) proteins which are involved in the regulation of tomato fruit and plant development. Another gene *TM8*, which encodes a MADS-box transcription factor and is expressed in tomato flowers can produce abnormal flowers when mis-expressed (Daminato *et al*., 2014). Plants overexpressing *TM8* produced flowers with splayed stamens similar to those observed in the epiRILs.

In many lines, not all of the flowers in one plant were abnormal. For example, in generation F_5, most of the flowers from lines 4.15 and 4.5 were abnormal and in the case of line 4.15, produced less than ten fruits and only two fruits having reached ripe stage at 4 months after planting.

Most of the lines still produced fruits even though flowers often showed a high abortion rate (Figure 2.9). Some lines were also lost because no fruits were formed. The flowers of these lines showed twisted stamens and even with manual pollination failed to produce any fruits. In generation F_7, the lines that failed to produce fruits or had fruits with no seeds were: 4.918, 4.66, 4.615, 4.5, 2.1 and 2.123 while lines 3.3, 4.69, 2.17, 2.18 and 2.117 still produced fruits with viable seeds. Thus only 111 lines were propagated further in F_8.

With some lines showing a higher flower abortion rate than others, the number and viability of pollen was investigated to determine if pollen viability was linked to flower abortion.

(A)



(B)

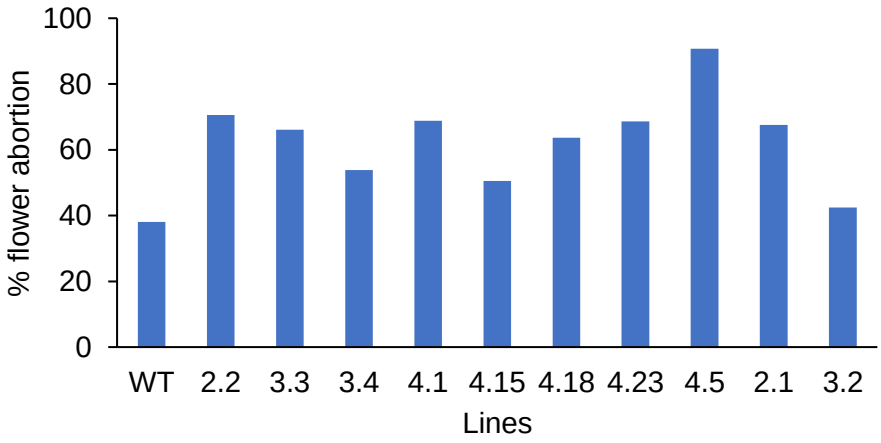


Figure 2. 9 Percentage of flower abortion. (A) The percentage of flowers failing to produce fruits in three generations and (B) is the average percentage of flower abortion from the three generations. Error bars indicate standard error. WT = wild type. Flower abortion was calculated from the mean of two plants per line.

Pollen viability

Pollen viability was investigated in lines showing abnormal flowers and high percentage of fruit abortion. The pollen was taken from generation F_5 because the flower abortion rate was still higher in the epiRILs. Additionally, in this generation, more than 90% of the population was homozygous (Reinders *et al*., 2009), so the abnormality in the reproductive organs was likely to be caused by epigenetic changes and not as the result of the environment. The iodine potassium iodide (IKI) method was used to evaluate pollen viability (Chang *et al*., 2014 and Abejide *et al*., 2014). Lines with abnormal flowers still contained viable pollen, shown by the dark staining with the iodide solution (Figure 2.10) except for line 4.5 in generation F_5. In this line, most of the pollen grains were only lightly stained compared to wild type indicating changes in pollen viability. Also the amount of pollen per line showed was variable (Table 2.5). Line 4.5 did not produce any seeds in generation F_5.

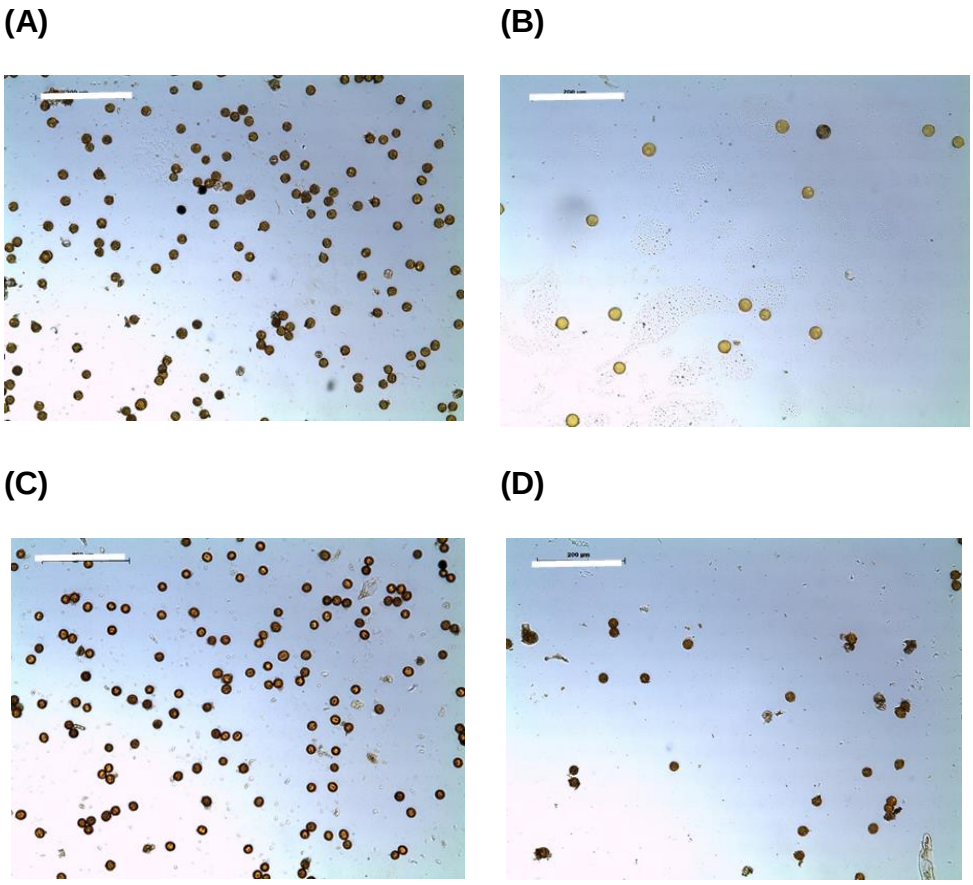


Figure 2. 10 Pollen grains from flowers of wild type and epiRILs in generation F_5. Pollen was extracted from (A) wild type, (B) Line 4.5 (C) line 2.17 and (D) line 2.1. The flowers of line 2.17 showed twisted stamens, but the pollen grains were viable and similar in number to wild type. Scale bars = 200µm

One cause of reduced pollen numbers could be differences in stamen cone morphology, e.g. closed vs open allowing pollen to escape more readily in the open form. However, observations on closed and open flowers of lines 3.3 and 4.5 (Appendix 16) showed that the amount of pollen obtained from closed flowers was not substantially higher than that from opened flowers. Further study is needed to investigate the cause of high flower abortion in some of the epiRILs.

Table 2. 6 Numbers of viable pollen in wild type and selected epiRILs in generation F_5

Lines	Number of pollen grains	Number of viable pollen grains	Percentage of viable pollens (%)
WT	411	407	99
2.2	588	568	96.6
3.3	604	564	93.4
3.4	329	266	80.9
4.1	445	423	95.1
4.15	315	281	89.2
4.18	511	468	91.6
4.23	262	241	91.98
4.5	174	5	2.8
2.1	270	254	94.1
3.2	420	406	96.7

Note: Pollen numbers were taken from three randomly selected areas of the microscope slide from four flowers per line.
WT = wild type.

2.3.4 Phenotyping fruit traits

Total Soluble Solids (TSS)

Tomato flavour is determined by the level of sugar, acid and volatile compounds in the fruits (Georgelis and Scott, 2004). Sugar levels are generally determined as total soluble solids (TSS) (Magwaza and Opara, 2015) and TSS is usually measured as degrees Brix (^{o}Brix) using a refractometer. Brix levels in tomato usually range from 4-6, but can reach 12 or 16 in some varieties (Stommel and Haynes, 1993; Harrill, 1998). The fruits from selected epiRILs and wild type showed values in the range 4.67-6.24 ^{o}Brix in generations F_2, F_3 and F_4 (Figure 2.17).

In the F_2 generation only line 4.23 showed a significant ($P<0.05$) difference in the degree Brix compared to the wild type. The mean TSS of the wild type was 6.22 while the mean for line 4.23 was 5.60. In generation F_3, there were significant differences ($P<0.05$) between wild type and lines 4.23, 4.15 and 3.4 while in F_4 between wild type and lines 4.23 and 3.4. In both generations values were lower than wild type (Figure 2.11).

The differences in TSS between wild type and line 4.2 was significant ($P < 0.05$) across three generations (Appendix18). This suggests a stable change in line 4.23 that was affecting the levels of TSS in the fruit. However, as there was limited time to conduct the investigation, the measurement of TSS was not undertaken after generation F_4 and the levels of TSS in fruits from 4.23 should be investigated again once the lines are fixed and replicate samples are available.

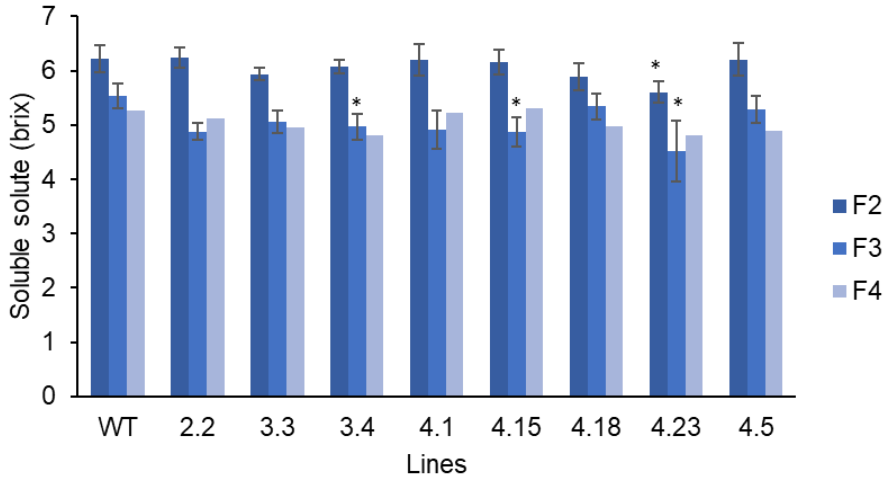


Figure 2. 11 Total soluble solids at fruit stage Breaker + 7 days. Data from generations F_2, F_3 and F_4. Error bars = standard error. Asterisks (*) show significant difference ($P<0.05$) to the wild type. WT = wild type. In the F_2, n = 3, in F_3, n = 5 and in F_4, n = 2. The s.e.d values are presented in Appendix 22.

Fruit Colour Index

Colour index (CI) was measured in three fruits from each plant of selected epiRILs and wild type. The CI was then determined according to Camelo and Gomez (2004). The results indicated that CI was generally lower in the epiRILs than in the wild type (Figure 2.12). However, statistical analysis showed that there were no significant ($P<0.05$) differences in the fruit CI between the epiRILs and the wild type and amongst the epiRILs at the F_2 generation.

At F_3, only line 4.1 showed a significantly ($P<0.05$) lower CI compared to the wild type. In F_4, there was a significant ($p<0.05$) difference in fruit CI between wild type and lines 4.15, 4.5, 4.1 and 2.2. However, we must be cautious in concluding that this difference is due to any genetic or epigenetic changes. Environmental factors such as temperature and light during the ripening process can affect CI (Camelo and Gomez 2004). Further analysis is required when the epiRILs have been fixed and can be grown as replicated lines. This should involve an assessment of the metabolites that are involved in fruit pigmentation. This should include determination of carotenoid and flavonoid levels (Arias *et al*., 2000; Adato *et al*., 2009).

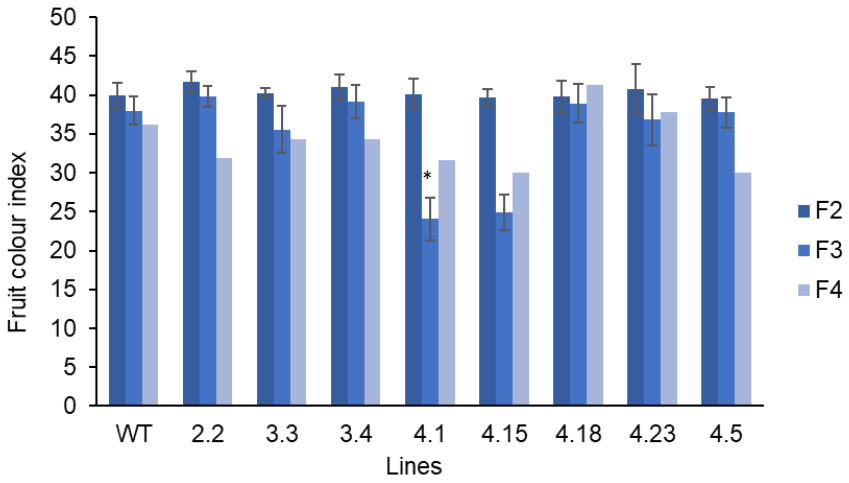


Figure 2. 12. Fruit colour index at stage Breaker + 7 days. Data from generations F_2, F_3 and F_4. Error bars = standard error. Asterisks (*) show significant difference ($P<0.05$) to the wild type. WT = wild type. In the F_2, n = 3, in F_3, n = 5 and in F_4, n = 2. The s.e.d values are presented in Appendix 22.

Fruit weight

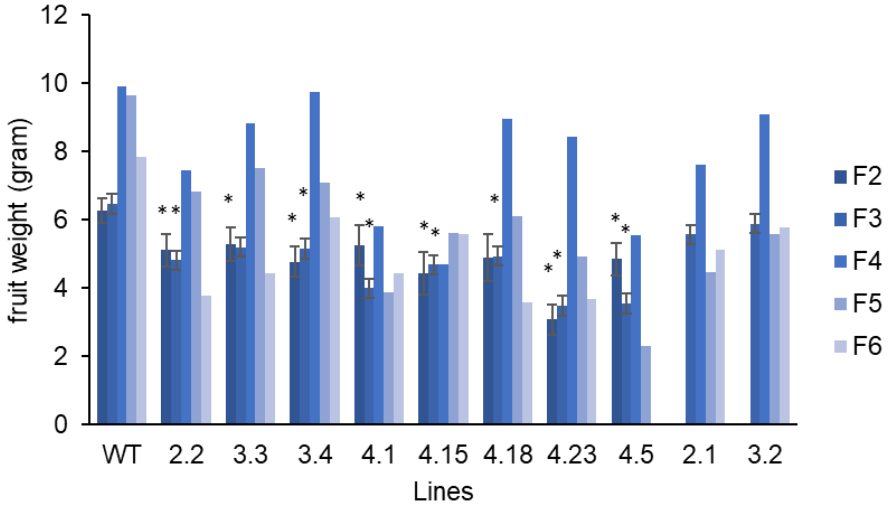
Fruit weight was determined in several generations of selected epiRILs. Variation in the fruit weight of ten lines was compared to the wild type from F_2 until F_6 (Figure 2.19). There was a fluctuation in fruit weight between each generation, which could be the result of environmental factors including the season when the plants were grown. The average weight of the ten lines from all the generations showed that the epiRILs generally had lower fruit weight compared to wild type.

Due to the effect of environment on the epigenetic information epigenome of an organism (Gallusci *et al*., 2017), we must consider that the difference in fruit weight between generations could be caused by environmental factors such as light and temperature. The different generations were planted in different seasons (Table 2.1). Data from the metoffice.com (Appendix 2) showed the difference in mean temperature between each season in 2015-2019. However, it must also be taken into account that the glasshouse temperature and light was also maintained by artificial light and heating during winter (Appendix 3). Differences between each generation such as a higher mean fruit weight in F_4 could therefore be caused by a difference in an environment factor compared to other generations. However, statistical analysis of the fruit weight data revealed significant differences ($P<0.05$) to wild type across several generations which was planted in different seasons.

In lines 2.2 and 4.15 the fruit weight was always significantly ($P<0.05$) lower than wild type in each generation. While for other lines such as 4.18 and 4.23, the average fruit weight was only significantly

(P<0.05) lower than wild type in some of the generations (Figure 2.13). This suggests that even though there may be changes in fruit weight due to alterations in DNA methylation and its impact on gene expression, environmental factors are still having an effect. This is likely due to the complex nature of fruit weight as a trait. It is known that there are at least six loci that are involved in the regulation of tomato fruit size (Tanksley, 2004; Cong *et al*., 2002) which shows the complexity of fruit size regulation. Images of typical fruits from the control and a range of epiRIL lines are shown in Figure 2.14.

(A)



(B)

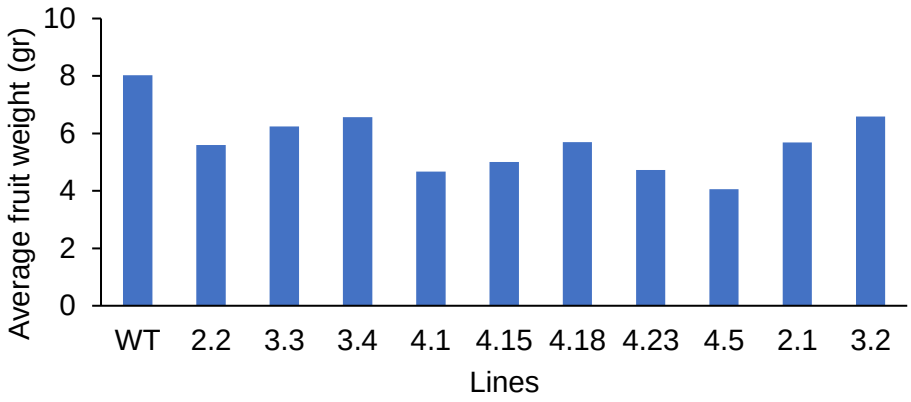


Figure 2. 13 Fruit weight at breaker+7 of several epiRILs. (A) Fruit weight of selected epiRILs and wild type fruits in generations F_2 until F_6 (B) The average fruit weight from all five generations. Error bars = standard error. Asterisks (*) show significant difference ($P<0.05$) to the wild type. WT = wild type. In the F_2, n = 3, in F_3, n = 5 and in F_4, n = 2. The s.e.d values are presented in Appendix 22.

Figure 2. 14 Comparison of fruit size from selected F_3 lines compared to the wild type. The fruits from several epiRILs (lines 4.5, 4.1, 2.2, 4.23, 4.15 and 4.18) were smaller compared to the wild type. This was confirmed using ANOVA and DMRT statistical tests (Appendix 17-20).

Genetic studies have shown that fruit size is a quantitative character and is regulated by several loci. Marker assisted QTL mapping and cloning have demonstrated a range of loci are involved in the regulation of fruit size in tomato. Six major QTLs *fw1.1, fw2.1, fw2.2, fw3.1, fw3.2,* and *fw11.3* have been mapped in several studies. They are referred as major QTL and account for a major portion of the fruit weight variation in tomato (Grandillo and Tanksley, 1999).

One of the QTL, *fw2.2*, was found to be responsible for 30% and 47% of the total phenotypic variance in the *L. pimpinellifolium* and *L. pennellii* populations, respectively, indicating that this is a major QTL controlling fruit weight. This QTL also maps to the same position in chromosome 2 in *L. pimpinellifolium* and *L. pennellii* (Alpert *et al.,* 1995).

The gene at the *fw2.2* locus has been cloned (Frary *et al*., 2000). and shown to be a plant-specific protein and regulates cell division in the fruit (Cong and Tanksley, 2006). The data indicated that there was a negative correlation between the expression of the *fw2.2* and fruit size (Frary *et al*., 2000). Investigation on the *fw2.2-like* (*PbFWL*) genes in pear also suggested that the genes might be negatively related to the cell division in pear fruit (Jia *et al*., 2016).

The data from the studies of the epiRILs indicates that changes in DNA methylation (hypomethylation) may be directly or indirectly affecting (increasing) the expression of the gene at the *fw2.2* locus. We describe how we have tested this hypothesis in Chapters 3 and 4. Chapter 3 presents the result from gene expression analysis and

Chapter 4 discuss enzyme analysis which investigates methylation status of selected genes.

Fruit trichomes

In generation F_5, fruit from some epiRILs showed a difference not only on the size, but also an abnormal fruit surface. This trait was again seen in generation F_6 and could be seen as a fixed trait in the epiRIL population. Two lines in the population showed trichomes on the fruit epidermis which were not present in wild type. In the wild type, the fruit surface is smooth, but in fruits from lines 2.1 and 4.69 the surface was sticky and when observed by microscope, showed a higher density of trichomes (Figure 2.15).

The lines with fruit trichomes also had smaller fruit, more locules and produced fewer seeds (Figure 2.16). The lines showing this abnormal trait were also lines showing abnormal flowers. Line 2.1 produced only two fruits in generation $F_{6,}$ but unfortunately did not produce seeds. Line 4.69 produced several fruits and some viable seeds.

The trichome phenotype was similar to fruits found in plants with reduced expression of the *SlEZ2* gene (Boureau *et al*., 2016) The *SLEZ2* gene is one of the genes involved in the production of the Enhancer of Zeste (E(z)) Polycomb group (PcG) proteins which regulate tomato fruit and plant development.The repression of this gene in genetically modified plants has been shown not only change leaf and flower phenotype, but also to produce fruits that were sticky due to a higher

density of trichomes. Furthermore, changes in the phenotype of several tomato epiRILs including subnormal anther cones were similar to those in transgenic lines with abnormal expression of the *SLEZ1* and *SlEZ2* genes (How Kit *et al*., 2010; Boureau *et al*., 2016)

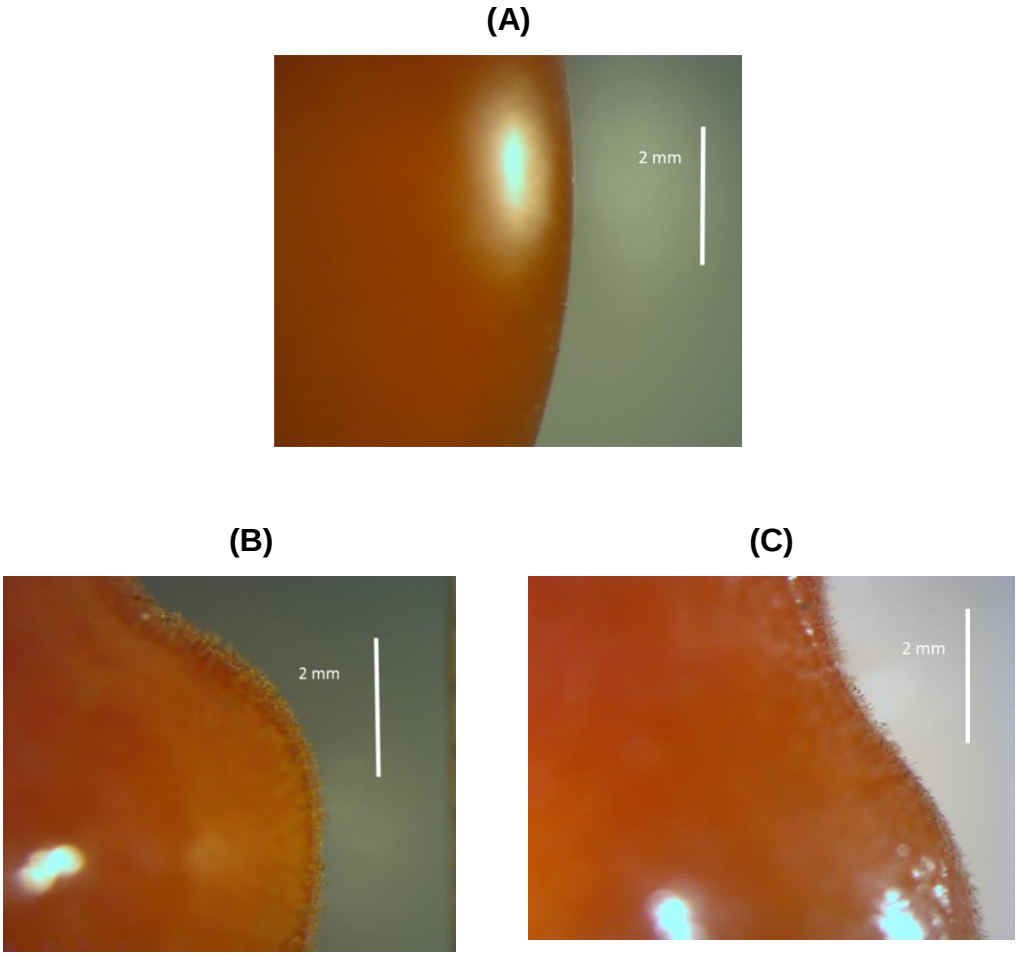


Figure 2. 15 Tomato fruit surface showing difference in trichome density. (A) Smooth fruit surface of wild type fruit. Trichomes on fruit surface of lines 4.69 (B) and 2.1(C) in generation F_6. Scale bar =2mm

(A)

(B)

Figure 2. 16 Altered fruit morphology. (A) Fruits of wild type and line 4.69. (B) Whole and cut fruits showing locule number and the seeds. Scale bar = 1cm

2.4 Conclusion

Differences in phenotype between the epiRILs and the wild type indicated a range of effects of the *MET1* knockdown on the growth and development in tomato. Changes in the stem and leaf shape might result from alterations in the *KNOX1* gene expression or other genes involved with the regulation of the stem apical meristem.

Changes in flower morphology also showed that a reduced in *MET1* expression affected the plant phenotype in several generations of the epiRILs. The abnormal flowers, fruits and some of the change in leaf shape are similar to those of the *EZ1* and *EZ2* knockdown plants and *TM8* mutants Effects on fruit weight in the epiRILs may be the result of hypomethylation of genes such as *FW2.2.* The molecular basis of these epiRIL phenotypes were investigated further by quantitative PCR and RNA sequencing and this work is described in Chapter 3

CHAPTER 3 INVESTIGATION ON THE GENE EXPRESSION OF THE epiRILs

3.1 Introduction

Changes in DNA methylation are known to affect gene expression (Zhang *et al*., 2018). In plants, DNA methylation can occur in promoter regions or gene bodies (Zhang *et al*., 2006). The effects of DNA methylation on transcription are mediated through a number of mechanisms including inhibition of the binding of transcription activators and indirectly by promoting repressive histone modifications (Zhang *et al*., 2018).

In this Chapter the experiments undertaken explored the effects of the *MET1* silencing on gene expression in selected epiRILs. Candidate genes likely to be involved in phenotypic changes in fruit and flower characteristics were nominated and their expression in wild type and the selected epiRILs was explored. There are reports on the epigenetic effects in the regulation of the selected genes (Gallusci *et al*., unpublished results; Liu *et al*., 2011) which makes them a good model in this study. The genes selected were (1) *FW2.2* (*FRUIT WEIGHT 2.2*) which is involved in the regulation of tomato fruit weight, *SlEZ1* and *SlEZ2* (*ENHANCER OF ZESTE*) involved in tomato plant and flower development and *TM8* which is involved in the regulation of tomato flower development. The expression of these genes was then investigated in several epiRILs in different generations.

Quantitative PCR (qPCR) was used to detect changes in the expression of specific genes. RNA sequencing was used to detect differentially expressed genes across the genome. RNA sequencing was also used to detect changes in expression of transposable elements (TEs). Movements of TEs have been found in Arabidopsis hypomethylated mutants and are likely to be partly responsible for changing plant phenotype (Mirouze and Vitte, 2014).

The hypothesis for this part of the project was that there would be an increased gene expression in some lines due to methylation changes in the promoter regions. Demethylation of TEs could cause their activation and result in increased expressions.

3.2 Materials and Methods

3.2.1 RNA extraction

The Spectrum Plant Total RNA Kit (Sigma®) was used for RNA extraction. The tissue samples (six week old leaves, flowers at anthesis or fruit at breaker +7) were first ground to a powder in liquid nitrogen. A maximum of 110mg sample was added to 2mL microtubes. Each sample was then mixed with 600µl of the lysis solution and 6 µL of 2-mercaptoethanol. The sample was vortexed immediately for at least 30 seconds and then incubated at 56 ^{0}C for 5 minutes before centrifuged at 13,000 g for 3 minutes. The supernatant was placed into a Filtration column (blue coloured ring) in a 2 mL collection tube. The column

containing the sample was centrifuged at 13,000 g for 1 minute and the clarified flow-through lysate was saved for the next step.

Binding Solution (750 µL) was then added to the lysate and this solution was mixed immediately by a brief vortex. From the mixture, 700 µL was placed into a binding Column (red coloured ring) seated in a 2mL collection tube. The column and sample were then centrifuged at 13,000 g for 1 minute to bind the RNA. The flow-through liquid was disposed and the collection tube tapped briefly on clean tissue to drain the residual liquid. The column was then returned to the collection tube and the remaining mixture (first lysate + Binding Solution) placed in the column and centrifuged. The flow-through was again disposed of and the tube dried briefly on tissue.

Wash Solution I was then added to the Binding Column and centrifuged at 13,000 rcf for 1 minute. For the removal of DNA, DNase I and DNase digestion buffer were used (Sigma®, Product Code 2816 and 1566). For each sample, 10µl of Dnase I and 70 µL of Dnase digestion buffer were needed. The solution was mixed by pipetting gently in a 2 mL microtube before dispensing into the centre of the filter in the Binding Column. The sample was then incubated at room temperature for 15 minutes. After incubation, 500 µL of Wash Solution I was added to the column and the sample was centrifuged at 13,000 rcf for 1 minute to remove the digested DNA. The flow-through liquid was removed and the tube dried briefly on clean tissue before placing it back into the same collection tube.

For the next step, it was ensured that the Wash Solution 2 concentrate had been diluted with ethanol before use. 500 µL of the diluted Wash Solution 2 was added into the column and centrifuged at 13,000 g for 30 seconds. The flow-through liquid was discarded and the tube dried briefly on clean tissue. This step was repeated and an additional centrifugation at 13,000 g for 1 minute was added to dry the column.

The column was removed and placed into a new collection tube and 50µLof Elution Solution was placed onto the centre of the filter inside the column. The tube was closed and left for 1 minute at room temperature. The sample was then centrifuged at 13,000 g for 1 minute to elute the RNA. The purified RNA sample was stored at -80 C.

3.2.2 cDNA Synthesis

Synthesis of cDNA was undertaken using the Protoscript II first Strand cDNA Synthesis Kit (New England Biolabs®). The RNA sample was first denatured by adding 2µL of oligo-dT primer [d(T)23VN] to every 1µg of RNA and adding nuclease free water up to a total reaction volume of 8µL. The sample was denatured at 65 ^{0}C for 5 minutes then centrifuged briefly and promptly on ice.

The denatured samples were then added with 10µL of ProtoScript II Reaction Mix (2X) and 2µL of ProtoScript II Enzyme Mix (10X). The sample was then incubated at 42^{0}C for 1 hour and inactivated by heating to 80^{0}C for 5 minutes. The cDNA can be directly used for analysis or kept in -20 ^{0}C.

3.2.3 Quantitative PCR Analysis

The samples used were from leaves, flowers and fruit tissues taken from two plants per line. There were three leaves/flowers/fruit used from each plant. Each plant represent segregating individuals in the early generations of the epiRILs so was not an actual biological replicates of each line. There were three technical replicates for each plant.

Quantitative PCR (QPCR) was undertaken to quantify single gene transcripts. The relative value of the amplified product in the epiRILs was compared to that of the wild type. The qPCR amplification used Luna® universal qPCR master mix (New England Biolab®).The reaction mix and PCR conditions are shown in Table 3.1. qPCR analysis was undertaken using the 480 II system lightcycler (Roche®).

Primers were designed using Primer3Plus software (Table 3.2). Primer sequences for the detection of *SlEZ1* and *SlEZ2* transcripts were from How-Kit *et al*. (2010). The *ELONGATION FACTOR 1-ALPHA* gene (*EF-1α*) was used as an internal control (Wang *et al*., 2018)

To calculate the relative quantity (RQ) of the gene expression, the 'delta-delta method' was used (Pfaffl, 2001). The relative expression of between each line with wild type was compared using the equation :

$$RQ = 2^{-[\Delta Ct\ epiRIL - \Delta Ct\ wild\ type\ ]} = 2^{-\Delta\Delta Ct}$$

Statistical analysis was not undertaken for the RQ data because biological replicates were not sufficient in the data (only two plants used as samples due to restricted glasshouse space). Regression analysis

was undertaken between the Relative Quantification (RQ) value of each gene from green fruit and flower tissues against fruit weight at B+7. A linear regression model was applied using the Genstat® software (19th edition).

Table 3. 1 Components and program used in the qPCR reactions

Component	Volume	Final concentration
Luna Universal qPCR Master Mix	5µL	1 x
upstream primer	1 µL	0.25 µM
downstream primer	1 µL	0.25 µM
cDNA	1.5 µL	40ng
Nuclease-free PCR-grade water	Up to 10µL	

Program Name	Cycles	Analysis Mode	Acquisition Mode	Ramp Rate (^{0}C/s)	Hold Time	Temp (^{0}C)
Pre-incubation	1	None	None	4.4	3 min	95
amplification	40	quantification	none	4.4	5 sec	95
			single	2.2	10 sec	60
Melting curve	1	Melting curves	none	4.4	5 sec	95
			none	2.2	1min	65
			Continuous (5 acquisition/sec)	0.11	-	97
Cooling	1	none	none	2.2	30sec	40

Table 3. 2 Primers for qPCR

Primer Name	Gene Target	Direction of replication	Sequence (5'-3')
ORFX1-F	*FW2.2*	Forward	ATGGGAATAGGGTGGCAAGC
ORFX1-R	*FW2.2*	Reverse	CCTGCATGATAAGGGGGCAT
EZ1-F (How-Kit *et al*., 2010)	*SlEZ1*	Forward	CCAATAATCTCATTGAGACAG
EZ1-R (How-Kit *et al*., 2010)	*SlEZ1*	Reverse	ACCAGACACCTTGTTCGGAC
EZ2-F (How-Kit *et al*., 2010)	*SlEZ2*	Forward	CCGAGGCATCTGGTACTACGAA
EZ2-R (How-Kit *et al*., 2010)	*SlEZ2*	Reverse	CTGGTGGTCGTCAATGATGATGAGTTG
TM8-F	*TM8*	Forward	GAGAATCAAACAAATAGGCAAGT
TM8-R	*TM8*	Reverse	CTTCAGCACATAGAATAGAGA
GA20ox2-F (Li *et al*., 2012)	*SlGA20ox2*	Forward	TTTCCATATTCTACCCTACAAG
GA20ox2-R (Li *et al*., 2012)	*SlGA20ox2*	Reverse	TCATCGCATTACAATACTCTT
EF1-F (Wang *et al*., 2018)	*EF1*	Forward	GACAAGAAGGACCCAACTGGTG
EF1-R (Wang *et al*., 2018)	*EF1*	Reverse	CAGAGTCTAGATAGCACACTCGATG

3.2.4 RNA sequencing

For RNA sequencing (RNA-Seq), the samples used were 6-week-old leaves RNA from the lines of wild type and two lines from generation F_7. Line 3.3 originated from the initial parental cross and line 4.69 from the backcross. RNA was extracted as described in sub section 3.2.1. Seven microgram of the RNA was sent to Polar Genomics (New York, USA) for sequencing.

The RNA-seq workflow consisted of several steps. (1) RNA-seq library construction; (2) sequencing and mapping; and (3) normalization and (4) statistical modeling to identify the differentially expressed genes (Li et al., 2017). Steps 1-3 were undertaken by Polar Genomics (New York, USA) with the construction of strand specific cDNA libraries with PolyA depletion of ribosomal RNA then pair end 150 bp sequencing on Illumina HiSeq2500.

The last step was undertaken by Dr Michael Wilson a Bioinformatician in the Future Food Beacon (University of Nottingham, UK). In the statistical analysis, DESeq2 software was used to analyse the FPKM (Fragment Per Kilobase per Million mapped fragment) of each line and obtain the value of log2foldchange and adjusted P values (Benjamini-Hochberg False Rate Discovery) for False Discovery Rate (FDR). Genes and transposable elements (TEs) that were differentially expressed with values at Padj<0.05 were used to indicate differences in expression between the epiRILs and wild type.

3.3 Results and Discussion

Investigations on gene expression in the tomato epiRILs were undertaken based on abnormal plant morphology observed in some lines.

3.3.1 Relative expression of *FW2.2 gene*

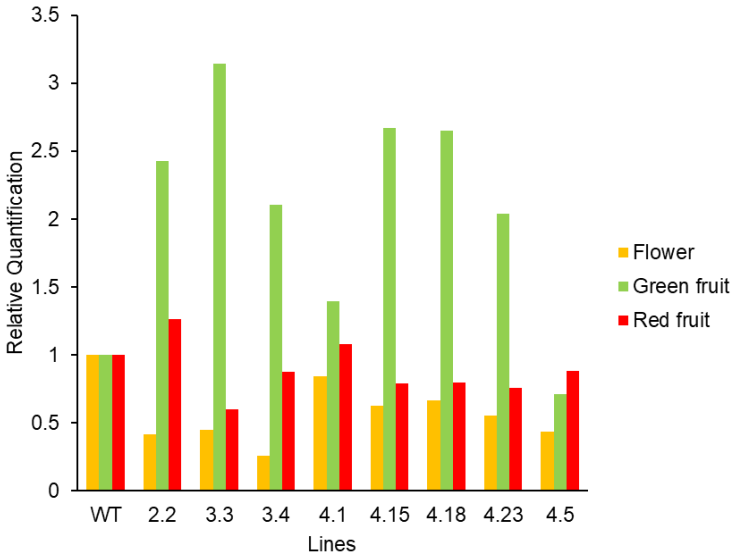
One of the quantitative trait loci (QTL) responsible for controlling fruit size in tomato is the *fw2.2* locus (Frary *et al*., 2000). In this investigation, the expression of the *FW2.2* gene was measured for three different generations of tomato epiRILs. A variety of lines showing smaller fruit were assessed for *FW2.2* expression at F_2, F_4 and F_6 using tissue from flowers, green and ripe fruit (Figure 3.1A).

Differences were apparent in the expression of the *FW2.2* in flowers, green fruits (20dpa) and red ripe fruits (stage breaker+7days) between wild type and the epiRIL lines. However, significant differences were found in most of the lines in the green stage. A further more detailed investigation was therefore undertaken using green fruits and expression between wild type and a range of epiRILs was compared across F_2, F_4 and F_6 generations. (Figure 3.1B). The relative expression of *FW2.2* gene in green fruit tissue of the seven epiRILs samples was higher than that of wild type in several lines, especially in generation F_4 where all of the sample lines showed a significant ($P<0.05$) difference to wild type.

In generation F_6, all of the sample lines showed the same or lower *FW2.2* expression than wild type except for line 3.3. Line 3.3 showed

abnormal morphology such as more side shoots during its early vegetative phase, abnormal flowers with twisted stamens and small fruits with few or no seeds. These abnormal phenotypes were present in Line 3.3 from generation F_4. However, due to its high flower abortion rate, this line did not produce seeds in generation F_6.

(A)



(B)

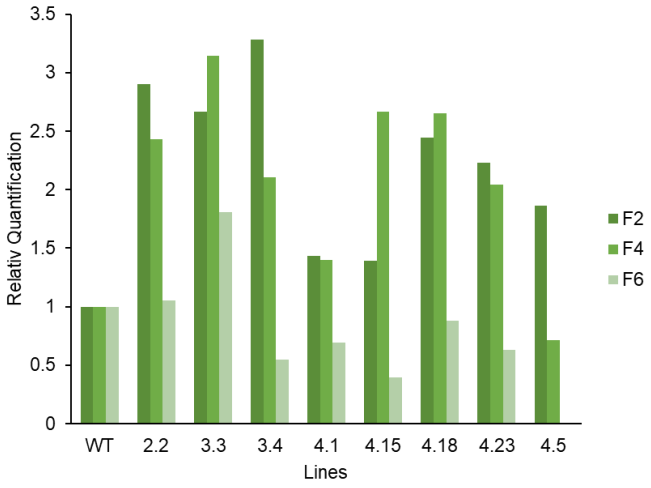


Figure 3. 1 Relative expression of *FW2.2* in epiRILs in comparison to wild type. (A) Relative expression of *FW2.2* in open flowers, green and B+7 fruit in generation F_4 and (B) relative expression of *FW2.2* in green fruits from generations F_2, F_4 and F_6. The data were analysed from two plants per line with three technical replicates for each plant. WT=wild type.

The data on *FW2.2* expression are consistent with the findings of Cong *et al.* (2002) where the expression at green fruit stage was higher than in red ripe fruit (Figure 3.1A). Cong *et al*. (2002) reported that the *FW2.2* transcript levels were significantly higher in small fruit varieties at 12-30 days post anthesis.

Previous investigations on this gene indicated a negative correlation between its expression and fruit weight (Frary *et al*., 2000; Cong *et al*., 2002; Jia *et al*., 2016). The transcript levels were found to be inversely correlated to the pericarp and placental cell mitotic activity in early fruit development. To determine if *FW2.2* expression and fruit weight were related across the selected epiRILs, an analysis was undertaken on the data from generations F_2, F_4 and F_6. There was no apparent correlation between the relative expression of *FW2.2* from green fruit stage and fruit weight at B+7 (Figure 3.2). The expression data from green fruit was used against fruit weight at B+7 because of the highest expression of the *FW2.2* gene was found to be highest at this stage (Figure 3.1A). The correlation was found not significant with $P>0.05$ ($P =0.053$).

With only nine lines used in the analysis, it does not represent all of the variation in the whole population of the epiRILs. However, it might also reflect the non-linear relationship between the expressions of *FW2.2* with tomato fruit weight. We know that the *FW2.2* locus is not the only one controlling fruit weight in tomato. Other fruit weight loci have also been described (Tanksley, 2004; Mu *et al*., 2017).

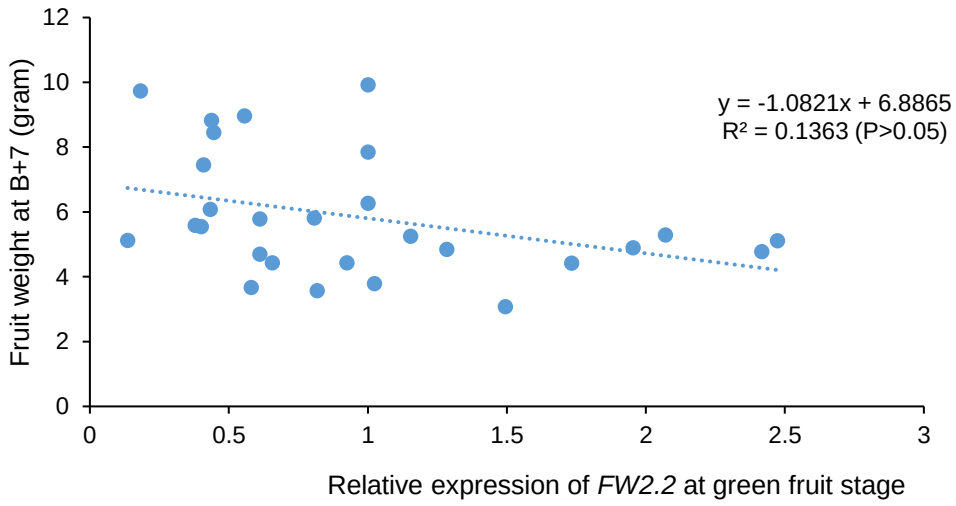


Figure 3. 2 Relationship between *FW2.2* expression and fruit weight. The equation shows the regression line equation and the correlation coefficient (R^2). The regression analysis showed no significant difference (P>0.05).

3.3.2 Relative expression of the *SLEZ1, SlEZ2, TM8* and *SlGA20ox2* genes

The analyses were undertaken on selected genes chosen for their roles in the vegetative and generative plant development.

The relative expression of *SLEZ1, SlEZ2* and *TM8* in tomato flowers and fruits

In several epiRILs, flowers showed abnormal phenotypes including twisted stamens and high flower abortion rates (Chapter 2). These traits are similar to transgenic tomato lines showing altered expression of *SlEZ1* and *SlEZ2* reported previously by How-Kit *et al*. (2010) and Boureau *et al.* (2016). Twisted stamens in tomato flowers were also seen in lines with over-expression of the MADS-BOX gene *TM8,* which is involved in the development of floral organs (Daminato *et al*., 2014).

Differences were apparent in the expression of *SlEZ1* and *SlEZ2* between several epiRILs and these showed a statistically significant ($P<0.05$) difference when compared to wild type (Figure 3.3) when tissues were examined at F_4. In the leaves, the expression of *SlEZ1* was lower in the epiRILs compared to wild type. While in the flowers, its expression increased up to 3.5 times compared to the wild type. The opposite was found for *SlEZ2,* where its expression was higher in leaves than in flowers.

SlEZ2 expression in the flowers was lower in the epiRIls compared to wild type and this was the opposite of the expression

pattern for *SlEZ1* and *TM8*. In the characterisation of *SlEZ1* RNAi lines by How-Kit *et al*. (2010), it was suggested that there might be some functional redundancy between *SlEZ1* and *SlEZ2*. Also that the *SlEZ1* might play a more specific role in flower development. This could explain the opposite level of expression between *SLEZ1* and *SlEZ2* in the tomato epiRILs in different tissues.

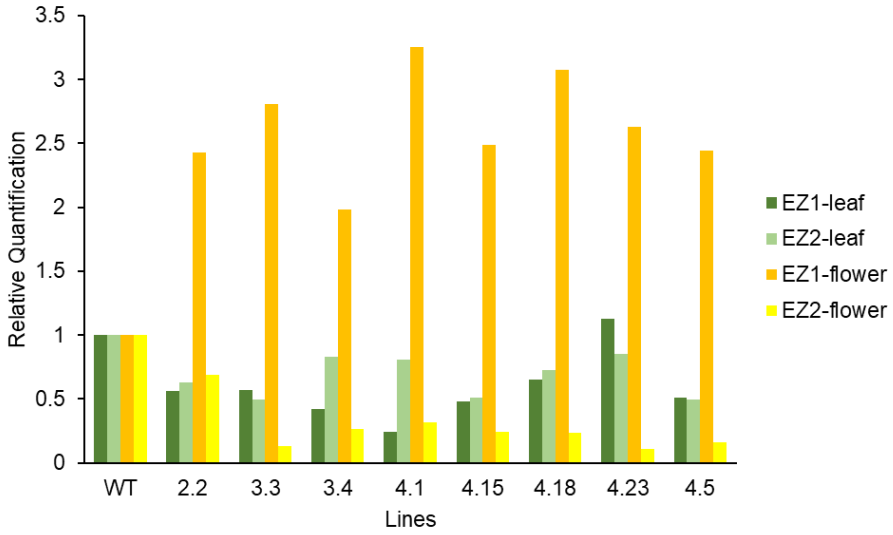


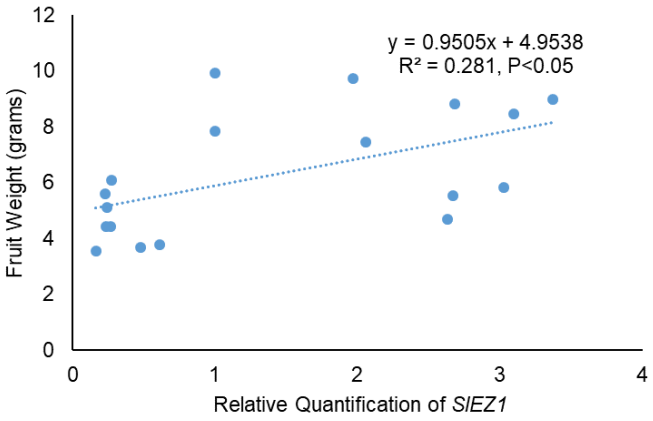
Figure 3. 3 Relative expression of *SlEZ1* and *SlEZ2* genes in generation F_4. The data were analysed from two plants per line with three technical replicates for each plant. WT=wild type.

Significant correlation was observed between each *SlEZ1*, *SlEZ2* and *TM8* expression and fruit weight (Figure 3.4). Inheritance of the altered expression of *SlEZ* and *TM8 genes* was examined by investigating their expression in the F_4 and F_6 epiRIL generations. The data indicated that in F_4 generation the expression of *SlEZ1* in the epiRILs was higher than in wild type (Figure 3.5A). However, by the F_6 generation the elevated levels of *SlEZ1* expression had declined in the majority of the lines (Figure 3.5A).

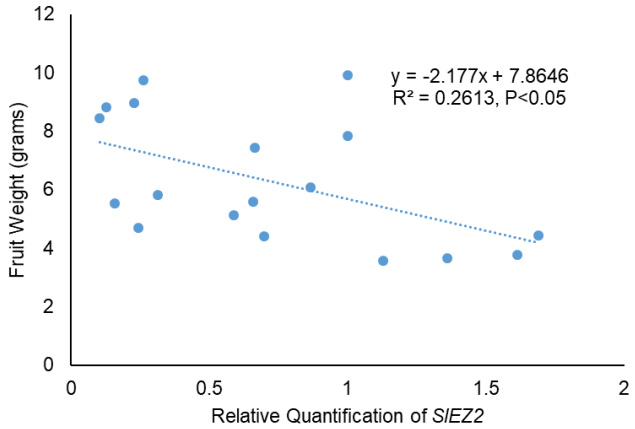
The opposite was found for the relative expression of *SlEZ2* which was lower in generation F_4 (Figure 3.5B) and increased in generation F_6. The relative expression of *TM8* was similar to that of *SlEZ1* with a higher expression in the epiRILs in generation F_4 (Figure 3.5C) and lower in generation F_6.

The *EZ* genes and *TM8* genes were found to have altered expression in the epiRILs compared to wild type during anthesis. Their role in altered flower development in the epiRILs is unclear, but *EZ* genes are part of the Polycomb Repressive Complex 2 (PRC2) which functions as a histone methyltransferase (Chanvivattana *et al*., 2004) and this complex can also affect flower development.

(A)



(B)



(C)

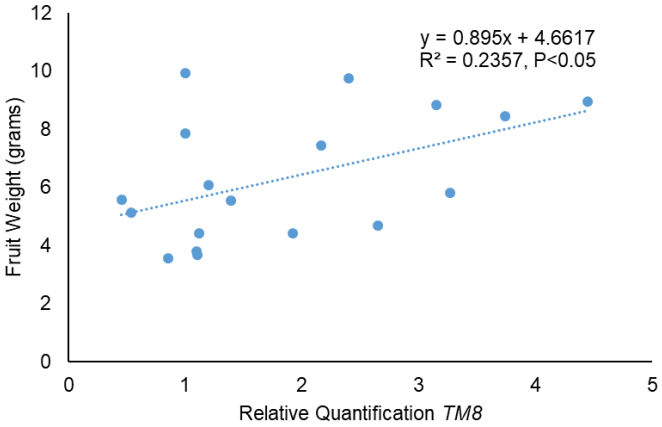
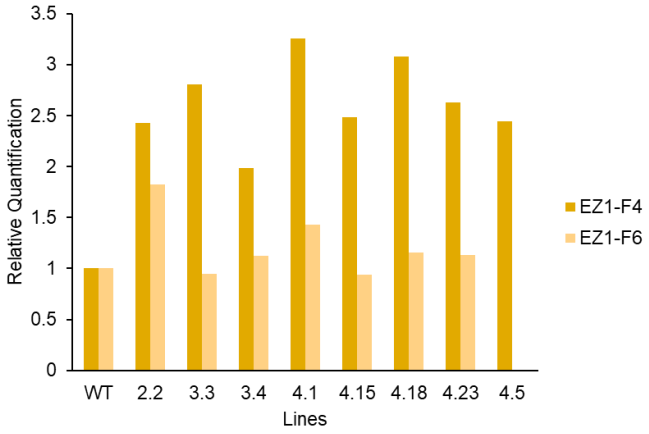
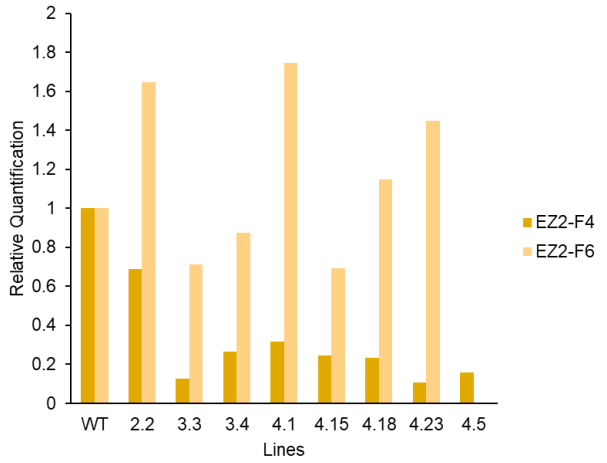


Figure 3. 4 Correlation between *SlEZ1, SlEZ2,* and *TM8* gene expression and fruit weight. (A) *SlEZ1*, (B) *SlEZ2* and (C) *TM8* in F_4 flowers. Equation showing the regression line equation and the correlation coefficient (R^2). Regression analysis showed significant difference ($P< 0.05$).

(A)



(B)



(C)

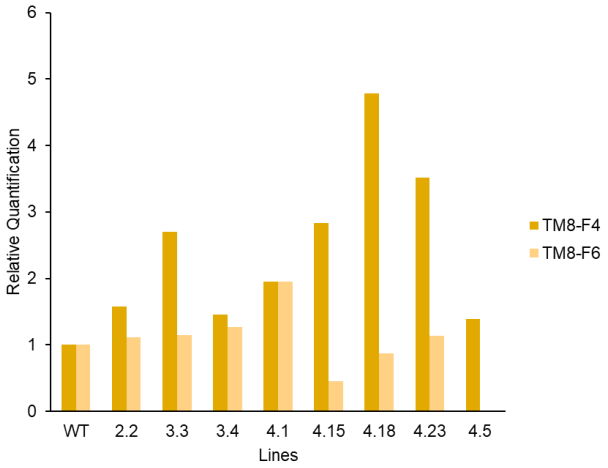


Figure 3. 5 *SlEZ* and *TM8* gene expression in flowers from epiRIL generations F_4 and F_6. Figures show results of (A) *SlEZ1*, (B) *SlEZ2* and (C) *TM8* expressions. The data were analysed from two plants per line with three technical replicates for each plant. WT=wild type.

The relative expression of *SLEZ1, SlEZ2* and *TM8* in tomato stem and leaf development

Some of the epiRILs also showed fasciated, twisted stems and abnormal leaves (Chapter 2). Similar phenotypes were also observed in *SlEZ2* silenced transgenic plants by earlier authors (Boureau *et al*., 2016). To investigate the expression of the *EZ* genes during the vegetative stage, *EZ* gene expression in leaves was examined for several epiRILs in two generations (Figure 3.6).

The expression of *SlEZ1* was lower than wild type in both generations with lower expression in generation F_6 than F_4 (Figure 3.6A). The expression of *SlEZ2* was also lower in the epiRILs. Some of the lines showed lower expression in generation F_6 (Figure 3.6B). A consistent abnormal leaf phenotype was seen in line 3.3 which had a higher level of *SlEZ2* expression in F_6. However, it should be noted that the leaves of lines 2.2 and 4.23 appeared normal but also showed higher levels of SlEZ2 expression.

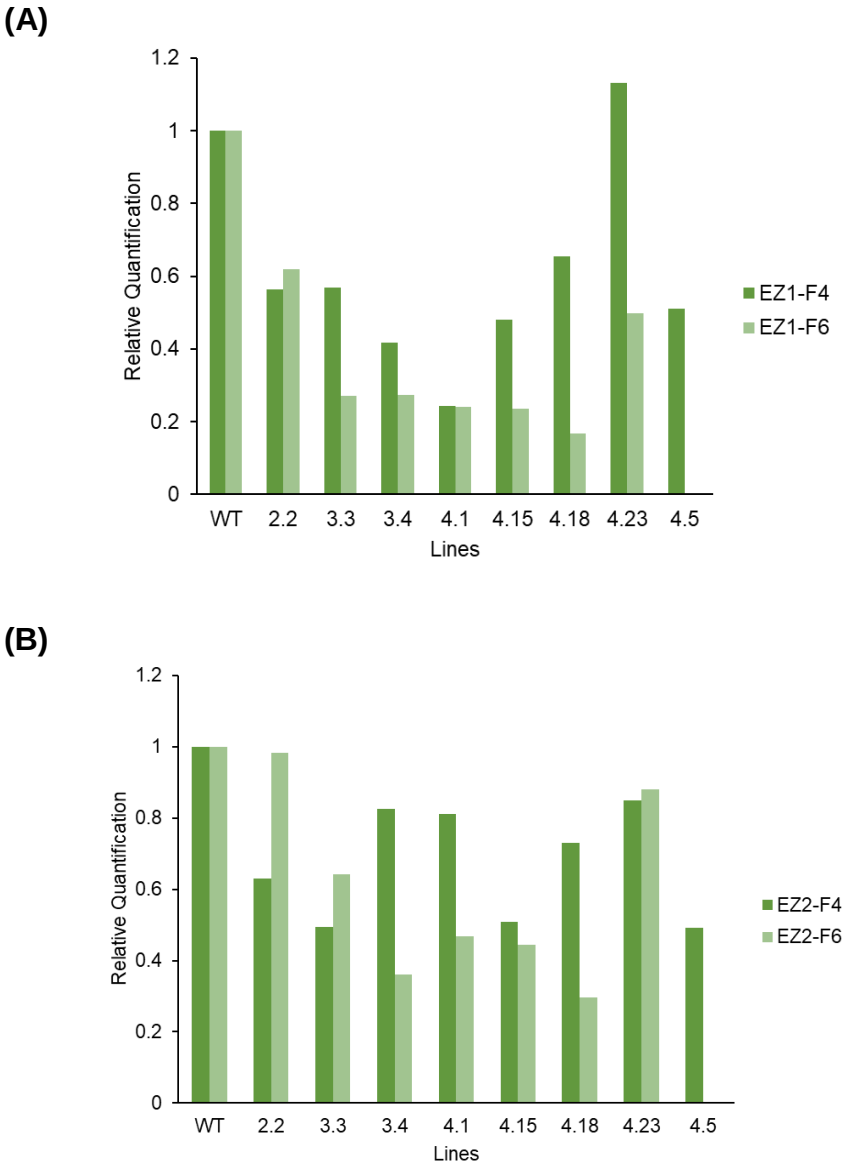


Figure 3. 6 *SlEZ1 and SLEZ2* gene expression in leaves from F_4 and F_6 epiRIL generations. (A) *SlEZ1* and (B) *SlEZ2* expression expression respectively. The data were analysed from two plants per line with three technical replicates for each plant. WT=wild type.

The relative expression of *SlGA20ox2* in tomato leaves and flowers

Plant hormones are known to be involved in plant growth and development. There are five classical plant hormones: auxins, cytokinins, gibberellins, abscisic acid and ethylene (Wang and Irving, 2011). A gene involved in the production of a gibberellin was investigated because this hormone has been found to affect plant height and fruit development such as found in the epiRIL population (Chapter 2). One of the gene investigated in this project was *SlGA20ox2* which is involved in the gibberellin biosynthetic pathway (Chen *et al.*, 2016) and was downregulated in a tomato dwarf phenotype *GA20ox2* is involved in the production of the precursors of bioactive GA synthesis (Li *et al*., 2012). The data showed that the relative expression of *SlGA20ox2* was lower in the leaves of the epiRILs (Figure 3.7)

The level of *SlGA20ox2* expression was higher in the flowers of the epiRILs than in the leaves, which has been reported previously in tomatoes (Chen *et al*., 2016). However, the expression was also higher in the epiRILs than the wild type. The expression in line 4.15 was more than 5 times that of wild type. When compared to fruit weight in the same generation, line 4.15 also showed lower fruit weight than wild type (Chapter 2) but the value was not significantly different ($P<0.05$).

Overexpression of gibberellin can cause parthenocarpy, which is the production of fruit without the process of fertilization and can produce fruits without seeds (Joldersma and Liu, 2018). Some fruits from the tomato epiRILs did not produce seeds such as line 4.5 in generation F_5. Also in lines 4.918, 4.66, 4.615, 2.123, 2.1 and 3.3 no seeds were

produced in some generations (Chapter 2). The differences in *SlGA20ox2* expression in the epiRILs may indicate DNA methylation changes in the *SlGA20ox2* gene and these could be affecting seed number.

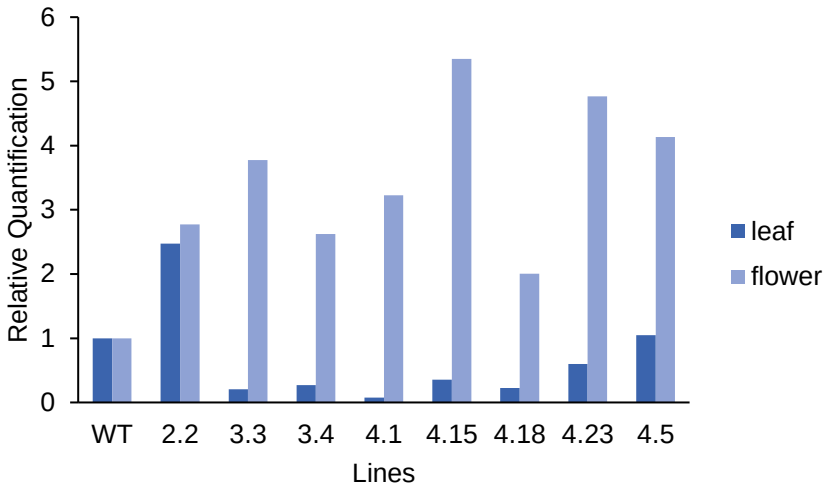


Figure 3. 7 *SlGA20ox2* gene expression in leaves and flowers of F_4 epiRILs. The data were analysed from two plants per line with three technical replicates for each plant. WT=wild type.

There are other genes which could be affecting the high abortion rate and the formation of parthenocarpy fruits in some of the epiRIls. Other than gibberellin, genes controlling auxin can also affect fruit parthenocarpy in tomato (de Jong *et al*., 2009). The auxin response factors are transcription factors that control auxin-dependent developmental processes and genes that encode these include *ARF7*. Plants with reduced levels of *SlARF7* produced parthenocarpic fruits that were heart-shaped and had thick pericarp (de Jong *et al*., 2009). These features were, however, not seen in the epiRILs. Homeotic genes such as *SlAGl6* can also have an effect on fruit development and genes from the AGAMOUS sub-clade of the MADS box genes are known to control carpel identity in plants. A mutation in *SlAGl6* caused parthenocarpic fruits in tomato (Klap *et al*., 2017) but the fruit shape and weight were unaffected unlike the epiRILs which showed changes in fruit weight

3.3.3 Global transcriptome analysis by RNA Sequencing

RNA sequencing was undertaken on 6 week old leaves of two epiRILs from generation F_7. At this generation, the population should be in an almost homozygous state (Reinders *et al*., 2009) and the majority of the epigenotypes observed in the lines will be fixed. Lines 3.3 and 4.69 were chosen for RNA sequencing due to several abnormal phenotypes that were observed during their development including abnormal leaves, twisted stamens, high flower abortion rates and low fruit weight (Chapter 2).

Differentially expressed genes and transposable elements (TEs)

The transcriptome between the epiRILs and wild type lines was compared including the expression of various types of TEs. Changes in gene expression were calculated from the FPKM values and a log2 fold change was determined. (Dündar *et al*., 2018; Li *et al*., 2017). The values were adjusted to detect false discovery rates in the data.

The data demonstrate that both genes and TEs were differentially expressed in the epiRILs in comparison with wild type and that the number of differentially expressed TEs was higher than the number of genes (Figure 3.8). There was also a higher number of upregulated genes and TEs compared to downregulated genes and TEs in both lines.

The total number of differentially expressed genes and TEs with Padj value < 0.05 was higher in line 3.3. There were 260 differentially expressed genes in line 3.3 with 182 that were upregulated and 78 that were downregulated. The number of differentially expressed TEs in line

3.3 were 356, with 310 that were differentially upregulated and 46 downregulated.

In line 4.69, there were 65 differentially expressed genes. From that number, 55 genes were upregulated and 10 downregulated. The number of differentially expressed TEs in line 4.69 was 223, with 216 that were differentially downregulated and 7 downregulated.

Changes in TEs mobility would be expected in the tomato epiRIls and have been observed in the Arabidopsis epiRILs. There was an increase of transposon reads in the *met1* Arabidopsis mutant with TEs being 47% of the reads compared to 15 % in wild type (Lister *et al*., 2008). In Arabidopsis epiRILs there was also evidence of movement for the CACTA transposons (Reinders *et al*., 2009) and accumulation of the TEs in several generations of the epiRILs in the *met1*-derived epiRIL. In the *ddm1*-derived epiRIL, TE movements were seen for CACTA and MULE TE families (Johannes *et al*., 2009).

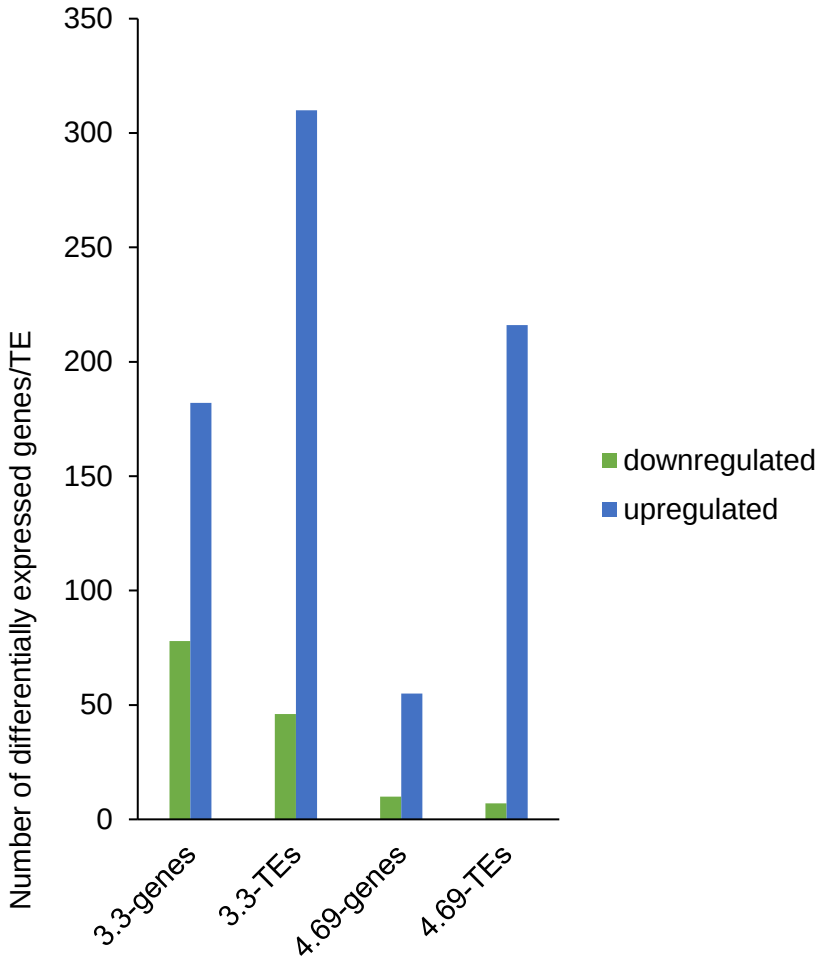


Figure 3. 8 The numbers of differentially expressed genes and TEs in lines 3.3 and 4.69 as determined by RNAseq. The cut off value was Padj<0.05.

Upregulated genes

In comparison to wild type, the number of genes with increased expression in both epiRILs 3.3 and 4.69 was higher than the number of downregulated genes. Applying a cut off Padj value of <0.05 and mean FPKM value > 5 for the epiRIL, twenty genes with the highest value of log2 fold change are presented in Tables 3.3 and 3.4.

The top three genes with highest fold change values in line 3.3 encoded for NAC domain containing protein 70, F-box family protein and basic helix-loop-helix (bHLH) DNA-binding superfamily protein (Table 3.3). The NAC domain containing protein 70 is a transcription factor involved in the regulation of root cap maturation in Arabidopsis (www.ncbi.nlm.nih.gov/gene/826627, 2019). The second most highly upregulated gene encoded an F-box family protein and these are known to play a role in protein ubiquitination and degradation (Lechner *et al*., 2006; Xu *et al*., 2008). The third most highly expressed gene encoded a bHLH DNA-binding superfamily protein. These are transcription factors with diverse functions ranging from light and hormone signalling to plant defense response (Carretero-Paulet *et al*., 2010).

In epiRIL 4.69 the three genes with highest fold change values were different from those of line 3.3. They encoded a Cytochrome b561 and DOMON domain-containing protein, a serine/threonine phosphatase-like protein and a BED zinc finger, hAT family dimerization domain containing protein (Table 3.4). The Cytochrome b561 protein is a transmembrane protein involved in electron transport (Verelst and Han Asard, 2003) and containing a DOMON (dopamine b-monooxygenase N

terminus) domain. The serine/threonine phosphatase belongs to the phosphoprotein phosphatase (PPP) family, which functions in auxin and brassinosteroid signalling, phototropism, regulation in rapamycin pathway and in cell stress responses (Uhrig *et al*., 2013). The zinc finger encoding gene functions as a transcription factor, but the function of this specific protein is unknown.

There were two genes that were found in the top twenty list in both lines. They were the ATP-dependent helicase/deoxyribonuclease subunit B and zinc-binding dehydrogenase family protein. Helicases are enzymes that are involved in the unwinding of DNA, dissociating the hydrogen bond between the nucleic acid duplex and other non-covalent bonds between the base pairs (Tuteja, 2003). Zinc-binding dehydrogenase family protein was also found highly expressed in both lines. Zinc finger homeodomain genes were found to be involved in tomato plant development and stress response (Khatun *et al*., 2017) , but further work will be needed to confirm the function of the encoded protein

Table 3. 3 List of 20 genes with highest increased expression in line epiRIL 3.3, generation F_7

N0.	Gene ID	Anotation	Mean FPKM	Mean FPKM	Fold change	Log2 FoldChange	Padj
			WT	3.3	3.3	3.3	
1	Solyc08g014355	NAC domain containing protein 70	0.01	5.69	854.00	9.98	1.5E-13
2	Solyc05g010367	F-box family protein	0.01	8.40	630.25	9.35	1.2E-11
3	Solyc05g032640	LOW QUALITY:basic helix-loop-helix (bHLH) DNA-binding superfamily protein	0.06	30.29	504.83	8.99	2.5E-10
4	Solyc00g147570	villin 2	0.05	15.74	314.87	8.34	2.3E-39
5	Solyc02g021170	LOW QUALITY:cyclic nucleotide gated channel 9	0.12	25.07	214.89	7.93	5E-09
6	Solyc05g021167	translocase of inner mitochondrial membrane 23	0.08	7.32	95.48	6.59	1.7E-06
7	Solyc10g045153	ATP-dependent helicase/deoxyribonuclease subunit B	0.15	10.81	70.52	6.13	8.1E-07
8	Solyc01g060020	beta-1,3-glucanase TOMB13GLUB	0.69	48.96	71.30	6.09	0.00057
9	Solyc01g059965	Beta-1,3-glucanase	0.56	39.54	70.18	6.08	0.00057
10	Solyc01g059980	Beta-1,3-glucanase	0.70	48.77	70.00	6.07	0.0006
11	Solyc10g045150	LOW QUALITY:oxidoreductase, zinc-binding dehydrogenase family protein	0.21	13.85	64.92	5.98	2.1E-06
12	Solyc07g006380	Defensin-like protein	0.48	24.66	51.38	5.64	0.02846
13	Solyc08g083500	Cytochrome P450 family protein	0.17	8.63	50.76	5.53	3.7E-27
14	Solyc02g063255	Enolase	0.46	16.74	36.65	5.24	1.5E-25
15	Solyc08g080640	Osmotin-like protein (Fragment) IPR017949 Thaumatin, conserved site IPR001938	0.54	20.58	38.35	5.20	0.00425
16	Solyc08g080650	Osmotin-like protein (Fragment) IPR001938 Thaumatin, pathogenesis-related	2.45	86.82	35.39	5.09	0.00175
17	Solyc04g009860	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	0.39	5.65	14.50	3.83	0.00147
18	Solyc08g079230	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	0.44	6.42	14.58	3.81	0.04743
19	Solyc01g106620	Pathogenesis-related protein 1	2.74	36.51	13.31	3.68	0.00323
20	Solyc01g106630	CBL-interacting protein kinase 23	2.77	36.63	13.21	3.67	0.00332

Table 3. 4 List of 20 genes with highest increased expression in epiRIL 4.69, generation F_7

N0.	Gene ID	Anotation	Mean FPKM	Mean FPKM	Fold change	Log2 FoldChange	Padj
			WT	4.69	4.69	4.69	
1	Solyc12g035223	Cytochrome b561 and DOMON domain-containing protein At5g35735	0.05	7.42	139.06	7.43	5.2E-08
2	Solyc00g005092	Serine/threonine phosphatase-like protein	0.09	12.36	132.43	7.09	9.1E-06
3	Solyc12g035225	BED zinc finger,hAT family dimerization domain, putative	0.08	6.68	83.50	6.61	5E-15
4	Solyc10g045153	ATP-dependent helicase/deoxyribonuclease subunit B	0.15	5.82	37.98	5.42	0.0001
5	Solyc05g047680	Cytochrome P450	0.17	5.82	34.94	5.41	0.0483
6	Solyc10g045150	LOW QUALITY:oxidoreductase, zinc-binding dehydrogenase family protein	0.21	7.34	34.39	5.25	0.0003
7	Solyc02g063255	Enolase	0.46	10.20	22.34	4.71	4.6E-20
8	Solyc03g033820	P-loop containing nucleoside triphosphate hydrolases superfamily protein	0.27	5.84	21.91	4.49	3.4E-14
9	Solyc12g040685	Sec14p-like phosphatidylinositol transfer family protein	2.63	26.80	10.18	3.46	1.2E-41
10	Solyc12g040690	fatty acid hydroxylase 2	3.30	23.97	7.27	2.99	1.5E-11
11	Solyc01g100490	chloronerva	1.60	9.96	6.21	2.86	0.0171
12	Solyc12g049613	Glycosyltransferase	4.29	19.31	4.50	2.28	1.6E-09
13	Solyc08g062290	Light-independent protochlorophyllide reductase subunit B	1.47	5.86	3.98	2.15	0.0081
14	Solyc07g009150	Protein kinase family protein	7.46	26.76	3.59	1.97	5.5E-07
15	Solyc00g011160	Glucan endo-1,3-beta-glucosidase, basic vacuolar isoform	4.20	12.01	2.86	1.59	0.0292
16	Solyc04g051720	LOW QUALITY:transmembrane protein	3.39	9.18	2.71	1.57	0.0168
17	Solyc01g017220	ATP synthase epsilon chain, chloroplastic	326.54	672.35	2.06	1.17	0.0171
18	Solyc10g044543	ATP synthase subunit b, chloroplastic	10.20	20.69	2.03	1.13	0.0041
19	Solyc08g076060	Serine/Threonine kinase family protein	8.94	16.02	1.79	0.93	0.0006
20	Solyc03g033825	Ubiquitin-related modifier 1 homolog	3.72	6.41	1.72	0.88	0.0033

Downregulated genes

The number of downregulated genes was less than the upregulated genes in both lines. Using a cut off value of log2 fold change< -2, only ten genes with adjusted Pvalue < 0.05 were downregulated in line 4.69 while in line 3.3 there was 78 genes that were downregulated. To compare the two lines, only ten genes are listed from line 3.3 (Table 3.5) and compared to those in line 4.69 (Table 3.6).

In line 3.3, the gene that was most downregulated encoded for cytochrome C oxidase subunit II-like, transmembrane domain-containing protein. The cytochrome C oxidase is involved in the respiratory chain in plant mitochondria (Schertl and Braun, 2014). Also a homeobox leucine zipper protein, which is a transcription factor involved in the negative regulation of cell elongation and specific cell proliferation processes (https://www.uniprot.org /uniprot/Q05466, 2019). The third gene most highly downregulated in line 3.3 encoded for a prolyl oligopeptidase family protein which may be involved in tolerance to abiotic factors (Tan *et al*., 2013).

In line 4.69 the top two genes most downregulated encodes for prolyl oligopeptidase family protein which was also found in line 3.3. The third most highly down regulated gene to be downregulated in line 4.69 encodes for proline-rich receptor-like protein kinase (PERK) family which is involved in the growth of stem and root (Borassi *et al*., 2016).

Table 3. 5 List of downregulated genes in line 3.3, generation F_7

N0.	Gene ID	Anotation	Mean FPKM	Mean FPKM	Fold change	Log2 FoldChange	Padj
			WT	3.3	3.3	3.3	
1	Solyc10g039393	Cytochrome C oxidase subunit II-like, transmembrane domain-containing protein	7.80	0.07	0.01	-6.27	3.8E-24
2	Solyc06g035940	Homeobox leucine zipper protein	41.73	2.30	0.06	-4.22	2.6E-59
3	Solyc02g050240	Prolyl oligopeptidase family protein	15.52	1.63	0.10	-3.20	1.1E-08
4	Solyc02g050207	Prolyl oligopeptidase family protein	5.92	0.76	0.13	-2.90	0.0074
5	Solyc08g067160	Acyl-protein thioesterase 2	7.38	0.99	0.13	-2.93	6.7E-08
6	Solyc03g111970	Cytochrome P450	6.65	1.06	0.16	-2.61	0.0059
7	Solyc07g007040	G-patch domain-containing protein	12.77	2.33	0.18	-2.50	4E-118
8	Solyc08g066260	Histidine decarboxylase	23.67	5.34	0.23	-2.20	0.0010

Table 3. 6 List of downregulated genes in line 4.69, generation F_7

N0.	Gene ID	Anotation	Mean FPKM	Mean FPKM	Fold change	Log2 FoldChange	Padj
			WT	4.69	4.69	4.69	
1	Solyc02g050240	Prolyl oligopeptidase family protein	15.52	0.21	0.01	-6.18	3.7E-24
2	Solyc02g050207	Prolyl oligopeptidase family protein	5.92	0.11	0.02	-5.46	1.1E-08
3	Solyc05g041750	LOW QUALITY:Proline-rich receptor-like protein kinase PERK10	13.99	2.38	0.17	-2.56	3.3E-05
4	Solyc09g008840	Pyruvate kinase family protein	18.65	5.22	0.28	-1.81	0.03098
5	Solyc12g035710	DUF21 domain-containing protein	2.48	0.71	0.29	-1.83	0.00045
6	Solyc02g050205	Prolyl oligopeptidase family protein	7.64	2.21	0.29	-1.75	2.7E-21
7	Solyc02g005115	Glycosyltransferase	5.03	1.62	0.32	-1.58	0.03248
8	Solyc02g005110	Homeobox leucine-zipper protein	2.71	0.88	0.33	-1.60	0.03098
9	Solyc03g062830	Nudix hydrolase-like protein	35.58	18.08	0.51	-0.94	0.00092
10	Solyc02g055467	Disease resistance protein (CC-NBS-LRR class) family	23.95	12.74	0.53	-0.86	0.00087

Differentially expressed transposable elements

There were 14 types of TEs that were differentially expressed in the epiRILs. These differentially expressed TEs are presented as percentage of all of the types of TEs found in each line (Figure 3.9). In both epiRILs, the highest percentage was found to be the gypsy elements which reached more than 60% of the total. In line 3.3, the second most differentially expressed TE was unable to be identified (unknown) while the third belonged to the copia group. In line 4.69, the second most differentially expressed TE was copia while the third group was unknown.

These data provide evidence that most of the differentially expressed TEs in the tomato epiRILs were from the Class I elements or LTR (Long Terminal Repeat) retrotransposons. Class I elements are normally found in tomato (Jouffroy *et al*., 2016) whereas in Arabidopsis Class II elements were more abundant (Le *et al*., 2000). Class I TEs generateRNA intermediates for a 'copy and paste' mechanism while class II TEs are known to transpose or move via a 'cut and paste' mechanism (Dubin *et al*., 2018)

In Arabidopsis, transposon movements have been observed in hypomethylated mutants. Both Class I and class II elements have been shown to remobilize in a *met1* Arabidopsis mutant (Mirouze *et al*., 2009; Mari-Ordonez *et al*., 2013) and in Arabidopsis epiRILs (Johannes *et al*., 2009; Reinders *et al*., 2009). Changes in TE movements were also linked to increased TE expression. Changes in methylation levels and mobilization of TEs in the tomato epiRIls, are investigated further in Chapter 4.

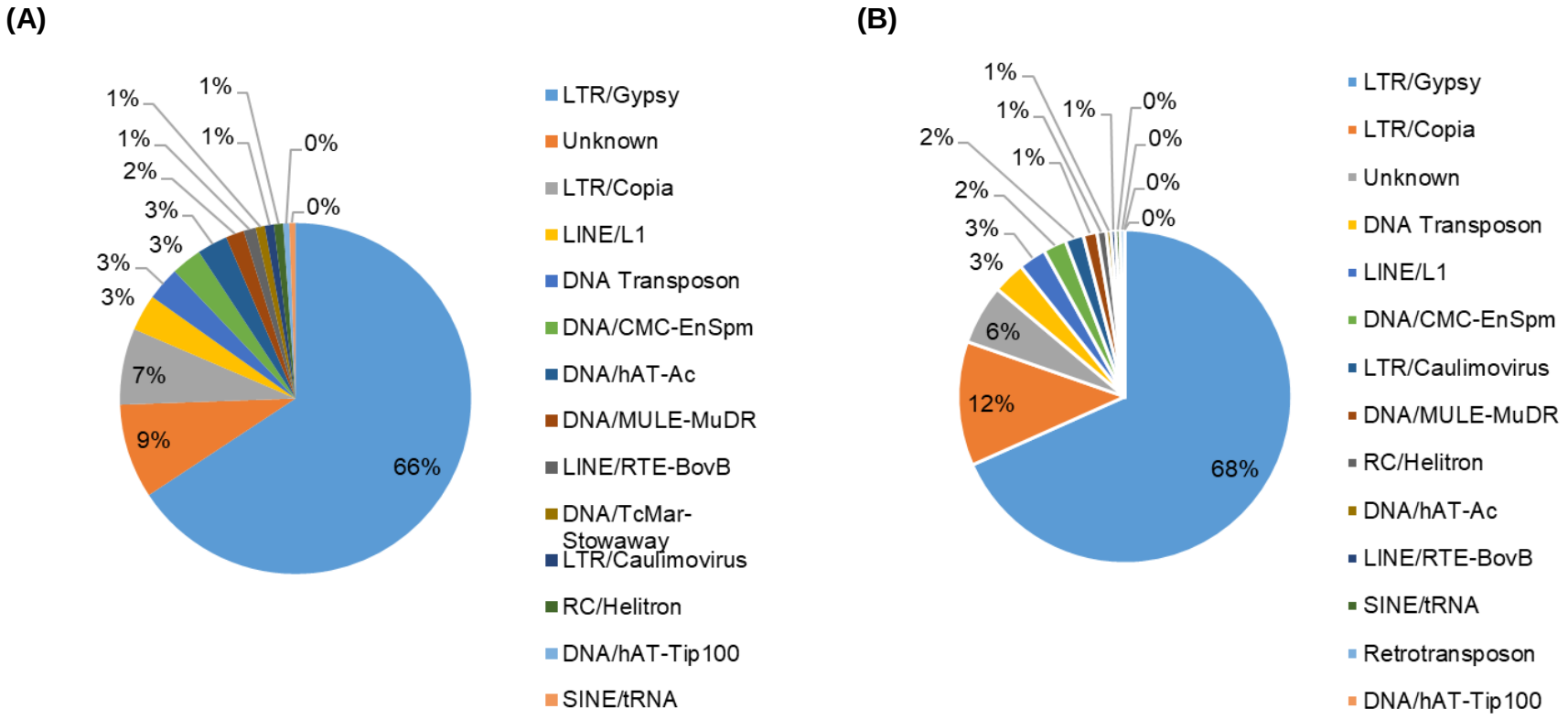


Figure 3. 9 The proportion of differentially expressed TEs.

Data from epiRILs (A) line 3.3 and (B) line 4.69. Padj <0.05 as threshold of significance

Downregulated TEs.

Using a cut off value of mean FPKM wild type > 5 to detect changes in expression, there were 23 TEs downregulated in line 3.3 (Table 3.7). The log2 fold change <-2 was not used as cut off value as not many of the downregulated TE had a lower value. Most of the TEs that were downregulated in line 3.3 were unknown types and those that belonged to the DNA transposons which are Class II TEs and replicates via a cut-and-paste mechanism (Feschotte and Pritham, 2007). Both the unknown group and DNA transposons were in the top 4 groups that were differentially expressed in the epiRILs.

Using the same cut off value in line 4.69 there were only 2 TEs that were downregulated (Table 3.8). Both belonged to the unknown and DNA transposon types. These results showed that the types of TEs downregulated belonged to a different types of TEs to those that were upregulated. The regulation of the different types of TEs could be different according to the mechanism of transposition. Gypsy and copia TEs are Class I elements while the DNA transposon is a Class II element. Further analysis is needed to investigate if differences in cytosine methylation between the epiRILs could contribute to the differences in the regulation of TEs.

Table 3. 7 List of downregulated TEs in line 3.3 in generation F_7

No.	Feature ID	repeat class	Mean FPKM WT	Mean FPKM 3.3	log2FoldChange	Padj
1	SL3.0ch01.83093047.83093100	Unknown	42.93	11.79	-2.00	0.0008007
2	SL3.0ch06.42150234.42150340	Unknown	185.65	52.88	-1.98	0.0002345
3	SL3.0ch08.59754929.59755912	DNA/hAT-Ac	5.14	1.52	-1.96	0.0032257
4	SL3.0ch01.95903960.95904034	DNA	38.93	11.56	-1.91	0.0024532
5	SL3.0ch01.95903976.95904088	Unknown	21.22	6.64	-1.89	0.0106141
6	SL3.0ch06.42150262.42150674	Unknown	34.73	10.70	-1.87	0.0001965
7	SL3.0ch09.70335532.70335697	Unknown	16.01	5.41	-1.79	0.0016835
8	SL3.0ch08.59754527.59754887	DNA/hAT-Ac	20.50	6.83	-1.75	0.016252
9	SL3.0ch01.6238418.6239914	DNA/hAT-Ac	1.67	0.58	-1.75	0.002095
10	SL3.0ch12.1023554.1023846	DNA	8.38	2.66	-1.74	0.0508965
11	SL3.0ch02.26696468.26696525	DNA/TcMar-Stowaway	49.33	16.22	-1.64	0.0095725
12	SL3.0ch05.59171675.59171757	DNA/TcMar-Stowaway	44.73	16.70	-1.59	0.0354423
13	SL3.0ch04.49001916.49002098	Unknown	25.52	9.85	-1.57	0.0043391
14	SL3.0ch08.568069.568317	DNA/TcMar-Stowaway	8.49	3.46	-1.54	0.0082755
15	SL3.0ch08.26773064.26773109	DNA/MULE-MuDR	989.48	400.12	-1.45	0.0034215
16	SL3.0ch05.59197623.59197863	LTR/Copia	27.82	11.65	-1.44	0.0346495
17	SL3.0ch05.1377485.1377576	Unknown	51.65	22.44	-1.43	0.0299197
18	SL3.0ch06.37682393.37683177	DNA/hAT-Ac	3.58	1.43	-1.42	0.0221979
19	SL3.0ch08.59753182.59753396	DNA/hAT-Ac	22.82	9.63	-1.41	0.0422374
20	SL3.0ch06.46733441.46733583	Unknown	17.91	7.70	-1.32	0.0411509
21	SL3.0ch12.66970764.66970844	LINE/RTE-BovB	23.89	10.07	-1.29	0.0440374
22	SL3.0ch02.37015423.37015875	DNA	6.41	2.85	-1.28	0.0307387
23	SL3.0ch01.89341905.89342064	Unknown	26.33	12.53	-1.17	0.0198855

Table 3. 8 List of downregulated TEs in line 4.69 in generation F_7

No.	Feature ID	Repeat class	FPKM WT	FPKM 4.69	log2FoldChange	Padj
1	SL3.0ch05.21537916.21539495	LTR/Gypsy	0.3466667	0	-6.826865978	8.82E-05
2	SL3.0ch08.7788330.7789478	LTR/Gypsy	0.6633333	0.01333333	-5.461185516	0.001311
3	SL3.0ch08.7789491.7790692	LTR/Copia	0.79	0.03333333	-4.472601133	0.000114
4	SL3.0ch08.7792855.7795582	LTR/Gypsy	0.21	0.01666667	-3.499733441	0.036333
5	SL3.0ch08.7786759.7788347	LTR/Gypsy	1.1133333	0.11666667	-3.40008769	0.0006
6	SL3.0ch01.95903960.95904034	DNA	38.926667	13.7433333	-1.497599803	0.046792
7	SL3.0ch01.83093047.83093100	Unknown	42.933333	14.7	-1.470127735	0.039816

CHAPTER 4 DETECTION OF DNA METHYLATION AND TRANSPOSON MOBILIZATION USING RESTRICTION ENZYMES

4.1 Introduction

The tomato epiRIL population would be expected to have reduced levels of cytosine DNA methylation across the genomes of the various lines based on down regulation of the *MET1* gene in one of the initial parents. In this Chapter the location of DNA methylation in the epiRILs is explored using a number of approaches chosen to maximise the ability to characterise the population, within the constraints of the funding available within this PhD project.

The three main objectives of the DNA methylation analysis described in this and the following chapter were to identify differentially methylated regions (DMRs) in:

(1) Targeted analysis of cytosine DNA methylation in specific genes of interest based on phenotypic and gene expression changes.

(2) Investigate potential transposon mobilization as a result of genome-wide hypomethylation.

(3) In a small number of epiRIL lines investigate the changes in cytosine DNA methylation at single nucleotide resolution and on a genome-wide scale (Chapter 5).

This chapter addresses objective 1 and 2. In objective 1 the targeted DNA methylation analysis is explored using the well-characterised McrBC enzyme assay, while transposon mobilization has been investigated by Southern blotting (objective 2).

McrBC recognizes two methylated or hemimethylated sites (RmC), located on one and/or the other strand of the DNA about 30 - >2000 base pairs apart. DNA digested with McRBC and PCR is used to amplify the region of interest from digested and undigested DNA. The presence of methylation within the target region results in its fragmentation by the McrBC enzyme and inhibits PCR amplification (Stewart *et al*., 2000; Bai *et al*., 2016). .

Southern blotting has been used to detect transposition in studies of epiRIL populations (Reinders *et al.*, 2009; Johannes *et al.*, 2009). The transposition of the CACTA transposon was detected in Arabidopsis epiRILs derived from a cross between *met1* mutant and wild type line (Reinders *et al*., 2009). In the epiRILs derived from *ddm1* mutant and wild type, two transposons (CACTA and MULE) were found to be mobilized in plants showing abnormal flowering (Johannes *et al*., 2009).

In the tomato epiRILs changes in DNA methylation in genes of interest was undertaken using the MrcBC assay and transposition was investigated using Southern blotting. The hypothesis for this part of the project was to detect changes in CG methylation at selected promoter regions and transposable elements using restriction enzymes.

4.2 Materials and Methods

4.2.1 DNA extraction I for the McrBC assays

This method was selected for its DNA purity and delivery of between 1-5 µg of DNA / extraction. It followed the protocol of ISOLATE II Plant DNA Extraction Kit from Bioline® (London, UK). DNA was extracted from 0.1 g leaf tissue from four week old plants. The leaf tissue was added to a 2 mL microfuge tube and this was then placed in liquid nitrogen. After being snap frozen, the leaf tissue was ground to a fine powder using a plastic grinding stick. Then 400 µL of Lysis buffer PA1 was added to the tissue and the mixture vortexed. The ISOLATE II filter (violet colour) was placed in a new Collection Tube and the lysate was added. The sample was centrifuged at 11,000 x g for 2 minutes. The Elution Buffer PG was prepared at this stage by placing on a 65 ^{0}C heating block.

The filter was discarded and to the flow-through was added 450 µL Binding Buffer and this mixture was then vortexed. The ISOLATE II Plant DNA Spin Column (green colour) was placed inside a new collection tube and a maximum of 700 µL of the sample was loaded onto the column. The sample was then centrifuged at 11,000 x g for 1 min. The step was repeated until all of the lysate was used. To wash the DNA, 400 µL of Wash Buffer PAW1 was added to the column and centrifuged at 11,000 x g for 1 minute. The flow-through was discarded. For the next wash, 700 µL of Wash Buffer PAW2 was added and centrifuged at 11,000x g for 1 minute. After the flow-through was discarded, 200 µL of Wash Buffer PAW2 was added and centrifuged again at 11,000 x g for 2 minutes.

The ISOLATE II Plant DNA Spin Column was then placed in a new 1.5 mL microfuge tube and the DNA eluted by adding 50 µL of Elution Buffer PG (which had been placed at 65 ^{0}C) and the sample incubated at 65 ^{0}C for 5 minutes. The sample was then centrifuged at 11,000 x g to elute the DNA.

4.2.2 DNA extraction II for Southern blotting

The Illustra Nucleon PhytoPure Kit from GE Healthcare® (Buckinghamshire, UK) was used to obtain at least 15 µg of pure DNA needed for Southern Blotting . Leaf samples (100mg) were frozen and ground to a fine powder. Reagent 1 (R1) was prepared by adding 0.7 µL of β-mercaptoethanol for every 1mL of R1 solution. For every sample, 900 µL of R1+mercaptoethanol solution was needed.

The 900µL of the R1 plus mercaptoetanol solution was added to the frozen leaf powder and mixed with a small spatula. A volume of 300 µL Reagent 2 (R2) was then added to the mixture and the tubes inverted 3-4 times until the mixture became homogenous and more viscous. The sample was then incubated at 65 ^{0}C for 30 minutes then placed on ice for 20 minutes.

A volume of 600 µL of cold chloroform and 200 µL of resin was added to the sample. The tube containing the sample was inverted several times at room temperature for 10 minutes. The sample was then centrifuged at 1300 x g for 10 minutes. The supernatant was transferred to a new 2mL tube and an equal volume of cold isopropanol was added. The tube was inverted again several times.

The sample was centrifuged at 4000 x g for 5 minutes and the supernatant was discarded. The pellet was then washed by adding 100 µL cold 70% ethanol and centrifuged at 4000 x g for 1 minute. The supernatant was discarded. Washing with ethanol was repeated twice and then the sample was dried at room temperature for a maximum of 5 min. The pellet was then resuspended in100 µL nuclease free water.

To remove RNA, 20 µL /mL of RNAse (Qiagen) was added to the sample and incubated at 37 ^{0}C for 30 min then the RNAse was inactivated at 65 ^{0}C for 10 min. The concentration of DNA was measured in a NanoDrop1000 Spectrophotometer (Thermo Fisher Scientific Inc.) .

4.2.3 McrBC enzyme Digestion

For McrBC digestion the McrBC digestion kit from New England Biolab® (Hertz, UK) was used. The digestion reaction mix consisted of 10 µL NEB Buffer (1X), 0.5 µL BSA (100µg/mL), 10µL GTP (1mM) and 2µL McrBC Enzyme (50U/µg gDNA), 1 µg DNA and nuclease free water was added until a total volume of 100 µL. The reaction mix was incubated at 37 ^{0}C for 5 hours and inactivated at 65 ^{0}C for 20 minutes. The digested DNA was then stored at -20 ^{0}C or could be used directly for downstream analysis. Two biological samples were used for the McrBc digestion taken from each line.

4.2.4 PCR of McrBC digested DNA

Polymerase chain reaction (PCR) was undertaken using the Promega® Gotaq Colourless Master Mix (Wisconsin, USA). The PCR mixture had a total volume of 12.5 µL. The reaction mix consisted of 6.25 µL Gotaq Colourless Master Mix (2X), 1.25 µL of forward and reverse primers (1µM), 50 ng DNA and nuclease free water added until the total volume was 12.5 µL . The PCR condition is presented in Table 4.1. The primers used in the amplification of McrBC digested DNA are presented in Table 4.2 and its position relative to the gene transcription start site is presented in Appendix 21.

Table 4. 1 PCR Condition for amplification of McrBC digested DNA

Step	Temperature	Time	Cycles
Initial denaturation	95°C	5 min	1
Denaturation	95°C	30 sec	30
Annealing	58°C	30 sec	30
Extension	72°C	30 sec	30
Final extension	72°C	5 min	1
Hold	4°C		

Table 4. 2 Primers used in amplification of McrBC digested DNA

Primer name	Target gene promoter region	Sequence (5'-3')
FW 0-500-F	*FW2.2*	CCCTCACTCTCTTTTCTTGGCA
FW 0-500-R	*FW2.2*	AAGCGCAAAAGGTTTAAATGGA
FW 500-1000-F	*FW2.2*	TTGAAGTGTTGAATAAAGTT
FW 500-1000-R	*FW2.2*	AAGAAATGTTGTAGTGTAAA
EZ1 0-500-F	*SlEZ1*	TGACATAGCGGGATTCTA
EZ1 0-500-R	*SlEZ1*	TAATGTGTGGGTATTGTG
EZ1 500-1000-F	*SlEZ1*	CGAGGAGGCAAAAGGATA
EZ1 500-1000-R	*SlEZ1*	ACATATAAGACTACAAAGTAACGA
EZ2 0-500-F	*SlEZ2*	AAGAGTCGTGTGTAAATC
EZ2 0-500-R	*SlEZ2*	CGTGTATTCAGATTCAAAGAT
EZ2 500-1000-F	*SlEZ2*	CGTATGGAGATTCTACATTTGA
EZ2 500-1000-R	*SlEZ2*	TTTGCTATTTATGACTGCTCTTT
TM8 0-500-F	*TM8*	GAATTATTGAGATTAAGTGAGT
TM8 0-500-R	*TM8*	AATAGATGGATTGAGTTGAA
TM8 500-1000-F	*TM8*	AACACAAATGAACGACCAAT
TM8 500-1000-R	*TM8*	AGAGATTGAATAAAAGCACCAT
CNR-F (Liu *et al*., 2015)	*CNR*	TGAGCATCAACCACTCCTAAT
CNR-R	*CNR*	CAGACTTAGTAATAACTCCGAT

4.2.5 Quantitative (Q)-PCR of McrBC digested DNA

The method used for qPCR was that described in Chapter 3. However, the samples used were McrBC digested leaf DNA. The samples were taken from two plants per line. The McrBC digested DNA was diluted to a 500ng solution and 40 ng was used for the qPCR reaction. Three technical replicates was used for each sample.

To calculate the relative quantity (RQ) of the gene expression, the 'delta-delta method' was used (Pfaffl, 2001). The Ct (Cycle threshold) value of each McrBC sample was calculated relative to the Ct value of gDNA (genomic DNA) sample to find ΔCt epiRIL. Then for each line, the ΔCt value was calculated relative to ΔCt wild type. The relative expression of between each line with wild type was compared using the equation :

$$RQ = 2^{-[\Delta Ct\ epiRIL - \Delta Ct\ wild\ type\ ]} = 2^{-\Delta\Delta Ct}$$

Percentage methylation was calculated using the formula: (1/RQ) x 100%. Regression analysis was undertaken between the McrBC-QPCR data of 0-500 bp and 500-1000 bp upstream of TSS of *FW2.2* gene with the cDNA-QPCR data of green fruit in Chapter 3. A linear regression model was applied using the Genstat® software (19th edition). Results from the regression analysis are presented in Appendix 22.

4.2.6 Restriction Enzyme Digestion for Southern Blotting

The leaf DNA was digested with a range of restriction enzymes to detect transposon methylation and movement. The enzymes were purchased from New England Biolabs® (Hertz, UK). For the reaction, 10-15 µg of DNA was needed. The reaction mix consisted of 10 µL buffer (10x), 5

µL enzyme, DNA and nuclease free water up to a total volume of 60 µL. The mixture of enzymes and DNA were incubated at 37 ^{0}C for 90 min. For EcoRI the reaction was stopped by heat inactivation at 65 ^{0}C for 20 min, while HindIII was inactivated at 80 ^{0}C and 20 minutes.

4.2.7 DIG Southern Blot

Probe design

The presence and movement of two transposon types were investigated in the epiRIls: the *copia* and *gypsy*- retrotransposons. The *copia* probe was designed from the complete sequence of the tomato retrotransposon Kielia (GenBank: EU195798.2). The *gypsy* probe was designed from the tomato mRNA for *gypsy*-like retrotransposon sequence (GenBank: Z95351.1). The probes were designed using the Primer 3 Plus software. The sequence of the copia probe sequences were: 5'-GTAGACGGAGGCAGCATCTC-3' (forward) and 5'-ATCTGCAGCCTCCCAAAGAC-3' (reverse). The gypsy probes sequences were: 5'-TCCTCTTTCTCGCATTGATGACT-3' (forward) and 5'-CACATGCCCCAAGAAGGACA-3' (reverse). The product size were 568 bp (copia) and 402 bp (gypsy).

Digoxigenin (DIG) Probe labelling

Probe labelling was carried out using the DIG Probe Synthesis Kit from Roche® (Indianapolis, USA). The following components were added to a microfuge tube: 5 µL PCR Buffer with MgCL2, 5 µL PCR DIG Mix, 0.75 µL Enzyme Mix, forward and reverse primers, 10 ng of DNA, and nuclease

free water up to a volume of 50 µL. The components were mixed and centrifuged briefly. The tubes were placed in a thermalcycler and the program used is presented in Table 4.3.

Gel electrophoresis and capillary transfer

For Southern blotting a 0.8% agarose gel was used. For this, 0.8g of agarose (Sigma®) was dissolved in 100 mL of 0.5x TBE buffer (Sigma®). The solution was cooled to 60 ^{0}C and 0.5 µg/mL of ethidium bromide was added and mixed before pouring the agarose into a gel cast. 15 µL of 10 µg digested DNA was mixed with 5 µL of blue dye (New England Biolab) and pipetted into the well. For the molecular marker, 5 µL of DNA Molecular Weight Marker III (Roche®) was used.

The gel was depurinated in 0.25M of HCl by gentle shaking at room temperature for 15 minutes. Then it was rinsed with sterile double distilled water. The gels was denatured by submerging in a solution of 0.5 m NaOH +1.5m NaCl and shaking gently for 2 x 15 minutes at room temperature. The gel was rinsed again in sterile double distilled water. The gel was then submerged in neutralization solution (0.5m Tris-HCl pH 7.5 and 1.5M NaCl) for 2 x 15 minutes at room temperature and then 20 x SSC (Invitrogen®) for 10 minutes.

To blot the gel (Figure 4.1), a long piece of Whatman 2 mm paper was soaked with 20 x SSC solution was placed on a 'bridge' that rested in a container with a shallow reservoir of 20 x SSC solution. The gel was placed face down on top of the soaked Whatman paper and any air bubbles were removed by rolling a sterile tip over the gel. A positively charged Nylon

membrane (Roche®) that was cut according to the size of the gel was placed on top of the gel. Air bubbles were again removed by rolling a sterile tip over the paper-gel sandwich.

Two pieces of Whatman paper cut into the size of the gel and placed over the membrane. A stack of paper towels were placed above the 'sandwich' and a weight of 250 g was placed on top of the assembly (Figure 4.1). The transfer was left overnight at room temperature. The DNA was fixed on the membrane the next day by placing the membrane face up on Whatman paper that has been soaked in 2 x SSC. The membrane was then exposed to UV light in a transilluminator (wavelength 302 nm) for 1 minute.

Table 4. 3 PCR Condition for DIG probe labelling

Step	Temperature	Time	Cycles
Initial denaturation	95°C	2 min	1
Denaturation	95°C	30 sec	30
Annealing	60°C	30 sec	30
Extension	72°C	40 sec	30
Final extension	72°C	7 min	1
Hold	4°C		

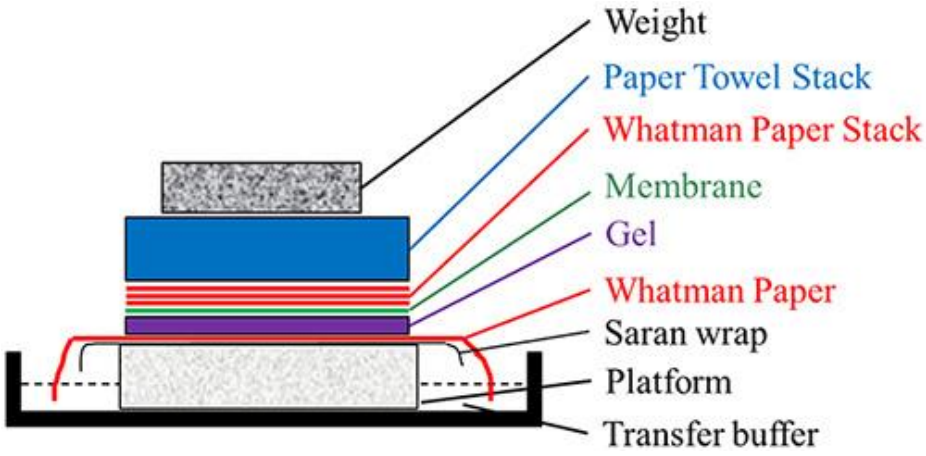


Figure 4. 1 Southern Blot transfer assembly. The gel was placed face down on the Whatman paper. Note: saran wrap was not used in the experiment. (Figure adapted from http://orbitbiotech.com, 2018)

Pre-hybridization and Hybridization of the DIG probe

After UV fixation, the membrane was placed in a plastic container with 25 mL of DIG Easy Hyb solution that had been warmed to 42 ^{0}C. The membrane was incubated in the solution for 30 minutes with agitation at 42 ^{0}C in a shaking incubator.

The labelled probe (50µL) which had already been prepared was placed in a microfuge tube with 50 µL of nuclease free water. The probe and water mix was placed in a thermal cycler at 95 ^{0}C for 5 min then held at 4 ^{0}C and placed in ice. The denatured probe was then placed in 25 mL of DIG easy Hyb and mixed by inversion to form the hybridization solution.

The pre-hybridization solution was poured out and replaced with a pre-warmed hybridization solution (containing the labelled probe). The plastic container was sealed using parafilm and placed in a shaking incubator at 42 ^{0}C overnight.

Stringency Wash

A 200mL Low Stringency Buffer (2x SSC containing 0.1% SDS) was prepared in a plastic container. The membrane from the hybridization procedure was placed in the container with the Low Stringency Buffer and shaken gently at room temperature for 5 minutes. The buffer was poured out and filled with fresh buffer and shaken gently again for 5 minutes.

A high stringency buffer solution (0.5 x SSC with 0.1% SDS) was prepared and placed in the wash temperature of 65 ^{0}C. The low stringency buffer was poured off and replaced with the pre-heated high stringency

buffer. The membrane was incubated in this buffer at 65 ^{0}C for 30 minutes with gentle agitation.

Chromogenic Detection

For chromogenic detection, the DIG Wash and Block Buffer Set (Roche) and the DIG-High Prime DNA Labelling and Detection Starter Kit I (Roche) were used. The membrane from the previous step was transferred to a plastic container with 100 mL of 1x Washing Buffer (DIG Wash and Block Buffer Set). It was incubated at room temperature for 2 minutes with shaking. The Wash Buffer was discarded and replaced by 100 mL of prepared Blocking Solution (10mL Blocking Solution + 90 mL maleic acid buffer). The membrane was shaken gently for 30 minutes. The Blocking Solution was discarded and replaced by Antibody Solution (50mL prepared Blocking solution + 10 µL antibody). The membrane was incubated for 30 minutes with gentle shaking. After the antibody solution was discarded, the membrane was washed twice (2 x 15 minutes) with 100 mL of 1x Washing Buffer. The membrane was then placed in 20 mL of 1 x Detection Buffer for 3 minutes.

For visualizing the result, the membrane was placed in a plastic container with 20 mL of the Colour Substrate (400 µL NBT/BCIP in 20mL of Detection Buffer). The container was covered so that the reaction occurred in the dark without shaking. The reaction was left overnight and the membrane photographed to record the result.

4.3 Results and Discussion

4.3.1 McrBC analysis

The detection of methylation changes was investigated in 1000 bp region upstream of selected gene transcription start site. The primers were in the regions 0-500 bp and 500-1000bp upstream of the coding sequences. Using the McrBC-qPCR analysis, promoters of four genes were investigated, these were *FW2.2*, *SlEZ1*, *SlEZ2* and *TM8*. Analysis with the McrBC enzyme was used to compare the percentage of methylation in generations F_2 and F_6.

To check the efficiency of the digestion with McrBC, the samples were amplified using a primer for the *CNR* gene (Solyc02g077920) promoter region which is known to be methylated in tomato leaf tissues and has been used as a control in McrBC digestion (Manning *et al*., 2006; Liu *et al*., 2015). The result showed that the enzyme was able to digest the methylated *CNR* gene DNA (Figure 4.2).

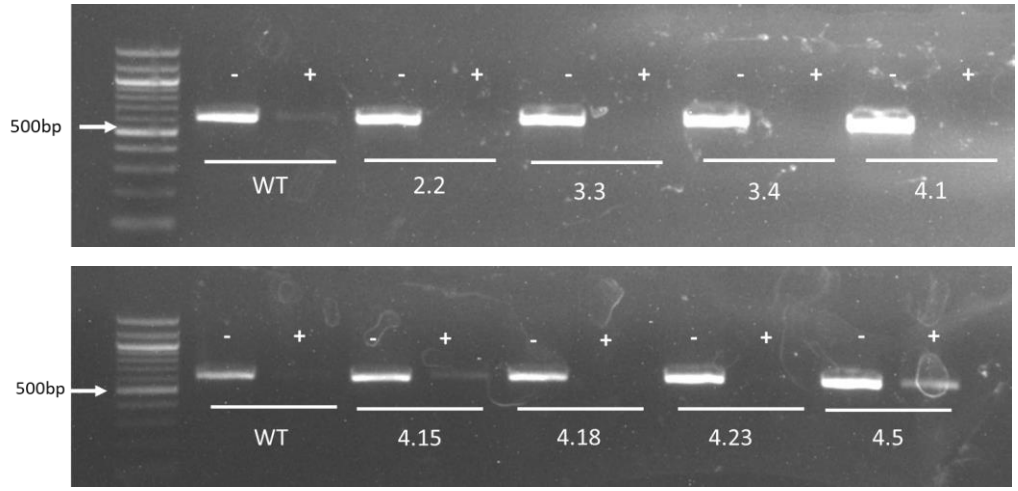


Figure 4. 2 Testing the digestion efficiency of McrBC restriction enzyme. Amplification of the *CNR* (Solyc02g077920) gene promoter was used as a control for enzyme efficiency. DNA samples not digested with McrBC showed amplicon at 650 bp while no band showed for the digested sample. (-) indicates amplified DNA not digested with McrBC and (+) are amplified McrBC digested DNA.

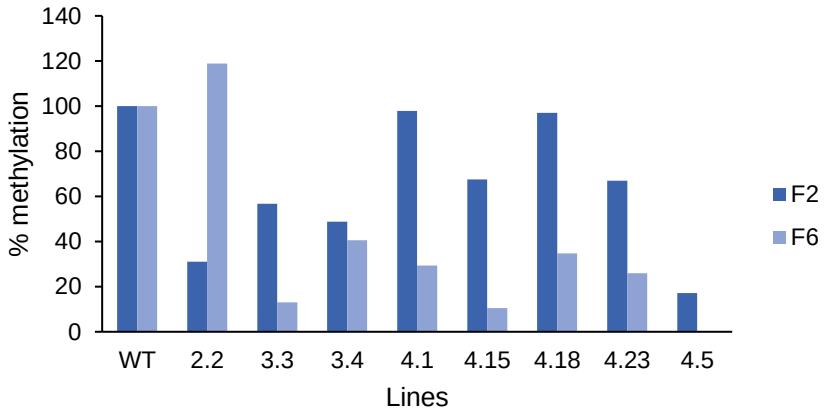
McrBC-qPCR analysis of *FW2.2* promoter region

The McrBC-qPCR analysis was undertaken for a selection of eight different epiRILs and wild type. The assay indicated that the epiRILs had a lower percentage of methylation compared to wild type (Figure 4.3) in the *FW2.2* regulatory region. In region 0-500 bp upstream of the gene, the percentage of methylation was lower in generation F_6 compared to F_2 except for line 2.2, which showed higher methylation in generation F_6. The greatest change was seen in line 4.15 with only 10.6 % of the levels of methylation of wild type. Line 4.5 did not produce seeds in F_5 so there was no data in generation F_6.

In the region 500-1000 bp upstream of *FW2.2*, all eight lines showed lower methylation than wild type in the F_2 generation. In generation F_6, the percentage of methylation changed. Some lines showed an increase while others showed a decrease in methylation.

The lower percentage of methylation in the epiRILs at the *FW2.2* locus was likely to have resulted from reduced methylation of the *MET1* parental line. The recombination between the parental epialleles can be stably inherited over several generations (Reinders *et al*., 2009). The level of methylation across selected whole epiRIL genomes is reported in Chapter 5.

(A)



(B)

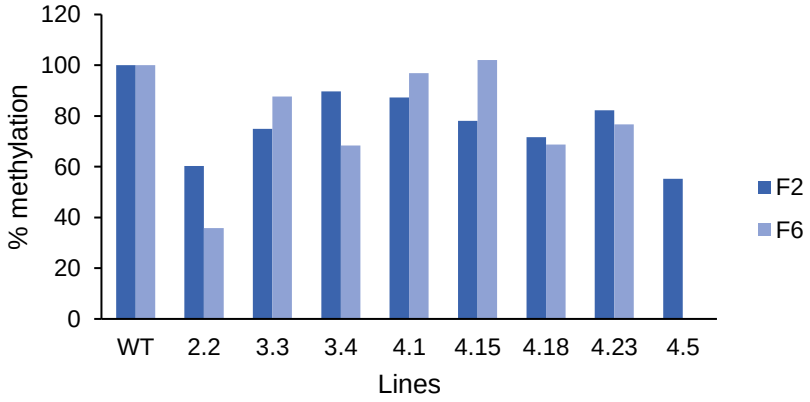


Figure 4. 3 Percentage of methylation for upstream region of *fw2.2* gene. The regions analysed were (A) 0-500 bp and (B) 500-1000 bp upstream of the gene. Percentage of methylation in wild type was 100% for the purpose of this calculation. Data was from generations F_2 and F_6.

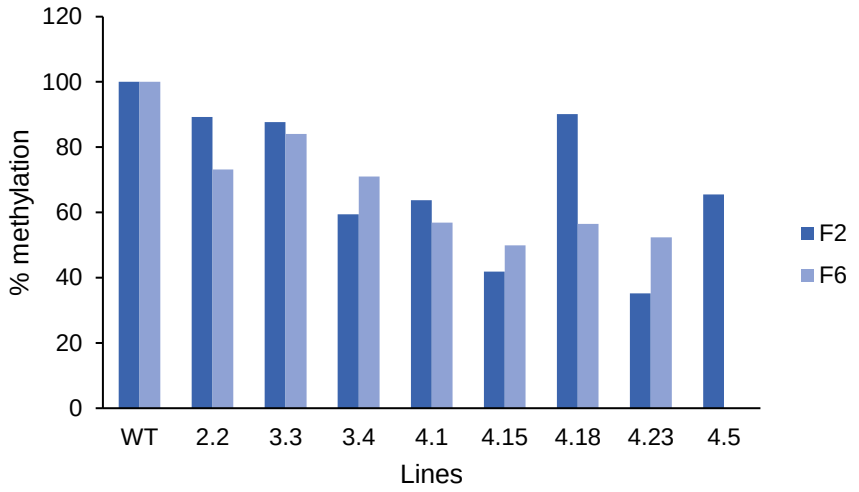
Frary *et al*. (2000) suggested that the lower fruit weight in tomato might be due to changes in the coding and upstream region of the gene at that locus based on sequence alignments of the *S. pennellii* and *S. lycopersicum fw2.2* regulatory regions. In the Frary *et al.* (2000) study, increased expression of the *FW2.2* gene was associated with smaller fruits. In the epiRILs, hypomethylation of the *FW2.2* regulatory region might result in increased expression of the gene.

McrBC-qPCR analysis of the Enhancer of Zeste promoter regions

The *SlEZ* promoter regions were also analysed with the analysis covering a small region in 0-500 bp upstream and 500-1000 upstream of the transcription start site of the *SlEZ1* and *SlEZ2* genes. These two genes have been found to play an important role in tomato plant development (How-kit *et al*., 2010; Boureau *et al*., 2016).

In the 0-500 bp upstream region of the *SlEZ1* regulatory region, the methylation ranged from 35 -90% in the epiRILs that were analysed (Figure 4.4). The percentage of methylation was compared in two generations: F_2 and F_6 with some differences in DNA methylation levels apparent in individual lines across the generations. The methylation levels were lower than wild type F_6 for lines 2.2, 3.3, 4.1 and 4.18 (Figure 4.4).In most of the selected epRILs the methylation levels were lower in the 500-1000 bp than in the 0-500 bp region.

(A)



(B)

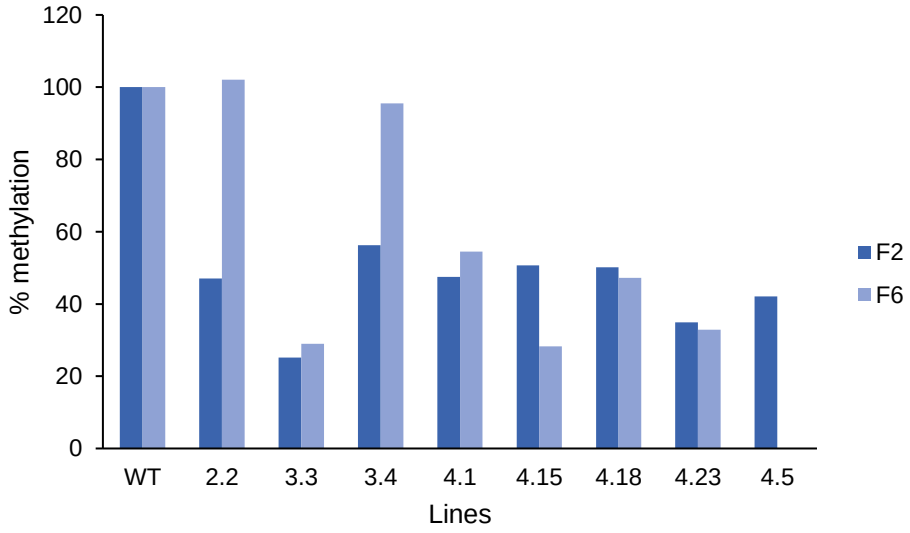
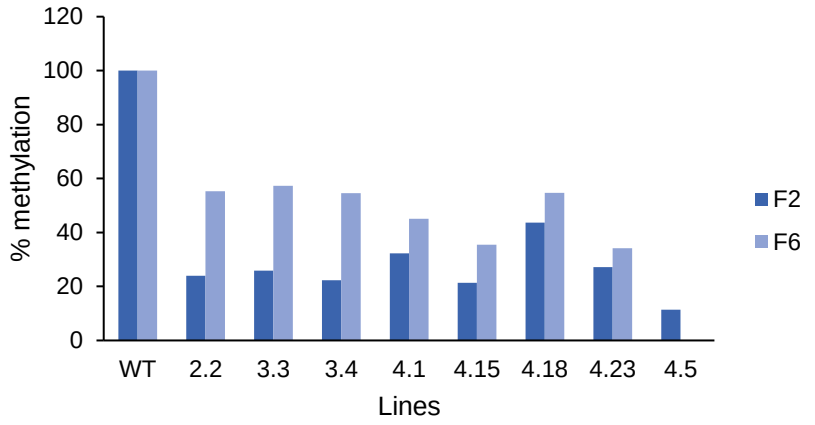


Figure 4. 4 Percentage of methylation for upstream region of *SlEZ1* gene. The regions analysed were (A) 0-500 bp and (B) 500-1000 bp upstream of the gene. Percentage of methylation in wild type was assumed 100%. Data was from generations F_2 and F_6.

In the regulatory region of the *SlEZ2* gene, comparison of wild type and epiRIL DNA methylation indicated reduced methylation levels in the 0--500 bp region (Figure 4.5). In the 500-1000 bp upstream region the levels of DNA methylation were lower than wild type in the F_6 generation for most of the lines except 2.2.

Alterations in the methylation in gene regulatory regions can result in changes in gene expression (Zhang *et al.*, 2018). The expression of the *SlEZ1* and *SLEZ2* genes was altered in the epiRILs in comparison to wild type (see Chapter 3). The genes encoded by *SlEZ1* and *SLEZ2* form the Polycomb Repressive Complex (PRC) which is a histone modifier along with histone acetyl transferase (HAT) and histone deacetylase (HDAC) (Gallusci *et al*., 2017). These genes have a major role in plant development.

(A)



(B)

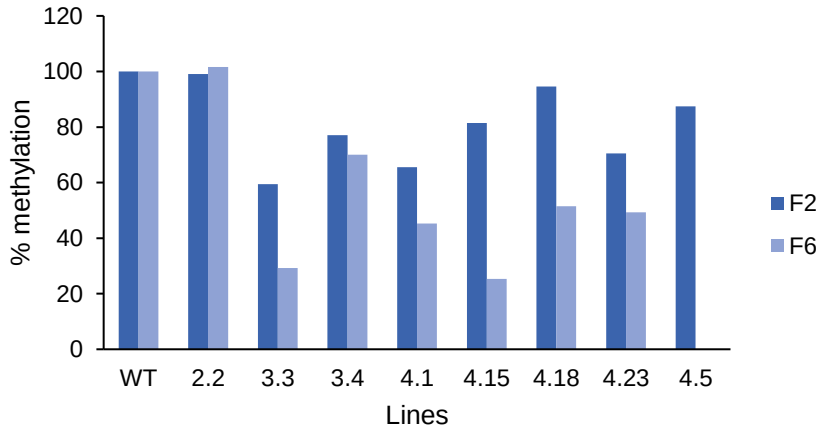


Figure 4. 5 Percentage of methylation for upstream region of *SlEZ2* gene. The regions analysed were (A) 0-500 bp and (B) 500-1000 bp upstream of the coding sequence of the gene. Percentage of methylation in wild type was assumed 100%. Data was from generations F_2 and F_6.

McrBC-qPCR analysis of the *TM8* promoter regions

Transgenic plant experiments indicate that the *TM8* gene is likely to be involved in anther, ovary and fruit development (Daminato *et al.*, 2014). The *TM8* gene regulatory region was investigated for alterations in DNA methylation using the McrBC assay (Figure 4.6). Analysis of the 0-500 bp region upstream of the coding sequence in the F_2 generation showed reduced levels of DNA methylation in all selected lines in comparison to wild type (Figure 4.6A). However, these differences were less pronounced by F_6. In the region 500-1000 bp upstream of the coding sequence many of the lines showed evidence of reduced DNA methylation levels in both the F_2 and F_6 generations (Figure 4.6B).

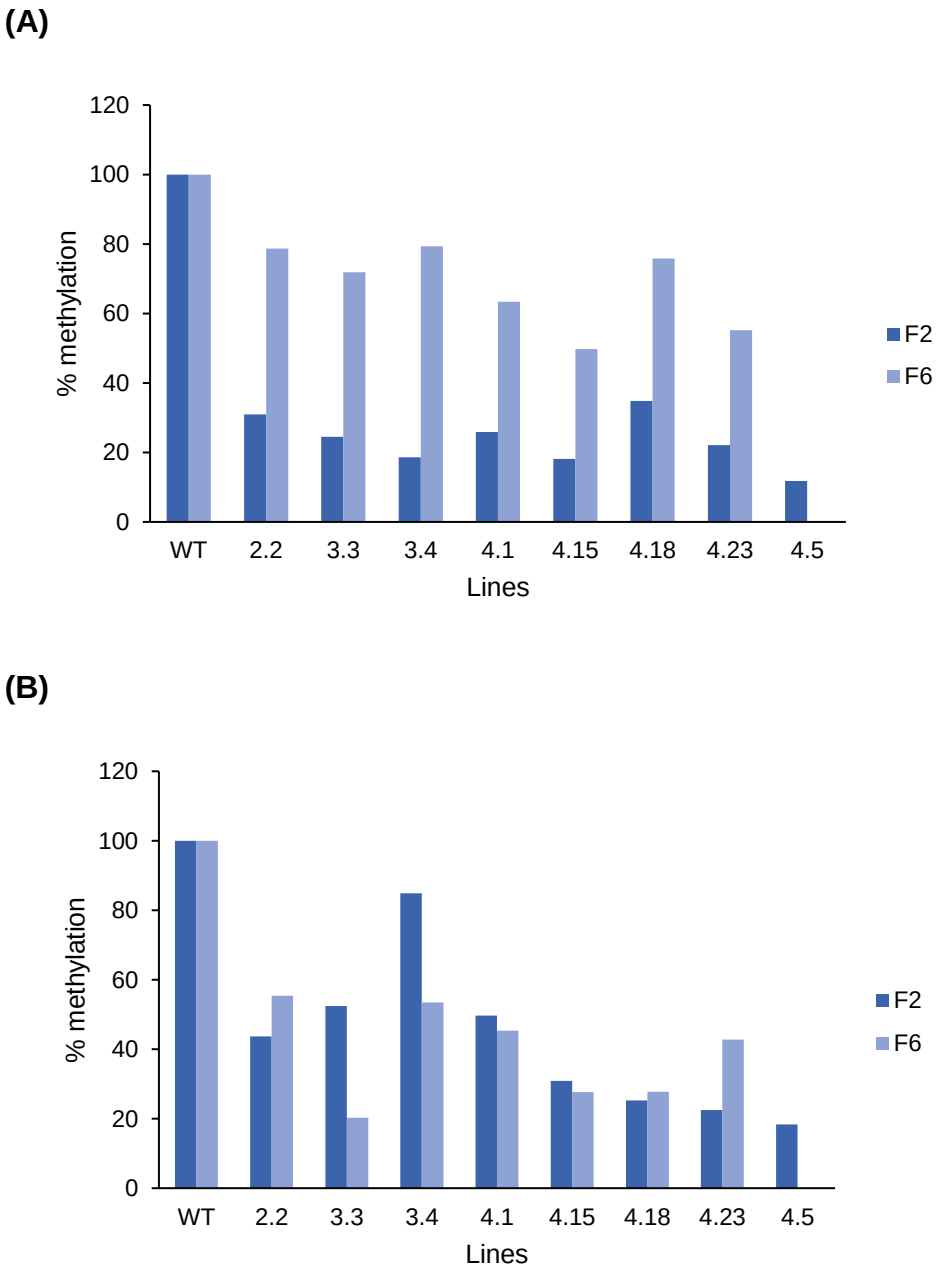


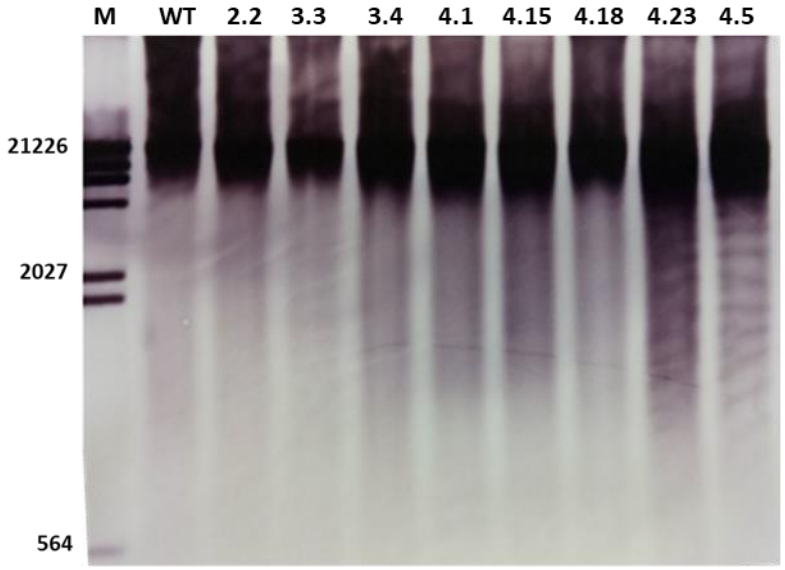
Figure 4. 6 Percentage of methylation for upstream region of *TM8* gene. The regions analysed were (A) 0-500 bp and (B) 500-1000 bp upstream of the gene. Percentage of methylation in wild type was assumed 100%. Data was from generations F_2 and F_6.

4.3.2 Investigating transposon methylation and movement by Southern analysis

The enzyme *Hpa*II was used to investigate the methylation status of tomato transposons. *Hpa*II cuts the sequence –C^1C^2GG-. The *Hpa*II enzyme is blocked when a single C or both are methylated while *Msp*I is blocked when C^1 is methylated (Teyssier *et al*., 2008). Southern analysis was used to investigate the methylation of *gypsy*-like retrotransposons in selected epiRILs in the F_2 generation.

To detect unequal volume during loading, gel pictures were taken after gel electrophoresis and before continuing the blotting process (Appendix 23). Using the *Hpa*II enzyme, the blots showed a strong band in the region of the 21226 bp marker, with evidence of smaller size bands in some of the epiRIL lines in comparison with wild type. This was especially evident in line 4.5 (Figure 4.7A). This result was consistent with the low methylation levels shown in the McrBC-qPCR analysis for candidate genes investigated in line 4.5. Enhanced digestion by *Hpa*II indicates reduced CG methylation (Teyssier *et al*., 2008).

(A)



(B)

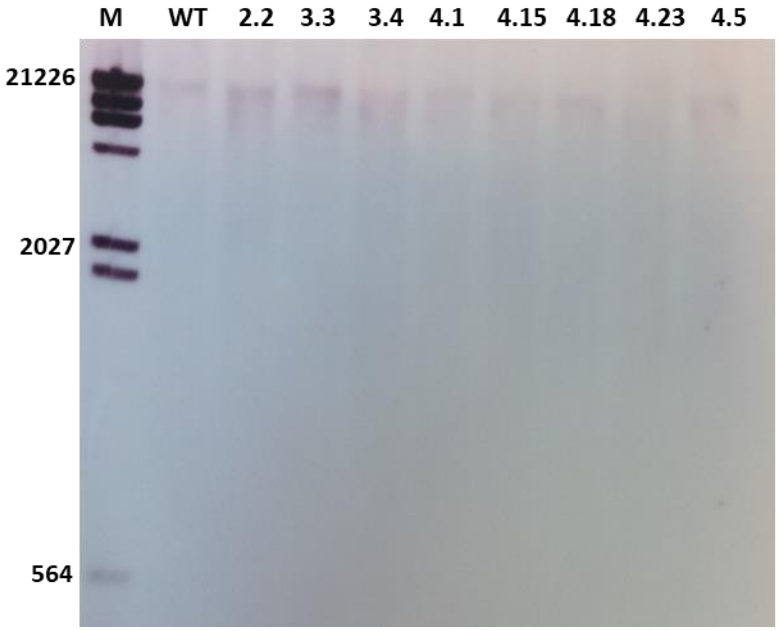


Figure 4. 7 DIG Southern Blot detection in generation F_2. Genomic DNA was digested with *Hpa*II and probed for (A) *gypsy*- and (B) *copia*-like retrotransposons. WT= wild type sample and the numbers above each lane indicate the lines from the epiRILs.

Probing the blot for *copia*-like retrotransposons (Figure 4.7B) produced a result that was more difficult to interpret. Only one band was apparent in wild type in the 21226 bp marker region although this was less intense than the result for the *gypsy*-like transposon. However, the banding pattern was weak and can be seen as a smear. The ability of *Hpa*II to digest the DNA seen in this result suggests reduced CG methylation in the CCGG sequence context. The smear pattern for the *copia* probe suggests that there is low copy number of the *copia* transposons or the probe design was not specific to the target sequence. The brighter smear pattern in the epiRILs could indicate a higher *copia* copy number compared to the wild type.

Detection of transposon movement using Southern analysis

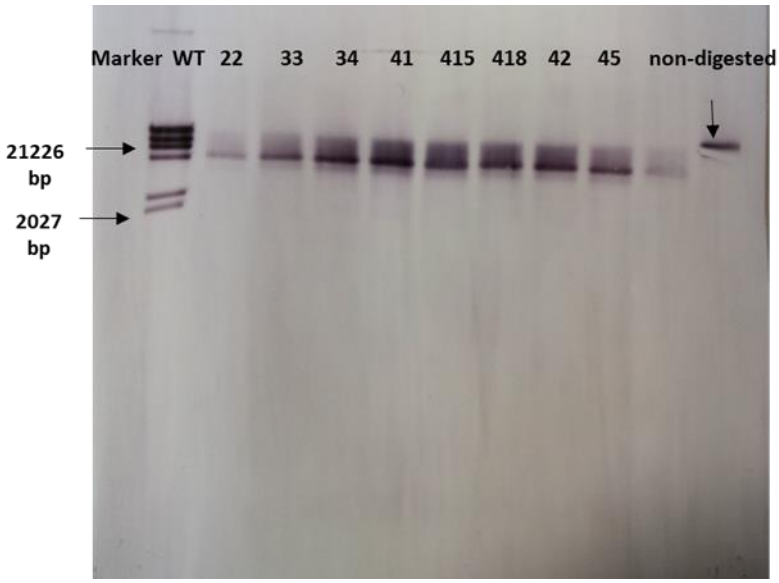
Gypsy and *copia* elements are retrotransposons and replicate via an RNA intermediate before reverse transcription to DNA and reintegration into the genome. The *copia* and *gypsy* elements are the two main families of LTR (Long Terminal Repeats) retrotransposon and the most abundant transposons found in plants (Kumar and Benetzen,1999; Galindo-Gonzalez *et al*., 2017). Hypomethylation in Arabidopsis is known to mobilize transposable elements and the first reported mobile transposon was a member of the *CACTA* family found in inbred *ddm1* (*Decrease in DNA Methylation1*) mutant (Ikeda and Nishimura, 2015).

For analysing transposon movement the wild type and epiRIL DNA was cut with either *Hind*III or *Eco*RI. *Hind*III digests generated a single visible band in the 21226 bp region when the blot was hybridised with the *gypsy* transposon probe in both the wild type and epiRIL lanes, but the band was

generally more intense in the epiRIL samples (Figure 4.8A). The *Hind*III / *gypsy* results indicated that only high molecular weight fragments of DNA contained sequences that hybridised with the gypsy probe, although there were more of these in the epiRILs

Using the *copia* probe, a smear was seen on the *Hind*III digest from approximately 5000 bp until 2000 bp (Figure 4.8B) and this was most apparent in the epiRIL lanes. The additional banding pattern / smear in the epiRIL lanes hybridised with the *copia* transposon probe could indicate a higher copy number of the *copia* transposon in the epiRILs.

(A)



(B)

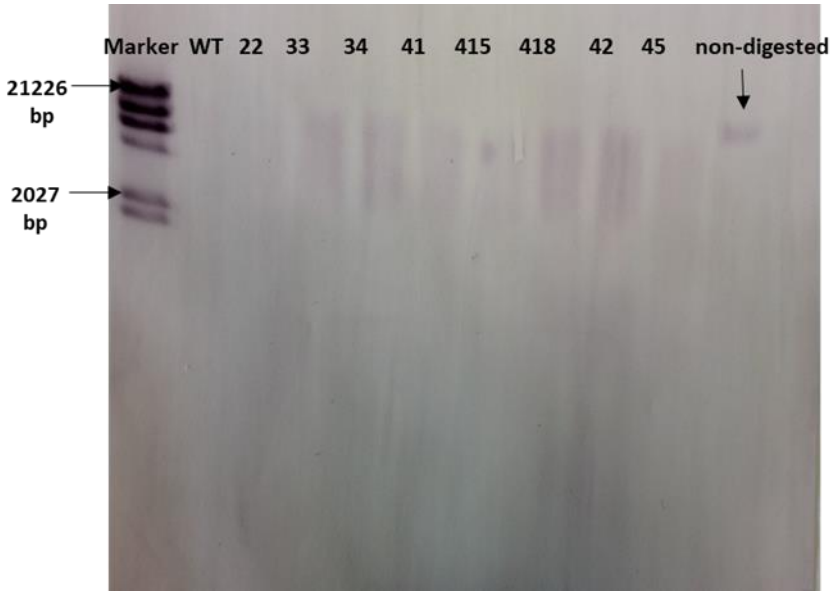


Figure 4. 8 DIG Southern blot using *Hind*III. The blots were probed for (A) *gypsy* and (B) *copia* transposons. The samples were from generation F_2. WT= wild type sample and the numbers above each lane indicate the lines from the epiRILs. Non-digested DNA was used as a control in the last lane.

Detection using *Eco*RI digested DNA samples was also used and probed with the *copia* transposon. The analysis was done to investigate if the *Eco*RI enzyme could produce bands with higher intensity and better digestion than using *Hind*II. However the result showed only smears in two regions (Figure 4.9). The first smear was in the region of approximately 5000 bp to 3000 bp. The second smear was in the region of approximately 2000 bp to 1500bp. The patterns in the epiRILs are more intense than in the wild type. However, more investigation is needed using replicated samples to confirm the effect of differences in the epigenome on transposons.

RNA analysis (Chapter 3) showed that the expression of transposable elements was higher in the epiRILs especially the gypsy and copia type. These two types of TEs are Class I TEs which use the ‘copy and paste’ mechanism. The increase in expression could indicate the hypomethylation of TEs which caused their activation. Further investigations on the changes in methylation of TEs and its effect on TE activation are needed for the tomato epiRILs. More sample replication is needed to confirm the changes.

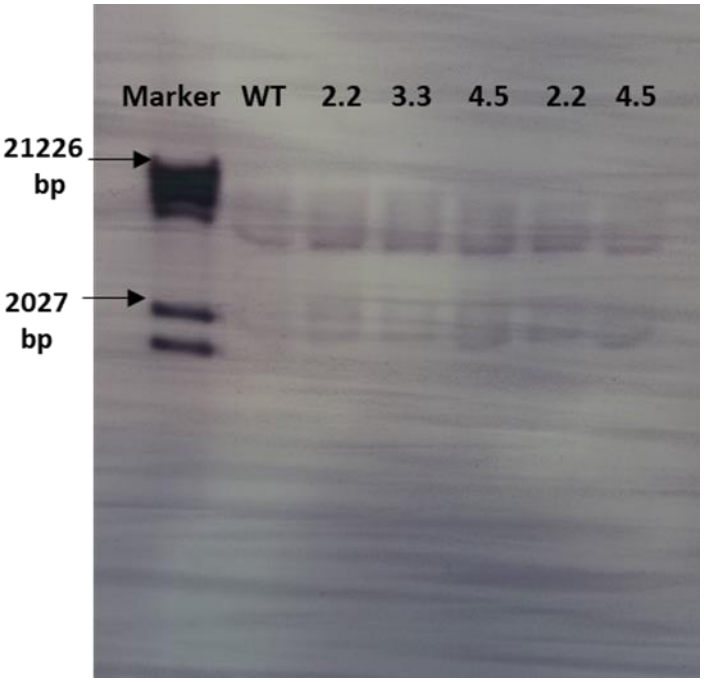


Figure 4. 9 DIG Southern blot of *Eco*RI digested DNA. The blot was hybridised with a *copia* transposon probe. Samples were from generation F_2. WT= wild type sample and the numbers above each lane indicate the lines from the epiRILs.

4.4 Conclusion

The McrBC assays provided evidence that the *MET1* silencing in the parental line of the tomato epiRILs had affected the levels of DNA methylation in the regulatory elements of the genes investigated. Using the McrBC-qPCR method, there was evidence of hypomethylation in at least five genes of the epiRILs compared to wild type. The lower percentage of methylation was found in two generations showing that the reduction of methylation was inherited from generation F_2 to F_6 in several of the epiRILs.

The transposon analysis indicated that DNA methylation of the selected *gypsy* and *copia* transposons region was not significantly different when analysed using *Hpa*II restriction enzyme. The CG methylation in the CCGG context (which is the *Hpa*II target site) did not show significant difference between some of the epiRILs and the wild type in the F_2 generation. Analysis using *Eco*RI and *Hind*III also did not show differences in the banding patterns of the selected epiRILs. Limitations on which transposons could be analysed in this project caused difficulty in choosing the correct target site and probe design in the Southern blot experiment. Further studies should be undertaken using mapping data from the generated epiRIL population to detect transposons movements.

Epigenetic control of retrotransposons has been investigated using epimutants of Arabidopsis. In the hybrid of *met1* and wild type plant, a *copia*-type retrotransposon showed reinsertion in the genome. Transcriptional suppression of the retrotransposon was due to CG methylation supported by RNA-directed DNA methylation (Mirouze *et al*., 2009).

In the Arabidopsis epiRIL population generated by wild type and *met1* parental lines, transposition of a *CACTA* transposon was observed in 28% of the epiRILs. There was also progressive accumulation of *CACTA* movements across the generations (Reinders *et al*., 2009). Mobilization of transposons has also been documented in *ddm1* Arabidopsis mutants. In the Arabidopsis epiRILs from *ddm1* mutant and wild type parental lines, showed the mobilization of *CACTA* and *MULE* transposons (Johannes *et al*., 2009).

The discovery of a *copia*-type transposon (named *EVD/Evade*) in Arabidopsis showed that epigenetic control was individualized for a particular type of transposon. Other transposons such as *CACTA*, non-LTR retrotransposons and other families remain immobile in the *met1* Arabidopsis mutant (Mirouze *et al*., 2009). In the tomato epiRILs, the change in methylation seems to affect both *gypsy* and *copia* retrotransposon.

McrBC digestion is an economical and effective way to detect methylated regions of DNA, but it cannot detect changes at single nucleotide resolution and for this is whole genome or targeted bisulfite sequencing is required. However, these experiments are very costly. We therefore, undertook bisulfite sequencing on only a small number of samples and these experiments are described in Chapter 5.

CHAPTER 5 DETERMINATION OF DNA METHYLATION CHANGES IN SELECTED epiRILS BY BISULFITE SEQUENCING

5.1 Introduction

In Chapter 4, analysis of the regulatory regions of selected genes using the McrBC enzyme assay and Southern blotting indicated a change in DNA methylation in the epiRILs. To confirm these results further work was undertaken to establish the patterns of DNA methylation in selected epiRILs. This involved two approaches, (1) targeted bisulfite sequencing and (2) whole genome bisulfite sequencing (WGBS). The aim of the work described in this Chapter was to link changes in the phenotype of the epiRILs with altered gene expression and changes in DNA methylation.

Bisulfite sequencing can be used to detect methylation changes in targeted regions of particular genes (Manning *et al*., 2006; Masser *et al*., 2013) or across whole genomes (Clark *et al*., 1994; Suzuki *et al*., 2018). Targeted sequencing was undertaken to detect methylation in the upstream regions of *FW2.2*, *SlEZ1*, *SlEZ2* and *TM8* genes selected in Chapters 3 and 4. PCR products were generated and either sequenced directly or cloned and then sequenced to detect changes in methylation (Figure 5.1).

In sequencing bisulfite treated DNA, the reverse strand of the original DNA can sometimes be sequenced. Demethylated cytosines in the original forward strand can then be detected by Adenine (A) showing in the aligned sequencing result (Figure 5.2). Demethylated cytosines in the 3' strand of the original DNA convert to Uracil (U) after the bisulfite reaction. After PCR amplification of both strands, the complementary strand of the demethylated

cytosines (converted to U) will be adenine (A). The differences in the final DNA strands are important in analysing sequencing results for bisulfite treated DNA.

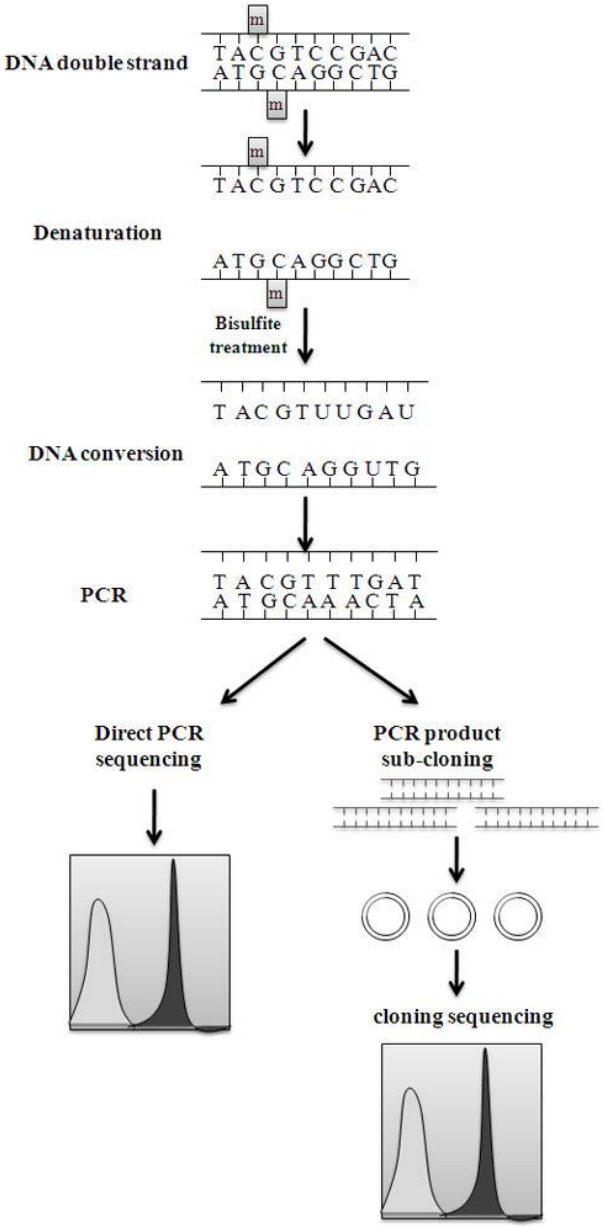


Figure 5. 1 Methylation Analysis using Bisulfite Sequencing. After treatment with sodium bisulfite, unmethylated cytosine residues are converted to uracil and 5-methylcytosine (5mC) remains unaffected. After PCR amplification, uracil residues are converted to thymine. DNA methylation status can be determined by direct PCR sequencing or cloning sequencing (Li and Tollefsbol, 2011).

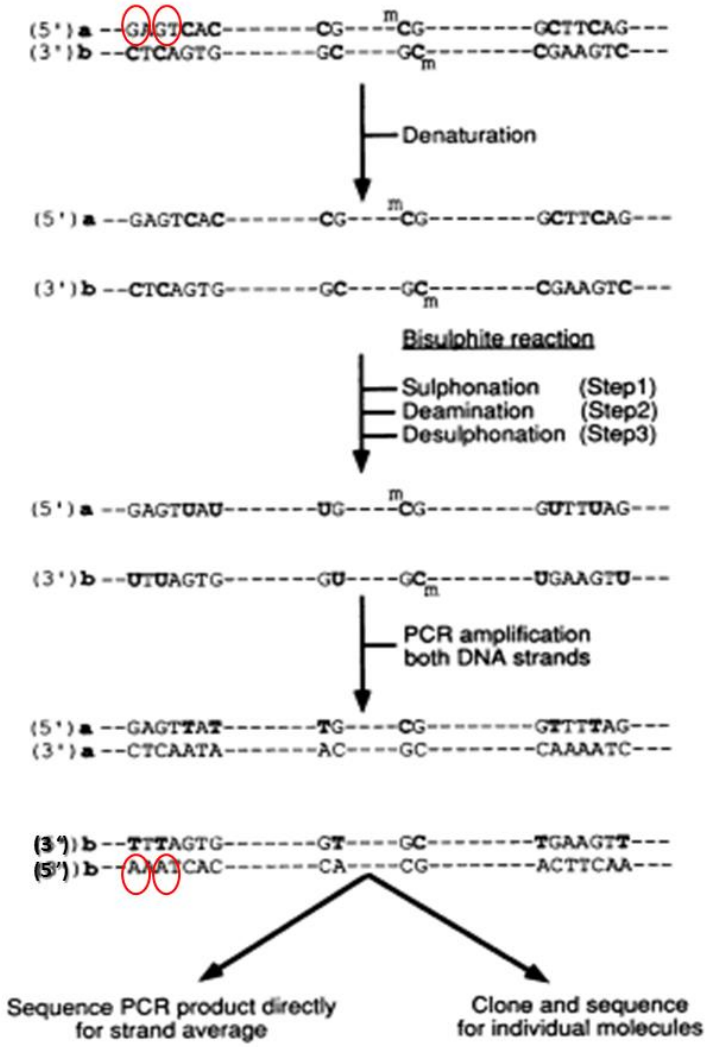


Figure 5. 2 Procedure of bisulfite genomic sequencing. The two complementary strands in the original DNA are labelled (a) and (b). Cytosine residues and their corresponding uracil and thymine conversion products are shown in bold type. After the bisulfite reaction, the two DNA strands (a) and (b) are no longer complementary and therefore can be amplified independently. The guanine base is circled red, which shows its conversion to adenine after PCR amplification. The cytosine in the forward strand is unmethylated, so changed to T after PCR (figure adapted from Clark *et al*., 1994)

WGBS is the gold standard for detecting changes in methylation at the single nucleotide level and has been used to map the epigenome of both Arabidopsis (Mathieu *et al*., 2007; Lister *et al*., 2008) and tomato (Zhong *et al*., 2013). In the tomato, we wanted to use WGBS to map the mosaics of DNA methylation in the epiRILs and the epigenetic changes that occurred due to reduced *MET1* expression in the initial transgenic lines. The financial constraints in the current project have restricted the number of epiRILs that could be subjected to WGBS to a very limited number. By using Sanger sequencing and WGBS, in this part of the investigation, the hypothesis was differentially methylated regions should be detected in selected promoter regions and in the genomes of selected epiRILs.

5.2 Materials and Methods

5.2.1 DNA extraction

DNA was extracted from leaf tissue of plants that were 6 weeks old using the ISOLATE II Plant DNA Kit (Bioline). A more detailed description of the method can be found in Chapter 4.

5.2.2 Bisulfite Conversion

The EpiTect® Bisulfite Kit (Qiagen) was used for bisulfite conversion following the manufacturer's protocol. Information on the buffers used in this protocol is presented in Appendix 24. The reactions were optimised for 1 ng to 2 µg of total DNA and undertaken in a total volume of 140 µL (in a 200 µL PCR tube). To each sample, 85 µL of bisulphite mix and 35 µL of DNA protect buffer were added and the total volume made to 140 µL with RNase free water.

The bisulfite thermal cycling program provides the incubation steps necessary for thermal DNA denaturation and subsequent sulfonation and cytosine deamination. The bisulfite conversion involved an initial denaturation for 5 min at 95^{o}C, incubation for 25 min at 60^{o}C and denaturation for 5 min at 95^{o}C. There was then a further round of incubation for 85 min at 60^{o}C, denaturation for 5 min at 95^{o}C and incubation for 175 min at 60^{o}C before returning to room temperature.

After the bisulfite conversion, the PCR tubes were centrifuged briefly and the reaction mix transferred to clean 1.5 mL microcentrifuge tube. Buffer BL was prepared by adding 6.2 µL of carrier RNA solution for each 620 µL of buffer BL. An amount of 560 µL of freshly prepared Buffer BL was then added to the bisulfite converted samples and these were vortexed and centrifuged briefly. The samples were then loaded onto the EpiTect spin columns, which were placed in new collection tubes.

The spin columns were then centrifuged at 13000 rcf or maximum speed for 1 minute. The flow through was discarded and the spin columns were placed back in the collection tubes. Then 500 µL Buffer BW was added to each spin column and they were centrifuged at 13000 rcf for 1 minute. The flow through was discarded and the spin columns placed back into the collection tubes. Then 500 µL Buffer BD (contained sodium hydroxide) was added to each spin column and the mixture incubated for 15 minutes at room temperature (15-25 ^{0}C). The spin columns were then centrifuged at 13000 rcf for 1 minute.

The flow through was discarded and the spin columns placed back into the collection tubes. The columns were then washed with 500 µL Buffer BW and centrifuged at 13000 rcf for 1 min. This step was repeated once. The columns were then spun without any buffer to remove residual liquid. For the final step, the spin columns were placed in clean 1.5 mL microcentrifuge tubes and 100 µL Buffer EB was placed in the centre of each of the tubes. The purified DNA was then eluted by centrifugation at 13000 rcf for 1 minute. The eluted DNA was stored at -20 ^{0}C.

5.2.3 Primer design

Primers used to amplify bisulfite treated DNA for target sequencing were designed from the 1000 bp region upstream of the target genes. The positions of the target regions were: >Solyc02g090730.2 SL2.50ch02: 52252557..52253347 (*FW2.2*), >SL2.50ch01: 78463681..78464680 (*SlEZ1*), >SL2.50ch03:8958098..8959097(*SlEZ2*) and SL2.50ch03:61436784..61440333 (*TM8*). Degenerate primers were designed by replacing cytosines with Y (in the forward primers) and replacing guanine with R (in the reverse primers) (How-Kit *et al*., 2010). This was done to reduce bias towards only methylated cytosines being amplified (Candiloro *et al*., 2017). The primers used for the amplification of bisulfite treated DNA are presented in Table 5.1.

5.2.4 PCR amplification of bisulfite treated DNA

PCR amplification was undertaken using the Gotaq Colorless Reaction Mix (Promega). The reaction mix consisted of 6.25 µL Gotaq Colourless Master Mix (2X), 1.25 µL of forward and reverse primers (1µM), 50 ng DNA and nuclease free water was added until the total volume was 12.5 µL. The PCR amplification involved one cycle of initial denaturation for 5 min at 95°C, 30 cycles of denaturation (95°C, 30 sec), annealing (58°C, 30 sec) and extension step (72°C, 30 sec). Then followed by a final extension of 72°C for 5 min and held at 4°C.

Table 5. 1 Primers used in amplification of bisulfite treated DNA

Primers	Direction	Gene target	Sequence (5'-3')
FW-F8	Forward	*FW2.2*	YATTTAYYTYTTGAATAGGAYAGTAA
FW-R8	Reverse	*FW2.2*	CATTTTTTAACTTTATCAACACTTC
FW-F11	Forward	*FW2.2*	GAAGAAAGTTTGATTAAATTG
FW-R11	Reverse	*FW2.2*	TCAARAAATRTTRTARTRTAAAA
EZ1-F2	Forward	*SlEZ1*	GGTTTAATTGGTTATTTTATA
EZ1-R2	Reverse	*SlEZ1*	ATAAAAAAAATATCAACATCTAAAACAAT
EZ1-F3	Forward	*SlEZ1*	ATTGTTTTAGATGTTGATATATTT
EZ1-R3	Reverse	*SlEZ1*	CATCTTATCATTTATCCTTC
EZ2-F2	Forward	*SlEZ2*	AGGTYGTATTAYTYATTTTTYTTGTATT
EZ2-R2	Reverse	*SlEZ2*	CTCTTTTRTCACCATTACTTACTACCCA
EZ2-F3	Forward	*SlEZ2*	TGGGTAGTAAGTAATGGTGAYAAAAGAC
EZ2-R3	Reverse	*SlEZ2*	TTCCARCTATRTTCCATCAATACRA
TM8-F1	Forward	*TM8*	YGATTTAAYTTATTTTAAGAGAATTTTA
TM8-R1	Reverse	*TM8*	TRARTTCTRATCTTTCCAARTTACTTT
TM8-F2	Forward	*TM8*	YGTTATYTGTTTTTAGTGTATTGGAG
TM8-R2	Reverse	*TM8*	TRATATTTTTAATTTCATTTARATTCR

5.2.5 PCR purification

The QIAquick PCR Purification Kit (Qiagen) was used for PCR product purification. Five volumes of Buffer PB was added to one volume of each PCR sample and the solution was mixed by pipetting. To bind the DNA, the mixed samples were placed in QIAquick spin columns within 2 mL collection tubes. The columns were then centrifuged for 1 minute at 13,000 rcf. The flow through was discarded and the columns placed back in the same collection tubes.

To wash the DNA, 0.75 mL Buffer PE was added to the columns and centrifuged at 13,000 rcf for 1 minute. The flow through was discarded and the columns placed back in the tubes. The centrifugation step was repeated to completely dry the samples. To elute the DNA, 50 µL Buffer EB was added to the centre of the column membrane and incubated at room temperature for 1 minute. The columns were centrifuged at 13,000 rcf for 1 minute. The NanoDrop1000 Spectrophotometer (Thermo Fisher Scientific Inc.) was used to measure the concentration of DNA. The purified PCR product was stored at -20 ^{0}C.

5.2.6 Sanger sequencing of PCR product

For DNA sequencing, the purified PCR product was diluted to 5 ng/µL and a volume of 15 µL was used for Sanger sequencing. This was undertaken by Eurofins Genomics (Germany). The results were then analysed using the Clustal Omega Multiple Sequence Alignment (www.ebi.ac.uk) program by comparing the sequence to the original genomic DNA sequence.

5.2.7 Cloning

For cloning of PCR products, the pGEM®-T Vector System (Promega) was used. For the ligation step, 2 µL of the purified PCR product (see section 5.2.5) was placed in a 200 µL tube with 5 µL of rapid ligation buffer, 1 µL of the PGEM-T Vector, 1 µL of the T4 DNA Ligase (3 Weiss Units/ µL), and nuclease free water to a final total volume of 10 µL. The reaction was mixed by pipetting and incubated at 4 ^{0}C overnight.

For the transformation protocol, LB plates with ampicillin/IPTG/X-Gal were prepared. A volume of 100 mL of LB media was added with 100 µL of ampicillin, 500 µL IPTG and 100 µL X-Gal. The volume of media was adjusted according to the number of plates needed. Three plates were needed for each ligation reaction.

The tubes containing the ligation reactions were centrifuged to collect the contents at the bottom of the tubes. A 1.5 mL microcentrifuge tube was prepared for each ligation reaction. Frozen tubes of JM109 High Efficiency Competent Cells (Promega) were placed in an ice bath until just thawed (about 5 minutes). A volume of 2 µL of ligation reaction was added to the 1 mL tube and also 50 µL of the thawed JM109 High Efficiency Competent Cells. The mixture was gently mixed by flicking the tube and then the mixture was placed on ice for 20 minutes. The reaction mix was then placed at 42 ^{0}C in a heating block for 50 seconds and then placed immediately in ice for 2 minutes. A volume of 600 µL LB broth was added to the tubes and incubated at 37 ^{0}C for 2 hours. From the reaction mix, 100 µL of each transformation was spread onto the LB/ampicillin/IPTG/X-Gal plates. The plates were incubated at 37 ^{0}C for 24 hours.

White colonies that formed after overnight incubation were isolated and placed on a new plate of LB/ampicillin/IPTG/X-Gal. Each selected colony was then screened for an insert by PCR amplification using MangoMix (Bioline) mastermix for the amplification. The PCR reaction mixture consisted of 12.5 µL of MangoMix Buffer, 1 µL of forward and reverse primers (1µM) and 9.5 µL nuclease free water.

The PCR amplification involved one cycle of initial denaturation for 10 min at 95°C, 40 cycles of denaturation (95°C, 40 sec), annealing (50°C, 40 sec) and extension (70°C, 60 sec). Then followed by a final extension of 72°C for 10 min and held at 4°C. The amplified products were analysed on a 1% (w/v) agarose gel run in 0.5X TAE buffer at 100 Volts for 60 minutes. Colonies with inserts were detected by the presence of a band at 534bp (F11 primer) and 526bp (EZ1-F2 primer). Five clones from each transformation were picked for plasmid extraction.

5.2.8 Plasmid extraction

For plasmid extraction, the colonies with the insert were picked using pipette tips and placed in 3 mL of LB Broth containing 3 µL ampicillin. The reaction was then incubated at 37 ^{0}C for 24 hours. The cells were harvested by pipetting 1 mL of the reaction into 1.5 mL microcentrifuge tube centrifuged at 13.000 rcf for 1 minute and the supernatant was discarded. This step was repeated until all of the LB Broth have been used from the incubation step.

The GenElute™ Plasmid Miniprep Kit (Sigma) was then used for the plasmid extraction. The bacterial pellet was resuspended in 200 µL of Resuspension Solution containing 1.3 µL of RNase-A. The pellets were

vortexed to thoroughly mix the samples. The cells were lysed by adding 200 µl of the Lysis Solution and the mixture was inverted gently (6–8 times) until the solution became clear and viscous. Cell debris was then precipitated by adding 350 µL of the Neutralization/Binding Solution. The tubes were again gently inverted 4–6 times. The cell debris was then pelleted by centrifugation at ≥12,000 rcf for 10 minutes. Cell debris, proteins, lipids, SDS, and chromosomal DNA were pelleted from the solution as a cloudy, viscous precipitate.

The GenElute Miniprep Binding Columns were placed in microcentrifuge tubes. A volume of 500 µL Column Preparation Solution was added to each miniprep column and centrifuged at ≥12,000 rcf for 1 minute. The flow-through liquid was then discarded. The cleared lysate from the previous step was then transferred to the column and centrifuged at ≥12,000 rcf for 1 minute. The flow-through liquid was discarded. A volume of 750 µL Wash Solution (diluted with 96% ethanol according to the manufacturer's instruction) was added to the column and centrifuged at ≥12,000 rcf for 1 minute. The columns were transferred to fresh collection tubes and 100 µL of nuclease free water pipetted into the centre of the column and centrifuged at ≥12,000 rcf for 1 minute. The eluted DNA was stored at –20 °C.

5.2.9 Sanger sequencing of clones

For the sequencing of the plasmid extract, 20 µL of the eluted DNA containing 50-100 µL of plasmid was sent for sequencing. Sequencing analysis was carried out by Eurofins Genomics (Germany). The universal M13 forward primer was used in the sequencing. The result was then

analysed using Clustal Omega Multiple Sequence Allignment (www.ebi.ac.uk) program by comparing the sequence to the original genomic DNA sequence.

5.2.10 Whole genome bisulfite sequencing

Whole genome bisulfite sequencing was undertaken working with colleagues at INRA, France, using leaves tissues from T_0 and F_5 generations. The samples were: wild type, lines 2.1 and 4.6. Samples from generation T_0 (wild type and RNAi lines) were provided by INRA and samples from F_5 were provided by Nottingham (wild type and epiRILs). Two biological replicates were used for each line. Bisulfite treatment, library preparation, and whole-genome sequencing were performed by BGI Genomics Company (Hong Kong) and the bioinformatics analysis undertaken by Professor Nicolas Bouché at Institut Jean-Pierre Bourgin, UMR1318 INRA-AgroParisTech-ERL3559 CNRS, Versailles Cedex France, in collaboration with Professor Philippe Gallusci, INRA, Bordeaux, France.

5.3 Results and Discussion

5.3.1 DNA methylation of selected gene regulatory regions

Enzymatic analysis using McrBC (Chapter 4) indicated changes in tmethylation in the promoter regions of *FW2.2*, *SlEZ1*, *SlEZ2* and *TM8*. In the next step, we focused in a region 1000 bp upstream of the transcription start site (TSS) of the target genes. Preliminary analysis of the regions by sequencing PCR products indicated changes in methylation in some target regions (Appendix 10 and 11). For some genes, the methylation changes in the promoter regions were confirmed by directly sequencing PCR products and cloning before sequencing.

The *fw2.2* Locus

For the upstream region of this locus (Solyc02g090730) several rounds of primer optimisation were needed and only two sets of primers were used for sequencing of this region. The two primers used in this investigation covered the regions -233 until -547bp (primer F8) and -97 until -431bp (primer F11) upstream of the *fw2.2* TSS. The region covered by primer F11 overlaps the region covered by primer F8 and the regions covered by the two sets of primers shows elevated levels of DNA methylation in wild type tomato (Figure 5.2).

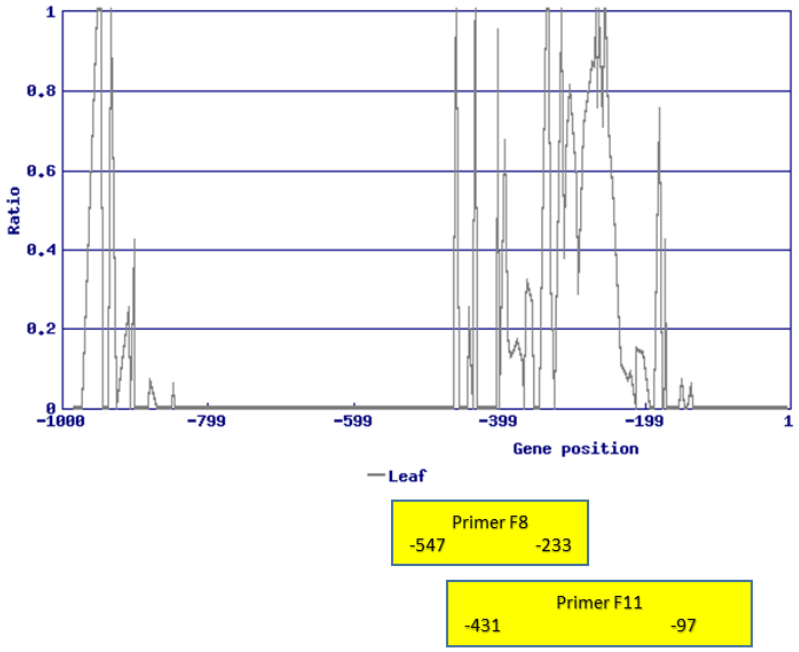


Figure 5. 2 Region covered by primers in 1000 upstream region of *fw2.2* (Solyc02g090730) locus. The yellow boxes represents the region covered by the primers. The ratio indicated the methylation ratio of that region (ratio= 5mC/(5mC+C)). Methylation ratio data obtained from http://ted.bti.cornell.edu/cgi-bin/epigenome for the wild type Ailsa Craig genome. The data was taken from leaf tissue.

Primer F8 yielded inserts that proved difficult to sequence due to the presence of a long poly T region (Appendix 25). The targeted bisulphite sequencing in the region covered by primers F11 was more successful. Sequence of PCR products indicated changes in methylation for lines 3.3 and 4.2 (Appendix 26). To detect methylation at single cell level, cloning of the PCR product was undertaken for lines 3.3 and 4.2 to confirm the change in methylation.

Analysis of the sequence data indicated some demethylation in the target region. This was demonstrated by a change in ‘G’ in the wild type sequence to ‘A’ in the epiRILs. In bisulfite treated DNA, guanines are sometimes read as adenine (Clark *et al*., 1994). This occurs when the reverse strands is read during sequencing (Figure 5.2). The guanine which converts to adenine after bisulfite treatment is from the reverse strand of an unmethylated cytosine.

There were 22 cytosine nucleotide positions which showed demethylation of the *FW2.2* regulatory region in line 3.3 (Figure 5.4) when PCR products and clones were aligned to the coding strand of the genomic DNA. The result indicated that the change in cytosine methylation was not in the CG context as was expected due to the reduced MET1 expression in the parental line.

Guanines in the wild type were detected as adenine in the clones, indicating the sequenced strand was the new DNA strand complementary to the original non-coding strand which was sequenced (Clark *et al*., 1994). The sequence alignment showed that not all of the nucleotides were the same sequence as the wild type. This might be caused by several limitations of the Sanger sequencing such as the poor quality of the first 15-40 bases due to the primer binding and some cloning vector sequence can find its way to the final sequence (Shaffer, 2019). Cytosine methylation in plants can be found in CG, CHG and CHH sequence context. In this result, the demethylation of of the *FW2.2* regulatory region in line 3.3 was shown to be mostly in CHH context (Table 5.2).

Misalignment of clones 2, 3, 5 and 6 in Figure 5.4 suggests the sequenced strand was the original non-coding strand. This was confirmed when alligment was undertaken using the original non-coding strand (Appendix 27).

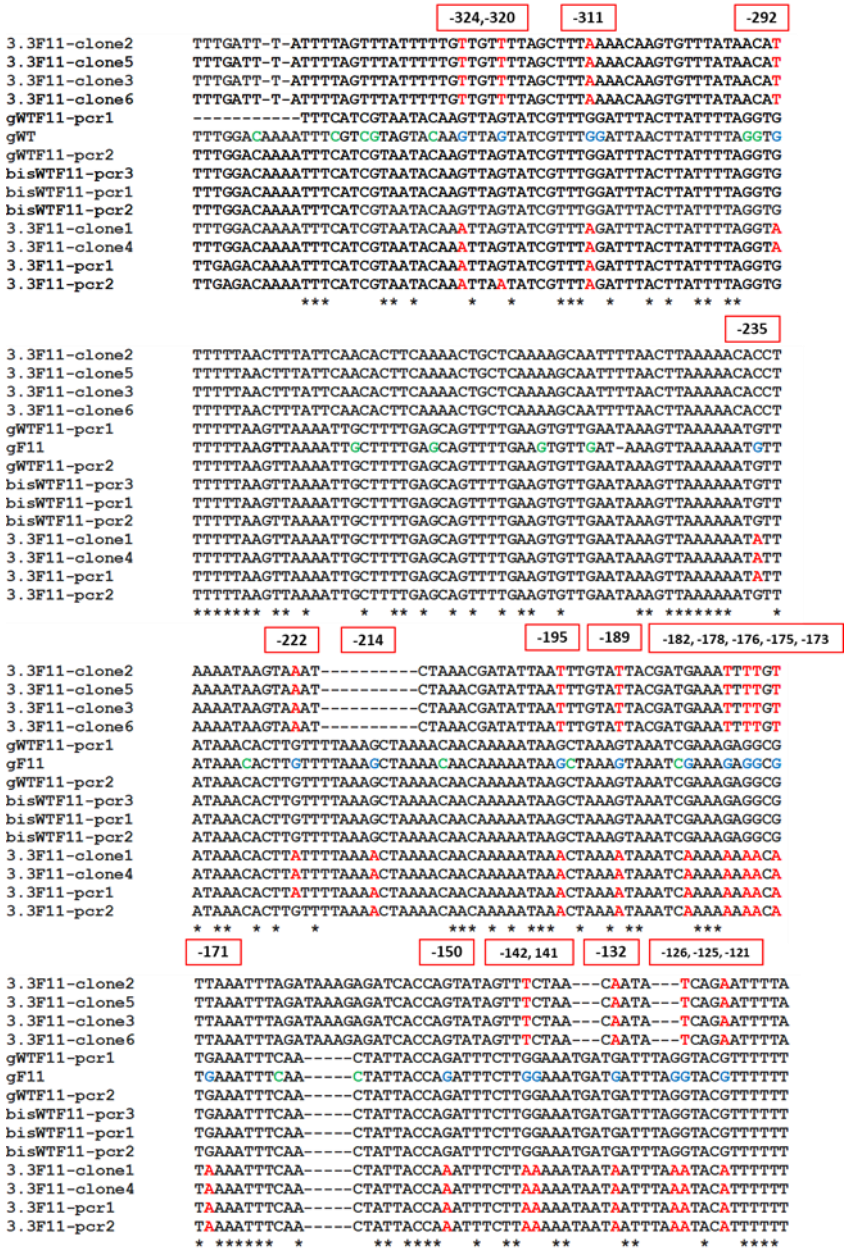


Figure 5. 3 Location of demethylation based on sequence alignment of wild type and clones of line 33-primer F11. The region covers -409 until -115 upstream of *fw2.2* transcription start site (TSS). The numbers in red box indicate the position of demethylation upstream of the gene TSS. Guanines (G) coloured blue show demethylation in both PCR product and clones. Green coloured G show demethylation only in clones suggesting it is the sequence of the reverse DNA strand. The rows show the sequence alignment of bases A (adenine), T (thymine), G (guanine) and C (Cytosine). (*) indicate the same base aligned in that position for all samples. WT= wild type

Table 5.2 shows the demethylated positions were most of the CHH context . This indicated that the changes due to *MET1* knockdown in the parental lines affected the methylation not only in CG sequence context. The effect of reduced MET1 on the methylation levels of CHH was also observed in Arabidopsis. However, investigations showed both increased and decreased of non-CG methylation. Analysis on the *met1* Arabidopsis mutant at the 5S rDNA showed an increased level of CHH methylation in the first generation (Mathieu *et al*., 2007) while at some repetitive sequence there was a lower methylation level for CHG sites (Kenkel *et al*., 2003).

Cloning of line 4.2 also showed demethylation in several sequence positions (Figure 5.5). However, the changes in methylation were seen in less than 5 clones which is the minimum number of clones for high confidence in the result (Li and Tollefsbol, 2011). The difficulty in sequencing line 4.2 using primer F11 was also indicated by the sequence of the wild type which showed missing regions between the bases.

Table 5. 2 The CG/CHG/CHH sequence context for the demethylation in the *fw2.2* upstream region in line 3.3

No.	Position upstream of *fw2.2* transcription start site	G position in the sequenced strand	Sequence in the reverse strand	CG/CHG/CHH context (H= A, C or T)
1	-324	GTA	CAT	CHH
2	-320	GTA	CAT	CHH
3	-311	GGA	CCT	CHH
4	-292	GTT	CAA	CHH
5	-232	GTT	CAA	CHH
6	-222	GTT	CAA	CHH
7	-214	GCT	CGA	CHH
8	-195	GCT	CGA	CHH
9	-189	GTA	CAT	CHH
10	-182	GAA	CTT	CHH
11	-178	GAG	CTC	CHH
12	-176	GGC	CCG	CHG
13	-175	GCG	CGC	CHH
14	-173	GTG	CAC	CHH
15	-171	GAA	CTT	CHH
16	-150	GAT	CTA	CHH
17	-142	GGA	CCT	CHH
18	-141	GAA	CTT	CHH
19	-132	GAT	CTA	CHH
20	-126	GGT	CCA	CHH
21	-125	GTA	CAT	CHH
22	-121	GTT	CAA	CHH

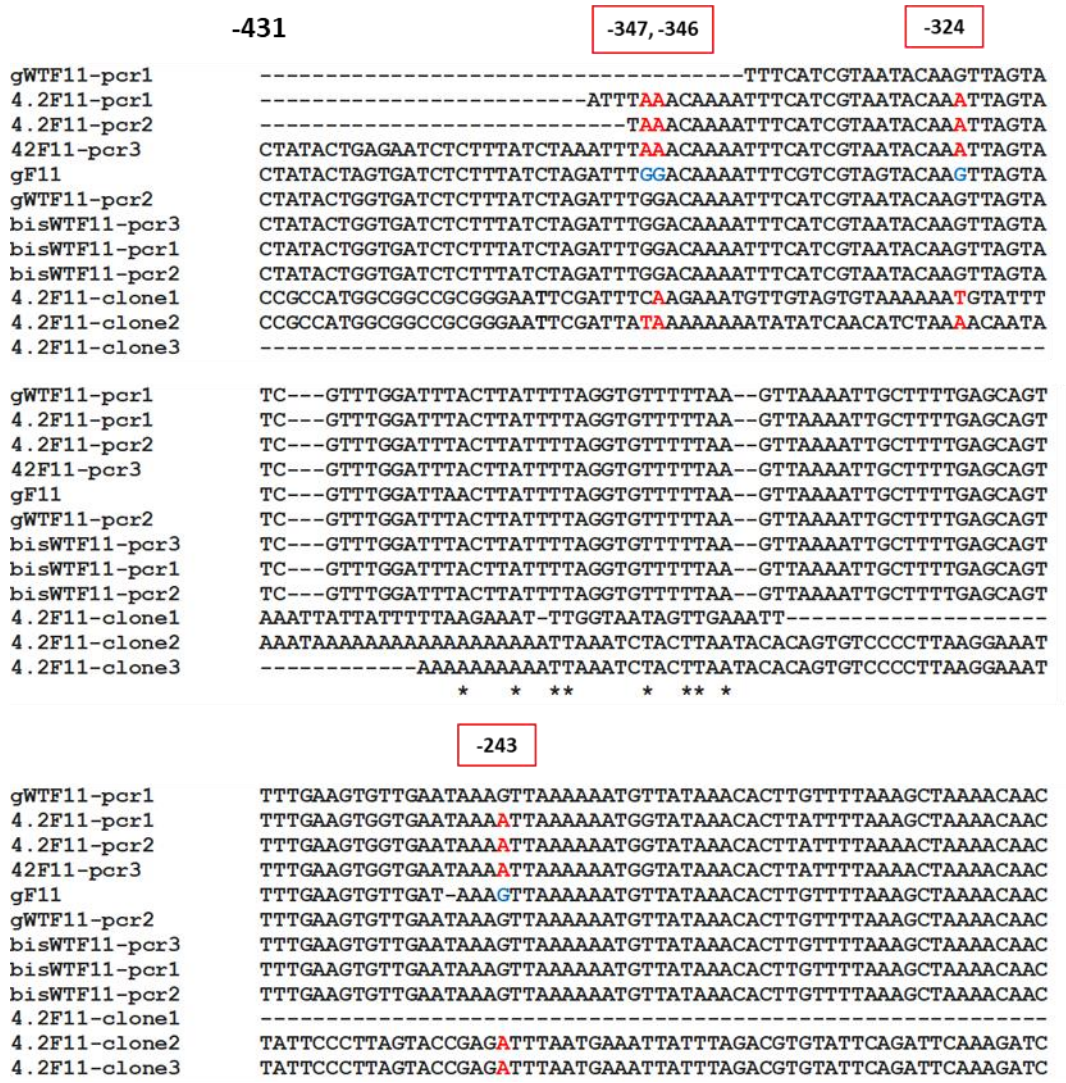


Figure 5. 4 Location of demethylation in clones of line 4.2-primer F11. The region covers -431 until -194 upstream of *FW2.2* transcription start site (TSS).The numbers in red box indicate the position of demethylation upstream of the gene TSS. Guanines (G) coloured blue in the original DNA sequence show demethylation in both PCR product and clones.The rows show the sequence alignment of the bases A (adenine), T (thymine), G (guanine) and C (Cytosine) for each sample. (*) indicate the same base aligned in that position for all samples. WT= wild type.

The *EZ1* and *EZ2* loci

Changes in DNA methylation were also investigated in the promoters of the *ENHANCER OF ZESTE* genes *SlEZ1* and *SlEZ2*. After primer optimization only two primer sets (EZ1-F2 and EZ1-F3) were used for further investigation (Figure 5.6).

Sequencing PCR products generated from both primer sets showed demethylation in the target region of EZ1-F2 in lines 3.3 and 4.2 (Appendix 28 and 29). It was difficult to obtain clones containing the EZ1-F2 PCR product. Six clones were obtained from the EZ1-F2 PCR product of line 3.3 while only three were obtained from line 4.2. Sequencing of line 3.3 indicated four changes in cytosine methylation in the region -760 until -822 upstream of the *SlEZ1* TSS (Figure 5.7). However, the changes were apparent in less than five clones at those sites (Figure 5.7). This could be due to lack of efficient sequencing in these regions in several clones. Sanger sequencing is limited in its quantitative accuracy, read length and sample throughput (Masser *et al*., 2015).

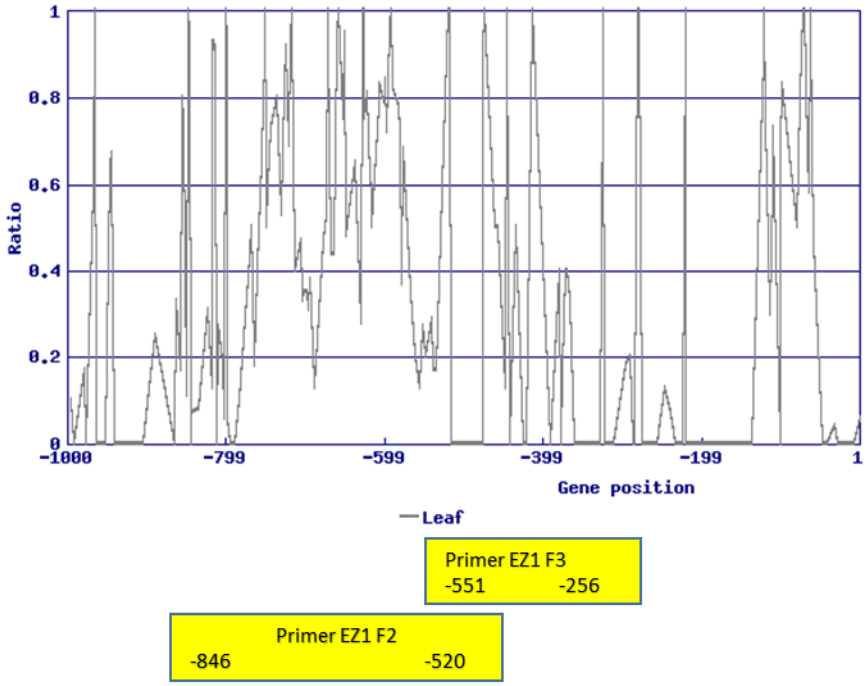
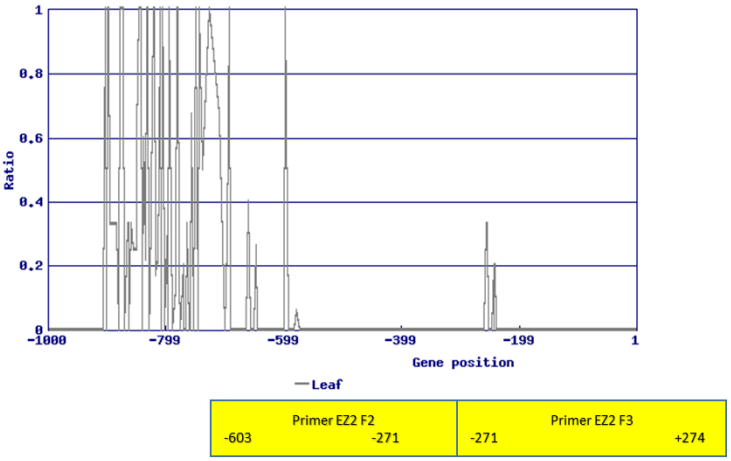


Figure 5. 5 Region covered by primers for bisulfite sequencing of *SlEZ1* (Solyc01g079390) promoter. The region covered is in 1000 upstream of *SlEZ1* translation start site. The yellow box represents the region covered by the primers. The ratio indicated the methylation ratio of that region (ratio= 5mC/(5mC+C)). Methylation ratio data obtained from http://ted.bti.cornell.edu/cgi-bin/epigenome based on data from Ailsa Craig The data was taken from leaf tissue.

Figure 5. 6 Location of demethylation in clones of line 3.3-primer EZ1-F2. The amplified region covers from -670 until -822 upstream of *SlEZ1* TSS. Guanines coloured blue show demethylation in less than 5 clones.The rows show the sequence alignment of the bases A (adenine), T (thymine), G (guanine) and C (Cytosine).(*) indicate the same base aligned in that position for all samples. WT= wild type.

The promoter of *SlEZ2* was also investigated for methylation changes. However, due to time constraints on the experimental work, the PCR products were not cloned and were sequenced directly. As with other genomic regions direct PCR sequencing gave results comparable with cloning. After several primers were tested, only two were found able to amplify the upstream region of *SlEZ2* transcription start site. These primers targeted the region +274 until -603 upstream of *SlEZ2* transcription start site (Figure 5.8A). Sequencing of EZ2-F2 PCR product showed no changes in methylation between the samples (Appendix 30) and sequencing of EZ2-F3 PCR product only confirmed the unmethylated region found in that region (Figure 5.8B).

(A)



(B)

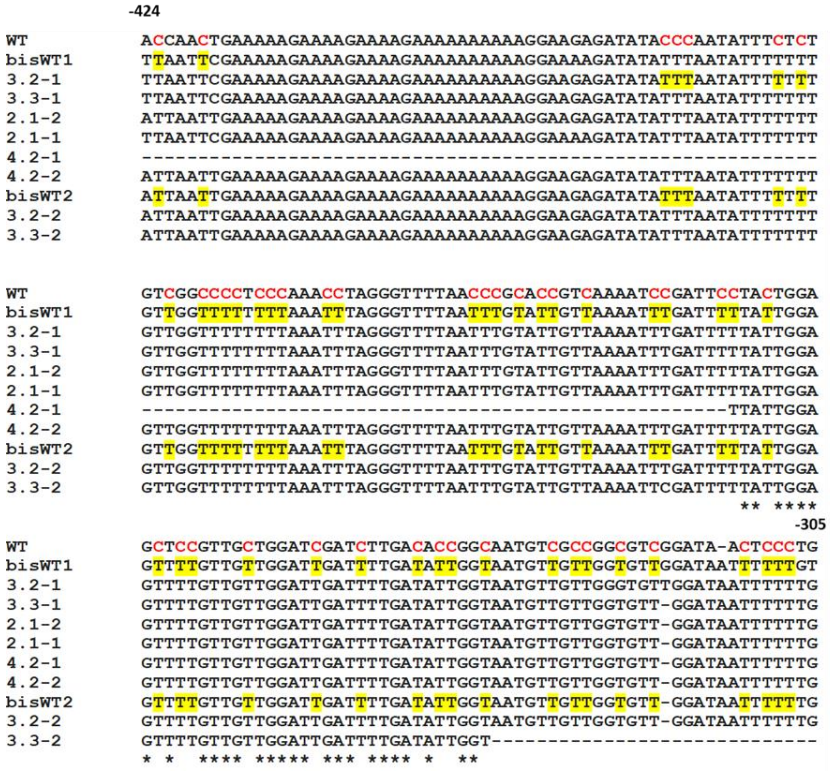


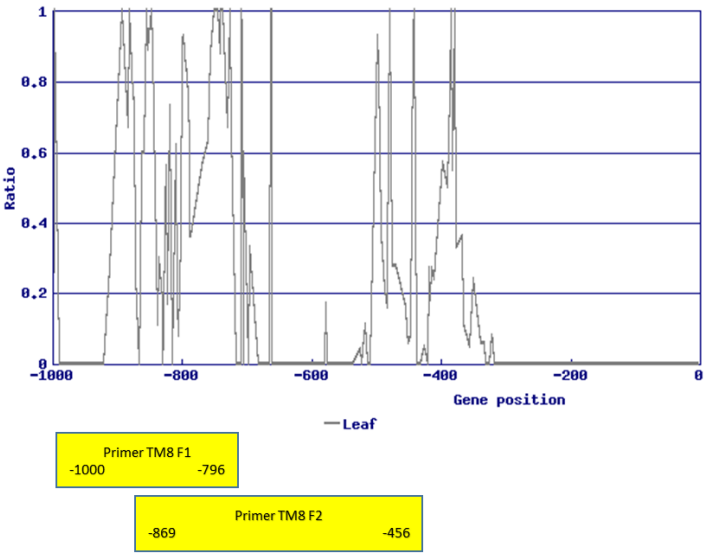
Figure 5. 7 Region covered by primers and demethylated cytosines in upstream region of *SlEZ2*. (A) The region covered is in 1000 upstream of *SlEZ2* (Solyc03g044380) transcription start site and (B) Location of demethylation in bisulfite treated wild type and several epiRILs using primer EZ2-F3. The yellow boxes shows the region covered by the primers. The ratio indicate the methylation ratio of that region (ratio= 5mC/(5mC+C)). Methylation ratio data obtained from http://ted.bti.cornell.edu/cgi-bin/epigenome. The data was taken from leaf tissue. Non-methylated cytosines are coloured red in wild type sample. Thymine due to demethylation are coloured yellow in bisulfite treated wild type.

The *TM8* locus

In this case again time constraints on the experimental work restricted experiment to clone the converted fragments and so PCR products were subject to direct sequencing. Primers used in the investigation of *TM8* promoter region covered -456 to -869 bp (primer TM8-F2) and -796 to -1000 bp upstream (primer TM8-F1) of the *TM8* transcription start site (Figure 5.9A). Sequencing of PCR product from primer TM8-F1 showed demethylated cytosines in lines 3.2, 2.1, 4.2 and 3.3 (Figure 5.9B).

The result from TM8-F2 primer showed no changes in DNA methylation in the epiRILs (Appendix 31), but sequencing from these amplicons was of poor quality which could be caused by the limitations of Sanger sequencing method for bisulfite treated DNA (Masser *et al*., 2015).

(A)



(B)

```
          -939
3.3-1     TTTAAATTTAAATTCGCGGT----------------------------------------     20
2.1-1     TTTAATTTAAATTCGTC-------------------------------------------     20
bisWT     TTTCATTTAAATTCGTGGTTATAATACTTAAATAAATAATATATTAGAATTATTGAGATT     62
WT        TTTAAATTTAAATTGTGGTTATAATACTTAAATAAATAATATATTAGAATTATTGAGATT     120
3.2-2     TTTAAATTTAAATTGTGGTTATAATATTTAAATAAATAATATATTAGAATTATTGAGATT     64
2.1-2     GTTTAATTTAAATTGTGGTTATAATATTTAAATAAATAATATATTAGAATTATTGAGATT     65
4.2       GTTTAATTTAAATTGTGGTTATAATATTTAAATAAATAATATATTAGAATTATTGAGATT     66
3.3-2     GTTTAATTTAAATTGTGGTTATAATATTTAAATAAATAATATATTAGAATTATTGAGATT     67
3.2-1     ----------------GTTATAATATTTAAATAAATAATATATTA--------------     29

                                                                  -820
3.3-1     ------------------------------------------------------------     20
2.1-1     ------------------------------------------------------------     20
bisWT     AAGTGAGTTTTGAGTGAATTA-----------------------------------------     83
WT        AAGTGAGTTTTGAGTGAATCAAAACGTTATCTGTTTTTAAGTGTATTGGAGTCCAAGTCC     180
3.2-2     AAGTGAGTTTTGAGTGAATTAAAACGTTATCTGTTTTTAAGCGTATTGGAGTTTAAGTTT     124
2.1-2     AAGTGAGTTTTGAGTGAATTAAAACGTTATCTGTTTTTAAGCGTATTGGAGTTCAAGTTC     125
4.2       AAGTGAGTTTTGAGTGAATTAAAACGTTATCTGTTTTTAAGCGTATTGGAGTTTAAGTTT     126
3.3-2     AAGTGAGTTTTGAGTGAATTAAAACGTTATCTGTTTTTAAGCGCATTGGAGTCCAAGTC-     126
3.2-1     ------------------------------------------------------------     29
```

Figure 5. 8 Region covered by primers and demethylated cytosines in upstream region of *TM8*. (A)The region covered is in 1000 upstream of *SlEZ2* (Solyc03g019710) TSS and (B) Location of demethylation in bisulfite treated wild type and several epiRILs. The yellow boxes shows the region covered by the primers. The ratio indicate the methylation ratio of that region (ratio= 5mC/(5mC+C)). Methylation ratio data obtained from http://ted.bti.cornell.edu/cgi-bin/epigenome. The data was taken from leaf tissue. Demethylated cytosines are coloured red in wild type sample. Thymine due to demethylation are coloured yellow in bisulfite treated wild type.

5.3.2 Global patterns of DNA methylation

Global patterns of DNA methylation were investigated using WGBS. Genomic DNA was extracted from the leaves of six week old wild type plants, two epiRILs (from generation F_5) and *MET1* RNAi transgenic lines. The epiRILs used for the analysis were lines 2.1 and 4.6. Two sets of wild type lines were used, one grown with the *MET1* transgenics at INRA and the other grown with the epiRILs at University of Nottingham. All lines were of an identical genetic background. The DNA was subject to bisulfite conversion and then sequenced. The WGBS reads were then aligned to the annotated *Solanum lycopersicum* tomato genome vSL2.5 (www.solgenomics.net).

Comparison between the wild type, *MET1* RNAi and epiRILs revealed differences in overall levels of DNA methylation (Figure 5.10A). Both RNAi transgenic lines and the epiRILs showed evidence of global hypomethylation in comparison to wild type especially in the CG context. RNAi lines showed reduced CG methylation of 25% and 35 % respectively while in the epiRILs there was only a 5 % of reduction.

The methylation levels in the CHG sequence context were lower (Figure 5.10B) in both the RNAi lines and EpiRILs, but the effects were less pronounced. However, the range of methylation level was wider in the CHG sequence, ranging from 40 – 85 % for almost all of the samples. The methylation levels in the CHH context were much lower than for CG or CHG, ranging only from 5 to 12 % for all the samples with no apparent global differences between treatments (Figure 5.10C).

The lower level of methylation in CG context compared to CHG and CHH in the tomato epiRILs confirmed the importance of *MET1* in tomato CG

methylation. The difference in the range of methylation levels in the CHG context showed that *MET1* might have some indirect role in the methylation of CHG such as found in Arabidopsis *met1* mutants (Kenkel *et al*., 2003; Mathieu *et al*., 2007)

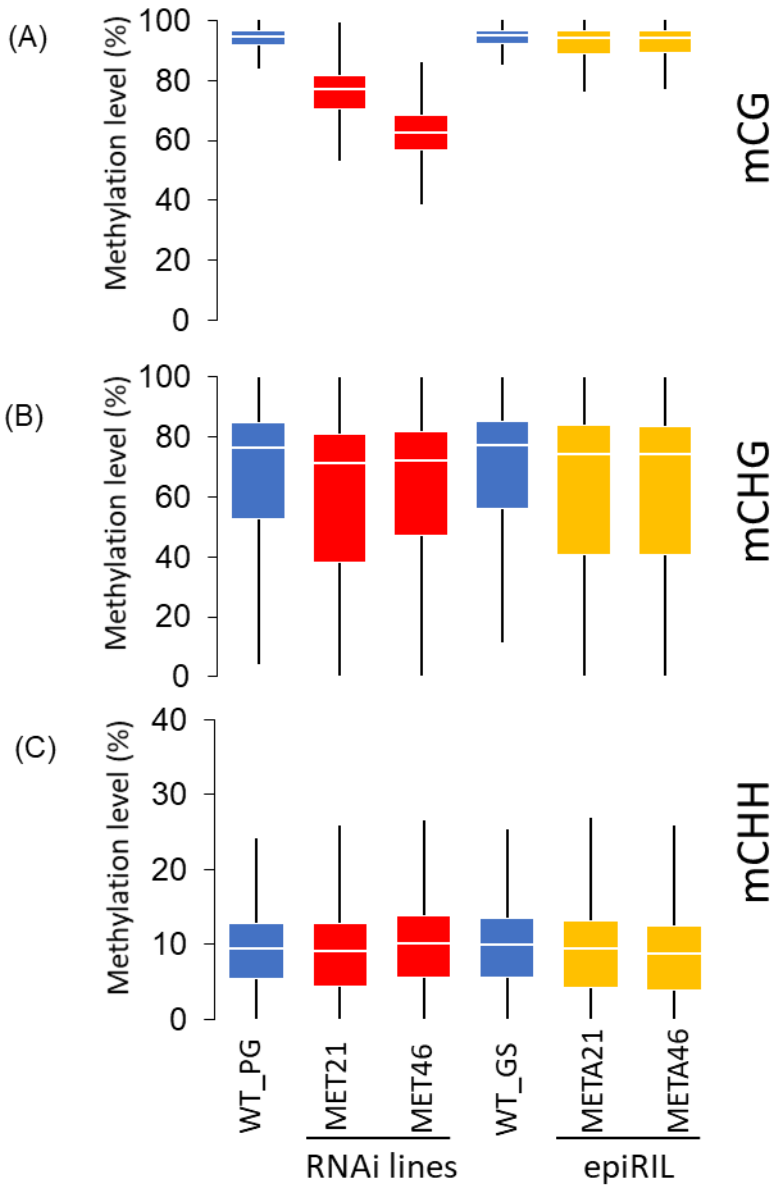


Figure 5. 9 Box plots showing the mean levels of DNA methylation in wild type, the RNAi lines and epiRILs. Methylation levels are shown for the (A) CG, (B) CHG and (C) CHH contexts. The samples were: wild type grown at INRA (WT_PG) and at Nottingham (WT_GS), *MET1* RNAi transgenic lines 2.1 and 4.6 and epiRILs 2.1 and 4.6 from the F_5 generation. The tomato genome was partitioned into 0.5 kb bins for the methylation analysis.

The global changes in DNA methylation that were observed in our *MET1-RNAi* lines and epiRILs were similar to those reported in related experiments undertaken in Arabidopsis and tomato. In *met1* Arabidopsis mutant (from ethylmethanesulfonate mutagenized seeds), digestion using the methylation sensitive enzyme HpaII showed loss of CG methylation and CHG methylation (Kenkel *et al*., 2003). The reduction in methyl cytosines were in both repetitive and single copy genes. Bisulfite sequencing of a *met1* Arabidopsis mutant also confirmed the dramatic loss of CG methylation but also changes in the methylation level of CHG and CHH sequences (Mathieu *et al*., 2007; Lister *et al*., 2008).

Arabidopsis hypomethylated mutants were found at the *DDM1* (*DECREASE IN DNA METHYLATION*) locus, designated the *ddm*1 mutants. The *ddm1* mutants differs in their effect on DNA methylation where it cytosines in all contexts were effected, but mostly in repetitive sequences (Kenkel *et al*., 2003; Catoni and Cortijo, 2018).

The construction of Arabidopsis epiRIL population used both *met1* (Reinders *et al*., 2009) and *ddm1* mutants (Johannes *et al*., 2009). The two epiRIL populations showed differences in their phenotypes and patterns of methylation. The level of 5-methyldeoxycytidine (mC) in several *met1*-derived epiRILs were found to be similar to wild type, with levels up to 80% of wild type. The mapping of DNA methylation also showed random distribution of non-parental epialleles in the epiRILs. In the *ddm1*-derived epiRILs, the inheritance of hypomethylated epi-alleles were also seen in the population with some remethylation of alleles after two to five generations in some lines.

In tomato, *ddm1* mutants were obtained using the CRISPR/Cas9 technology using the tomato cultivar M82. A reduction in CG and non-CG methylation was observed in the mutants (Corem *et al*., 2018). The levels of methylation ranged from 85-95% for the *Slddm1* mutant. This level was higher than in our tomato *met1*-RNAi line with only 60-80% methylation.

In the tomato epiRILs, the DNA methylation levels were nearer to those of wild type than in the RNAi lines. This likely reflects how the epiRILs were constructed by backcrossing parental lines (RNAi lines) with wild type, which introduced the wild type epigenome. Similar effects on epiRIL hypomethylation were also seen in Arabidopsis (Johannes *et al*., 2009; Reinders *et al*., 2009). The restoration of the *MET1* expression in the early generations of the epiRILs would maintain CG methylation in the genome and also the RNA dependent DNA Methylation (RdDM) pathway that is involved in de novo methylation (Zhang *et al*., 2018).

5.3.3 Patterns of methylation in genes and transposable elements

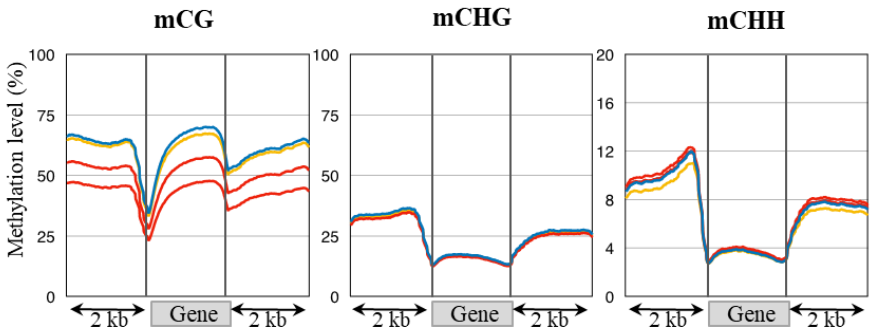
Compared to wild type, the CG methylation in RNAi lines was 10 - 25 % lower in gene bodies and in the 2kb of the upstream and downstream regions (Figure 5.11A). In the CHG context, the levels of methylation were almost the same between all the samples with a range of only 5 – 10 %. In the CHH context, the RNAi lines showed the highest methylation level while the epiRILs showed the lowest in the upstream and downstream regions. In gene coding regions the methylation levels were almost the same for all treatments (Figure 5.11A).

In transposable elements (TEs), the RNAi lines showed the lowest CG methylation (Figure 5.11B) and these were 15 - 25 % lower than wild type. In the CHG context, the methylation levels of TEs in the RNAi lines was also the lowest. However, in the CHH context, the RNAi lines showed the highest methylation level, while the epiRILs showed the lowest level (Figure 5.11B).

The higher methylation levels in the CG context indicated that methylation in tomato mostly occurred in this sequence context. This concurs with findings by Zhong *et al*. (2013) where CG methylation in tomato leaf can reach 85.51 % compared to CHG and CHH methylation which only reached 56.15 and 8.63% respectively.

The lower levels of CG methylation in the RNAi lines showed that the *MET1* knockdown affected the CG methylation in gene bodies and methylation associated with tomato TEs. Higher level of CHH methylation in RNAi lines agrees with findings by Mathieu *et al*. (2007) in Arabidopsis where reduced CG methylation was associated with an increase in CHH methylation.

(A)



(B)

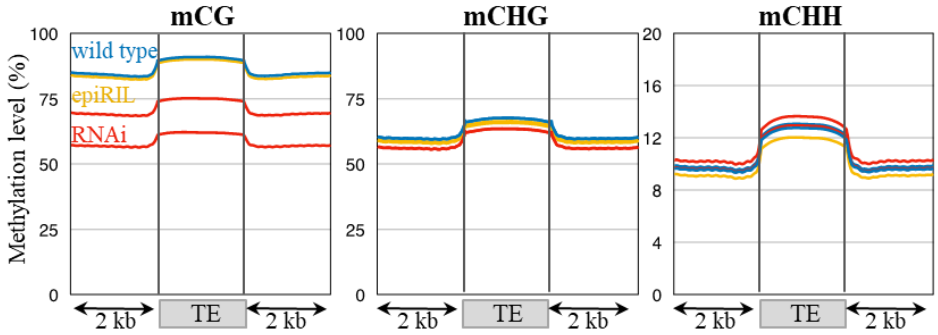


Figure 5. 10 Patterns of methylation in genes and TEs. Wild type samples are shown by a blue line, RNAi lines are indicated by a red line and epiRILs as yellow line. Two biological replicates were analysed for each line. TE = transposable elements, mCG = methylation in CG context, mCHG = methylation in CHG context, mCHH = methylation in CHH context.

5.3.4 Differentially Methylated Regions (DMRs) in RNAi Lines and epiRILs

The presence of differentially methylated regions (DMRs) was analysed by comparing methylated regions of RNAi and epiRILs with wild type.

DMRs in all CG sequence context

The number of DMRs was analysed for all the different methylation contexts (CG, CHG and CHH) (Figure 5.12). In line 2.1, the number of DMRs in the CG context was higher than the number of DMRs in the CHG or CHH context. In comparison to wild type, the number of hypomethylated DMRs in CG context was 12268 in the RNAi line 2.1, while in the related epiRIL 2.1 it was 3685. However, the number of hypomethylated DMRs in CHG and CHH context was higher in the epiRIL 2.1, than in either the RNAi or wild type lines (Figure 5.12A).

Hypermethylated DMRs in CG and CHH context for line 2.1 were higher in the epiRIL compared to the RNAi line. In the CHG context, the number of hypermethylated DMR was almost the same in both lines. The number of hypomethylated CG DMRs in the epiRIL suggests that some of hypomethylated CG was transgenerationally inherited in the epiRIL. A higher number of hypermethylated DMR in CG and CHH context in the epiRIL indicated that there was an increase of CG and CHH methylation after successive selfing in the population.

Analysis of CG methylation in line 4.6 showed that the RNAi 4.6 line had a greater number of hypomethylated CG DMRs than WT (Figure 5.12B). In comparison with WT there were 286,636 additional DMRs. In the epiRIL 4.6 there were only 3724 additional DMRs in comparison to WT. For line 4.6,

hypomethylated CHG and CHH DMRs were found to be higher in the epiRIL than in the RNAi line.

A higher number of hypomethylated CG DMRs in the RNAi lines of both line 2.1 and 4.6 indicated that the initial *MET1* knockdown caused a significant loss of CG methylation. After successive rounds of selfing, the CG methylation was restored and was seen in the reduced number of hypomethylated CG DMR in the epiRILs. The higher number of hypomethylated CHG and CHH DMR in the epiRIL indicated that loss of CG methylation in the parental line indirectly affected CHG and CHH methylation in successive generations.

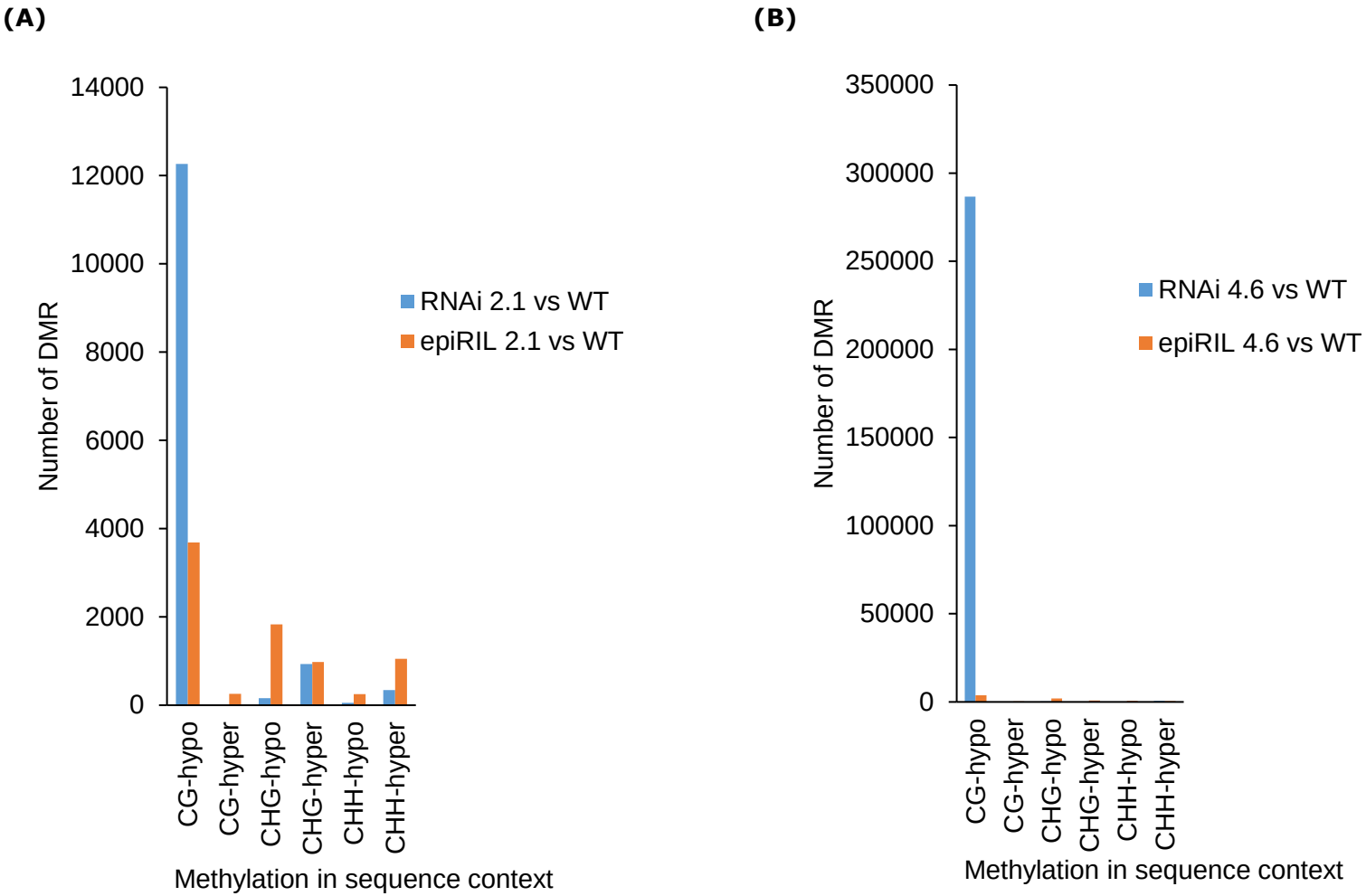


Figure 5. 11 Number of hypo-and hyper- methylated DMRs between RNAi and epiRILs with wild type.

(A) The number of DMRs in line 2.1 and (B) The number of DMR in line 4.6

DMRs in CG sequence context

With the main role of *MET1* in CG methylation, the analysis was focused on the DMRs in the CG context on each of the 12 tomato chromosomes. The results indicated that in comparison to wild type the RNAi line 2.1 showed CG hypomethylation almost evenly in all of the 12 chromosomes (Figure 5.13). The DMR density showed an even distribution in almost all of the chromosomes with some peaks of hypomethylation on chromosomes 2 and 9.

In comparison to the parent RNAi line the epiRIL 2.1 showed less hypomethylation, although clusters of of DMRs were more apparent on chromosomes 2 and 3. However, in comparison to wild type and the RNAi line, hypermethylation at CG sites was present in the epiRIL line as indicated by the presence of blue dots on the circle plot (Figure 5.13) showing differences in hypermethylated DMRs between the epiRIL vs wild type. These mosaics of methylation in the tomato epiRILs are consistent with those seen by other workers when epiRILs were generated in the model plant Arabidopsis (Reinders *et al*., 2009; Johannes *et al*., 2009).

Changes seen in epiRIL line 4.6 were broadly similar to those seen for epiRIL 2.1. Focusing on the DMR of the CG context, line 4.6 showed a high density of hypomethylation in RNAi vs wild type (Figure 5.14). The density of DMRs was almost evenly distributed in all of the chromosomes. In the epiRIL 4.6 there were clusters of DMRs on chromosomes 3, 5 and 7. More hypermethylated CG DMRs were also found in the epiRIL 4.6 vs wild type in comparison with the RNAi 4.6 line vs wild type (Figure 5.14).

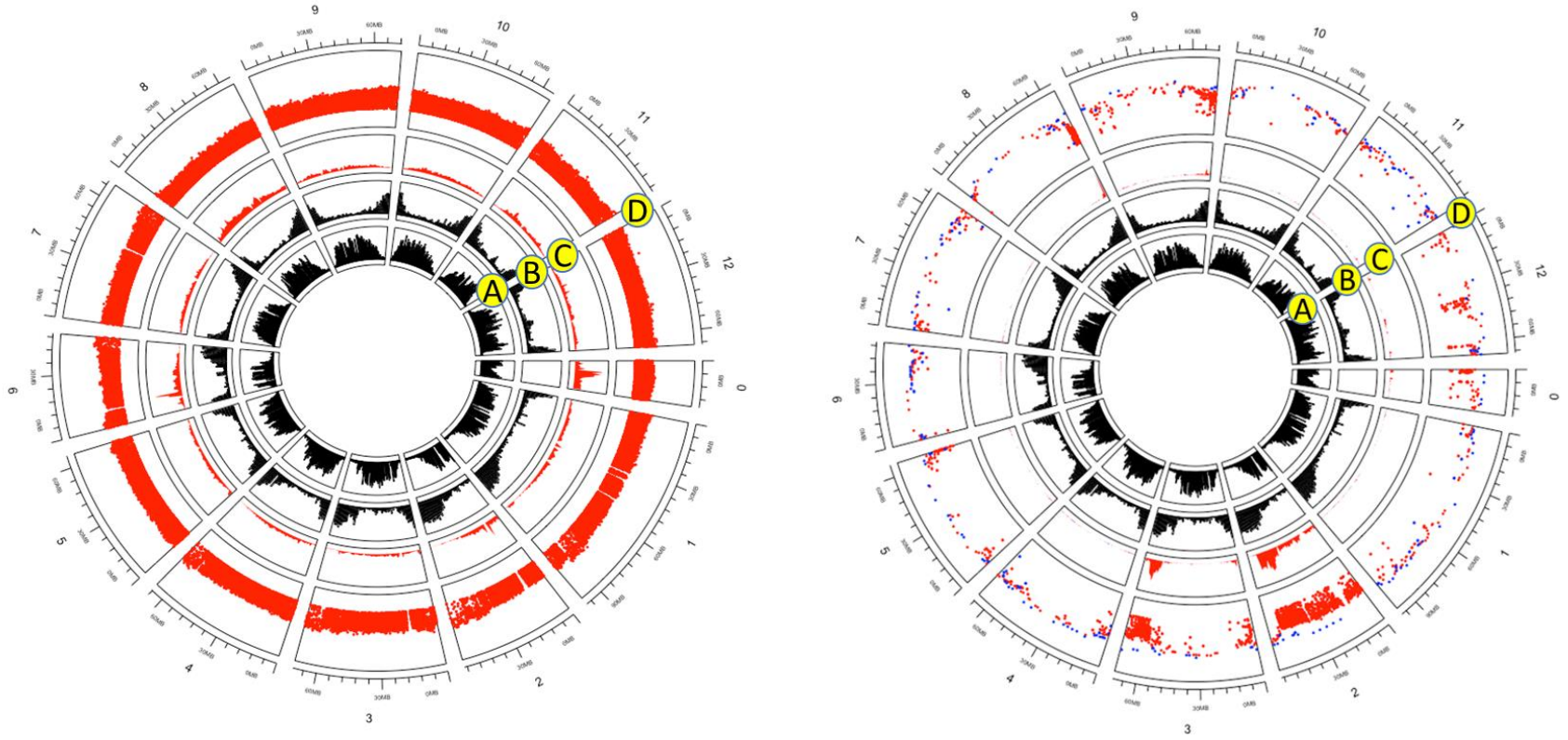


Figure 5. 12 Differentially Methylated Region (DMR) in the CG context compared between wild type with RNAi line 2.1 (left) and epiRIL 2.1 (right). (A)Transposable element density, (B) Gene density, (C) density of DMRs and (D) dots representing DMRs where one dot = one DMR. Red dots represent hypo CG DMR and blue dots represent hyper CG DMR. The numbers in the outer circles (1-12) represent the chromosome number.

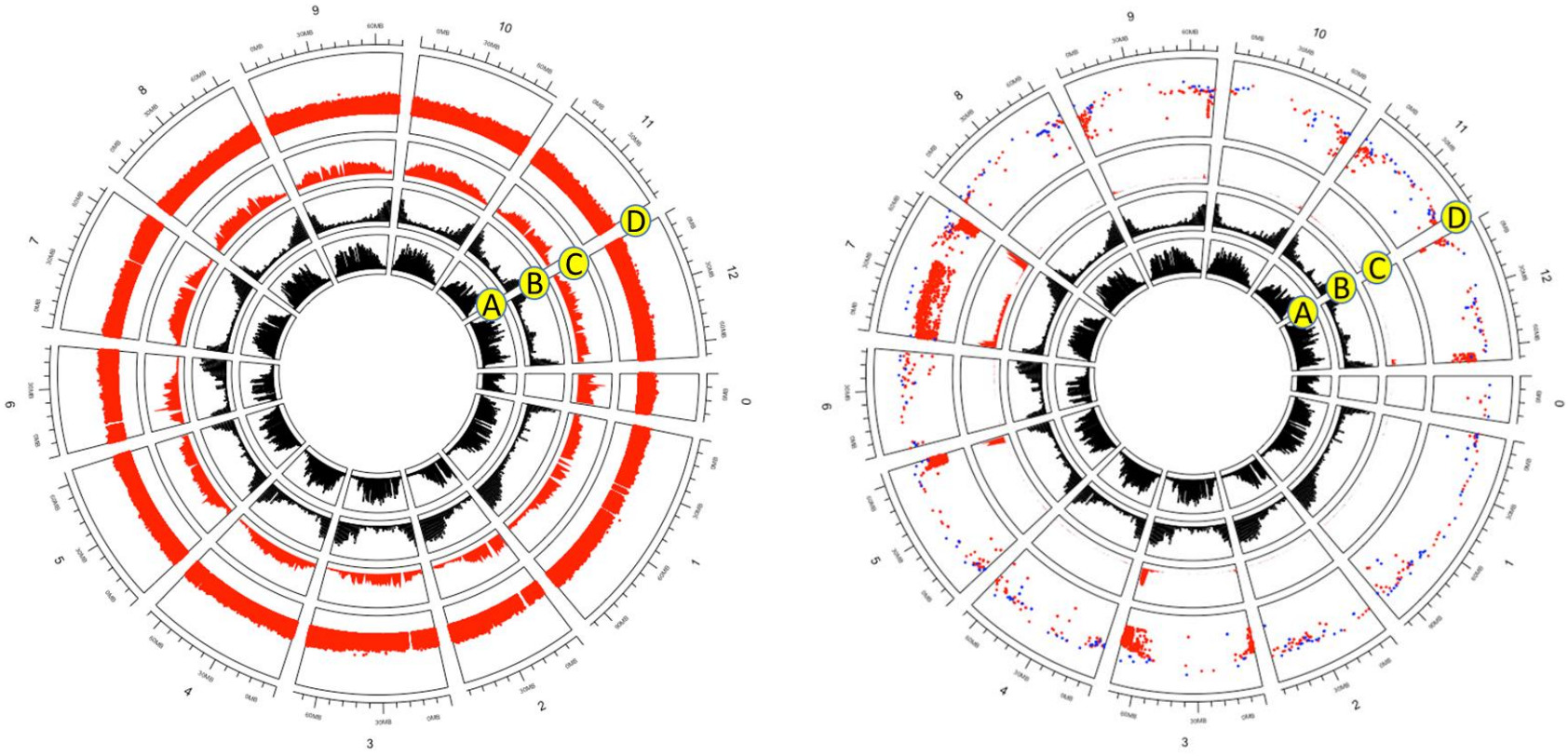


Figure 5. 13 Differentially Methylated Region (DMR) in the CG context compared between wild type with RNAi line 4.6 (left) and epiRIL 4.6 (right). (A)Transposable element density, (B) Gene density, (C) density of DMR and (D) dots representing DMRs where one dot = one DMR. Blue dots represent hyper CG DMR and red dots represent hypo CG DMR. The numbers in the outer circles (1-12) represent the chromosome numbe

The change in hypomethylated regions suggests that new epi-alelles were formed in the parental lines or in early generations and through successive inbreeding became fixed differently across the chromosomes. The difference in the pattern of DMRs between epiRILs 2.1 and 4.6 showed that active methylation which occurred during each cycle of regeneration can create specific patterns of methylation. Non-parental methylation polymorphism were also observed in Arabidopsis epiRILs (Reinders *et al*., 2009).

From the results of both the target and whole genome bisulfite sequencing, we have observed differences in methylation between the tomato epiRILs and wild type and amongst individual epiRILs in the population. In this study it is difficult to compare the data from the targeted bisulfite sequencing and WGBS because different tissues were used for each study Also the DNA came from different plants and different generations. However, the WGBS data does indicate major changes in the epigenome in the epiRILs consistent with the results from the targeted sequencing of individual genes. From the limited number samples analysed we also see evidence of new methylation patterns that are specific to different epiRILs from the difference in hypo- and hyper- methylated regions on different chromosomes.

5.4 Conclusion

In this chapter, bisulfite sequencing provided evidence that removal of MET1 caused global methylation changes in the RNAi lines and epiRILs. Sequencing specific target regions showed that hypomethylation was found in the upstream region of the *fw2.2* and *SlEZ1* transcription start site in epiRILs 3.3 and 4.2 Demethylations were found in the CHH context in line 3.3.

Cloning the PCR product of primers for the *SlEZ1* upstream region proved to be difficult. Less than five clones were obtained for the sequencing, making the result low in confidence level. For the *SlEZ2* and *TM8* genes, sequencing of the upstream region was undertaken from PCR products due to the time restriction for this project. Sequencing the upstream region of *SlEZ2* proved the presence of the unmethylated region that naturally exists in the wild type. While sequencing of *TM8* upstream region showed some demethylation that corresponded to changes from cytosines in the wild type to thymine in some of the epiRILs.

Whole genome sequencing of two epiRILs and RNAi lines showed a lower CG methylation in RNAi lines globally and also in gene bodies and TEs. The methylation levels in CHG and CHH context were not always lowest for RNAi lines. However, the level of CHH methylation in genes and TEs was often lowest in the epiRILs.

When methylation levels of RNAi and epiRILs were compared to wild type, there were differences in demethylated regions. Changes in the density of DMR were seen in both RNAi lines vs WT and epiRILs vs WT. Several chromosomes showed more hypomethylated region in the epiRILs. In line 2.1, hypomethylated DMR was found to be maintained in the epiRIL and can be seen by the number of

hypomethylated DMR in chromosomes 2,3 and 7. While in line 4.6 this was found in chromosomes 3, 5 and 7.

Reduced methylation in non-CG sequence context indicated reduced *MET1* expression in parental lines could change methylation in other contexts than CG. The effect of loss in CG methylation on CHG and CHH was also found in Arabidopsis (Mathieu *et al*., 2007). Brocklehurst *et al*. (2018) also stated that the effect of MET1 depletion in dense methylated region could cause both CG and non-CG marks due to the involvement of MET1 in a multiprotein complex. The complex which contains chromomethylases could lose function without MET1. Another explanation is the indirect effect of MET1 on other epigenetic factors such as histone regulators. This was seen in *met1* mutant where the loss of H3K9 methylation patterns resulted in the loss of CHG and CHH methylation.

The difference in hypomethylated DMR in the chromosomes of the two lines (2.1 and 4.6) indicated that the methylation patterns between the two epiRILs were unique. Irregular and stochastic inheritance of methylation was also observed in early generations of inbreeding (Mathieu *et al*., 2007) and in the epiRIL population of Arabidopsis (Reinders *et al*., 2009; Johannes *et al*., 2009).

CHAPTER 6 GENERAL DISCUSSION AND FUTURE WORK

DNA methylation has been shown to be a stable and heritable epigenetic mark in plants. Evidence from work on the model plant species, *Arabidopsis thaliana,* has demonstrated that epigenetic variation can be stably inherited across several generations (Johannes *et al*., 2009; Reinders *et al*., 2009).

The stability of DNA methylation and its inheritance in plants has been used to create novel populations with altered epigenomes. In Arabidopsis, a population of epigenetic Recombinant Inbred Lines (epiRILs) have been generated (Johannes *et al*., 2009) that combined phenotypic and epigenomic data into a quantitative genetics framework (Reinders *et al*., 2009). The development of epiRILs from a cross between a RNAi knockdown line and wild type and the cycles of self-fertilization from each generations stabilised the novel epigenomes in each line.

In this project, the reduced *MET1* expression in transgenic tomato parental lines was used as a starting point to generate a tomato epiRIL population. Tomato was used for several reasons: (1) Its importance as a model system for fleshy fruit, (2) epiRILs have so far only been generated in the simple genome of Arabidopsis and tomato is a complex genome with many repeat elements, (3) an epimutation has been shown to govern tomato ripening (Manning *et al*., 2006) indicating this process can be under epigenetic control.

The aim of this project was to generate and characterize a novel population of tomato epiRILs and identify the effects of the epigenetic variation on tomato development and ripening.

6.1 Morphological changes of the tomato epiRILs

In Arabidopsis epiRILs, transgenerational inheritance was apparent for several phenotypic characters such as flowering time, stress response and plant height. There were other phenotypes, but these appeared in only one or two generations (Johannes *et al*., 2009; Reinders *et al*., 2009). The unstable nature of new/abnormal phenotypes was also observed in our tomato epiRILs (Chapter 2). Only stable phenotypes seen in several generations were investigated using gene expression analysis, restriction enzymes and sequencing.

Different methods have been used in generating the epimutant parental lines. In Arabidopsis, EMS mutation and T-DNA have been used in creating the demethylated parental line. In this project, RNAi knockdown was used to reduce MET1 expression thus reducing CG methylation in the genome. The effect of a reduced expression of MET1 which is an enzyme involved in the maintenance of CG methylation showed that it is vital in the embryo development of tomato. In the early generation, it was observed that the germination rate of T_1 seeds obtained from the cross between *MET1* knockdown and wild type was very low. From 4000 seeds sown, only 21 of seeds germinated (Gallusci *et al*., unpublished result). The effect of low *MET1* expression on the development of embryo have been observed in other species. Frequent seed abortion, which might correspond to early

embryonic lethality, was found in *met1* Arabidopsis (Saze *et al*., 2003) while in mice, low levels of Dnmt3a (a methyltransferase found in mammals) caused death at early age and embryonic lethality (Jin *et al*., 2011).

Once the epiRIL population had been established, a range of phenotypes that were different from the wild type control became apparent. These included twisted and flat stems, non-serrated and chimeric leaves, flowers with twisted stamens, high flower abortion rate and low fruit weight. Analogous developmental phenotypes were also seen in Arabidopsis hypomethylated *met1* mutant (Kankel *et al*., 2003; Saze et al., 2003; Reinders *et al*., 2009). Both vegetative and generative stage were affected by loss of *MET1* expression in Arabidopsis resulting in a delay in flowering time, increase numbers of rosette leaves prior to flowering, increased thickness of the inflorescence stem and shorter plants.

Recent work in Arabidopsis has indicated that these types of developmental effects can result from both CG methylation and other methylation changes. For example, in the *Slddm1a Slddm1b* double mutant generated from a knockout of the *DECREASE IN DNA METHYLATION 1* (*DDM 1*) gene showed abnormal phenotypes in both vegetative and generative development (Corem *et al*., 2018). The double mutant showed hypomethylation in the CHG and CHH context, but showed abnormal variegated leaves, abnormal flower and small fruits (which are similar to the phenotypes found in the tomato epiRILs).

To analyse the effects of reduced MET1 expression in the RNAi parental line on the epiRIL population, several approaches were taken. The first was to detect changes in gene expression in selected lines in different

generations and correlate the expression to an abnormal phenotype that was stable in several generations of the epiRILs. The techniques used were QPCR and RNA sequencing. Changes in methylation were detected using restriction enzymes to detect changes in promoter regions and transposable elements which can alter gene expression. Changes in methylation were also analysed using bisulfite sequencing in selected promoter regions and whole genome to detect hypomethylated cytosines. Lower level of CG methylation was expected in the epiRILs but this was not always the case as discussed in Chapter 5. The findings showed that the changes in the tomato epiRILs phenotypes were not caused by direct control of methylation on one specific gene. Reduced MET1 expression in the parent line was also found to affect other than CG methylation in the epiRILs.

6.2 Changes in genes and TEs expression

Low fruit weights were seen in some of the tomato epiRILs and although this trait is controlled by several genes (Tanksley, 2004; van der Knapp, 2014), one gene was chosen for the investigation because it is well characterised. The *FW2.2* gene controls 30% of tomato fruit weight variance (Frary *et al*., 2000) and acts as a negative regulator of cell division in tomato and pear fruits (Jia *et al*., 2016). The findings in this investigation showed the expression of *FW2.2* was different between several epiRILs and wild type and varied in different tissues and in different generations of epiRILs. However, the correlation between *FW2.2* expression with fruit weight was not significant (Chapter 3). The results are consistent with findings which showed that regulation of tomato fruit weight is not due to expression of a

single gene and further investigation on *FW2.2* promoter region is still needed to discover how it is involved in the regulation of tomato fruit weight.

In the tomato epiRILs, the expression of the Two *Enhancer of Zeste* (*SlEZ1* and *SlEZ2*), *TM8* and *SlGA20ox2* genes were different to that of wild type. The difference in expression varied depending on the tissue and the generation of the sample. The result confirmed that the expression of the *EZ* genes was differentially regulated during plant development (How-Kit *et al*., 2010) and that there might be functional redundancy between the two genes.

A number of genes selected in this study were analysed to observe the correlation of their expressions with fruit weight in several lines. The selected genes (*FW2.2*, *SlEZ1*, *SlEZ2* and *TM8*) are known to either indirectly affect fruit weight by regulating flower and carpel development or by regulating flower morphology which can affect fruit development. The results showed that the correlation between the genes expression at anthesis or green fruit stage was not significant.

If changes in methylation in the epiRILs caused changes in gene expression, it could be involved in a complex pathway such as shown in Figure 6.1. Figure 6.1 shows an example of the master regulator *KNOX* (*KNOTTED1-LIKE HOMEOBOX*) genes which involves the action of *EZ* and *GA20ox2* (two of the genes investigated in this project). The *EZ* genes encode proteins which form the POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) which is involved in the regulation of *KNOX* genes and which in turn regulates the expression of *GA20ox2*. *KNOX* genes produce transcription factors which are important in producing pluripotent cells in the shoot apical meristems (Hay and Tsiantis, 2010). The target genes then regulate

hormone homeostasis in the plant meristem, which in turn have an effect on the development of plant morphology.

Using RNA sequencing, it was discovered that changes in gene and TE expression were specific to the different epiRILs used in the analysis. Both increased and reduced expression of genes and TEs were found in lines 3.3 and 4.69 in generation F_7 of the epiRILs. This findings concurs with other species where both increased and reduced gene expressions were also observed in *met1* mutant lines (Garg *et al*., 2015; Chen *et al*., 2008; Brocklehurst *et al*., 2018).

Reduced *MET1* expression in one parental line should increase gene expression if the reduced MET1 only affected silencing of genes and TEs via CG methylation of promoter regions. However, the reduced expression of some genes and TEs in the epiRILs showed that there is a more complex regulation of methylation by *MET1*.

Immobilization of transposable elements (TEs) were expected in the epiRILs due to changes in methylation. Loss of methylation have been associated with immobilization and increase expression of TEs. Both Class I and Class II TEs were found to be differentially expressed in the tomato epiRILs. The two classes were also found in the Arabidopsis epiRILs (Johannes *et al*., 2009; Reinders *et al*., 2009; Mirouze *et al*., 2009) showing that both classes of TEs are affected by reduced MET1 expression.

Investigation on gene expression did not show how methylation could affect the differences in the morphological data of the epiRILs. To further investigate the correlation between changes in methylation, gene

expression and plant phenotype analysis using a restriction enzyme was undertaken (Chapter 4).

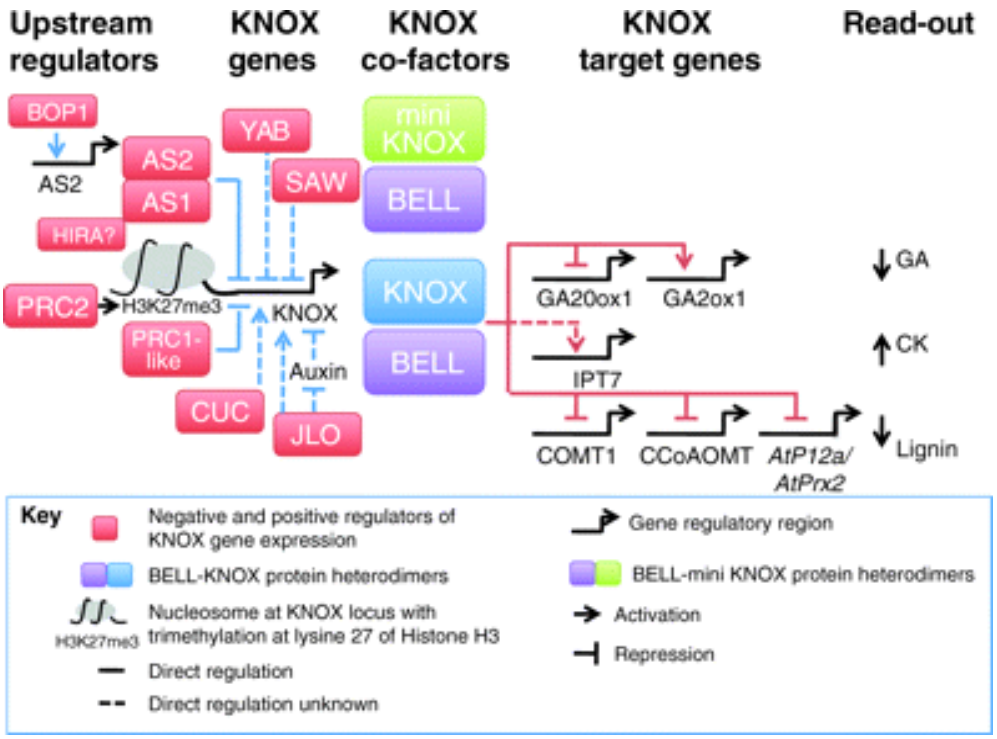


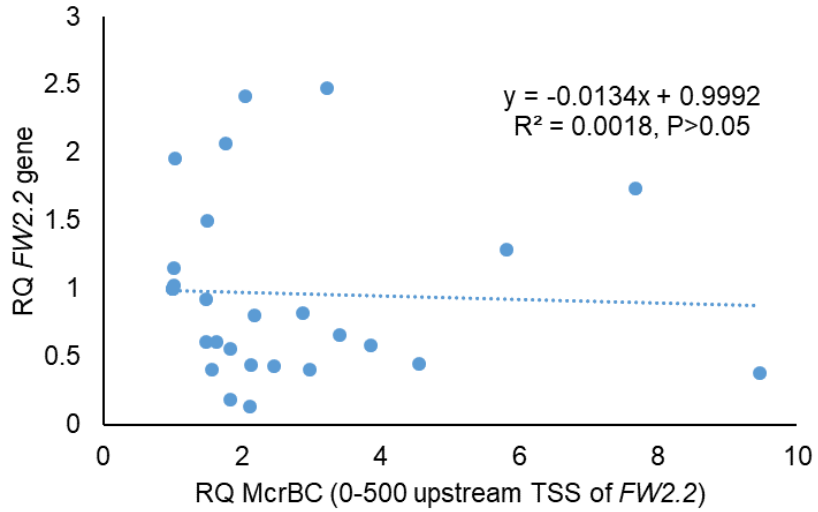
Figure 6. 1 KNOX gene regulatory network. Upstream regulators control the expression of *KNOTTED1-LIKE HOMEOBOX* (*KNOX*) genes (Hay and Tsiantis, 2010). Abbreviations: AS1, ASYMMETRIC LEAVES1; BELL, BEL1-like homeodomain family; BOP1, BLADE-ON-PETIOLE1; CUC, CUP-SHAPED COTYLEDON; H3K27me3, trimethylation of histone H3 at lysine 27; HIRA, histone regulatory protein A; JLO, JAGGED LATERAL ORGANS; PRC2, polycomb repressive complex 2; SAW, SAWTOOTH; YAB, YABBY.

6.3 DNA Methylation patterns of the epiRILs

The McrBC analysis of the tomato epiRILs showed changes in DNA methylation in regions upstream of the coding sequences of the selected genes. There were differences in methylation between epiRIls and wild type and between different generations of epiRILs. The McrBC enzyme recognizes two methylated or hemimethylated RC sites (R^mC) at approximately 30 to >2000 base pairs between the sites and cleaves the DNA close to one of the site (Piper *et al*., 2002). Therefore the McrBC analysis covered changes in methylation in CG as well as CHG and CHH contexts. Changes in methylation could be caused by the activities of different methyltransferases, chromatin remodellers and small interfering RNAs (Zhang *et al*., 2018). The variation in methylation between generations indicates that the epi-marks in the selected sites are actively changing across generations.

To investigate the correlation between methylation marks, gene expression and fruit weight in the epiRILs, data from three generations were used. Correlation between McrBC-QPCR and cDNA-QPCR data was analysed and showed no significance (Figure 6.2). These results suggests that the McrBC-QPCR data could not explain the direct relationship between difference in methylation of the target regulatory regions and the gene expression.

(A)



(B)

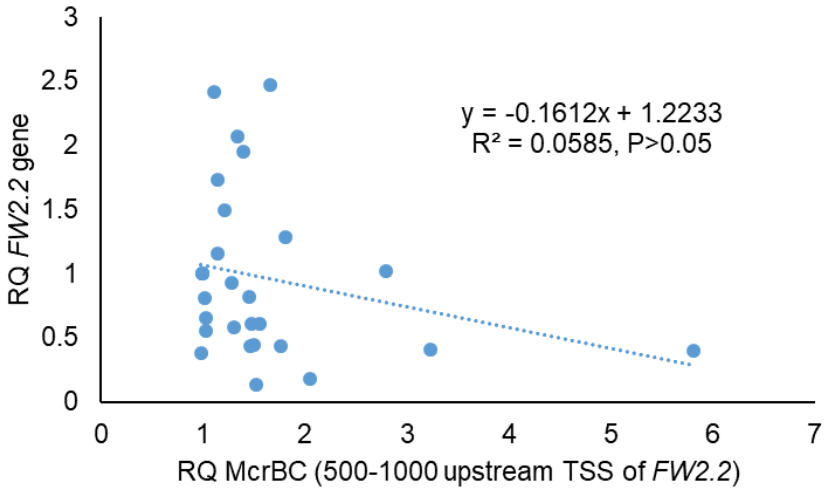


Figure 6.2 Relationship between the relative quantification of *FW2.2* gene and McrBC target regions. The McrBC-QPCR analysis used leaves tissue while cDNA-QPCR analysis used green fruits. The regression equation and correlation coeffiecient (R^2) of the line is shown. Statistical analysis showed no significance in the correlation (P>0.05).

Early analysis in Chapter 3 showed that the correlation between the relative expression of *FW2.2* with fruit weight was not significant. This could be caused by the limited number of samples used in the analysis or due to the difference in tissue samples used in the McrBC-QPCR and the cDNA-QPCR analysis. However, all the correlation analysis that was undertaken showed a negative trend for the relationship.

To further test the effect of methylation and gene expression on fruit weight, a multiple linear regression model was used (Appendix 22). The result showed the relationship between the three variables were : Y = 7.367 -1.105 cDNA-RQ - 0.173 McrBC-RQ (region 0-500 bp upstream of TSS) and Y = 7.221 -1.144 cDNA-RQ - 0.17 McrBC-RQ (region 500-1000 bp upstream of TSS), where Y = fruit weight at B+7. Both equations showed that the relationship between RQ values of gene expression and McrBC with fruit weight is indirectly proportional which means that higher level of methylation and higher expression of *FW2.2* leads to lower fruit weight. This concurs with results showing a negative correlation between *FW2.2* and tomato fruit weight (Cong *et al*., 2002) and that lower level of methylation will cause higher gene expression (Zhang *et al*., 2018) and thus lower fruit weight. Further investigation using more lines and replicates and using the same tissue samples is needed to confirm this finding.

Another investigation on the changes in methylation was using the Southern Blot method. The CG methylation of gypsy and copia transposons were detected using the *Hpa*II restriction enzyme. The results indicated slight changes in methylation and were consistent with the WGBS data presented

in Chapter 5 that showed relatively small changes in global CG methylation between wild type and epiRILs. Limitations on the data were due to the limited number of samples and time for the analysis. It is also important to note that transposons are often regulated by changes in CHH and CHG methylation (Corem *et al*., 2018; Gouil and Baulcombe, 2016) and this will not be detected by the *Hpa*II analysis. A more accurate method would be to first map the population and identify possible insertions due to mobilization of TEs. Once identified, the design of the probes will then be more accurate to detect the change in methylation and movement of the TEs.

6.4 Differentially methylated regions of selected epiRILs

Higher levels of hypomethylation were found in the regions upstream of the TSS of *FW2.2*, *SlEZ1* and *TM8* in line 3.3 compared to line 4.23 at F_6. Line 3.3 exhibited several abnormal phenotypes and RNA sequencing in generation F_7 also showed hundreds of differentially expressed genes in this line. Line 3.3 was generated from the F_2 lines from the initial cross between wild type with the RNAi knockdown line.

In the whole genome bisulfite sequencing (WGBS), two lines (line 2.1 and 4.61) from the BC1-F2 lines was used. WGBS showed a lower CG methylation in RNAi lines and lower CHH/CHG methylation in the epiRILs in comparison to wild type. The reduced methylation in non-CG sequence context was a likely consequence of reduced *MET1* expression in parental lines and there was evidence that re-methylation in later epiRIL generations involved CHG and CHH contexts. The effect of loss in CG methylation on CHG and CHH was also found in Arabidopsis (Mathieu *et al*., 2007).

Brocklehurst *et al*. (2018) also stated that the effect of MET1 depletion could cause both changes in both CG and non-CG marks due to the involvement of MET1 in a multi-protein complex; where a complex which contain chromomethylases could lose full function without MET1. Another explanation is the indirect effect of MET1 on other epigenetic factors such as histone regulators. This was seen in *met1* mutant where the loss of H3K9 methylation patterns resulted in the loss of CHG and CHH methylation (Stroud *et al*., 2013; Brocklehurst *et al*., 2018).

The WGBS result confirms that remethylation of CG sequence occurred after several selfings. Low CG methylation in the RNAi parental line was not seen in the F_5 generation of epiRILs. The selection against the transgene in the early generations could cause the normal function of MET1 causing methylation of CG.

Analysis on differentially methylated regions (DMRs) between the epiRILs and wild type confirms that the remethylation of CG sequence did not occur on all of the chromosomes. Some chromosomes showed specific patterns of methylation in the two different epiRILs at generation F_5 (Chapter 5). In line 2.1, hypomethylated DMR was found to be maintained in the epiRIL and can be seen by the number of hypomethylated DMRs on chromosomes 2,3 and 7. While in line 4.6 a different distribution of DMRs was observed with clusters on chromosomes 3, 5 and 7. These different pattens of methylation support the idea that mosaics of methylation in individual epiRILs are unique. Irregular and stochastic inheritance of methylation were also observed in early generations of inbreeding (Mathieu *et al*., 2007) and in the

epiRIL population of Arabidopsis (Reinders *et al*., 2009; Johannes *et al*., 2009).

6.5 Future Work

The next step in the investigation of tomato epiRIL population will be to undertake a full phenotypic analysis on a properly replicated population under a range of environmental conditions. This will now be possible as all lines are at F_8 and essentially fixed and seed has been bulked for such an analysis. The analysis should be performed at the level of visible phenotypes as well as analysis of the transcriptome and metabolome. The epigenomes of each of the epiRILs need to be analysed by WGBS. Replicated trials with the lines would then indicate lines that could be followed in more detail to link epigenetic changes with phenotypes.

In the *ddm1*-derived Arabidopsis epiRILs, differentially methylated regions (DMRs) were found to act as epigenetic quantitative loci (epiQTL) and could account up to 60-90 % of the heritability of flowering time and primary root length (Cortijo *et al*., 2014). Interval mapping was undertaken for the two traits and linkage analysis detected highly significant epiQTL on several chromosomes. A more recent study on the impact of epigenetic variation on plant metabolic composition using the *ddm1*-derived Arabidopsis epiRILs discovered epiQTLs linked to plant growth and morphology (Kooke *et al*., 2019). The detection of epiQTL suggests that the metabolic variation can partly be explained by the variation in DMRs.

The tomato epiRIL population can also be used to find epiQTLs for stable phenotypic variation discovered in this project. Further analysis of the metabolome and transcriptome of the tomato epiRILs can be used to advance our understanding of epigenetic control in plant primary and secondary metabolism. The understanding of the nutritional quality in tomato is important because it is one of the goals in tomato breeding besides achieving higher yield, longer shelf-life and better taste (Bai and Lindhout, 2007). A better understanding of both genetic and epigenetic control of plant secondary metabolites could help to achieve the goal in producing a tomato variety with better nutritional quality. The epiRIL population provides a unique resource for investigating the epigenetic basis of trait variation in tomato which could also be applied in other crop species.

References

Abejide, D.R, Falusi, O.A.Daudu, O.A.Y., Gado, A. Dangana, M.C. and Lateef, A.A. (2013) Assessment of Pollen Production, viability and germinability in three Sesame cultivars. *International Journal of Applied Biological Research*, 5(10):62-71

Adato, A.,Mandel,T., Mintz-Oron, S., Venger, I., Levy, D., Yativ, M., Dominguez, E., Wang, Z., De Vos, R.C.H., Jetter, R., Schreiber, L., Heredia, A., Rogachev, I. and Aharoni, A. (2009) Fruit-surface flavonoid accumulation in tomato is controlled by a SlMYB12-regulated transcriptional network. *PLOS Genetics,* 5 (12): e1000777

Alberts, B., Johnson,A.,Lewis,J., Morgan,D., Raff,M., Roberts, K. and Walter, P. (2015) *Molecular Biology of The Cell*. Garland Science, Taylor and Francis Group. New York.

Alpert, K.B., Grandillo,S., Tanksley, S.D. (1995) *fw2.2*:a major QTL controlling fruit weight is common to both red- and green-fruited tomato species. *Theoretical and Applied Genetics,* 91:994-1000

Arias, R., Lee,T.,Logendra, L. and Janes, H. (2000) Correlation of lycopene measured by HPLC with the L^*, a^*, b^* color readings of a hydroponic tomato and the relationship of maturity with color and lycopene content. *Journal of Agricultural and Food Chemistry,* 48(5):1697-1702

Bai, Y. and Lindhout, P. (2007) Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? *Annals of Botany,* 100: 1085-1094

Bai, S., Tuan, P. A., Saito, T., Honda, C., Hatsuyama, Y., Ito, A. and Moriguchi, T. (2016). Epigenetic regulation of MdMYB1 is associated with paper bagging-induced red pigmentation of apples. *Planta,* 244**:** 573-86.

Bassel, G.W., Mullen, R.T. and Bewley, J.D. (2008) *Procera* is a putative *DELLA* mutant in tomato (*Solanum lycopersicum*): effects on the seed and vegetative plant. *Journal of Experimental Botany*, 59(3): 585–593,

Bergougnoux, V. (2014) The History of Tomatoes: From Domestication to Biopharming. *Biotechnology Advances,* 32: 170-189

Bertin, N. (2005). Analysis of the tomato fruit growth response to temperature and plant fruit load in relation to cell division, cell expansion and DNA endoreduplication. *Annals of Botany,* 95(3):439-47.

Bolger, A., Scossa, F., Bolger, M. E., Lanz, C., Maumus, F., Tohge, T., Quesneville, H., Alseekh, S., Sørensen, I., Lichtenstein, G., Fich, E. A., Conte, M., Keller, H., Schneeberger, K., Schwacke, R., Ofner, I., Vrebalov,

J., Xu, Y., Osorio, S., Aflitos, S. A., Schijlen, E., Jiménez-Goméz, J. M., Ryngajllo, M., Kimura, S., Kumar, R., Koenig, D., Headland, L. R., Maloof, J. N., Sinha, N., van Ham, R. C. H. J., Lankhorst, R. K., Mao, L., Vogel, A., Arsova, B., Panstruga, R., Fei, Z., Rose, J. K. C., Zamir, D., Carrari, F., Giovannoni, J. J., Weigel, D., Usadel, B. & Fernie, A. R. (2014) The genome of the stress-tolerant wild tomato species *Solanum pennellii*. *Nature Genetics*, 46: 1034.

Boureau, L., How-Kit, A., Teyssier, E., Drevensek, S., Rainieri, M., Joubes, J., Stammitti, L., Pribat, A., Bowler, C., Hong, Y. & Gallusci, P. (2016) A CURLY LEAF homologue controls both vegetative and reproductive development of tomato plants. *Plant Molecular Biology*: 1-17

Borassi, C., Sede, A. R., Mecchia, M. A., Salgado Salter, J. D., Marzol, E., Muschietti, J. P. and Estevez, J. M. (2015). An update on cell surface proteins containing extensin-motifs. *Journal of Experimental Botany,* 67**:** 477-487.

Bowman, J.L,Smyth, D.R. and Meyerowitz, E.M. (2012) The ABC flower model of flower development : Then and now. *Development,* 139: 4095-4098

Brocklehurst, S., Watson, M., Carr, I. M., Out, S., Heidmann, I. and Meyer, P. (2018). Induction of epigenetic variation in Arabidopsis by over-expression of DNA METHYLTRANSFERASE1 (MET1). *PLoS One,* 13**:** e0192170.

Broman, K.W. (2005) The Genomes of Recombinant inbred Lines. *Genetics,* 169: 1133–1146

Busch, B.L.,Schmitz,G., Rossmann,S., Piron,F., Ding,J., Bendahmane, A. and Theres, K. (2011) Shoot branching and leaf dissection in tomato are regulated by homologous gene modules. *The Plant Cell*, 23: 3595–3609

Cao,K., Cui,L., Zhou,X., Ye,L., Zou,Z. and Deng,S. (2016) Four tomato flowering locust-like proteins act antagonistically to regulate floral initiation. *Frontiers in Plant Science*, 6: 1213

Camelo, A.F.L. and Gomez, P.A. (2004). Comparison of Color Indexes for Tomato Ripening*. Horticultura Brasileira*, 22 (3): 534-537

Candiloro, I. L. M., Mikeska, T. and Dobrovic, A. (2017). Assessing alternative base substitutions at primer CpG sites to optimise unbiased PCR amplification of methylated sequences. *Clinical epigenetics,* 9**:**31-31.

Carretero-Paulet, L., Galstyan, A., Roig-Villanova, I., Martínez-García, J. F., Bilbao-Castro, J. R. and Robertson, D. L. (2010). Genome-Wide Classification and Evolutionary Analysis of the bHLH Family of Transcription Factors in Arabidopsis, Poplar, Rice, Moss, and Algae. *Plant Physiology*, 153: 1398-1412.

Cassler, M.D. (2015) Fundamentals of experimental design: Guidelines for designing successful experiments. *Agronomy Journal*, 107(2): 692-705

Catoni, M. and Cortijo, S. (2018). EpiRILs:Lessons From Arabidopsis. In M.Mirouze, E.Bucher and P.Gallusci. *Plant Genetics Coming of Age For Breeding Applications*. London, UK. Academic Press (pp:87-116).

Chaturverdi, P. and Tyag,S.C. (2014). Epigenetics mechanisms underlying cardiac degeneration and regeration. *International Journal of Cardiology*, 173(1): 1–11

Chang, F., Zhang,Z., Jin, Y. and Ma, H. (2014) Cell biological analyses of anther morphogenesis and pollen viability in Arabidopsis and Rice. In J.L. Riechmann and F.Wellmer (Eds) *Methods in molecular biology: Flower Development-methods and protocols*. New York. Springer Science and Business Media.

Chen, M., Ha, M., Lackey, E., Wang, J. and Chen, Z. J. (2008). RNAi of met1 reduces DNA methylation and induces genome-specific changes in gene expression and centromeric small RNA accumulation in Arabidopsis allopolyploids. *Genetics,* 178**:** 1845-58.

Chen, W. W., Kong, J. H., Qin, C., Yu, S., Tan, J. J., Chen, Y. R., Wu, C. Q., Wang, H., Shi, Y., Li, C. Y., Li, B., Zhang, P. C., Wang, Y., Lai, T. F., Yu, Z. M., Zhang, X., Shi, N. N., Wang, H. Z., Osman, T., Liu, Y. L., Manning, K., Jackson, S., Rolin, D., Zhong, S. L., Seymour, G. B., Gallusci, P. and Hong, Y. G. (2015) Requirement of CHROMOMETHYLASE3 for somatic inheritance of the spontaneous tomato epimutation Colourless non-ripening. *Nature Scientific Reports,* 5:9192

Chen, S., Wang, X., Zhang, L., Lin, S., Liu, D., Wang, Q., Cai, S., El-Tanbouly, R., Gan, L., Wu, H. and Li, Y. (2016) Identification and characterization of tomato gibberellin 2-oxidases (GA2oxs) and effects of fruit-specific SlGA2ox1 overexpression on fruit and seed growth and development. *Horticulture Research,* 3**:** 16059.

Cheng, X., Zhang, D., Cheng,Z., Keller,B. and Ling, H. (2009). A new family of Ty1-*copia*-like retrotransposons originated in the tomato genome by a recent horizontal transfer event. *Genetics* 181:1183-1193

Chuang, J.C. and Jones, P.A. (2007) Epigenetics and microRNAs. *Pediatric Research,* 61(5).

Clark,S.J., Harrison, J., Paul,C.L. and Frommer, M. (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Research* Vol. 22, No. 15: 2990-2997

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B. & Pang, E. C. K. (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-

assisted selection for crop improvement: The basic concepts. *Euphytica,* 142**:**169-196.

Cong, B., Liu, J. and Tanksley, S.D. (2002). Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations*. PNAS*, 99 (21): 13606-13611

Corem, S., Doron-Faigenboim, A., Jouffroy, O., Maumus, F., Arazi, T. and Bouché, N. (2018). Redistribution of CHH Methylation and Small Interfering RNAs across the Genome of Tomato *ddm1* Mutants. *The Plant Cell,* 30**,** 1628-1644.

Cortijo, S., Wardenaar, R., Colomé-Tatché, M., Gilly, A., Etcheverry, M., Labadie, K., Caillieux, E., Hospital, F., Aury, J.-M., Wincker, P., Roudier, F., Jansen, R. C., Colot, V. and Johannes, F. (2014) Mapping the Epigenetic Basis of Complex Traits. *Science,* 343**:** 1145-1148.

Costa, J.M and Heuvelink, E. (2004) Introduction: The tomato crop and industry. In E. Heuvelink (Ed). *Tomatoes*. Oxfordshire. UK. CABI Publishing (pp: 1-20).

Daminato, M., Mariero,S., Resentini,F., Lovisetto, A. and Casadoro, G. (2014) Characterization of *TM8*, a MADS-box gene expressed in tomato flowers. *BMC Plant Biology,* 14: 319

de Jong, M., Mariani, C. and Vriezen, W. H. (2009). The role of auxin and gibberellin in tomato fruit set. Journal of Experimental Botany, 60: 1523-1532.

Dreni,L. and Zhang, D. (2016) Flower development : the evolutionary history and functions of the AGL6 subfamily MADS-box gene. *Journal of Experimental Botany,* 67(6):1625-1638

Dubin, M. J., Mittelsten Scheid, O. and Becker, C. (20180. Transposons: a blessing curse. *Curr Opin Plant Biol,* 42**:** 23-29.

Elkind. Y, Gurnick, A. and Kedar, N. (1991) Genetics of semideterminate growth habit in tomato. *Hortscience,* 26(8): 1074-1075

Fambrini, M., Bonsignori,E., Rapparini,F., Cionini,G., Michelotti, V., Bertini,D., Baraldi, R., and Pugliesi, C. (2006) *Stem fasciated*, A recessive mutation in Sunflower (*Helianthus annuus*), alters plant morphology and auxin level. *Annals of Botany,* 98: 715–730

FAOSTAT (2014) Food and agricultural commodities-2012. *faostat.fao.org*

Felsenfeld, G. (2014) A Brief History of Epigenetics. *Cold Spring Harbor Perspective in Biology,* 6 (1):a018200

Feschotte, C. and Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annual review of genetics,* 41**:** 331-368.

Frary, A., Nesbitt, T. C., Frary, A., Grandillo, S., Knaap, E. v. d., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K. B. and Tanksley, S. D. (2000) *fw2.2*: A Quantitative Trait Locus Key to the evolution of tomato fruit size. *Science,* 289**:** 85-88

Finnegan, E.J. and Kovac, K.A. (2000) Plant DNA methyltransferases. *Plant Molecular Biology* 43:189-201

Galindo-Gonzalez, L., Mhiri,C., Deyholos, M.K. and Grandbastien, M. (2017) LTR-retrotransposons in plants: Engines of evolution. *Gene*, 626:14-25

Gallusci, P., Dai, Z., Genard, M., Gauffretau, A., Leblanc-Fournier, N., Richard-Molard, C., Vile, D. and Brunel-Muguet, S. (2017) Epigenetics for Plant Improvement: Current Knowledge and Modeling Avenues. *Trends in Plant Science,* 22**:** 610-623.

García-Hurtado, N., Carrera, E., Ruiz-Rivero, O., López-Gresa, M. P., Hedden, P., Gong, F. and García-Martínez, J. L. (2012) The characterization of transgenic tomato overexpressing gibberellin 20-oxidase reveals induction of parthenocarpic fruit growth, higher yield, and alteration of the gibberellin biosynthetic pathway. *Journal of Experimental Botany,* 63**:** 5803-5813.

Garg, R., Narayana Chevala, V., Shankar, R. and Jain, M. (2015). Divergent DNA methylation patterns associated with gene expression in rice cultivars with contrasting drought and salinity stress response. *Sci Rep,* 5**:** 14922.

Gebhardt, C. (2016) The historical role of species from the Solanaceae plant family in genetic research. *Theoretical and Applied Genetics,* 129 (12): 2281-2294

Georgelis, N. and Scott, J.W. (2004) Relationship of tomato fruit sugar concentration with physical and chemical traits and linkage of RAPD markers. *Journal of the American Society for Horticultural Science*, 129 (6): 839-845

Giorno, F., Wolters-Arts,M., Mariani, C. and Rieu, I. (2013) Ensuring reproduction at high temperatures: The heat stress response during anther and pollen development, *Plants* 2(3):489-506

Giovannoni, J.J. (2004) Genetic regulation of fruit development and ripening. *The Plant Cell*, 16:S170-S180

Galindo-González, L., Mhiri, C., Deyholos, M.K. and Grandbastein, M. (2017) LTR-retrotransposons in plants: Engines of evolution. *Gene,* 626: 14–25

Gómez-Díaz E, Jordà, M., Peinado,M.A., and Rivero, A. (2012) Epigenetics of host–pathogen interactions: The road ahead and the road behind. *PLOS Pathogens*, 8(11)

Grandillo, S., Ku, H.M. and Tanksley, S.D. (1999). Identifying the loci responsible for natural variation in fruit size and shape in tomato. *Theoretical and Applied Genetics*, 99 (6): 978–987

Grierson, D. (2013) Ethylene and the Control of Fruit Ripening. In G.B. Seymour, M.Poole, J.JGiovannoni and G.A.Tucker (Eds). *The Molecular Biology and Biochemistry of Fruit Ripening*. Oxford. Wiley-Blackwell (pp: 43-74)

Grierson, D. (2016) Identifying and silencing tomato ripening genes with antisense genes. *Plant Biotechnology Journal*, 14: 835-838

Guo, Q., Liu,Q., Smith, N.A., Liangi, G. and Wanga, M. (2016) RNA silencing in plants: Mechanisms, technologies and applications in horticultural crops. *Current Genomics,* 17: 476-489

Harrill, R. (1998) Using a refractometer to test the quality of fruits and vegetables. *www.nutritionsecurity.org*

Hay, A. and Tsiantis, M. (2010). KNOX genes: versatile regulators of plant development and diversity. *Development,* 137**:** 3153-65.

He, X., Chen, T. and Zhu, J. (2011) Regulation and function of DNA methylation in plants and animals. *Cell Research,* 21:442-465

He, Y., Yan,H., Hua, W., Huang, Y. and Wang, W. (2016). Selection and validation of reference genes for quantitative Real-time PCR in *Gentiana macrophylla*. *Frontiers in Plant Science*, 7: 1-13

How-Kit, A., Boureau, L., Stammitti-Bert, L., Rolin, D., Teyssier, E. and Gallusci, P. (2010). Functional analysis of *SlEZ1* a tomato enhancer of zeste (*E*(*z*)) gene demonstrates a role in flower development. *Plant Molecular Biology,* 74:201-213.

How-Kit, A., Daunay, A., Mazaleyrat, N., Busato, F., Daviaud, C., Teyssier, E., Deleuze, J. F., Gallusci, P.and Tost, J. (2015) Accurate CpG and non-CpG cytosine methylation analysis by high-throughput locus-specific pyrosequencing in plants. *Plant Molecular Biolology,* 88 (4)**:** 471-85

Hrdlickova, R., Toloue, M. and Tian, B. (2017) RNA-Seq methods for transcriptome analysis. *Wiley Interdisciplinary Review RNA,* 8.

Ikeda, Y. and Nishimura, T. (2015) The role of DNA Methylation in transposable element silencing and genomic imprinting. In Pontes, O. and H. Jin (Eds.). *Nuclear functions in plant transcription, signaling and development.* New York. Springer Science and Business Media.

Irish, V. (2017) The ABC model of flower development. *Current Biology* 27: R887-R909

Jasinski, S., Kaur, H., Tattersall, A. and Tsiantis, M. (2007) Negative regulation of KNOX expression in tomato leaves. *Planta,* 226: 1255-1263

Jia, T., Zeng, B., Luo, S., Li, X., Wu, B. and Li, J. (2016) Cloning, localization and expression analysis of two *fw2.2-like* genes in small- and large fruited pear species. *Journal of Integrative Agriculture*, 15(2):282-294

Jin, B., Li, Y. and Robertson, K. D. 2011. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer,* 2**:** 607-17.

Jin, H., Ma, Y., Shen, Q. and Wang, X. (2012) Circulating methylated DNA as biomarkers for cancer detection. In A. Dricu (Ed) *Methylation - From DNA, RNA and Histones to Diseases and Treatment*. Rijeka, Croatia. InTech Publisher.

Johannes, F., Porcher, E., Teixeira, F. K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuisson, J., Heredia, F., Audigier, P., Bouchez, D., Dillmann, C., Guerche, P., Hospital, F. and Colot, V. (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet,* 5**,** e1000530

Joldersma, D. and Liu, Z. (2018) The making of virgin fruit: the molecular and genetic basis of parthenocarpy. *Journal of Experimental Botany,* 69**:** 955-962.

Jouffroy, O., Saha, S., Mueller, L., Quesneville, H. and Maumus, F. (2016). Comprehensive repeatome annotation reveals strong potential impact of repetitive elements on tomato ripening. *BMC Genomics,* 17**:** 624.

Kakutani, T., Jeddeloh, J.A. and Richards, E.J. (1995) Characterization of an *Arabidopsis thaliana* DNA hypomethylation mutant. *Nucleic Acid Research,* 23(1):130-137

Kankel, M. W., Ramsey, D. E., Stokes, T. L., Flowers, S. K., Haag, J. R., Jeddeloh, J. A., Riddle, N. C., Verbsky, M. L. and Richards, E. J. (2003) Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics,* 163**:** 1109-1122

Kanayama, Y. 2017. Sugar Metabolism and Fruit Development in the Tomato. The Horticulture Journal, 86: 417-425.

Kanazawa, A. Inaba, J., Shimura, H., Otagaki, S., Tsukahara, S., Matsuzawa, A., Kim, B.M., Goto, K. and Matsuta, C. (2011). Virus-mediated efficient induction of epigenetic modifications of endogenous genes with phenotypic changes in plants. *The Plant Journal,* 65: 156–168

Kato, M., Miura, A., Bender, J., Jacobsen, S.E. and Kakutani, T. (2003) Role of CG and Non-CG methylation in immobilization of transposons in Arabidopsis. *Current Biology,* 13: 421-426

Khatun, K., Nath, U. K., Robin, A. H. K., Park, J.-I., Lee, D.-J., Kim, M.-B., Kim, C. K., Lim, K.-B., Nou, I. S. and Chung, M.-Y. (2017). Genome-wide analysis and expression profiling of zinc finger homeodomain (ZHD) family genes reveal likely roles in organ development and stress responses in tomato. *BMC Genomics,* 18**:** 695.

Kimura, S. and Sinha, N. (2008) *Tomato (Solanum lycopersicum): A Model Fruit-Bearing Crop*. Cold Spring Harbor Protocols. 3(11): doi:10.1101/pdb.emo105

Klap, C., Yeshayahou, E., Bolger, A. M., Arazi, T., Gupta, S. K., Shabtai, S., Usadel, B., Salts, Y. and Barg, R. (2017). Tomato facultative parthenocarpy results from SlAGAMOUS-LIKE 6 loss of function. Plant biotechnology journal, 15, 634-647.

Knapp, S., Bohs,L., Nee, M. and Spooner, D.M. (2004) Solanaceae-a model for linking genomics with biodiversity. *Comparative and Functional Genomics,* 5: 285 – 291.

Knapp, S. and Peralta, I.E. (2016) The Tomato (*Solanum lycopersicum* L., Solanaceae) and Its Botanical Relatives. In M. Causse, J.Giovannoni, M.Bouzayen and M.Zouine (Eds.). *The Tomato Genome, Compendium of Plant Genomes*. Verlag Berlin Heidelberg. Springer.

Kooke, R., Morgado, L., Becker, F., van Eekelen, H., Hazarika, R., Zheng, Q., de Vos, R. C. H., Johannes, F. and Keurentjes, J. J. B. (2019). Epigenetic mapping of the Arabidopsis metabolome reveals mediators of the epigenotype-phenotype map. Genome research, 29: 96-106.

Koornneef, M. and Meinke, D. (2010). The development of Arabidopsis as a model plant. The Plant Journal, 61 (6): 909-912

Kukurba, K. R. and Montgomery, S. B. (2015) RNA Sequencing and Analysis. *Cold Spring Harbor Protocol,* 2015**:** 951-69.

Kumar, A. and Benetzen, J.L. (1999) Plant retrotransposons. *Annual Review of Genetics*, 33:479-532

Laird, P.W. (2010) Principles and challenges of genome-wide DNA methylation analysis. *Nature Reviews Genetics,* 11:191-203

Lam, J.K.W., Chow, M.Y.T., Zhang, Y. and Leung, S.W.S. (2015) siRNA versus miRNA as Therapeutics for Gene Silencing. *Molecular Therapy—Nucleic Acids. www.moleculartherapy.org/mtna.*

Lauria, M. and Rossi, V. (2011) Epigenetic control of gene regulation in plants. *Biochimica et Biphysica Acta,* 1809: 369-378

Le, Q. H., Wright, S., Yu, Z. and Bureau, T. (2000). Transposon diversity in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America,* 97**:** 7376-7381

Lechner, E., Achard, P., Vansiri, A., Potuschak, T. and Genschik, P. (2006). F-box proteins everywhere. Current Opinion in Plant Biology, 9: 631-638.

Lepikhov, K., Tchernov, A., Zheleznaja,L., Matvienko, N., Walter, J. and Trautner, T.A. (2001). Characterization of the Type IV restriction modification system *Bsp*LU11III from *Bacillus* sp. LU11. *Nucleic Acid Research*, 29(22): 4691-4698

Li, Y. and Tollefsbol, T.O. (2011) DNA methylation detection: Bisulfite genomic sequencing analysis. *Methods in Molecular Biology,* 791:11-21

Li, J., Sima, W., Ouyang, B., Wang, T., Ziaf, K., Luo, Z., Liu, L., Li, H., Chen, M., Huang, Y., Feng, Y., Hao, Y. and Ye, Z. (2012) Tomato *SlDREB* gene restricts leaf expansion and internode elongation by downregulating key genes for gibberellin biosynthesis. *Journal of Experimental Botany,* 63**:** 6407-20.

Liu, X., Kim, Y. J., Müller, R., Yumul, R. E., Liu, C., Pan, Y., Cao, X., Goodrich, J. and Chen, X. (2011). *AGAMOUS* Terminates Floral Stem Cell Maintenance in *Arabidopsis* by Directly Repressing *WUSCHEL* through Recruitment of Polycomb Group Proteins. *The Plant Cell,* 23**:** 3654-3670.

Liu, R. E., How-Kit, A., Stammitti, L., Teyssier, E., Rolin, D., Mortain-Bertrand, A., Halle, S., Liu, M. C., Kong, J. H., Wu, C. Q., Degraeve-Guibault, C., Chapman, N. H., Maucourt, M., Hodgman, T. C., Tost, J., Bouzayen, M., Hong, Y., Seymour, G. B., Giovannoni, J. J. and Gallusci, P. (2015) A DEMETER-like DNA demethylase governs tomato fruit ripening. *Proceedings of the National Academy of Sciences of the United States of America,* 112**:** 10804-10809.

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H. and Ecker, J. R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell, 133: 523-36.

Livak, K.J. and Schmittgen, T.D. (2001). Analysis of relative gene expression data using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ method*. Elsevier Science Methods,* 25: 402-408

Lopez, C.M.R. and Wilkinson, M.J. (2015) Epi-fingerprinting and epi-interventions for improved crop production and food quality. *Frontiers in Plant Science*, 6 (397).

MacAlister, C.A., Park, S.J., Jiang,K., Marcel, F., Bendahmane, A., Izkovich, Y., Eshed, Y. and Lippman, Z.B. (2012). Synchronization of the flowering transition by the tomato *TERMINATING FLOWER* gene. *Nature Genetics*, 44 (12):1393-1398

Magwazaa, L.S. and Opara, U.L. (2015). Analytical methods for determination of sugars and sweetness of horticultural products. *Scientia Horticulturae,* 184: 179-192

Manning, K., Tor, M., Poole, M., Hong, Y., Thompson, A. J., King, G. J., Giovannoni, J. J. and Seymour, G. B. (2006) A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nature Genetics,* 38 (8)**:** 948-52.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F. C., McCarroll, S. A. and Visscher, P. M. 2009. Finding the missing heritability of complex diseases. Nature, 461: 747-753.

Margueron, R. and Reinberg, D. (2010) Chromatin Structure and the inheritance of epigenetic information. *Nature Reviews Genetics*, 11(4) : 285-296

Mari-Ordonez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V. and Voinnet, O. 2013. Reconstructing de novo silencing of an active plant retrotransposon. *Nat Genet,* 45**:** 1029-39.

Martienssen, R.A. and Colot, V. (2001) DNA methylation and Epigenetic Inheritance in Plants and Filamentous Fungi. *Science*, 293 (5532): 1070-1074.

Masser, D.R, Berg, A.S. and Freeman, W.M. (2013) Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenetics & Chromatin,* 6:33

Mathieu, O., Reinders, J., Caikovski, M., Smathajitt, C. and Paszkowski, J. (2007) Transgenerational stability of the Arabidopsis epigenome is coordinated by CG methylation. *Cell,* 130**:** 851-62.

Matzke, M.A. and Mosher, R.A. (2014) RNA-directed DNA Methylation: An epigenetic pathway of increasing complexity. *Nature Reviews,* 15 (15): 394-408

Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J. and Mathieu, O. (2009) Selective epigenetic control of retrotransposition in Arabidopsis. *Nature,* 461**:** 427-30.

Miura, A., Yonebayashi, S., Watanabe, K., Toyoma, T., Shimada, H. and Kakutani, T. (2001) Mobilization of Transposons by a Mutation Abolishing Full DNA Methylation in Arabidopsis. *Nature,* 411: 212-214

Mocellin, S. and Provenzano, M. (2004) RNA Interference: Learning gene knock-down from cell physiology. *Journal of Translational Medicine,* 2: 39

Moore, L.D., Le, T. and Fan, G. (2013) DNA methylation and its basic function. *Neuropsycopharmacology*, 1-16

Mu, Q., Huang, Z., Chakrabarti, M., Illa-Berenguer, E., Liu, X., Wang, Y., Ramos, A. and van der Knaap, E. (2017) Fruit weight is controlled by Cell Size Regulator encoding a novel protein that is expressed in maturing tomato fruits. *PLOS Genetics,* 13**:** e1006930.

Muller, F., Xu, J., Kristensen, L., Wolters-Arts, M., de Groot, P. F., Jansma, S. Y., Mariani, C., Park, S. and Rieu, I. (2016). High-Temperature-Induced Defects in Tomato (*Solanum lycopersicum*) Anther and Pollen Development Are Associated with Reduced Expression of B-Class Floral Patterning Genes. *PLoS One,* 11**,** e0167614.

Nesbitt, T.C., and Tanksley, S.D. (2001). *fw2.2* directly affects the size of developing tomato fruit, with secondary effects on fruit number and photosynthate distribution. *Plant Physiol*logy, 127: 575–583.

Nobel, D. (2015) Conrad Waddington and the origin of epigenetics. *The Journal of Experimental Biology,* 218: 816-818

Nosenko, T., Bondel, K. B., Kumpfmuller, G. and Stephan, W. (2016) Adaptation to low temperatures in the wild tomato species Solanum chilense. *Moecular Ecology,* 25**:** 2853-69.

Paane, D., Muller, S.A., Wirtz, S., Engel, A. and Bickle, T.A. (2001). The McrBC restriction endonuclease assembles into a ring structure in the presence of G nucleotide. *The EMBO Journal,* 20 (12): 3210-3217.

Pailles, Y., Ho, S., Pires, I. S., Tester, M., Negrão, S. and Schmöckel, S. M. (2017) Genetic Diversity and Population Structure of Two Tomato Species from the Galapagos Islands. *Frontiers in Plant Science,* 8: 138.

Paran, I. and van der Knaap, E. (2007) Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *Journal of Experimental Botany*, 58 (14): 3841–3852.

Pesaresi, P., Mizzotti, C., Colombo, M. and Masiero, S. (2014) Genetic regulation and structural changes during tomato fruit development and ripening. *Frontiers in Plant Science*, 5 (24).

Peralta, I.E, Knapp, S. and Spooner, D.M. (2005) New Species of Wild Tomatoes (Solanum Section Lycopersicon: Solanaceae) from Northern Peru. *Systematic Botany,* 30(2):424-434

Pieper, U., Groll, D. H., Wünsch, S., Gast, F.-U., Speck, C., Mücke, N. and Pingoud, A. (2002). The GTP-Dependent restriction enzyme McrBC from *Escherichia coli* forms high-molecular mass complexes with DNA and produces a cleavage pattern with a characteristic 10-base pair repeat. *Biochemistry,* 41: 5245-5254

Pikaard, C.S. and Scheid, O.M. (2014) Epigenetics regulation in plants. *Cold Spring Harbor Perspective in Biology,* (6): 1-33.

Prudent, M., Causse, M., Genard, M., Tripodi, P., Grandillo, S. and Bertin, N. 2009. Genetic and physiological analysis of tomato fruit weight and composition: influence of carbon availability on QTL detection. J Exp Bot, 60: 923-37.

Quinet, M. and Kinet, J. (2007) Transition to flowering and morphogenesis of reproductive structures in tomato. *International Journal of Plant Developmental Biology*, 1 (1): 64-74.

Ranjan, A., Ichihashi, Y. and Sinha, N.R. (2012) The tomato genome: implications for plant breeding, genomics and evolution. *Genome Biology,* 13 (8): 167.

Reinders, J., Wulff, B. B., Mirouze, M., Mari-Ordonez, A., Dapp, M., Rozhon, W., Bucher, E., Theiler, G. and Paszkowski, J. (2009) Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. *Genes and Development,* 23**:** 939-50.

Relton C.L.and Davey Smith, G. (2010) Epigenetic epidemiology of common complex disease: Prospects for prediction, prevention, and treatment. *PLOS Medicine*, 7(10).

Richardson, B. (2007). Primer: epigenetics of autoimmunity. *Nature Clinical Practice Rheumatology,* 3: 521-527

Samach, A. and Lotan, H. (2007) The Transition to Flowering in Tomato. *Plant Biotechnology,* 24: 71–82

Saze, H., Scheid, O.M. and Paszkowski, J. (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nature Genetics,* 34: 65-69

Schertl, P. and Braun, H.-P. (2014). Respiratory electron transfer pathways in plant mitochondria. *Frontiers in plant science,* 5**:** 163-163.

Sehgal, D., Singh, R. and Rajpal, V.R. (2016) Quantitative Trait Loci Mapping in Plants: Concepts and Approaches. In V.R. Rajpal, S.R.Rao and S.N.Raina (Eds.). *Molecular Breeding for Sustainable Crop Improvement, Sustainable Development and Biodiversity*. Switzerland. Springer International Publishing (pp: 31-59).

Serrani, J. C., Ruiz-Rivero, O., Fos, M. and García-Martínez, J. L. (2008) Auxin-induced fruit-set in tomato is mediated in part by gibberellins. *The Plant Journal,* 56**,** 922-934.

Shaffer, C. (2019) Challenges with Sanger sequencing. www.news-medical.net.

Shah, V., Shah, N.D. and Patrekar, P.V. (2013) Medicinal Plants from Solanaceae Family. *Research Journal of Pharmacy and Technology*, 6(2): 143-151.

Shani, E., Burko, Y., Ben-Yaakov, L., Berger, Y., Amsellem, Z., Goldshmidt, A., Sharon, E. and Ori, N. (2009) Stage-Specific Regulation of Solanum lycopersicum Leaf Maturation by Class 1 KNOTTED1-LIKE HOMEOBOX Proteins. *The Plant Cell,* Vol. 21: 3078–3092

Snape, J. W. and Riggs, T. J. 1975. Genetical consequences of single seed descent in the breeding of self-pollinating crops*. Heredity*, 35: 211-219.

Stokes, T.L., Kunkel, B.N. and Richards, E.J. (2002) Epigenetic variation in Arabidopsis disease resistance. *Genes and Development,* 16:171-182

Stommel, J.R. and Haynes, K.G. (1993). Genetic control of fruit sugar accumulation in a *Lycopersicon esculentum* X *L. hirsutum* Cross. *Journal American Society of Horticultural Science,* 118(6):859-863.

Schwarz.D, Thompson, J. and Klaring, H.P. (2014) Guidelines to use tomato in experiments with a controlled environment. *Frontiers in Plants Sciences*, 5: 625

Stewart, F. J., Panne, D., Bickle, T. A. and Raleigh, E. A. (2000). Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J Mol Biol,* 298**:** 611-22.

Strachan, T. and Read, A. (2011) *Human Molecular Genetics* 4th Edition. New York. Garland Science, Taylor and Francis Group.

Stroud, H., Greenberg, Maxim V. C., Feng, S., Bernatavichute, Yana V. and Jacobsen, Steven E. (2013). Comprehensive Analysis of Silencing Mutants Reveals Complex Regulation of the Arabidopsis Methylome. Cell, 152: 352-364.

Suzuki, M., Liao, W., Wos, F., Johnston, A. D., DeGrazia, J., Ishii, J., Bloom, T., Zody, M. C., Germer, S. and Greally, J. M. (2018) Whole-genome bisulfite sequencing with improved accuracy and cost. *Genome Research,* 28**:** 1364-1371

Tan, C.-M., Chen, R.-J., Zhang, J.-H., Gao, X.-L., Li, L.-H., Wang, P.-R., Deng, X.-J. and Xu, Z.-J. (2013). OsPOP5, a prolyl oligopeptidase family gene from rice confers abiotic stress tolerance in Escherichia coli. *International journal of molecular sciences,* 14**:** 20204-20219.

Tanksley, S.D. (2004) The genetic, developmental and molecular bases of fruit size and shape variation in tomato. *The Plant Cell,* 16: S181-S189

Teixeira, F. K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M. F., Voinnet, O., Wincker, P., Esteller, M. and Colot, V. (2009) A role for RNAi in the selective correction of DNA methylation defects. *Science,* 323**:** 1600-4.

Teyssier, E., Bernacchia, G., Maury, S., How Kit, A., Stammitti-Bert, L., Rolin, D. and Gallusci, P. (2008) Tissue dependent variations of DNA methylation and endoreduplication levels during tomato fruit development and ripening. *Planta,* 228 (3)**:** 391-9.

The Tomato Genome Consortium (2012) The tomato genome sequence provides insight into fleshy fruit evolution. *Nature,* 485: 635-641

Thorson, J. L. M., Smithson, M., Beck, D., Sadler-Riggleman, I., Nilsson, E., Dybdahl, M. and Skinner, M. K. 2017. Epigenetics and adaptive phenotypic variation between habitats in an asexual snail. Scientific Reports, 7: 14139.

Tieman, D., Zhu, G., Resende, M. F. R., Lin, T., Nguyen, C., Bies, D., Rambla, J. L., Beltran, K. S. O., Taylor, M., Zhang, B., Ikeda, H., Liu, Z., Fisher, J., Zemach, I., Monforte, A., Zamir, D., Granell, A., Kirst, M., Huang, S. and Klee, H. (2017) A chemical genetic roadmap to improved tomato flavor. *Science,* 355**:** 391-394

Top,O., Bar, C., O¨kmen, B., O¨zer, D.Y., Ruscxuklu,D., Tamer, N., Frary, A. and Doganlar, S. (2014) Exploration of three Solanum species for improvement of antioxidant traits in tomato. *HortScience,* 49(8): 1003-1009.

Tost, J. and Gut, I.G. (2007). DNA methylation analysis by pyrosequencing. *Nature Protocols*, 2 (9) : 2265-2275

Tuteja, N. 2003. Plant DNA helicases: the long unwinding road*. *Journal of Experimental Botany,* 54**:** 2201-2214.

Tsukaya, H. (2005) Leaf shape: genetic controls and environmental factors *International Journal of Developmental Biology*, 49: 547-555

Uhrig, R. G., Labandera, A.-M. and Moorhead, G. B. (2013). Arabidopsis PPP family of serine/threonine protein phosphatases: many targets but few engines. *Trends in Plant Science*, 18, 505-513.

Uniprot.org. (2019). Homeobox-leucine zipper protein HAT14. www.uniprot.org/uniprot/P46665

University of California - Davis (2014) Genetic history of tomatoes revealed by new sequencing. *ScienceDaily*.

van der Knaap, E. and Tanksley, S. D. 2003. The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in Yellow Stuffer tomato. Theoretical and Applied Genetics, 107: 139-147.

van der Knaap, E., Chakrabarti, M., Chu, Y. H., Clevenger, J. P., Illa-Berenguer, E., Huang, Z., Keyhaninejad, N., Mu, Q., Sun, L., Wang, Y. and Wu, S. (2014). What lies beyond the eye: the molecular mechanisms regulating tomato fruit weight and shape. *Frontiers in plant science,* 5**:** 227-227.

Verdin, E. and Ott, M. (2015) 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nature*, 16:258-264

Verelst, W. and Asard, H. 2003. A phylogenetic study of cytochrome b561 proteins. *Genome Biology*, 4: R38.

Vogel, K.E. (2009) Backcross Breeding. In M.P. Scott (Ed.). *Methods in Molecular Biology: Transgenic Maize*, vol. 526. New York. Humana Press, Springer Science + Business Media.

Vongs,A., Kakutani, T., Martienssen, R.A. and Richards, E.J. (1993) *Arabidopsis thaliana* DNA methylation mutants. *Science,* 260: 1926-1928

Vrebalov, J., Ruezinsky, D., Padmanabhan, V., White, R., Medrano, D., Drake, R., Schuch, W. and Giovannoni, J. (2002) A MADS-box gene necessary

for fruit ripening at the tomato ripening-inhibitor (rin) locus. *Science,* 296 (5566): 343-6.

Waalwijk, C. and Flavell, R. A. (1978). *Msp*I, an isoschizomer of *Hpa*II which cleaves both unmethylated and methylated *Hpa*II sites. *Nucleic Acids Research,* 5: 3231-3236

Waddington, C.H. (2012) The Epigenotype - Reprints and Reflections. *International Journal of Epidemiology*, 41:10–13

Wang, Y. H. and Irving, H. R. 2011. Developing a model of plant hormone interactions. Plant Signaling & Behavior, 6: 494-500.

Wang, D., Samsulrizal, N. H., Yan, C., Allcock, N. S., Craigon, J., Blanco-Ulate, B., Ortega-Salazar, I., Marcus, S. E., Bagheri, H. M., Perez Fons, L., Fraser, P. D., Foster, T., Fray, R., Knox, J. P. and Seymour, G. B. (2019) Characterization of CRISPR Mutants Targeting Genes Modulating Pectin Degradation in Ripening Tomato. *Plant Physiology,* 179**:** 544-557.

Weigel, D and Colot, V. (2012) Epialleles in Plant Evolution. *Genome Biology,* 13:249

Xu, G., Ma, H., Nei, M. and Kong, H. (2009). Evolution of F-box genes in plants: Different modes of sequence divergence and their relationships with functional diversification. Proceedings of the National Academy of Sciences, 106: 835-840.

Xu, Y. *Molecular Plant Breeding* (2010) Oxford. CAB International (pp: 417-457).

Yadav, V., Dwivedi, V.P., Bhattacharya, D., Mittal, A., Moodley, P. and Das, G. (2015) Understanding the Host Epigenetics in Mycobacterium tuberculosis Infection. *Journal of Genetics and Genome Research,* 2:1.

Yan, Y., Chen, H. and Costa, M. (2004) *Chromatin Immunoprecipitation Assays.* In T.O Tollefsbol (Ed). *Epigenetics Protocols*. New Jersey. Humana Press (pp: 9-20).

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W., Chen, H., Henderson, I. R., Shinn, P., Pellegrini, M., Jacobsen, S. E. and Ecker, J. R. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell,* 126**:** 1189-201.

Zhang, H., Lang, Z. and Zhu, J. (2018) Dynamics and Function of DNA Methylation in Plants. *Nature Reviews Molecular Cell Biology,* 19**:** 489-506

Zhong, S., Fei, Z., Chen, Y. R., Zheng, Y., Huang, M., Vrebalov, J., McQuinn, R., Gapper, N., Liu, B., Xiang, J., Shao, Y. and Giovannoni, J. J. (2013) Single-base resolution methylomes of tomato fruit development reveal

epigenome modifications associated with ripening. *Nature Biotechnology,* 31 (2): 154-159

Zielinski, Q.B. (1948) Fasciation in Lycopersicon. I.Genetic anlaysis of dominance modification. *Genetics,* 33:405-428

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. *Nature Genetics,* 39**:** 61-

Appendices

Appendix 1. List of lines used in the epiRIL population

Generation	Lines																		
T0	2		3				4												
T1	2.1*	2.A2	3.1	3.2*	3.A3	3.4A	4A1	4.2*	4.3*	4.4*	4.A5	4.6*	4.7*	4.8*	4.9*	4.10*	4.11*	4.A15	4.18*	4.A23
BC1-S1	2.11A			3.21A					4.32A			4.61A			4.94A	4.102A	4.111A			
	2.13A			3.22A					4.33A			4.62A			4.95A	4.103A	4.113A			
	2.17A			3.23A					4.34A			4.63A			4.96A	4.106A	4.114A			
	2.18A			3.24A					4.35A			4.64A			4.97A	4.107A	4.115A			
	2.19A			3.27A					4.36A			4.65A			4.98A	4.108A	4.116A			
	2.111A			3.28A					4.37A			4.66A			4.99A	4.109A	4.117A			
	2.113A			3.29A					4.38A			4.67A			4.910A	4.1011A	4.118A			
	2.115A			3.210A					4.310A			4.69A			4.916A	4.1012A	4.1113A			
	2.116A			3.213A					4.312A			4.610A			4.917A	4.1013A	4.1114A			
	2.117A								4.314A			4.611A			4.918A	4.1015A	4.1115A			
	2.118A											4.612A			4.919A	4.1017A	4.1117A			
	2.119A											4.613A			4.923A	4.1019A	4.1119A			
	2.120A											4.614A			4.924A	4.1024A	4.1122A			
	2.121A											4.615A			4.925A	4.1025A	4.1125A			
	2.122A											4.616A			4.926A	4.1026A	4.1126A			
	2.123A											4.617A			4.928A	4.1031A	4.1127A			
	2.124A											4.618A			4.931A		4.1128A			
	2.125A														4.932A		4.1130A			
	2.126A														4.934A		4.1132A			
	2.127A																			

T1 generated by crossing T0 transgenic met1 with WT. The azygous lines in T1 was planted to produce F2 which was used for the generation of the epiRILs.

Transgenic T1 (*) were backcrossed to WT and produced BC1-S1. 109 lines from the backcross was also used for the generation of the epiRILs. A= azygous

Appendix 2. Climate data from https://www.metoffice.gov.uk for year 2015-2019

Allowances have been made for topographic, coastal and urban effects where relationships are found to exist.
Seasons: Winter=Dec-Feb, Spring=Mar-May, Summer=June-Aug, Autumn=Sept-Nov. (Winter: Year refers to Jan/Feb).
Monthly values are ranked and displayed to 1 dp and seasonal/annual values to 2 dp. Where values are equal, rankings are based in order of year descending.
Data are provisional from January 2018 and Winter 2018. Last updated 01/10/2019.

England (the Midlands) Max Temperature (Degrees C)

Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	WIN	SPR	SUM	AUT	ANN
2015	7.2	6.8	10.0	14.3	14.8	19.2	20.5	20.3	17.1	14.3	12.0	12.1	7.29	13.01	19.97	14.45	14.07
2016	8.1	8.0	9.2	11.6	17.0	19.1	21.2	21.4	19.8	14.0	8.5	8.8	9.42	12.61	20.59	14.10	13.92
2017	6.5	8.3	12.1	13.5	17.7	20.3	21.3	19.8	17.3	15.2	9.8	7.3	7.89	14.45	20.47	14.13	14.14
2018	7.7	5.5	7.8	13.1	18.7	21.5	25.4	21.4	18.0	14.3	10.6	9.1	6.88	13.22	22.79	14.33	14.50
2019	6.6	11.0	11.3	13.9	16.1	18.4	22.3	21.7	18.5				8.82	13.75	20.81		

England (the Midlands) Min Temperature (Degrees C)

Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	WIN	SPR	SUM	AUT	ANN
2015	1.0	0.6	2.3	3.7	6.5	8.7	10.9	11.4	7.7	6.9	6.0	6.5	1.16	4.15	10.35	6.86	6.05
2016	2.1	1.2	1.5	2.8	7.2	10.8	12.1	12.2	11.7	6.8	2.2	2.5	3.30	3.86	11.74	6.90	6.12
2017	0.7	3.1	4.4	4.0	8.2	11.5	11.9	11.1	9.3	8.9	3.1	1.6	2.07	5.55	11.50	7.14	6.52
2018	2.1	-0.5	1.1	5.8	7.5	10.5	12.7	12.0	9.0	6.2	4.7	3.8	1.13	4.79	11.72	6.66	6.29
2019	0.9	1.7	3.9	3.6	6.0	9.7	12.5	12.4	9.5				2.16	4.53	11.53		

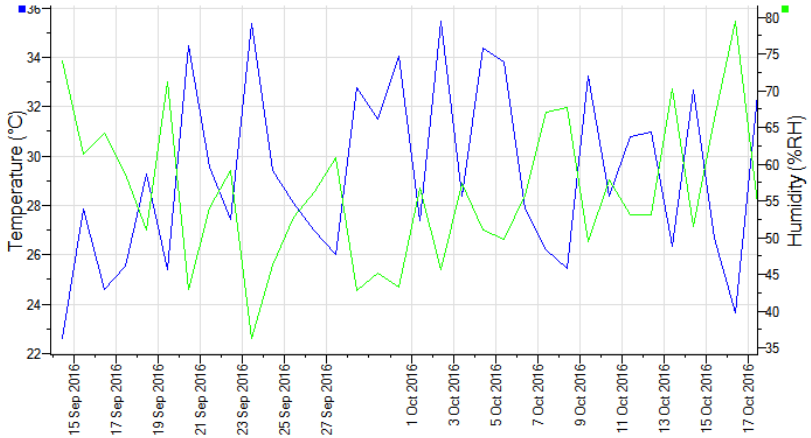
England (the Midlands) Mean Temperature (Degrees C)

Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	WIN	SPR	SUM	AUT	ANN
2015	4.1	3.7	6.1	9.0	10.6	13.9	15.7	15.8	12.4	10.6	9.0	9.3	4.22	8.58	15.16	10.66	10.06
2016	5.1	4.6	5.3	7.2	12.1	14.9	16.7	16.8	15.7	10.4	5.4	5.7	6.35	8.23	16.16	10.50	10.01
2017	3.6	5.7	8.3	8.7	12.9	15.9	16.6	15.5	13.3	12.1	6.5	4.5	4.97	10.00	15.99	10.64	10.33
2018	4.9	2.5	4.5	9.4	13.1	16.0	19.1	16.7	13.5	10.3	7.7	6.5	4.01	9.01	17.27	10.51	10.40
2019	3.9	6.3	7.6	8.7	11.0	14.0	17.4	17.0	14.0				5.52	9.14	16.17		

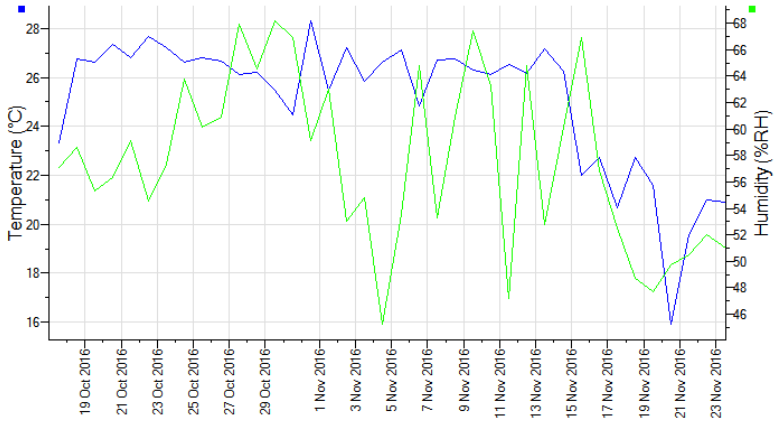
England (Midlands) Sunshine (Total hours)

Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	WIN	SPR	SUM	AUT	ANN
2015	72.5	77.4	119.8	218.5	173.4	220.3	186.1	148.2	169.3	85.8	35.3	34.0	225.6	511.7	554.6	290.4	1540.7
2016	44.6	99.5	117.2	165.2	206.2	125.5	188.3	203.9	133.1	107.9	81.3	53.0	178.2	488.7	517.8	322.3	1525.9
2017	51.3	48.9	125.0	175.6	198.0	174.1	177.2	175.8	114.3	83.3	82.3	55.8	153.2	498.6	527.1	280.0	1461.6
2018	53.6	94.3	76.4	115.7	254.0	243.4	274.2	167.8	151.2	123.9	63.7	46.7	203.7	446.1	685.4	338.8	1665.0
2019	50.1	118.6	132.3	168.2	197.3	143.7	187.1	198.1	167.3				215.5	497.8	528.9		

Appendix 3. Changes in temperature and humidity of the glasshouse during F3 generation (Sept 2016-January 2017)



Dates



Dates

= temperature
= humidity

Artificial light and heating were turned on at the end of British Summer Time at the end of October. The glasshouse temperature was maintained at 24-26 ^{0}C during the day and 18 ^{0}C at night.

Appendix 4. Randomized Complete Block Design at Generation F_2

Block	I	II	III
L I N E S	WT	4.15	4.1
	4.18	3.3	4.23
	3.4	4.5	2.2
	4.15	4.1	WT
	3.3	4.23	4.18
	4.5	2.2	3.4
	4.1	WT	4.15
	4.23	4.18	3.3
	2.2	3.4	4.5

Appendix 5. Randomized Complete Block Design at Generation F_3

Block	I	II	III	IV	V
L I N E S	2.2	4.1	BC2.1	3.3	BC1
	4.23	4.15	4.5	3.4	2.2
	BC3.2	4.23	4.15	2.2	4.18
	4.1	4.5	4.1	4.15	3.4
	3.3	BC3.2	3.3	WT	4.15
	4.5	3.4	2.2	BC2.1	4.23
	4.15	WT	4.18	4.5	BC3.2
	BC2.1	4.23	3.4	4.1	3.3
	4.18	BC2.1	WT	4.18	4.5
	3.4	4.18	BC3.2	4.23	4.1
	WT	2.2	4.1	BC3.2	WT

Appendix 6. Glasshouse layout in glasshouse A2 using complete randomized design for generation F_4 (193 plants)

497	4932	4932	4931	497	4925	496	496	4923	4923	494	494	4612	2116	4610
321														2119
495	418	418	3.2	3.2	4126	34	WT	41	499	41017	41019	4611		4611
495	41024	4926	34	41026	41031	WT	423	33	4916	498	4928	469		467
4618														2117
4113	4910	41127	41114	4114	41126	41122	4118	433	41017	499	2118	2125		419
434	4910	4102	41115	41113	41126	41122	4113	327	WT	498	4612	2123		467
4114 4111														466
4916	4917	4109	4919	4918	41024	4106	41130	41025	33	4117	2123	2121		465
41019	4107	4107	4919	4917	4918	4116	41113	41115	4116	4117	2125	2126		464
436														463
4108	41136	4628	415	435	323	41119	218	41127	41128	4106	41125	2124		462
324	4102	41132	4109	22	437	4103	217	4103	41132	4108	41119	2122		415
423														2.1
328	3210	22	438	4312	329	4924	4925	4934	4931	41011	41012	2120		4617
324	432	45	4314	4310	2111	4928	4924	4934	4926	41011	41012	2115		4618
45														4617
41015	41015	2.1	41117	41117	2123	2123	218	217	321					4616
41025	41031	41	4118	41125	211	211	219	219	4111					41115
														461
														4614
														2127

Note: empty columns represent pathways in the glasshouse. Lines 2.2, 3.3, 3.4, 4.1, 4.15, 4.18, 4.23 and 4.5 originate from F2 lines. All other lines were from BC1-S1 lines.

Appendix 7. Glasshouse layout in glasshouse F11 using complete randomized design for generation F_4 (44 plants)

463	3210	432
464	465	435
461	466	434
462	2126	4610
4115	2127	433
2122	2124	329
4115	2121	2111
2118	4613	322
4615	4613	327
4615	4614	328
4616	438	437
323	323	436
2119	2120	4310
2115	2117	4314
2116		4312

Total number of plants in generation F4 = 193 + 44 = 237

Lines 2.2, 3.3, 3.4, 4.1, 4.15, 4.18, 4.23 and 4.5 originate from F2 lines. All other lines were from BC1-S1 lines.

Appendix 8. Glasshouse layout in glasshouse A2 using complete randomized design for generation F_5 (172 plants)

435	42	434	432	418	434	437	4118	499	495	4310	22	4103	465	2123
2124														41114
4312	41019	4924	41012	4916	4910	437	466	4312	497	4924	34			2126
494	41015	4925	4108	4916	495	497	4925	496	494	496	22			435
2118														438
4118	499	4917	41012	4919	4923	2.1	4931	4113	41125	41026	4611			33
41025	4917	498	4111	4919	498	41113	4117	41125	4115	4113	41130			WT
462														418
4106	461	4114	45	41011	4109	4103	41017	2121	433	2125	4613			4611
41017	4314	3210	41019	41011	41031	41015	2120	322	433	2124	4614			2116
4618														327
328	323	324	2126	321	463	4928	41115	4931	41114	41117	4618			4614
327	2125	321	323	329	461	4928	4932	4107	4102	41119	4616			4616
4926														4612
324	3210	2122	2127	219	2111	467	463	41128	41119	436	4102			4115
328	2121	2116	2119	2120	213	2122	41128	41132	41130	4926	41132			41122
2115														4109
4314														41117
415	213	34	4613	41126	41	211	3.2	41126	4612	2118				41024
														462

Note: empty columns represent pathways in the glasshouse. . Lines 2.2, 3.3, 3.4, 4.1, 4.15, 4.18, 4.23 and 4.5 originate from F2 lines. All other lines were from BC1-S1 lines.

Appendix 9. Glasshouse layout in glasshouse D9 using complete randomized design for generation F_5 (62 plants)

4923	432	4114	4934	415 F4	438
211	41031	3.2	4107	45	45 F4
219	41024	4934	4108	41113	BC2.1 F4
465	4116	41122	41115	469	467
33	2111	322	4610	4111	4617
	329	42	4117	2.1	41025
	436	2117	4910	41026	217
	41	2119	41127	469	41127
	2127	2115	4617	4610	4106
	WT	2117	4932	217	464
	2123	415	218	4116	45
					218
					41025

Total number of plants = 172 + 62 = 234 plants.

Lines 2.2, 3.3, 3.4, 4.1, 4.15, 4.18, 4.23 and 4.5 originate from F2 lines. All other lines were from BC1-S1 lines.

Appendix 10. Glasshouse layout in glasshouse A2 using complete randomized design for generation F_6 (118 plants)

41024 2	41019 3	324 2	42 2	4910 2	4116 2	41025 2	4117 2	41 2	4118 2	4917 2	
41026 1	41114 2	3210 1	4925 2	329 2	42 3	4109 2	WT 3	217 2	2111 2	219 1	
4919 1	41113 2	4926 1	321 2	322 2	328 2	BC1 4	415 3	BC2 1	22 3	2118 2	
2314 2	418 3	323 2	4928 2	4106 3	4932 1	4102 2	4103 2	433 1	435 2	33 1	
2119 1	327 1	2116 2	2117 2	4616 2	4614 2	415 2	4934 2	22 1	4115 1	33 3	
34 2	438 2	WT 2	469 2	467 2	4611 1	4115 2	498 2	495 1	4931 1	418 1	
437 1	436 2	434 1	4610 2	41117 2	499 2	496 2	497 2	494 2	4923 1	4924 2	
432 1	41031 2	41015 2	4916 1	41128 2	41130 2	462 2	213 1	4618 2	465 1	41012 2	
4107 2	4613 1	41011 2	4113 2	4617 1	41132 2	41017 1	211 1	4114 2	4111 2	41119 2	
2125 1	2123 1	2127 2	4310 2	41122 2	41125 2	41126 1	41127 2	461 3	218 2		
463 2	2126 2	2115 1	4612 1	2120 1	2121 2	2124 2	2122 2	4612 1			
										DOOR	

Note : The last digit represent the replicate number of each line (1 or 2). Lines 2.2, 3.3, 3.4, 4.1, 4.15, 4.18, 4.23 and 4.5 originate from F2 lines. All other lines were from BC1-S1 lines.

Appendix 11. Glasshouse layout in glasshouse A6 using complete randomized design for generation F_6 (117 plants)

				341	332	41031	45 F5	411	4182	BC11	49341
	4971	4981	4991	49171	49492	49232	49241	49281	41091	41071	45 F4
	41015 1	41017 2	41011 1	41012 1	4952	4961	4941	49101	49162	49251	49262
	222	3241	3291	41119 1	41132 1	41025 1	41122 1	41113 1	41114 1	41151	41117 1
	41152	3272	3281	32102	41130 1	41024 1	41019 1	41123 1	41126 2	41127 1	41125 1
		21152	21241	3211	41161	41141	41171	41111	41131	41026 2	41031 1
	43141	21232	4621	2192	21111	2171	2181	41181	3231	2112	2132
46122	4631	46161	4361	43122	4342	4322	4641	21271	4691	46112	21211
4611	46141	21192	46181	4351	46132	46172	43101	BC24	WT1	21202	21221
		421	413	BC22	49312	41021	21161	21261	4671	46101	21181
		41061	343	4151	49322	41082	4381	4652	21171	4372	4332
	DOOR										

Note : The last digit represent the replicate number of each line (1 or 2). Lines 2.2, 3.3, 3.4, 4.1, 4.15, 4.18, 4.23 and 4.5 originate from F2 lines. All other lines were from BC1-S1 lines.

Appendix 12. Glasshouse layout in glasshouse A2 using complete randomized design for generation F_7 (112 plants)

21251	21261										
41181	411132	411142	411151	4111171	411191	411222	411251	411262	411271	411281	
21242	21222	21211	21201	21192	21182	21171	21162	21152	411321	411301	
46122	46111	46101	4691	4672	4651	4642	4632	4621	4611	21272	
221	WT1	BC2.1 F6	2123 F6	45 F5	41152	46181	46171	46161	46141	46131	
2181	2171	2132	2111	BC3.2 1	421	4181	4151	412	341	331	
3271	3241	3231	3222	3211	2111	2191	4342	4332	4322	4372	
4361	4352	43121	43101	4382	4951	4942	43141	4982	4971	4962	
49162	49102	4992	49232	49192	49171	49262	49252	49242	41172	41162	
41142	41132	41112	410312	410262	410254	410242	41091	410174	410152	410124	
49282	49312	49322	49341	41022	41032	41061	41072	41081	41092	410111	
											door

Note : The last digit represent the replicate number of each line (1 or 2). Lines 2.2, 3.3, 3.4, 4.1, 4.15, 4.18, 4.23 and 4.5 originate from F2 lines. All other lines were from BC1-S1 lines.

Appendix 13. Glasshouse layout in glasshouse A6 using complete randomized design for generation F_7 (126 plants)

				4694	BC1F6	333	2174	4184	423	BC24	4153
				414	32102	3292	3282	344	224	WT3	334
	2172	2184	21112	3212	3221	3242	3272	3281	3291	2123F6	32101
	BC1F6	WT2	222	332	342	411	4182	422	BC22	2112	2131
	46112	46121	46132	46142	46162	46172	46182	41151	45F5	2123F6	4152
	21252	21262	21271	4612	4622	4631	4641	4652	4671	4692	46102
	411194	411261	411272	411282	411302	21272	21181	21191	21202	21212	21221
	49234	49241	49251	49263	41182	411131	411141	411151	411172	411221	411252
	43103	4941	4954	4961	4972	4984	4991	49103	49164	49172	49191
	41143	41174	4321	4331	3444	4351	4362	4371	4381	43124	43142
	410121	410151	410172	410193	410241	410251	410264	410311	41111	41134	41161
	410112	41091	41082	41071	41062	41031	41023	49342	49324	49313	49281
door											

Note : The last digit represent the replicate number of each line (1 or 2). Lines 2.2, 3.3, 3.4, 4.1, 4.15, 4.18, 4.23 and 4.5 originate from F2 lines. All other lines were from BC1-S1 lines.

Appendix 14. Glasshouse layout in glasshouse A2 using complete randomized design for generation F_8 (120 plants)

		41130-2	497-2	4910-2	4917-2	41019-2	22-3	WT-4	22-4	415-4	465-2	438-2
		437-2	495-2	41122-2	2116-2	3210-2	2121-2	4109-2	2111-2	499-2	462-2	41125-2
		41128-2	2118-2	461-2	498-2	4103-2	2115-2	2126-2	4926-2	4916-2	2124-2	2117-2
		34-3	4919-2	494-2	41024-2	41026-2	4118-2	4924-2	4102-2	4923-2	4113-2	41031-2
		4934-2	41119-2	41126-2	41117-2	4931-2	41017-2	496-2	211-2	WT-3	41132-2	217-2
		34-4	327-2	BC32-2	BC32-2	BC21-3 F7	434-2	41-4	418-3	42-3	BC21-4 F7	41025-2
	41115-2	463-2	4613-2	2127-2	2119-2	323-2	4932-2	4117-2	41114-2	4108-2	2120-2	4106-2
		432-2	321-2	329-2	433-2	324-2	328-2	4107-2	4111-2	21125-2	2122-2	4115-2
		4925-2	42-4	41012-2	418-4	41011-2	4614-2	4610-2	218-2	4116-2	219-2	213-2
		41-3	322-2	33-3 F7	4617-2	4618-2	45-3 F6	41127-2	4612-2	469-2	4314-2	4616-2
									4928-2	415-3	4312-2	467-2
								4611-2	436-2	41015-2	435-2	4310-2
door												

Note : The last digit represent the replicate number of each line (1 or 2). Lines 2.2, 3.3, 3.4, 4.1, 4.15, 4.18, 4.23 and 4.5 originate from F2 lines. All other lines were from BC1-S1 lines.

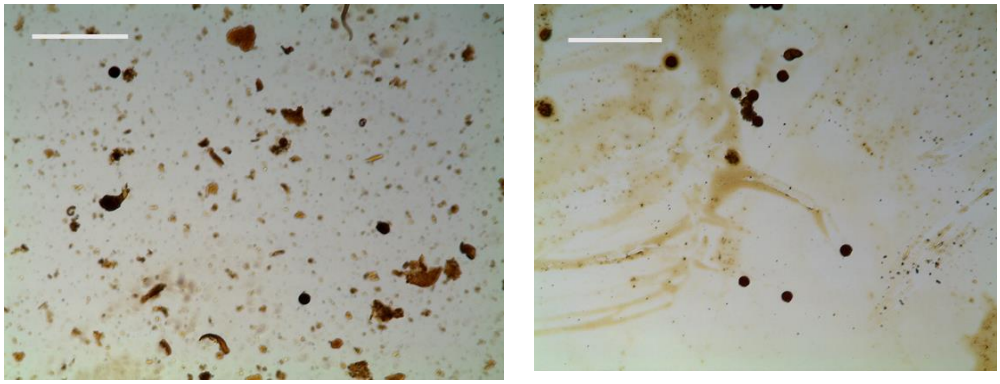
Appendix 15. Glasshouse layout in glasshouse A6 using complete randomized design for generation F_8 (120 plants)

											22-2	327-1	2111-1
											438-1	469-1	WT-2
			467-1	4613-1	41019-1	4111-1	2121-1	41-1	WT-1	2118-1	2122-1	211-1	2120-1
			462-1	41127-1	4614-1	4106-1	435-1	4117-1	2117-1	2116-1	415-1	34-1	41-2
			495-1	2126-1	415-2	4102-1	22-1	4116-1	4118-1	4314-1	BC32-1	494-1	418-1
			322-1	2125-1	BC21-1 F7	42-1	4312-1	2124-1	436-1	437-1	BC32-1	42-2	434-1
	4115-1	3210-1	2119-1	433-1	2115-1	4611-1	4113-1	41017-1	41024-1	4916-1	4934-1	41015-1	41012-1
	4926-1	496-1	498-1	4919-1	461-1	41132-1	41026-1	499-1	41025-1	41031-1	4928-1	4924-1	4910-1
			4612-1	33-1 F7	4114-1	2123-1 F7	41130-1	4108-1	4103-1	213-1	34-2	4923-1	4931-1
			464-1	218-1	321-1	324-1	323-1	41126-1	4107-1	41115-1	418-1	328-1	4109-1
			4932-1	463-1	2127-1	41125-1	41011-1	41119-1	41114-1	BC21-2 F7	4917-1	4925-1	497-1
			329-1	219-1	41122-1	41128-1	4617-1	4610-1	465-1	432-1	217-1	4618-1	4616-1
door													

Note : The last digit represent the replicate number of each line (1, 2, 3 or 4). Lines 2.2, 3.3, 3.4, 4.1, 4.15, 4.18, 4.23 and 4.5 originate from F2 lines. All other lines were from BC1-S1 lines.

Appendix 16. Pollen grains extracted from closed and open flowers of line 4.5

(A) **(B)**



(A) Pollen from closed flowers (B) Pollen from open flowers. The pollen were taken from generation F_5 plants. Scale bars = 200 µm

Appendix 17. Duncan Multiple Range Test (DMRT) of plant height at 4 weeks after planting at generations F_2 until F_7. Mean plant height with the same letter (a,b,c,d) shows no significant difference ($p<0.05$). Standard error of difference (s.e.d.) is presented for each ANOVA.

F_2		F_4		F_5		F_6		F_7	
Line	**Mean Plant Height (cm)**	**Line**	**Mean Plant Height (cm)**	**Line**	**Mean Plant Height (cm)**	**Line**	**Mean Plant Height (cm)**	**Line**	**Mean Plant Height (cm)**
4.5	20.33a	4.23	18.38a	4.5	6a	2.1	11.67a	3.3	13.23a
4.15	22.83ab	2.1	19.67ab	2.2	13.6a	4.18	15.33ab	4.15	15.5ab
4.23	24.67ab	4.15	20.88ab	4.15	14.33ab	3.3	17bc	2.2	17.25ab
3.3	24.83ab	2.2	21.17ab	2.1	17.43bc	WT	21cd	4.2	19.25abc
2.2	26.33ab	4.18	21.88ab	3.3	17.73bc	3.4	22d	4.1	21abcd
3.4	27.83ab	4.5	22ab	4.1	17.73bc	3.2	22d	WT	22.38bcd
4.1	27.83ab	3.3	22.88abc	4.23	18.07bc	4.23	22.33d	4.18	26.67cd
WT	28.33ab	WT	24.25bc	WT	18.93cd	4.15	22.67d	3.4	27.33d
4.18	30.17b	4.1	24.75bc	3.4	19.13cd	2.2	23.67d		
		3.4	27.88cd	4.18	20.23cd	4.1	23.67d		
		3.2	30.62d	3.2	22.07d				

Appendix 18. Duncan Multiple Range Test (DMRT) of the B+7 fruit soluble solids from azygous lines and wild type at generations F_2, F_3 and F_4. Degrees Brix with the same letter (a,b,c,d) shows no significant difference ($p<0.05$). Standard error of difference (s.e.d) for each generation was calculated.

F_2		F_3		F_4	
Line	Fruit soluble solids (^{0}Brix)	Line	Fruit soluble solids (^{0}Brix)	Line	Fruit soluble solids (^{0}Brix)
4.23	5.60 a	4.2	4.67 a	4.2	4.8 a
4.18	5.89 ab	4.15	4.79 a	3.4	4.82 a
3.3	5.94 ab	3.4	4.95 abc	4.5	4.9 ab
3.4	6.08 b	2.2	4.98 abcd	3.3	4.95 abc
4.15	6.16 b	3.3	5.07 abcd	4.18	4.97 abc
4.1	6.20 b	4.1	5.10 abcd	2.2	5.12 abc
4.5	6.21 b	4.5	5.22 bcd	4.1	5.22 bc
WT	6.22 b	4.18	5.33 cde	WT	5.27 bc
2.2	6.24 b	WT	5.51 def	4.15	5.30 c
s.e.d	0.1975	s.e.d	0.2317	s.e.d	0.1627

Appendix 19. Duncan Multiple Range Test (DMRT) of the B+7 fruit colour index from azygous lines and wild type at generations F_2, F_3 and F_4. Mean fruit colour index with the same letter (a,b,c,d) shows no significant difference ($p<0.05$). Standard error of difference (s.e.d) for each generation was calculated.

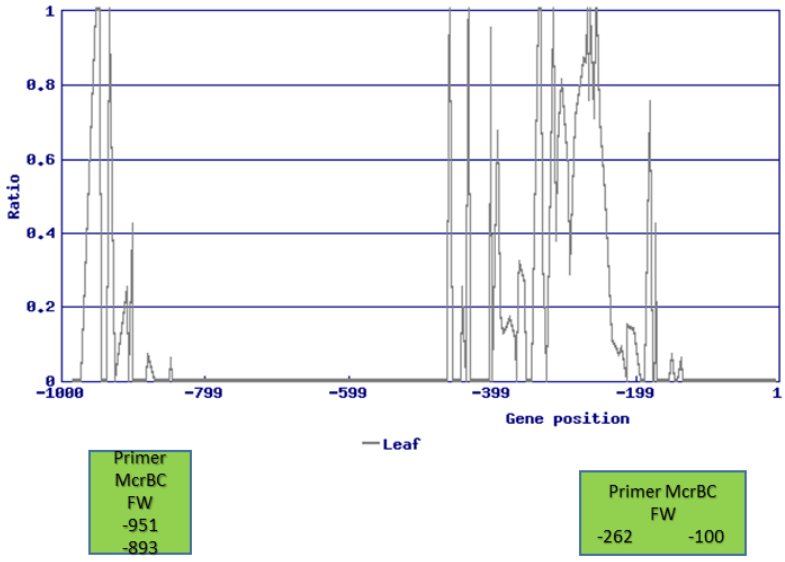
F_2		F_3		F_4	
Line	Mean Fruit Colour Index	Line	Mean Fruit Colour Index	Line	Mean Fruit Colour Index
4.5	39.54 a	4.1	34.06 a	4.15	29.95 a
4.15	39.69 a	4.15	34.88 ab	4.5	30.06 a
4.18	39.81 a	3.3	35.56 abc	4.1	31.67 ab
WT	39.91 a	4.2	36.8 abcd	2.2	31.89 ab
4.1	40.12 a	4.5	37.78 bcd	3.4	34.31 bc
3.3	40.26 a	WT	37.95 bcd	3.3	34.47 bc
4.23	40.77 a	4.18	38.92 cd	WT	36.17 cd
3.4	40.96 a	3.4	39.1 cd	4.2	37.8 cd
2.2	41.68 a	2.2	39.83 d	4.18	41.35 e
s.e.d	1.645	s.e.d	1.137	s.e.d	1.426

Appendix 20. Duncan Multiple Range Test (DMRT) of the B+7 fruit weight from azygous lines and wild type at generations F_2, F_3 and F_4. Mean weight with the same letter (a,b,c,d) shows no significant difference ($p<0.05$). Standard error of difference (s.e.d) for each generation was calculated.

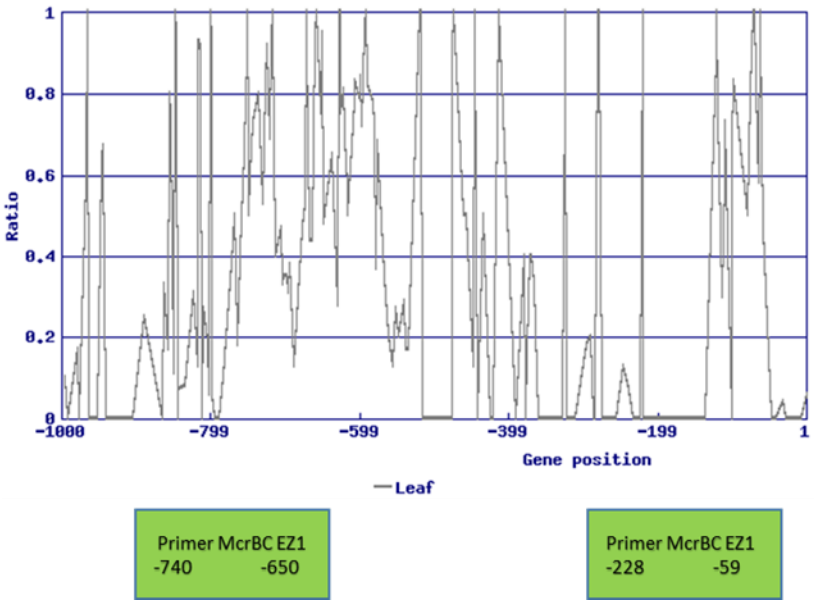
F_2		F_3		F_4		F_5		F_6	
Line	Mean Fruit Weight (gram)	Line	Mean Fruit Weight (gram)	Line	Mean Fruit Weight (gram)	Line	Mean Fruit Weight (gram)	Line	Mean Fruit Weight (gram)
4.23	2.67 a	4.23	2.47 a	4.15	4.73 a	4.5	2.336a	4.18	3.567a
4.15	4.35 b	2.2	2.52 a	4.5	5.55 a	4.1	3.86ab	4.23	3.667a
3.4	4.39 b	4.5	3.00 a	4.1	5.81 a	2.1	4.45bc	2.2	3.783a
4.5	4.62 b	4.1	3.18 ab	2.2	7.45 a	4.23	4.933bc	3.3	4.417a
3.3	5.00 b	4.15	3.32 ab	2.1	7.6a	3.2	5.583bcd	4.1	4.433ab
4.1	5.05 b	3.4	4.85 bc	4.23	8.45 bc	4.15	5.617bcd	2.1	5.117ab
2.2	5.38 b	4.18	4.90 bc	3.3	8.82 bc	4.18	6.1cde	4.15	5.583b
4.18	5.46 bc	3.3	5.33 cd	4.18	8.96 bc	2.2	6.817de	3.2	5.783b
WT	6.44 c	2.1	5.6cd	3.2	9.1bc	3.4	7.083de	3.4	6.083b
		3.2	5.9cd	3.4	9.74 c	3.3	7.55e	WT	7.85c
		WT	6.87 d	WT	9.92 c	WT	9.633f		
s.e.d	0.438	s.e.d	0.3841	s.e.d	0.746	s.e.d	0.827	s.e.d	0.771

Appendix 21. Position of the primers for McrBC analysis**.** (A) primer for *FW2.2*, (B) *SlEZ1*, (C) *SlEZ2* and (D) *TM8* analysis. The region is shown for 1000 bp upstream of TSS of each gene. . The ratio is methylated cytosine. The methylation ratio data is from the tomato epigenome data (http://ted.bti.cornell.edu/)

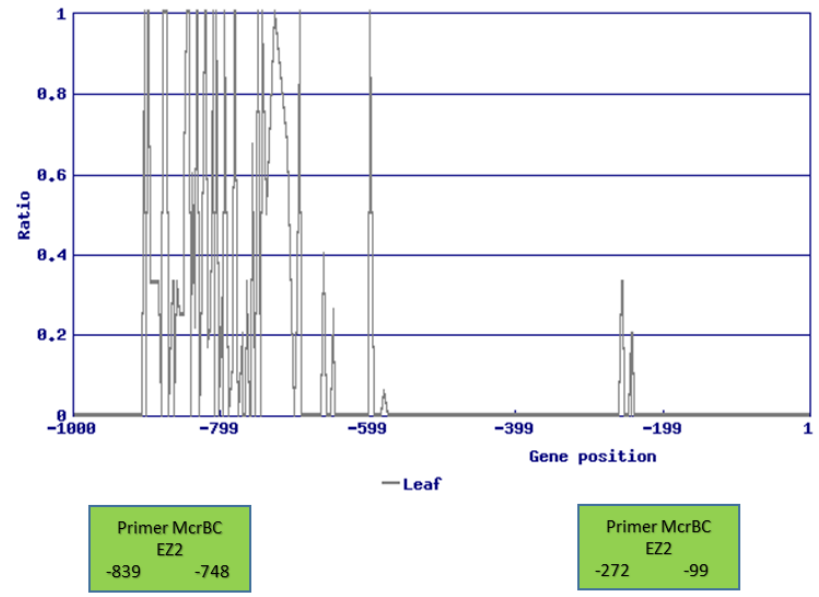
(A)



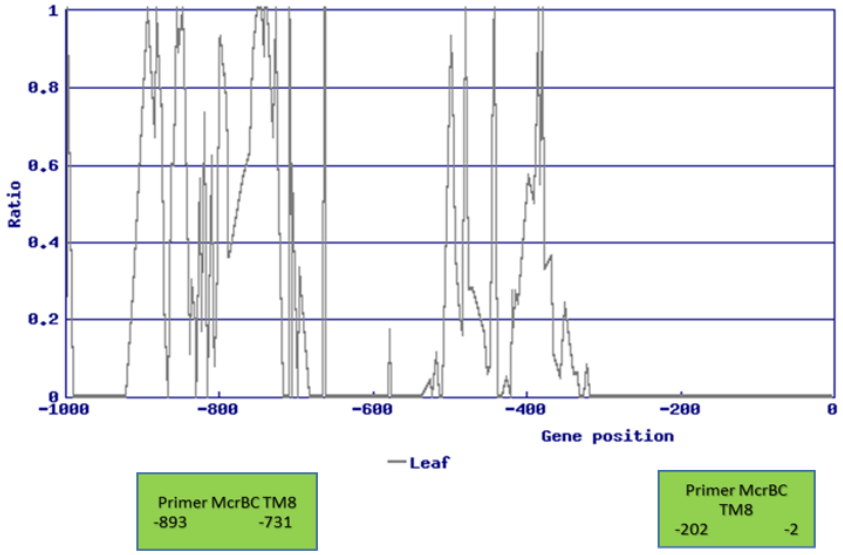
(B)



(C)


1
0.8
0.6
0.4
0.2
0
Ratio
-1000
-799
-599
-399
-199
1
Gene position
Leaf
Primer McrBC EZ2 -839 -748
Primer McrBC EZ2 -272 -99

(D)


1
0.8
0.6
0.4
0.2
0
Ratio
-1000
-800
-600
-400
-200
0
Gene position
Leaf
Primer McrBC TM8 -893 -731
Primer McrBC TM8 -202 -2

Appendix 22. Regression analysis between gene expression, fruit weight and methylation levels of *FW2.2*, *SlEZ1*, *SlEZ2* and *TM8* genes

A. Estimates of parameters for linear regression analysis between fruit weight (B+7) and RQ of *FW2.2* gene

Parameter	estimate	s.e.	t(26)	t pr.
Constant	6.886	0.618	11.15	<.001
RQ	-1.082	0.534	-2.03	0.053

B. Estimates of parameters for linear regression analysis between RQ of *SlEZ1* gene with fruit weight

Parameter	estimate	s.e.	t(16)	t pr.
Constant	4.954	0.708	7	<.001
Ez1	0.95	0.38	2.5	0.024

C. Estimates of parameters for linear regression analysis between RQ of *SlEZ2* gene with fruit weight

Parameter	estimate	s.e.	t(16)	t pr.
Constant	7.865	0.789	9.97	<.001
EZ2	-2.177	0.915	-2.38	0.03

D. Estimates of parameters for linear regression analysis between RQ of *TM8* gene with fruit weight

Parameter	estimate	s.e.	t(16)	t pr.
Constant	4.662	0.88	5.3	<.001
TM8	0.895	0.403	2.22	0.041

E. Estimates of parameters for linear regression analysis between McrBC-RQ (0-500 bp upstream TSS of *FW2.2*) and cDNA-RQ *FW2.2*

Parameter	estimate	s.e.	t(26)	t pr.
Constant	0.999	0.206	4.84	<.001
mcrbc_A	-0.0134	0.0621	-0.22	0.831

F. Estimates of parameters for linear regression analysis between McrBC-RQ (500-1000 bp upstream TSS of FW2.2) and cDNA-RQ *FW2.2*

Parameter	estimate	s.e.	t(26)	t pr.
Constant	1.223	0.237	5.15	<.001
mcrbc_B	-0.161	0.127	-1.27	0.215

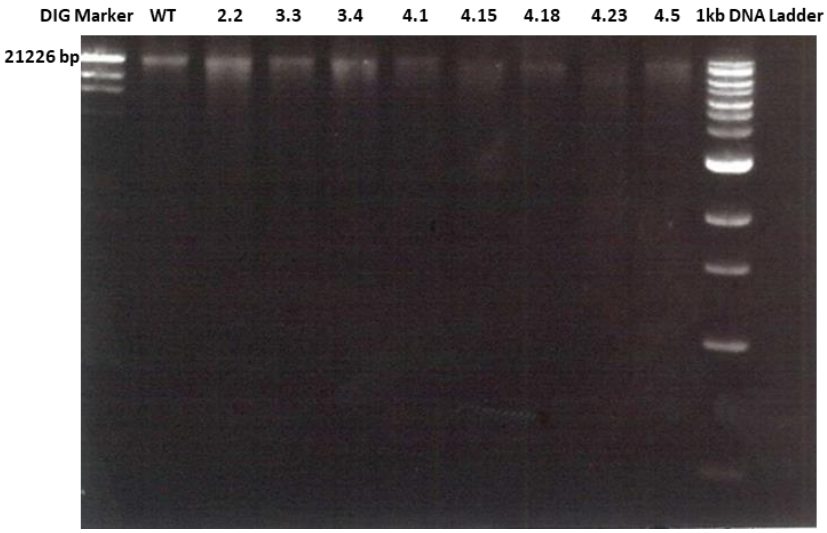
G. Multiple Regression between fruit weight, McrBC-RQ (0-500 bp upstream TSS of *FW2.2*) and cDNA-RQ

Parameter	estimate	s.e.	t(25)	t pr.
Constant	7.367	0.775	9.5	<.001
MCRBC	-0.173	0.169	-1.02	0.316
RQ	-1.105	0.534	-2.07	0.049

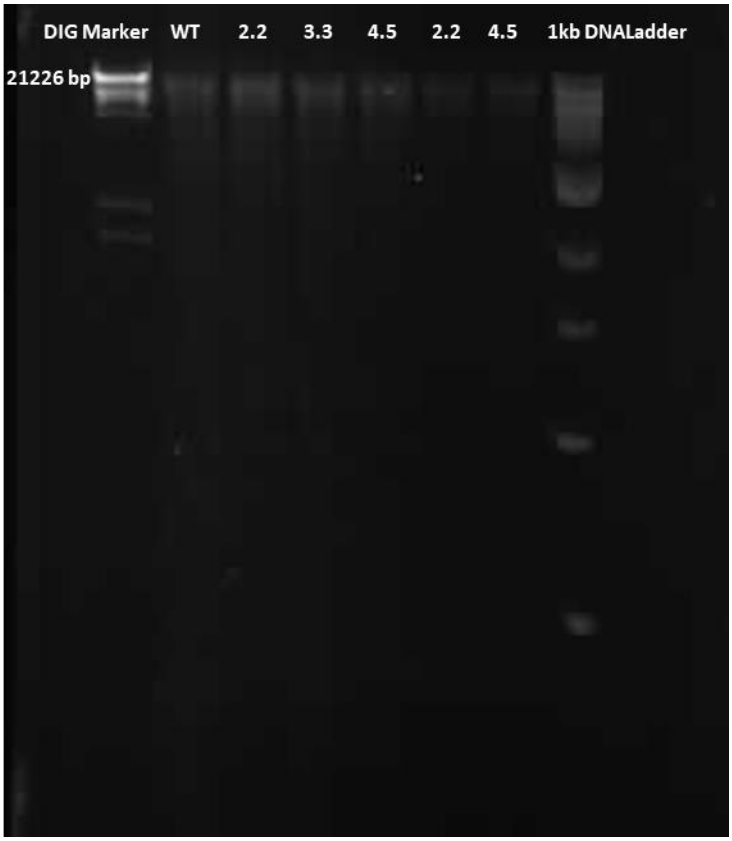
H. Multiple Regression between fruit weight, McrBC-RQ (500-1000 bp upstream TSS of *FW2.2*) and cDNA-RQ

Parameter	estimate	s.e.	t(25)	t pr.
Constant	7.221	0.963	7.5	<.001
RQ	-1.144	0.559	-2.05	0.051
mcrbc	-0.17	0.372	-0.46	0.651

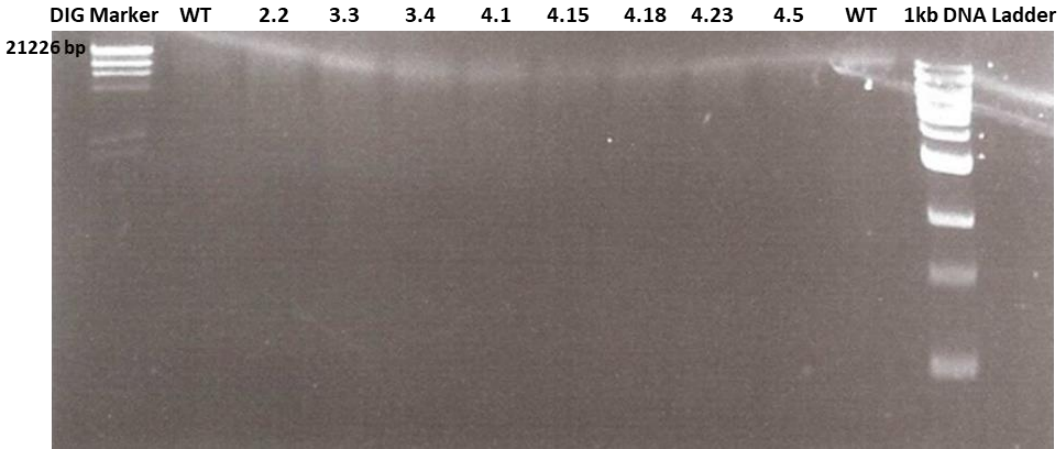
Appendix 23. Gel picture of restriction enzyme digested DNA for Southern blotting.



HpaII digested DNA



EcoRI digested

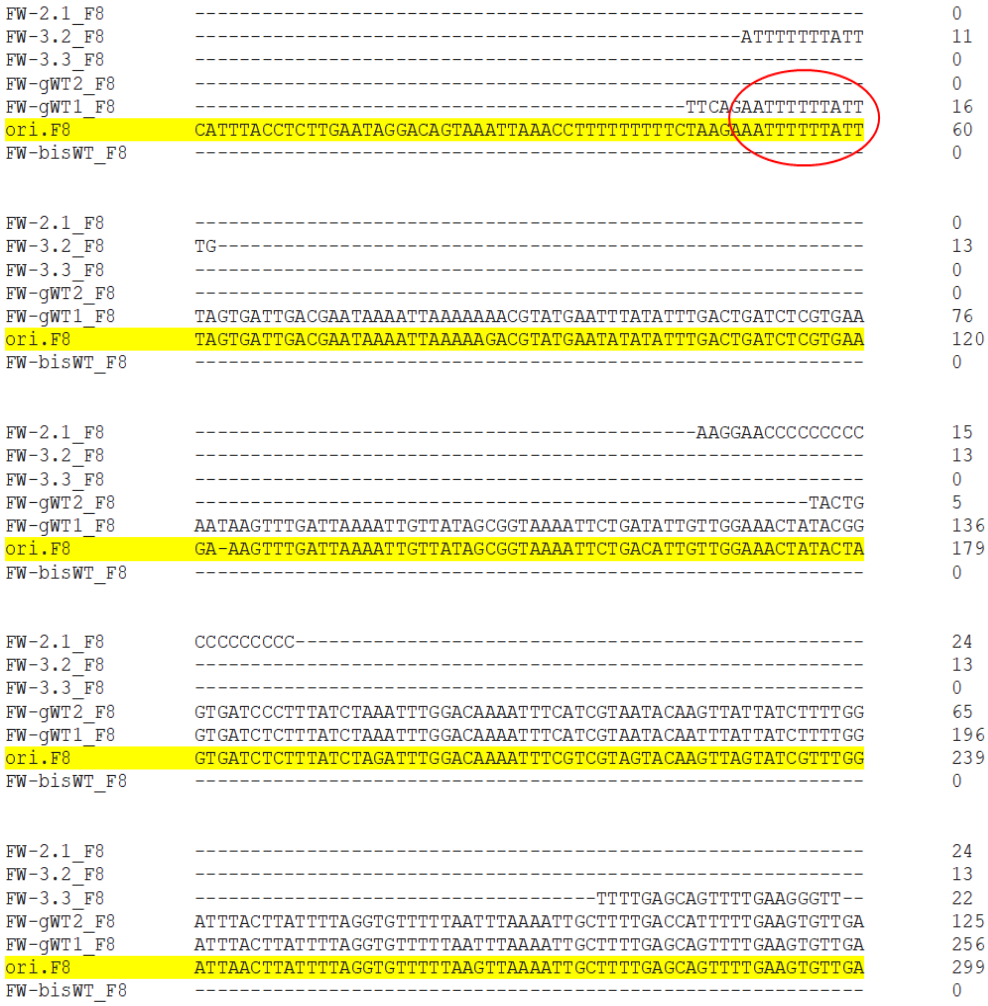


Hind III digested DNA

Appendix 24. Buffers for bisulfite conversion from Epitect® Bisulfite Kit (Qiagen)

Buffer Name	Content	Function
Bisulfite Mix	Sodium bisulfite	Mediate the conversion of unmethylated cytosines.
DNA protect buffer	Tetrahydrofurfuryl alcohol	Prevent fragmentation associated with bisulfite treatment
Buffer BL	Guanidine thiocyanate	Promotes binding of the converted single stranded DNA to the epiTect spin column
Buffer BW	Ethanol 96 %	Wash of membrane-bound DNA
Buffer BD	Sodium hydroxide	Desulfonation
Buffer EB	Information not available	Elution of DNA from the spin-column

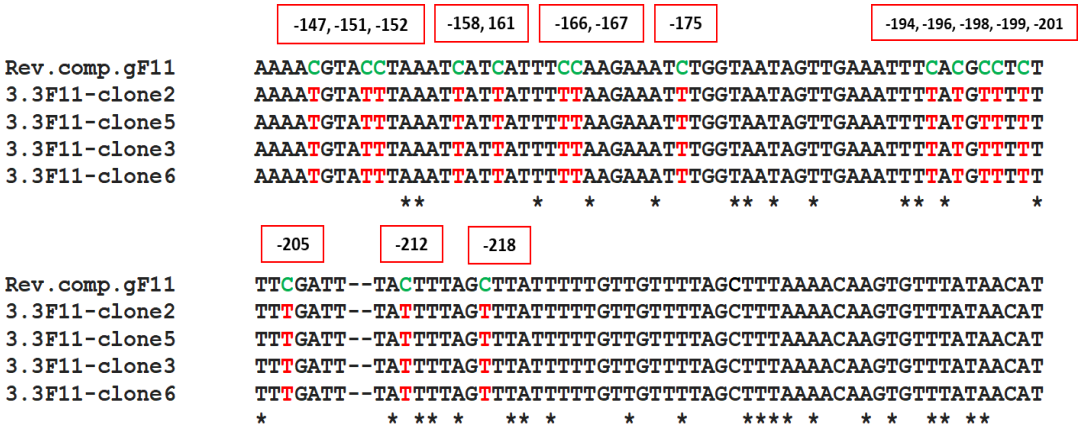
Appendix 25. Sequencing Result using Primer F8. The result was alligned to the original sequence (ori.F8). The other samples were: genomic non-bisulfite treated wild type DNA (gWT), bisulfite treated wild type DNA (bisWT) and epiRILs 2.1, 3.2, and 3.3 from generation F_6. The poly-T region is shown in the red circle. The original sequence is coloured yellow.The left column indicates the sample names, the right column shows the position of the sequence. The rows show the sequence alignment of the bases A (adenine), T (thymine), G (guanine) and C (Cytosine).



Appendix 26. Sequencing Result using Primer F11. The result was alligned to the original sequence (ori.F8). The other samples were : genomic non-bisulfite treated wild type DNA (gWT), bisulfite treated wild type DNA (bisWT) and epiRILs 2.1, 3.2, and 3.3 from generation F_6. The original sequence is coloured yellow and the change in nucleotide is coloured blue. The left column indicate the sample names, the right column shows the position of the sequence. The rows show the sequence alignment of the bases A (adenine), T (thymine), G (guanine) and C (Cytosine). (*) indicate the same base aligned in that position for all samples.

```
3.3_F11          ------------------------------------------------------------ 0
4.2-1_F11        ------------------------------------------------------------ 0
4.2-2_F11        ------------------------------------------------------------ 0
ori.F11          GAAGAAAGTTTGATTAAAATTGTTATAGCGGTAAAATTCTGACATTGTTGGAAACTATAC 60
3.2-1_F11        ------------------------------------------------------------ 0
2.1_F11          ----------------------------------------------TGTTGGAACTATAC 14
bisWT1_F11       ---------------------------------------------TTGTTGGAACTATAC 15
bisWT2_F11       -------------------------------------CTGATATTGTTGGAACTATAC 21
gWT_F11          ----------------------------------ACTCTGATATTGTTGGAACTATAC 24
3.2-2_F11        ------------------------------------------------------------ 0

3.3_F11          -----TTCTTTTATCTAGATTTGAGACAAAATTTCATCGTAATACAAATTAATATCGTTT 55
4.2-1_F11        ------------------ATTTAAACAAAATTTCATCGTAATACAAATTAGTATCGTTT 41
4.2-2_F11        --------------------TAAACAAAATTTCATCGTAATACAAATTAGTATCGTTT 38
ori.F11          TAGTGATCTCTTTATCTAGATTTGGACAAAATTTCGTCGTAGTACAAGTTAGTATCGTTT 120
3.2-1_F11        ------TCTCTTTATCTAGATTTGGACAAAATTTCATCGTAATACAAAT----------- 43
2.1_F11          TGGTGATCTCTTTATCTAGATTTGGACAAAATTTCATCGTAATACAAGTTAGTATCGTTT 74
bisWT1_F11       TGGTGATCTCTTTATCTAGATTTGGACAAAATTTCATCGTAATACAAGTTAGTATCGTTT 75
bisWT2_F11       TGGTGATCTCTTTATCTAGATTTGGACAAAATTTCATCGTAATACAAGTTAGTATCGTTT 81
gWT_F11          TGGTGATCTCTTTATCTAGATTTGGACAAAATTTCATCGTAATACAAGTTAGTATCGTTT 84
3.2-2_F11        ---------CTTTATCTAGATTTGGACAAAATTTCATCGTAATACAAATTA--------- 42
                                        ********** ***** ***** *

3.3_F11          AGATTTACTTATTTTAGGTGTTTTTAAGTTAAAATTGCTTTTGAGCAGTTTTGAAGTGTT 115
4.2-1_F11        GGATTTACTTATTTTAGGTGTTTTTAAGTTAAAATTGCTTTTGAGCAGTTTTGAAGTGGT 101
4.2-2_F11        GGATTTACTTATTTTAGGTGTTTTTAAGTTAAAATTGCTTTTGAGCAGTTTTGAAGTGGT 98
ori.F11          GGATTAACTTATTTTAGGTGTTTTTAAGTTAAAATTGCTTTTGAGCAGTTTTGAAGTGTT 180
3.2-1_F11        ------------------------------------------------------------ 43
2.1_F11          GGATTTACTTATTTTAGGTGTTTTTAAGTTAAAATTGCTTTTGAGCAGTTTTGAAGTGTT 134
bisWT1_F11       GGATTTACTTATTTTAGGTGTTTTTAAGTTAAAATTGCTTTTGAGCAGTTTTGAAGTGTT 135
bisWT2_F11       GGATTTACTTATTTTAGGTGTTTTTAAGTTAAAATTGCTTTTGAGCAGTTTTGAAGTGTT 141
gWT_F11          GGATTTACTTATTTTAGGTGTTTTTAAGTTAAAATTGCTTTTGAGCAGTTTTGAAGTGTT 144
3.2-2_F11        ------------------------------------------------------------ 42
```

Appendix 27. Sequencing result of clones from PCR product of line 3.3, primer F11. The result was alligned to the reverse complement of the target site coding strand. The numbers in red box is the position of hypomethylated cytosines (coloured green). The change in methylation is indicated by the red coloured thymine. The left column indicate the sample names. The rows show the sequence alignment of the bases A (adenine), T (thymine), G (guanine) and C (Cytosine). (*) indicate the same base aligned in that position for all samples.



Appendix 28. Sequencing Result using (A) Primer EZ1-F2 and (B) EZ1-F3. The result was alligned to the original sequence (ori.EZ1-F2 and ori.EZ1-F3). The other samples were genomic non-bisulfite treated wild type DNA (gWT), bisulfite treated wild type DNA (bisWT) and epiRILs 2.1, 3.2, 3.3 and 4.2 from generation F_6. The original sequence is coloured yellow and the change in nucleotide is coloured blue. The left column indicate the sample names, the right column shows the position of the sequence. The rows show the sequence alignment of the bases A (adenine), T (thymine), G (guanine) and C (Cytosine).

(A)

```
gWT_EZ1F2-2     ATGAGGTCTAAGACACTTTAATTGGACGATGTTCACTTTAGATAACTTATATA-------      165
42_EZ1F2-2      ------------TATCTAAAGTGAATAGGGT--------------------------      19
bisWT_EZ1F2     ---------------------------------------------------------      0
gWT_EZ1F2-1     --TCTATATAAGTTATCTAAAGTGAACATCGTCCAATTAAAGTGTCTTAAACCTCATTTG      92
33_EZ1F2-2      ------------------------------------------------AGATTTCATTTG      12
42_EZ1F2-1      ------------------------------------------------AGATTTCATTTG      12
ori.EZ1F2       --TCTATATAAGTTATCTAAAGTGAACATCGTCCAATTAAAGTGTCTTAGACCTCATTTG      158
33_EZ1F2-1      ------------------------------------------------AGATCTCATTTG      12

gWT_EZ1F2-2     ---------GAATCCCGCTATGTCATTTTGACGAC-CTTTTAAATATAAACAAAAA----      211
42_EZ1F2-2      ---------------------------------------------------------      19
bisWT_EZ1F2     ------------------------------ACACGTCTAAATAATTTCATTAAATCT      27
gWT_EZ1F2-1     TTTTTATTAAAATTGCGATCTTTGAATCTGAATACACGTCTAAA----------------      136
33_EZ1F2-2      TTTTTATTAAGATTGCGATCTTTGAATCTGAATACACGTCTAAATAA------------      59
42_EZ1F2-1      TTTTTATTAAGATTGCGATTTTTGAATCTGAATACACGTCTAAATAATTTCATTAAATTT      72
ori.EZ1F2       TTTTTATTAAGATTGCGATCTTTGAATCTGAATACACGTCTAAATAATTTCATTAAATCT      218
33_EZ1F2-1      TTTTTATTAAGATTGCGATCTTTGAATCTGAATACACGTCTAAATAATTTCATTAAATTT      72
```

(B)

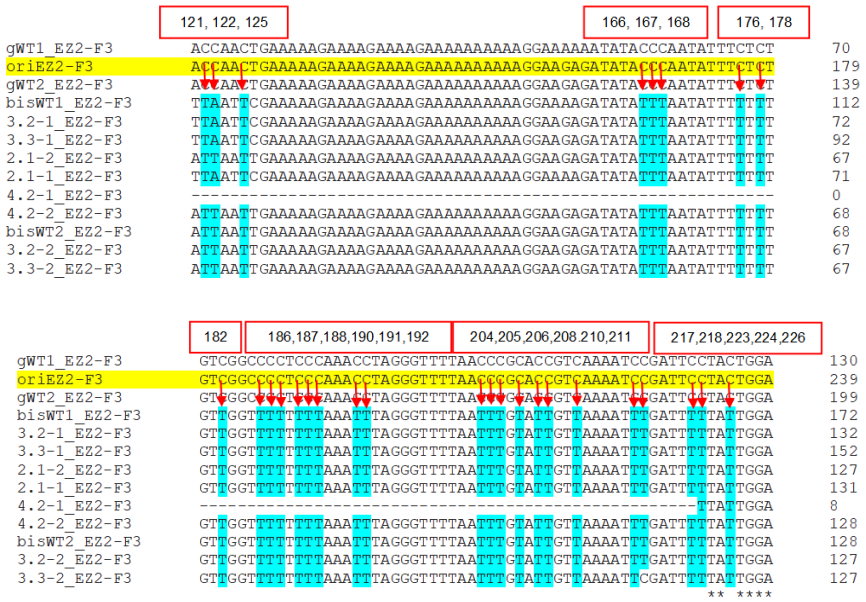
```
2.1_EZ1F3       ---------------------------------------------------------      0
3.3_EZ1F3       TAAATGTTTGTTA-----------------------------------------------      27
3.2_EZ1F3       TAAATGTTTGTTAATACGTATAAATATTATTTATTTATTAGAAAATTTAATGTTTTATGT      68
bisWT_EZ1F3     TAAATGTTTGTTAATACGTATAAATATTATTTATTTATTAGAAAATTTAATGTTTTATGT      127
gWT_EZ1F3       TAAATGTTTGTTAATACGTATAAATATTATTTATTTATTAGAAAATTTAATGTCTCATGT      132
ori.EZ1F3       TAAATGTTTGTTAATACGTATAAATATTATTTATTTATTAGAAAATTTAATGTCTCATGT      180
4.2_EZ1F3       TAAATGTTTGTTAATACGTATAAA------------------------------------      26

2.1_EZ1F3       ---------------------------------------------------------      0
3.3_EZ1F3       ---------------------------------------------------------      27
3.2_EZ1F3       TGATGTTTATTTAATTTTTTAAAAATATTAAATAATAATATTATTTCGAGGAGGTAAAAG      128
bisWT_EZ1F3     TGATGTTTATTTAATTTTTTAAAAATATTAAATAATAATATTATTTCGAGGAGGTAAAAG      187
gWT_EZ1F3       TGATGTTTATCTAATTTTCCAAAAATATCAAATAACAATACTACTTCGAGGAGGCAAAAG      192
ori.EZ1F3       TGATGTTTATCTAATTTTCCAAAAATATCAAATAACAATACTACTTCGAGGAGGCAAAAG      240
4.2_EZ1F3       ---------------------------------------------------------      26
```

Appendix 29. Sequencing result of PCR product and clones using primer EZ1-F2. The result was alligned to the original sequence (ori.F11). The other samples were genomic non-bisulfite treated wild type DNA (gWT), bisulfite treated wild type DNA (bisWT) and lines 3.3 and 4.2 from generation F_6. The original sequence is coloured yellow and the change in nucleotide is coloured blue. The numbers in red box is the position of hypomethylation. The left column indicate the sample names, the right column shows the position of the sequence. The rows show the sequence alignment of the bases A (adenine), T (thymine), G (guanine) and C (Cytosine).

Appendix 30. Sequencing Result using (A) Primer EZ2-F2 and (B) EZ2-F3. The result was alligned to the original sequence (ori.EZ2-F2 and ori.EZ2-F3). The other samples were genomic non-bisulfite treated wild type DNA (gWT), bisulfite treated wild type DNA (bisWT) and epiRILs 2.1, 3.2, 3.3 and 4.2 from generation F_6. The original sequence is coloured yellow and the change in nucleotide is coloured blue. The numbers in red box is the position of hypomethylation.The left column indicate the sample names, the right column shows the position of the sequence. The rows show the sequence alignment of the bases A (adenine), T (thymine), G (guanine) and C (Cytosine).

(A)

(B)



Appendix 31. Sequencing Result using (A) Primer TM8-F1 and (B) TM8-F2. The result was alligned to the original sequence (ori.TM8-F1 and ori.TM8-F2). The other samples were genomic non-bisulfite treated wild type DNA (gWT), bisulfite treated wild type DNA (bisWT) and epiRILs 2.1, 3.2, 3.3 and 4.2 from generation F_6. The original sequence is coloured yellow and the change in nucleotide is coloured blue. The left column indicate the sample names, the right column shows the position of the sequence. The left column indicate the sample names, the right column shows the position of the sequence. The rows show the sequence alignment of the bases A (adenine), T (thymine), G (guanine) and C (Cytosine).

(A)


```
                                  87
gWT_TM8F1       CTT----------------TTGCTAATACCCGGAAGTGGATACCATGATTCTT------      57
3.3-1_TM8F1     TTTAAATTTAAATTCGCGGT-------------------------------------      20
2.1-1_TM8F1     TTTAATTTAAATTCGTC----------------------------------------      20
bisWT_TM8F1     TTTCATTTAAATTCGTGGTTATAATATTTAAATAAATAATATATTAGAATTATTGAGATT    62
oriTM8F1        TTTAAATTTAAATTGTGGTTATAATACTTAAATAAATAATATATTAGAATTATTGAGATT    120
3.2-2_TM8F1     TTTAAATTTAAATTGTGGTTATAATATTTAAATAAATAATATATTAGAATTATTGAGATT    64
2.1-2_TM8F1     GTTTAATTTAAATTGTGGTTATAATATTTAAATAAATAATATATTAGAATTATTGAGATT    65
4.2_TM8F1       GTTTAATTTAAATTGTGGTTATAATATTTAAATAAATAATATATTAGAATTATTGAGATT    66
3.3-2_TM8F1     GTTTAATTTAAATTGTGGTTATAATATTTAAATAAATAATATATTAGAATTATTGAGATT    67
3.2-1_TM8F1     -----------------GTTATAATATTTAAATAAATAATATATTA--------------      29

                          140                                     180
gWT_TM8F1       ------------------------------------------------------------      57
3.3-1_TM8F1     ------------------------------------------------------------      20
2.1-1_TM8F1     ------------------------------------------------------------      20
bisWT_TM8F1     AAGTGAGTTTTGAGTGAATTA-----------------------------------------      83
oriTM8F1        AAGTGAGTTTTGAGTGAATCAAAACGTTATCTGTTTTTAAGTGTATTGGAGTCCAAGTCC    180
3.2-2_TM8F1     AAGTGAGTTTTGAGTGAATTAAAACGTTATCTGTTTTTAAGCGTATTGGAGTTTAAGTTT    124
2.1-2_TM8F1     AAGTGAGTTTTGAGTGAATTAAAACGTTATCTGTTTTTAAGCGTATTGGAGTTCAAGTTC    125
4.2_TM8F1       AAGTGAGTTTTGAGTGAATTAAAACGTTATCTGTTTTTAAGCGTATTGGAGTTTAAGTTT    126
3.3-2_TM8F1     AAGTGAGTTTTGAGTGAATTAAAACGTTATCTGTTTTTAAGCGCATTGGAGTCCAAGTC-    126
3.2-1_TM8F1     ------------------------------------------------------------      29
```

(B)

```
3.2-2_TM8-F2        ------------------------------------------------------------      0
3.2-1_TM8-F2        ------------------------------------------------------------      0
bisWT-2_TM8-F2      ------------------------------------------------------------      0
3.3-2_TM8-F2        ------------------------------------------------------------      0
4.2-1_TM8-F2        ---------------------------AAGTATATCTCTAATTTCAACTCAATCCATCTA     33
4.2-2_TM8-F2        ---------------AGTTATTTAACATATTATATCTCTAATTTCAACTCAATCCATCTA     45
bisWT-1_TM8-F2      ------------------------------------CTATTTCAACTCAATCCATCTA     22
3.3-1_TM8-F2        ----------------------------TAAGATCTCTAATTTCAACTCAATCCATCTA     31
2.1-2_TM8-F2        ACTCACCCAATTTTTATTTATTTAACATATTATATCTCTAATTTCAACTCAATCCATCTA     64
oriTM8-F2           CTCAACCCAATTTTTATTTATTTAACATATTATATCTCTAATTTCAACTCAATCCATCTA    120
2.1-1_TM8-F2        CTCAACCCAATTTTTATTTATTTAACATATTATATCTCTAATTTCAACTCAATCCATCTA     69


3.2-2_TM8-F2        ------------------------------------------------------------      0
3.2-1_TM8-F2        ------------------------------------------------------------      0
bisWT-2_TM8-F2      ------------------------------------------------------------      0
3.3-2_TM8-F2        ------------------------------------------------------------      0
4.2-1_TM8-F2        TTTAACATCGCTATTTACAAATAATCCGACCTAAAACATCATTTTTTTTTATAATTTTAA     93
4.2-2_TM8-F2        TTTAACATCGCTATTTACAAATAATCCGACCTAAAACATCATTTTTTTTTATAATTTTAA    105
bisWT-1_TM8-F2      TTTAACATCGCTATTTACAAATAATCCGACCTAAAACATCATTTTTTTTTATAATTTTAA     82
3.3-1_TM8-F2        TTTAACATCGCTATTTACAAATAATCCGACCTAAAACATCATTTTTTTTTATAATTTTAA     91
2.1-2_TM8-F2        TTTAACATCGCTATTTACAAATAATCCGACCTAAAACATCATTTTTTTTTATGATTTTAA    124
oriTM8-F2           TTTAACATCGCTATTTACAAATAATCCGACCTAAAACATCATTTTTTTTTATGATTTTAA    180
2.1-1_TM8-F2        TTTAACATCGCTATTTACAAATAATCCGACCTAAAACATCATTTTTTTTTATGATTTTAA     129
```