DISRUPTION OF DNA METHYLATION INDUCES GENOME-SPECIFIC CHANGES IN GENE EXPRESSION IN *Arabidopsis* ALLOTETRAPLOIDS

A Thesis

by

MENG CHEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2005

Major Subject: Genetics

DISRUPTION OF DNA METHYLATION INDUCES GENOME-SPECIFIC CHANGES IN GENE EXPRESSION IN *Arabidopsis* ALLOTETRAPLOIDS

A Thesis

by

MENG CHEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Co-Chairs of Committee,	Jeffery Chen
	David Stelly
Committee Member,	Dorothy Shippen
Head of Department,	James Wild

December 2005

Major Subject: Genetics

ABSTRACT

Disruption of DNA Methylation Induces Genome-Specific Changes in Gene Expression in *Arabidopsis* Allotetraploids.

(December 2005)

Meng Chen, B.S., Beijing University, PRC

Co-Chairs of Advisory Committee: Dr. Z. Jeffrey Chen
Dr. David M Stelly

Allopolyploids are formed by the combination of evolutionarily-diverged genomes, the union of which leads to dynamic changes in gene expression and genome organization. Expression patterns of orthologous genes are rapidly and stochastically established in newly created allotetraploids, where gene silencing is maintained by microRNAs, DNA methylation, and other chromatin modifications. Among them, DNA methylation has been known as an important mechanism of epigenetic regulation of gene expression and chromatin structure. However, it is unclear how DNA methylation affects genome-wide expression of homoeologous genes in the natural polyploid *Arabidopsis suecica* that contains genome of both *A. thaliana* and *A. arenosa*.

To understand the role that DNA methylation plays in the polyploidization process, a comparative analysis was performed comparing up- or down-regulated genes in *met1*-RNAi *A. suecica* lines with the non-additively expressed genes in the synthetic allotetraploids, i.e., different from the mid-parent value. The previous studies indicated that decreased DNA methylation in *A. suecica* induces *A. arenosa*-specific demethyla-

tion in centromere regions and differentially alters expression of >200 genes encoding many transposons, unknown proteins and some other functional proteins that are located along chromosomes, whereas >1,300 non-additively expressed genes in the synthetic allotetraploids are distributed randomly along the chromosomes and encode various proteins in metabolism, energy, cellular biogenesis, cell defense and aging, and hormonal regulation.

The origins of the progenitors of the genes whose expressions are altered in both *met1*-RNAi *A. suecica* and resynthesized allotetraploid were analyzed with single strand conformation polymorphism (SSCP) analysis. Reactivated genes in *met1*-RNAi *A. suecica* lines were predominately derived from the *A. thaliana* genome in euchromatic regions, whereas the suppressed genes were mainly derived from the *A. Arenosa* genome, indicating that changes in DNA methylation are genome-sensitive. The data suggest that allotetraploids incidentally display chromosome-specific changes and genome-dependent regulation of homoeologous genes in response to DNA methylation perturbations.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my primary advisor, Dr. Jeffrey Chen, for his guidance, instruction and encouragement. I am thankful for being given the wonderful opportunity to work in his lab.

I would like to extend my sincere gratitude to co-chair of my committee, Dr. David Stelly, and committee member, Dr. Dorothy Shippen, for their valuable advice and great support.

I would like to thank Dr. Jianlin Wang for providing *met1*-RNAi transgenic plants and the microarray data, as well as his generous help and great patience all the time. I would like to thank Dr. Zhongfu Ni for his great help with SSCP.

I thank Dr. Lu Tian for her honest friendship and constructive advice throughout my study and work in the lab.

I also want to thank our computerists, Joe Wang and Edward Wei, for helping me finish all the computer analysis for microarray data.

Finally, I would like to express my thanks to other Chen lab members, Dr. Hyeon-Se Lee, Dr. Sukhwan Yang, Paulus Fong, Sheetal Rao, Jenny Lee, Grace Kim, Meghan Roche, Misook Ha, and Leticia Nogueria, for their friendship and help in the research, and our undergraduate students, Jose Trejo, Ancy Alexander, Lauren Williamson, and Lauren Wells, for their hard work to provide a supportive working circumstance in the lab.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

POLYPLOIDIZATION AND POLYPLOIDY

Polyploidization is an evolutionary process that doubles the chromosome number and DNA content in the nucleus. It usually occurs via modified sexual reproduction. "autopolyploidization" and "allopolyploidization" allude to situations where the contributed genomes of the new polyploid are meiotically recombinant or nonrecombinant, respectively, where the meiotic behavior reflects genome "divergence". Polyploidy is a prominent phenomenon among plants. Around 95% of angiosperms and 70% of ferns went through polyploidization process at lease once during their evolution (Bennett and Leitch, 1997; Comai et al., 2000a). Moreover, most of cultivated crops, including wheat, cotton, sugarcane, and canola, are polyploids. Polyploidy is generally rare in animals, except some cases of ancient polyploidization in insects, fish, amphibians, and reptiles (Otto and Whitton, 2000), and thelytokous species. However, it is common in the plant kingdom and is associated with many advantages and benefits. It's believed that the success of polyploid plants may be due to the increased adaptability to the environment. One possibility to explain the success of polyploidy is "forced heterozygosity". In allopolyploids, preferential pairing of homologous chromosomes instead homoeologous

This thesis follows the style of Plant Cell.

chromosomes ensures the transmission of entire sets of ancestral genetic material (genomes) each generation, thereby allowing for fertility and the "permanent hybrid vigor" of successful sexual allopolyploids. The increased gene dosage of polyploids provides a larger capacity to mask and buffer effects of deleterious single-locus mutations. Nevertheless, variations in expression levels may be increased due to changes in gene expression associated with increased numbers of orthologous alleles, which gives polyploids better adaptability to different circumstances, compared to the progenitors (Soltis and Soltis, 2000). Thus, detailed analysis of polyploids could reveal general molecular and cellular basis of genome response to increased genome size, an understanding of which would help us understand the biological ramifications of polyploidization, and its impact on crop adaptability and yield.

Although polyploidy is a popular topic in plant molecular genetic research because of its prominence in plant evolution, the ramifications of polyploidization on gene regulation and expression remains unclear. Several scenarios have been proposed (Osborn et al., 2003a, Figure 1.1). The first hypothesis is that increased copy numbers of dosage-related genes may introduce more variations in the possibility of gene expression and lead to novel gene expression in polyploid progenies in comparison with progenitors. Moreover, since the expression of most of genes depends on complex regulatory networks, the regulatory factors from both progenitors may interact with each other in a more complicated way in polyploid progeny, giving another possibility to vary gene expression in polyploids. Thirdly, genetic and epigenetic changes can happen rapidly in resynthesized polyploid and vary gene expression sporadically and genome-widely. It

has been reported that polyploidization brings about both rapid and stochastic changes to plant genomes (Wang et al, 2004; Feldman and Levy, 2005) and causes extensive novel gene expression (Osborn et al, 2003b; Wang et al, 2005). The novel alterations may be due to genomic changes, such as elimination or rearrangement of certain DNA fragments in polyploids (Liu et al., 1998; Shaked et al., 2001; Kashkush et al., 2002), or epigenetic changes, such as histone modification and DNA methylation (Song et al., 1995; Chen and Pikaard, 1997; Liu et al., 1998; Pikaard, 2000a; Pikaard, 2000b).

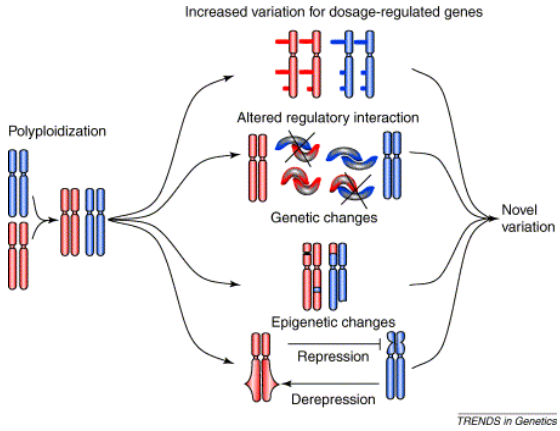


Figure 1.1 Novel gene expression in polyploid.

Diagram adopted from the paper from Osborn et al. (2003a).

Epigenetic phenomena are alternative states of gene activities which are heritable but can not be explained by DNA sequence variations. One of the best–studied phenomena of epigenetic regulation of gene expression in polyploids is nuclear dominance, referring to the preferential expression of rDNA from one progenitor in allopolyploid formed through interspecific hybridization (Chen and Pikaard, 1997). In *Arabidopsis* and *Brassica* allotetraploids, silencing of rRNA genes from one progenitor is associated with hyper-methylation of DNA and hypo-acetylation of histone (Chen and Pikaard, 1997). There are reports on transcriptional reactivations of transposable elements in newly formed *A. suecica* and wheat polyploids (Comai et al., 2000b; Madlung et al, 2002; Kashkush et al, 2003), which are also associated with DNA methylation level alterations in polyploids.

Arabidopsis polyploids and their relatives serve as a powerful model in analysis of regulatory mechanisms that govern gene expression in polyploids. The subspecies *Arabidopsis thaliana* and *A. arenosa* diverged from each other ~5.8 Mya and share more than 95% similarity in genomic sequence (Koch et al., 2001). Interspecific hybridization can occur between the two sub-species and give rise to an allotetraploid hybrid offspring, *Arabidopsis suecica*. The hybridization that has taken place in nature can be approximated in the laboratory by synthesizing autotetraploid *A. thaliana* from diploid *A. thaliana* by colchicine treatment and making crosses to the natural autotetraploid *A. arenosa* (Comai et al., 2000b). It provides an excellent system for polyploidy study given the fully sequenced genome information and relatively good understanding of genome structure and gene behavior in diploid *A. thaliana*.

DNA METHYLATION AND METHYLTRANSFERASES IN *Arabidopsis thaliana*

Cytosine methylation plays an important role in epigenetic gene regulation in eukaryotic genomes. Although DNA methylation is considered to be stable and conserved, changes in the status of cytosine methylation can be observed in response to external stress, including chilling (Steward et al., 2002), pathogen invasion (Wada et al., 2004), different development stages (Ko et al., 2005; Ruiz-Garcia et al., 2005), and different tissues or cell types (Imamura et al, 2001; Shiota et al. 2002; Song et al, 2004). DNA methylation is considered to be an important evolutionary mark and defensive machinery, as well as a regulatory mechanism to establish certain gene expression pattern to ensure normal growth and development in different organisms. DNA methylation is strongly associated with gene silencing and heterochromatin formation. In plants, animals, and fungi, symmetrical cytosine methylation at CG sites can be widely found (Finnegan and Kovac, 2000; Bird, 2002). In higher plants, cytosine methylation occurs at CNG sites and in asymmetrical contexts such as CN (Chen and Li, 2004). However, some organisms, such as *C. elegans* and *S. cerevisiae,* lack detectable methylation signs in their genomes (Colot and Rossignol, 1999). In addition, the biological significance of DNA methylation and methyltransferase varies from organism to organism. In mammals, DNA methylation involves crucial processes, for instance, imprinting and X-chromosome inactivation. In mouse, loss of methylation causes severely aberrant embryonic development or lethality (Li et al., 1992). In contrast to mammals, plants have a higher tolerance for DNA methylation defects, although it also leads to severe pleiotropic phenotype and ste-

rility in some occasions.

At least ten genes that encode DNA methylases have been identified in *Arabidopsis thaliana*. Based on their various functions and sequence homologies to mammalian DNA methyltransferases, those genes can be grouped into three (Tariq and Paszkowski, 2004) or four (Chan et al., 2005) major families. Extensive study has been done utilizing *Arabidopsis* mutants that are defective in those methyltransferase activities. The first class, represented by *DOMAINS REARRANGED METHYLTRANSFERASES* (*DRM*), which are the homologous genes of mammalian *DNA* (cytosine-5-)-*METHYLTRANSFERASE* 3 (*DNMT3*). *DRM1* and *DRM2* take charge of *de novo* CG and non-CG methylation (Cao et al., 2000; Cao and Jacobsen, 2002). *Arabidopsis drm1 drm2* double mutants were shown not to be able to establish *de novo* methylation at *FWA* and *SUPERMAN* loci (Cao and Jacobsen, 2002). Recently evidence shows that *drm* mutations block *de novo* methylation in RNA-directed DNA methylation (RdDM) pathway (Cao et al., 2003). The *DNMT2*-class of methyltransferase genes is conserved among eukaryotes, but the function for Arabidopsis *DNMT2* equivalent gene is not yet clear (Finnegan and Kovac, 2000), so some people group it to *DNMT3*-class of methyltransferases genes (Tariq and Paszkowski, 2004). The plant-specific group of methyltransferase genes, *CHROMOMETHYLASE3* (*CMT3*) share no similarities with animal methyltransferase genes (Lindroth et al., 2001). *CMT3* is responsible for CNG methylation which has only been found existing in plants. Mutations in *CMT3* result in loss of CNG methylation at centromeric repeats and many transposons (Bartee et al., 2001; Lindroth et al., 2001; Tompa et al., 2002). The last group comprises members of plant ME-

THYLTRANSFERASEs (METs), which are homologues to the DNA METHYL-TRANSFERASEs (DNMTs) in animals. Among them, *MET1* is the best characterized gene. *MET1* is widely considered to be a maintenance gene for methylation in symmetrical CG dinucleotides (Finnegan and Kovac, 2000), but there is evidence showing that *MET1* also contributes to *de novo* CG methylation in the presence of RNA signal (Aufsatz et al., 2004) and the maintenance of it after the RNA signal is removed (Jones et al., 2001). Different loss-of-function *Arabidopsis* mutants (*met1-1~4*) have been identified that have severe reduction of CG methylation, as well as moderate loss of non-CG methylation in genome (Kankel et al., 2003; Saze, et al., 2003). In the plants which have single mutation of either *met1* or *cmt3*, transposons show weak reactivation of transpositional activity, while in *met1 cmt3* double mutants, high frequencies of transposition can be observed (Kato et al., 2003), indicating *MET1* and *CMT3* are partially redundant in gene function. In *met1* mutants, transcriptional gene silencing (TGS) cannot be established by siRNA signal through RdDM pathway (Jones et al., 2001; Aufsatz et al., 2004).

HISTONE MODIFICATIONS

Besides DNA methylation, another major chromatin structure remodeling effort is histone modification, especially acetylation and methylation. In the primary chromatin fiber, DNA is packaged into nucleosomes with histone cores, which may limit the access of transcription factors to the regulatory DNA elements. Core histones are subject to various modifications, including acetylation, methylation, phosphorylation, and ubiquiti-

nation (Spencer and Davie, 1999). Those modifications are known as "histone code" (Jenuwein and Allis, 2001). Modified histone tails provide binding sites for chromatin-associated protein factors, which in turn causes alterations in chromatin structure and therefore leads to downstream transcriptional changes (Strahl and Allis, 2000; Jenuwein and Allis, 2001). Histone cores can be acetylated or deacetylated by HISTONE ACETYLTRANSFERASEs (HATs) and HISTONE DEACETYLASEs (HDs or HDACs) in eukaryotes (Allfrey et al., 1964). Unlike DNA methylation, histone modification is more dynamic and reversible throughout the development. Acetylation of specific lysine residues in histone cores, for instance, Lysine 9 residue of histone 3 (H3K9), is correlated with gene activation, which relaxes the association of positively charged histone cores and the negatively charged DNA to allow the transcription factors to access the gene and facilitate transcription, while deacetylation of certain lysine residues in histone cores is associated with transcriptional repression (Hassig and Schreiber, 1997; Struhl, 1998; Workman and Kington, 1998).

RNA INTERFERENCE

RNA-mediated gene silencing or RNA interference (RNAi) is present in both multi- and unicellular eukaryotes and performs a very important role in gene regulation. It refers to the phenomenon in which short RNA molecules trigger the repression of homologous sequences. There are a group of small RNAs known as small interfering RNAs (siRNAs) which are 21-26nt in length (Hamilton and Baulcombe, 1999; Elbashir et al., 2001a).

siRNA was firstly characterized in *C. elegance* (Lee and Ambros, 2001). In *Drosophila*, siRNAs are processed from double-stranded RNAs (dsRNAs) (Zamore et al., 2000) by the action of an RNaseIII-like enzyme named Dicer (Bernstein *et al*., 2001). dsRNAs can be derived from replication intermediates of plant RNA viruses, transgenic inverted repeats or products of RNA-dependent RNA polymerases (RdRps). siRNAs mediate gene silencing to defend the plant genome from viral attack and transposon activities. There are two sub-classes of siRNAs, short (21–22nt) and long (24–26nt), divided based on their size ranges and functions (Hamilton et al., 2002). The short siRNAs correlate with mRNA degradation while the long siRNAs serve as the triggers for RNA-directed DNA methylation (RdDM).

Another group of small RNA are called microRNAs (miRNAs) which are generated by Dicer-like 1 (DCL1) protein from endogenous stem-looped RNA precursors that are transcribed from miRNA genes (Reinhart et al., 2000; Grishok et al., 2001; Hutvagner et al, 2001). The major function of miRNAs is to regulate endogenous gene expression. siRNAs and miRNAs are incorporated into RNA-induced silencing complex (RISC) complexes containing a particular subset of Argonaute proteins and direct a sequence-specific gene repression. In addition to complementary-dependent mRNA degradation, which is known as a post-transcriptional gene silencing (PTGS) mechanism, RNAi in plants acts at various other levels, including RNA-directed DNA methylation (RdDM), known as transcriptional gene silencing (TGS) mechanism (Wassenegger and Pelissier, 1998; Mette *et al*., 2000; Jones *et al*., 2001), pre-mRNA processing (Mishra and Handa, 1998) and translation level gene silencing (VanHoudt *et al*., 1997).

CROSS-TALK BETWEEN DNA METHYLATION, HISTONE MODIFICATIONS AND RNA-INTERFERENCE

In addition to DNA methyltransferases, *Arabidopsis* also encodes other genes involved in DNA methylation and chromatin remodeling. *DECREASE in DNA METHYLATION1* (*DDM1*) encodes a chromatin remodeling factor of the SWI2/SNF2 family. As *MET1, DDM1* is also a DNA methylation maintenance gene, and is required for RNA directed DNA methylation (RdDM) (Morel et al., 2000). The *ddm1* mutants cause global reduction of cytosine methylation at both CG and CNG sites, as well as activation of *Mu*-like transposable element (MULE) (Singer et al., 2001), and loss of methylation at the lysine-9 residue of histone 3 (H3K9) in heterochromatic regions (Gendrel et al., 2002). In *Arabidopsis*, DNA methylation and histone modification, especially histone methylation, appear to relate closely to each other and to jointly determine chromatin structure and consequent gene expression. H3K9 methylation in *Arabidopsis* is also dependent on a histone methyltransferase gene named *KRYPTONITE* (*KYP),* which is similar to yeast *SU (VAR) 3-9* (Jackson et al., 2002). Moreover, *kyp* mutants are associated with DNA hypo-methylation and transpositional activation of retrotransposons *Ta2* and *Ta3* (Johnson et al., 2002). Moreover, Dnmt1 is also associated with histone deacetylase activity *in vivo* (Fuks et al., 2000). Its transcriptional repressive domain functions by recruiting histone deacetylase activity. Densely methylated DNA associates with transcriptionally repressive chromatin characterized by the presence of deacetylated histones (Antequera, et al., 1989; Eden, et al., 1998). The methyl-CpG-binding protein MeCp2 appears to re-

side in a complex with histone deacetylase activity present (Nan, et al., 1998; Jones, et al., 1998). MeCP2 can direct formation of transcriptionally repressive chromatin on methylated promoter templates *in vitro*, and this process can be reversed by trichostatin A (TSA), which is a chemical inhibitor of histone deacetylase (Nan, et al., 1998; Jones, et al., 1998; Yoshida, et al., 1995). This implies that DNA methylation and histone modification are coupled together in function.

When double-stranded RNA (dsRNA) targets a promoter region, RNAi machinery will direct a target-specific cytosine methylation to the region which is complementary to the dsRNA trigger (Sijen et al., 2001). A model is suggested in which RNAi first directs *de novo* DNA methylation through DRM-class methyltransferase, possibly involving MET1 and DDM1. Cytosine methylation (5mC) will be recognized by METHYL CYTOSINE BINDING DOMAIN (MBD) proteins to attract HISTONE DEACETYLASE (HDAC), such as HDAC6, and HISTONE METHYLTRANSFERASE (HMT), such as KYP, for subsequent histone deacetylation and methylation. After that, CMT3 or DRMs might be recruited for further non-CG methylation and form the heterochromatin region (Tariq and Paszkowski, 2004, Figure 1.2).

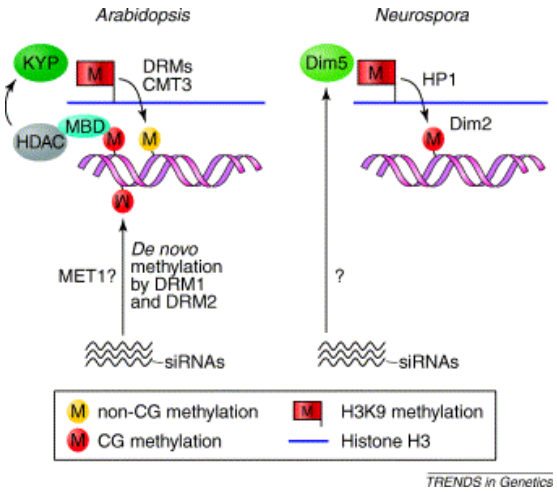


Figure 1.2 Cross talk between RNAi, DNA methylation and histone modification.

Diagram adapted from the paper from Tariq and Paszkowski (2004).

All of the above suggest interactions between all the pathways and regulatory machineries for chromatin structure remodeling and gene regulation. The ultimate goal for our research is to understand genome evolution and mechanisms that control gene expression upon polyploidization. In this study, we will focus on the impact of DNA methylation on gene expression regulation in *A. suecica* natural lines and resynthesized allopolyploid lines. By examining how gene expression pattern will be altered in the hypo-methylation background, and how they are related to genome polyploidization, we will be able to illustrate how DNA methylation affects gene expression, especially the expression of homoeologous genes in the allotetraploid, to get a better understanding of the role that epigenetic modification is playing in polyploids.

CHAPTER II

CHARACTERIZATION OF *MET1*-RNAi *A. suecica* LINES

INTRODUCTION

As one of the major epigenetic regulatory mechanisms, DNA methylation plays a key role in formation of chromatin structures and control of gene activities. An important experimental strategy for studying the impact of DNA methylation on polyploidization has been to reduce DNA methylation using chemical or genetic methods that result in genome-wide effects. It was reported that the resynthesized *A. suecica* lines are susceptible to the treatment of 5'-aza 2'-deoxycytosine (aza-dC; Madlung et al. 2002) which is an inhibitor of DNA methyltransferases (Haaf, 1995). In aza-dC-treated *Arabidopsis* and *Brassic*a, silenced rRNA genes as well as protein-coding genes were reactived (Chen and Pikaard, 1997; Lee and Chen, 2001; Madlung et al. 2002), which implicated DNA methylation as a significant mechanism for maintenance of gene silencing in polyploid plants.

In *Arabidopsis*, centromere satellite repeats represent another major target for DNA methylation, H3mK9 methylation and heterochromatin formation (Kankel et al., 2003; Tariq et al., 2003; Soppe et al., 2005). In the centromeric regions, the H3 normally found in the chromatin histone core is substituted with a unique centromere-

specific histone, CenH3, which is essential for the normal function of centromeres (Zhong et al., 2002; Nagaki et al., 2003). In *Arabidopsis*, siRNAs corresponding to centromere sequences have been detected (Xie et al, 2004), implying that a subset of the centromere repeats might get transcribed to provide siRNA triggers for RdDM. Demethylation of centromere repeats has been reported in a *cmt3* mutant background (Cao and Jacobsen**,** 2002). In addition, centromere satellite repeats in a *kyp* mutants are only partially demethylated at CNG sites in comparison with those in a *cmt3* mutant background (Jackson et al., 2002; Malagnac et al., 2002), suggesting that only part of the repeats are transcriptionally silent because of methylation of H3K9 residues. Although normal condensed heterochromatin structure is important for proper centromere behavior such as mitotic disjunction and protection from meiotic recombination, plants show high tolerance of defects in methylation even though centromeric repeats can undergo decondense extensively (Soppe et al., 2002; Probst et al, 2003).

Transposable elements are considered as a threat to the integrity of the host plant genome because their ability to translocate and insert can disturb or disrupt normal gene behavior (McClintock, 1965). To defense against transposons mediated disturbance of gene behavior, heterochromatin formation has evolutionarily become targeted at transposons as a means to suppress their transcription. The silencing mechanisms include DNA methylation, histone H3K9 deacetylation and methylation, and RNA interference (RNAi) (Rea et al. 2000; Gendrel et al. 2002; Johnson et al. 2002; Schotta et al. 2002). These chromatin modifications are interrelated to each other as described above (Martienssen and Colot 2001; Selker 2002; Sleutels and Barlow 2002). The accumulation of *CACTA* transcripts

is induced in *met1-1* mutation in the Columbia (Col) background, as well as in transgenic Col plants that express antisense *MET1* gene, but no transposition is detected in either of the mutants, despite the transcript accumulation (Lippman et al, 2003).

To avoid the toxic effects of aza-dC on cellular growth and plant development (Haaf, 1995), and to study the causal relationship between DNA methylation and gene silencing, Wang et al. (2004) generated transgenic lines of natural *A. suecica* that down-regulate *MET1* and *DDM1* by RNAi. Transgenic *met1*-RNAi *A. suecica* plants with only <5% of expression of endogenous *MET1* gene, show severe phenotypic abnormalities. In addition, microarray experiments that compare gene expression profiles of *met1*-RNAi and wild-type *A. suecica* have shown more than 200 genes differential expression in the hypo-methylation background (Wang et al., data unpublished). Further characterization of these effects in *met1*-RNAi *A. suecica* lines has been the focus of this work, the goal of which has been to enhance our understanding of the altered chromatin structures, centromeres, and behavior of transposons.

RESULTS

Phenotypic Abnormalities of *met1*-RNAi *A. suecica* Plants

Two met1-RNAi transgenic lines were obtained from the *Agrobacterium*-mediated transformation process, *met1-1* and *met1-12* (Wang et al., 2004). Abnormalities were severe in both transgenic lines, but more so in *met1-1*. Both RNAi knock-down lines were

dwarf, slow to grow and develop, and less viable. Flowering time varied among the transgenic individuals, but was delayed by 2-3 months compared to the flowering time of 3-3.5 months for wild type (WT) plants (Figure 2.1). Seed production was dramatically decreased to less than 5% of that of WT. The mutants also showed morphological alterations in rosette leaves and inflorescence organs, including narrower and smaller leaves, lack of chlorophyll in vegetative organs and smaller flowers. In addition, the severity increased across selfing generations. T3 *met1*-1 selfing plants were barely viable and produced no seeds, indicating that inability to establish DNA methylation may disturb meiosis or perturb embryo development in plants, as it does in animals (Li et al., 1992).

Figure 2.1 Developmental abnormalities in *met1*-RNAi *A. suecica* lines.

Photo was taken for the plants grown side by side. *met1*-1 plants are in their 3rd generation of selfing. Morphology changes can be observed in *met1*-RNAi *A. suecica* plants. As shown, the *met1*-1 line displays a more severe phenotype than *met1*-12 line.

Disturbance of *MET1* Gene Causes Changes in DNA Methylation Status and Up-regulation and Transcriptional Reactivation of Transposable Elements

Phenotypic abnormalities were generally more severe among of *met1*-1 plants than *met1-12* (Figure 2.1), so only the former were used in the microarray-based comparisons used to assess RNAi knock-down effects on gene expression (Wang et al., data unpublished). Around 200 genes were identified to have expression altered by more than 2-fold in *met1* lines compared to WT, ~50% genes up-regulated and ~50% down-regulated *met1*. RT-PCR was also conducted for some genes to verify the microarray result (Wang et al., 2004). Many transposons were found to be up-regulated in transcriptional level as well (Wang et al., 2004), which correlated with the expectation that the suppression of transposable elements are kept by DNA methylation (Martienssen and Colot, 2001; Soppe et al, 2000; Steimer, 2000; Johnson, 2002). To directly test if loss of DNA methylation is sufficient for mobilization of transposons, comparisons were made of DNA restricted with methylation-sensitive versus methylation-insensitive enzymes by Southern blotting and then probing for selected transposon genes which showed high level of up-regulation in *met1* line based on microarray data. EcoRI (5'-G^AATTC-3') and NdeI (5'-CA^TATG-3') are two restriction enzymes that are insensitive to cytosine methylation status at their digestion sites (Kessler and Manta, 1990), thus they were used to completely fragmentize plant genome to examine the transpositions of transposons. HpaII and MspI are isoschizomers (5'-C^CGG-3') which recognize the same DNA sequence as digestion site, but they are different in terms of sensitivity to cytosine methylation. MspI

will not cut if the external cytosine is methylated, while HpaII will not cut if the internal or external cytosine is methylated (Mann and Smith, 1977), thus they can be used to distinguish the methylated and unmethylated cytosine residues at CG dinuleotide sites. A repeated gene *SERINE CARBOXYPEPTIDASE* (*SCP*, At5g36180) and two CACTA-like transposase family (En/Spm) genes (At4g08010 and At1g44070) were also subjected to Southern analysis, as they showed huge transcriptional up-regulation by 30-, 12-, and 9-fold changes respectively in *met1*-1 based on previous microarray data (Wang et al., data unpublished). A general decrease of CG methylation was detected in Southern for all the three genes (Figure 2.2C, Figures 2.3C and F), which indicated that the suppression of expression of those genes was directly associated with CG methylation. In the hypomethylation genotype, the silencing was relieved and genes expression was increased because DNA methylation was reduced. Reactivation and transposition were detected in both *met1*-1 line and *met1*-12 line for the two selected transposons (Figures 2.3B, E and F), suggesting that in this case transcriptional up-regulation reactivated transposition.

Up-regulation of gene expression is known to common occur upon demethylation (Cao et al., 2000; Cao and Jacobsen, 2002). To examine the methylation status changes for the genes that were down-regulated in *met1* lines, At5g25610 was subjected to Southern analysis, which was characterized as a dehydration-responsive protein *RD22* precursor and was down-regulated by 2.5-fold change in *met1*-1 line based on microarray data. An increase of CG methylation was detected in the 5' prime region (shown as asterisk in Figure 2.4A), indicated by the appearance of larger fragments (about 1.2kb, Figure 2.4B). This might be associated with the down-regulation of this gene in an over-

all demethylation background, which correlated with the observation of dense methylation around the start of transcript and within the coding region of *SUPERMAN* (*SUP*) gene in *A. thaliana* diploid that expressed antisense *MET1* gene (Kishimoto et al., 2001). We observed methylation level at a specific site (shown as asterisk). In natural *A. suecica* 9502 and 9502-*met1*, this site is considerably CG methylated. However, in *A. thaliana* diploid and tetraploid and resynthesized allotetraploid *A. suecica* (alloF2 and allo701), CG methylation at the designated site is either missing (At2, At4 and allo701) or at very low level (alloF2). It suggests that this particular site might go through dynamic methylation status changes during allopolyploid evolution. *A. arenosa* lanes didn't hybridize well due to the sequence divergence in this region (Figure 2.4B).

Preferential Demethylation at *A. Arenosa* Centromeres in *met1*-RNAi *A. Suecica* Lines Is Associated with siRNA Accumulation in Centromere Regions

Centromeres of *Arabidopsis* are heavily methylated and packed as condensed heterochromatin. To study the consequence of hypo-methylation on centromeres, Southern blots of genomic DNA were probed using centromere-specific sequences. *A. thaliana* and *A. arenosa* share ~60% similarity in centromere satellite repeat (Figure 2.5). Primers were designed to amplify the genome-specific centromere satellite from *A. thaliana* and *A. arenosa* respectively. Southern analysis revealed that decrease of CG methylation occurred in both *thaliana* and *arenosa* originated centromeres in *met1 A. suecica*, while *arenosa*-specific centromeres were more responsive to hypo-methylation compared to

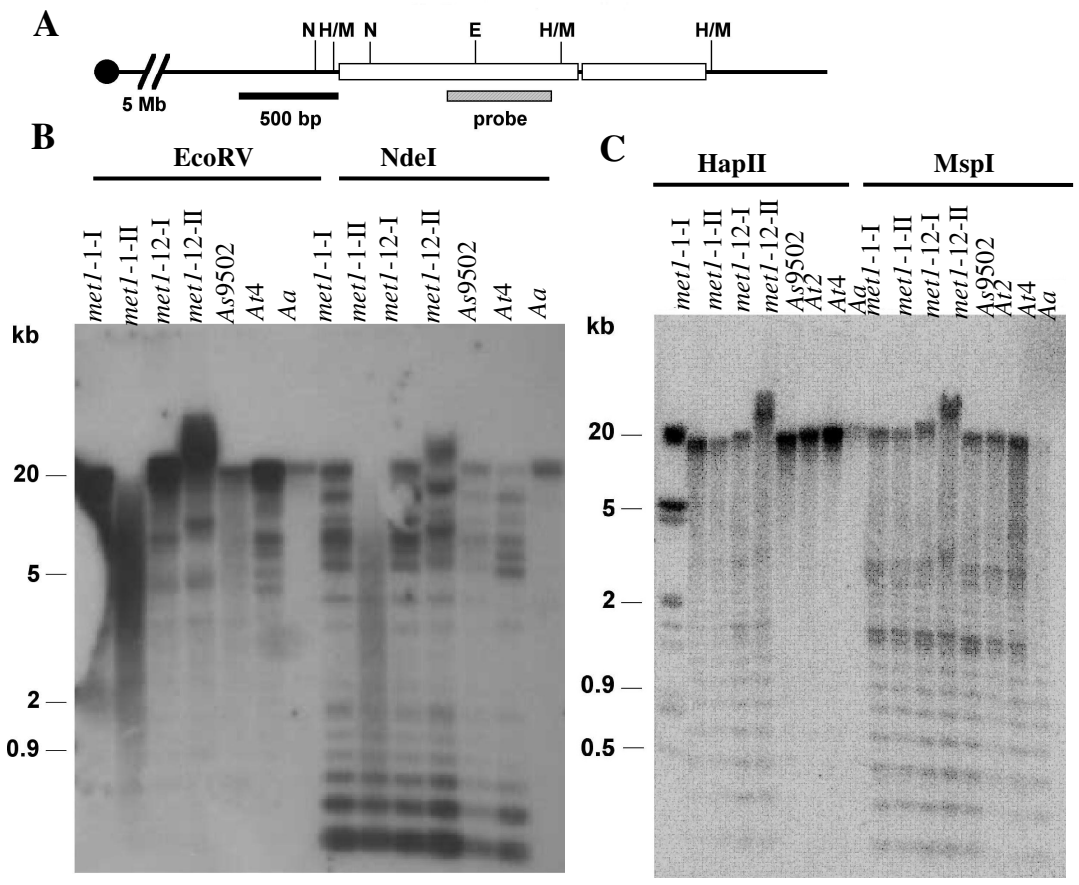


Figure 2.2 Demethylation of *SCP* gene in *met1* lines.

(A) Schematic diagram of *SCP* (At5g36180) gene structure, labeled with restriction digestion sites. E: EcoRV site, N: NdeI site, H/M: HpaII/MspI site. **(B)** and **(C)** show the Southern blots hybridized with probes derived from SCP. Demethylation was observed in *met1* lanes. Lanes: **1**-*met1*-1-I, **2**-*met1*-1-II, **3**-*met1*-12-I, **4**-*met1*-12-II, **5**-*As*9502 (natural *A. suecica* line 9502), **6**-*At*2 (*A. thaliana* diploid), **7**-*At*4 (*A. thaliana* tetraploid), **8**-*Aa* (*A. arenosa*).

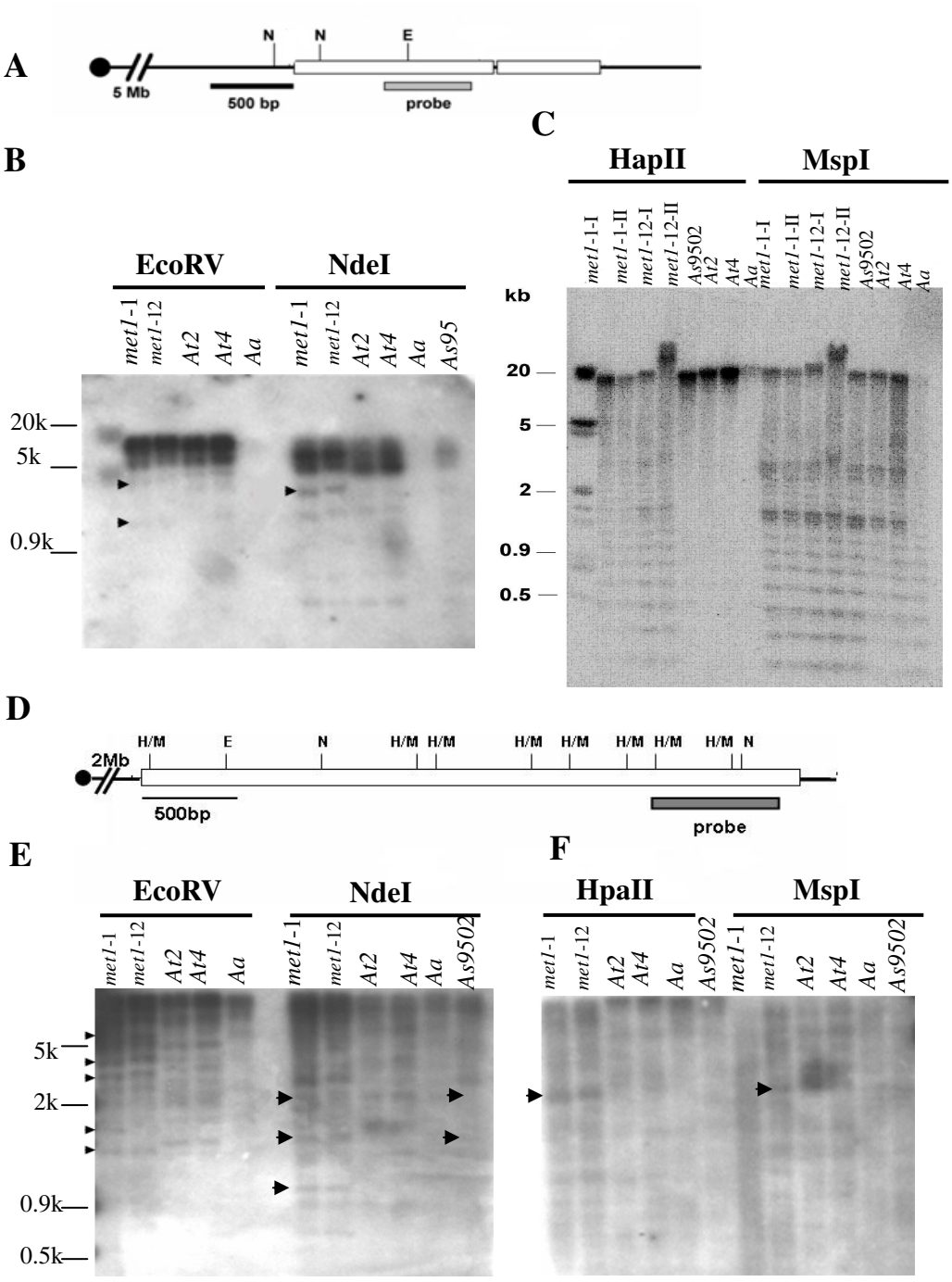
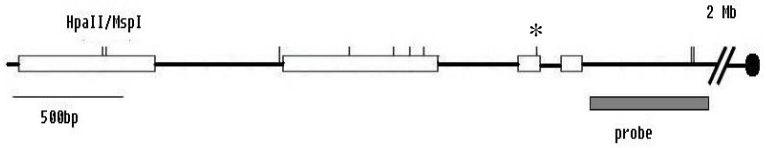


Figure 2.3 Demethylation and transpositional reactivation of transposons in *met1* lines.

Schematic structures for transposon genes At4g08010 and At1g44070 are shown in **(A)** and **(D)** respectively. **(B)**, **(C)** and **(E)**, **(F)** show the Southern blots hybridized with At4g08010 and At1g44070 probes respectively. The transpositional reactivations of transposons are shown as the bands indicated with black arrow heads.

A



B

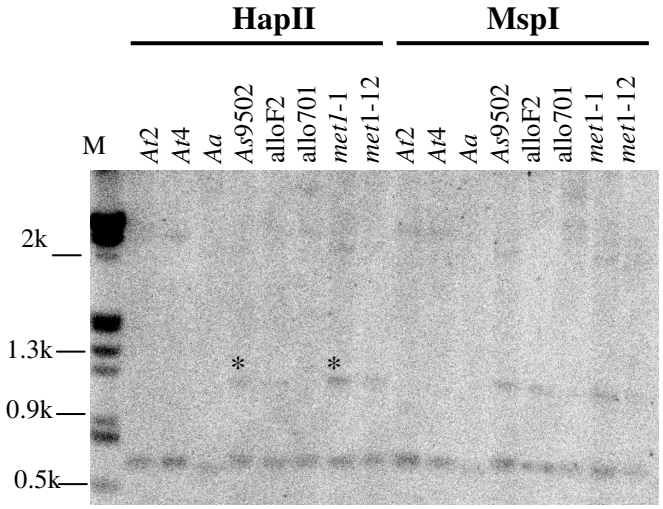


Figure 2.4 Increased methylation occurs in the 5' region of a down-regulated gene At5g25610 in *met1* lines.

Schematic gene structure and HpaII/MspI digestion sites are shown in **(A)**. Southern blot is shown in **(B)**. Asterisk shows the CG-methylation increased site and the presence of larger bands corresponding to it. Lanes: **1-***At*2, **2-***At*4, **3-***Aa*, **4-***As*9502, **5-**alloF2 (resynthesized *A. suecica* allotetraploid lines, 2nd generation), **6**-allo701 (resynthesized *A. suecica* allo lines, 5th generation), **7-***met1*-1 (transgenic RNAi *A. suecica*9502, *met1*-1 line), and **8-***met1*-12 (transgenic RNAi *A. suecica*9502*, met1*-12 line).

```
At_centromere      AAGCTTCTTCTTGCTTCTCAAAGCTTTGATGGTTTAGCCGGAGTCCATATGAGTCTTTGT 60
Aa_centromere      --GGATCCGGTTGCGGTTCTA-GTTCTTATA-CCCAACCATAAACACGAGATCTAGTCAT 56
                     *  **   ****   ** * * * * **     * **  *  *   *    *  *  *

At_centromere      CTTTGTATCTTCTAACAAGAAA-ACACTACTTAG-GCTTTTAGGAT----AAGATTGCGG 114
Aa_centromere      ATTTGACTCCAAAAACAAAAACCAAGCTTCTTATTGCTTCTCAATTCTTTGTGGGTGTGG 116
                    ****  **    ***** **  *  ** ****  **** *    *      *  ** **

At_centromere      TTTAAGTTCTT---ATACTCAATCATACACATCAAATCAAGTTATATTCGACTCCAAAAC 171
Aa_centromere      CCGAAGTCCTATGAGTTTTCGGTTTTGGAGCTTCTAAACGGAAAAATACTACTTTAGCAT 176
                      **** **     *  **  *  *  *  *   *    *  * ** * ***  *  *

At_centromere      ACTAACC 178
Aa_centromere      TTG---- 179
```

Figure 2.5 Sequence alignment of centromere satellite repeats originating from *A. thaliana* and *A. arenosa*.

Sequences were obtained form NCBI database and alignment was done by ClustalW analysis.

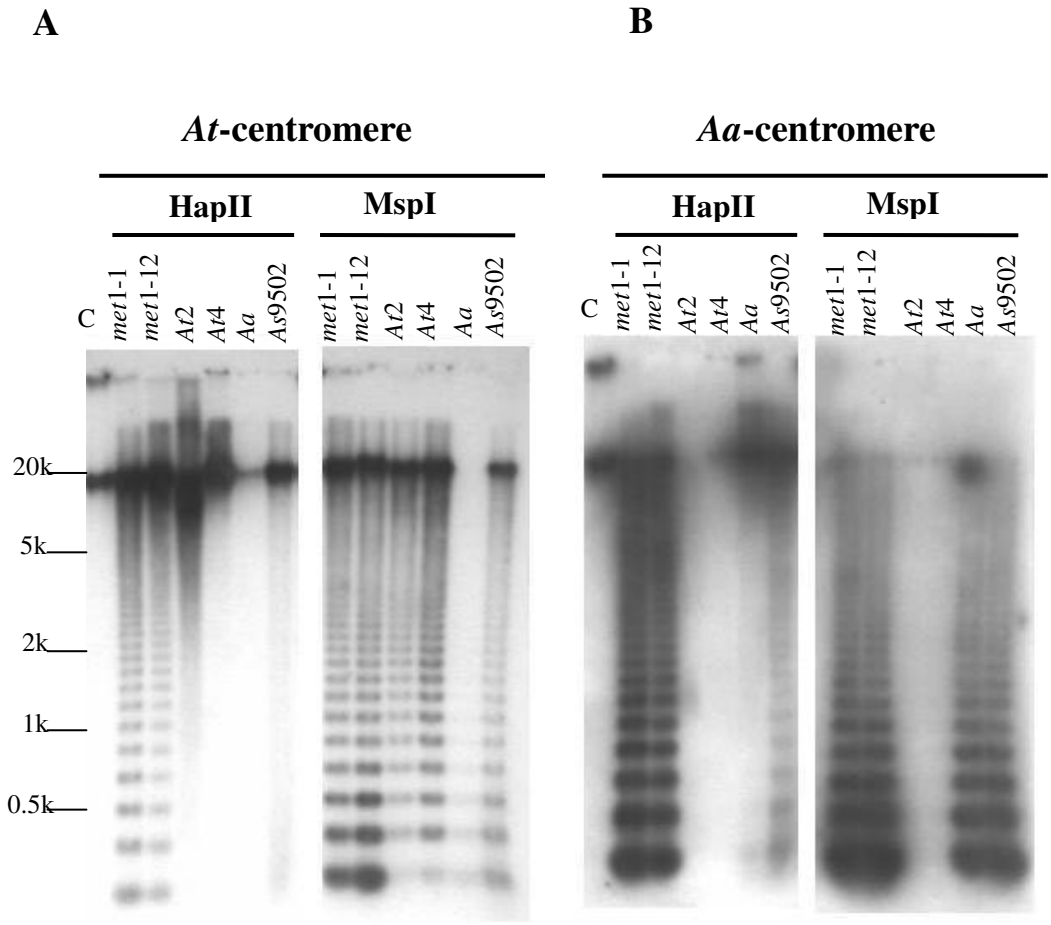


Figure 2.6 Preferential demethylation at *A. arenosa* centromeres in *met1*-RNAi *A. suecica* lines.

(A) and **(B)** is exactly the same membrane blotted with same amount of *A. thaliana*-specific and *A. arenosa*-specific centromere probes, respectively. Ladders show the centromere satellite single repeat. No cross hybridization happens as shown in lane 5 in **(A)** and lane 3 and 4 in **(B)**. Lanes: **C**. exposure control, **1-***met1*-1, **2**-*met1*-12, **3**-*At*2, **4-***At*4, **5**-*Aa*, and **6-***As*9502

thaliana-specific centromeres, by showing stronger hybridization signals for single centromere repeats (Figure 2.6, B shows the same membrane as in A, reprobed with a different probe). This suggests that *A. thaliana* genome and *A. arenosa* genomes might have different sensitivity to DNA methylation changes, at least in centromere heterochromatin region. In addition, *arenosa* originated centromeres were shown moderately demethylated even in wild type nature A. suecica 9502 line (Figure 2.6B), which suggested that the polyploidization process may alter the chromatin structure during evolution.

To investigate that whether siRNA or miRNA also play a role in centromere heterochromatin structure in polyploids as they do in diploids (Lippman and Martienssen, 2004), Northern blots for small RNA detection were conducted. RNAs were separated in 15% polyacrylamide gels for a better seperation of RNAs in a low size rage (20~70nt). Figure 2.7A showed that in *met1* lines, an overall increase of low size range RNAs was observed, maybe due to the increase of siRNAs derived from hypomethylated transposons in *met1* lines, as proposed previously (Lippman et al., 2003). A 70-mer DNA oligo was synthesized based on the *Arabidopsis U6* gene, and was run as size marker and blotted to the membrane to show hybridization control. After hybridized with *thaliana*-specific and *arenosa*-specific centromere probes separately, siRNA accumulations associated with centromere satellite repeats were detected (Figure 2.8A and B). In addition, by using PhosphorImager BAS1800II image analysis software (Fuji), a 10-fold higher signal was identified for *thaliana*-specific centromere siRNA accumulation, compared to *arenosa*-specific centromere siRNA (Figure 2.8C). This might explain the

genome-specific differential demethylation in the centromere region in *met1* lines by the mechanism of RNA directed DNA methylation (RdDM) (Hamilton et al, 2002). In other words, the DNA demethylation levels of genome-specific centromeres are inversely related to the abundance of the siRNAs derived from centromere regions. A DNA oligo was synthesized complementary to microRNA Mir173, which is 22nt in length, and hybridized to the membrane to show a size marker. The siRNAs associated with appeared to be about 24-26nt in size, which corresponded with the report that a subset of relatively larger siRNAs (24~26nt) are responsible for RNA dependent DNA methylation (Hamilton et al, 2002). In addition, transcripts around 70nt and 100nt were observed in *met1* lanes in the northern blots hybridized with both centromeres (Figure 2.9A and B), which were absent or much less abundant in other lanes. The 70nt bands may be processing intermediates of siRNAs derived from centromeres, similar to dsRNA intermediates which were defined when potato spindle tuber viroid (PSTVd) RNA was expressed in tomato (Denti et al., 2004). Transcripts with sizes close to single centromere satellite repeat appeared to accumulate in *met1*-1 and *ddm1*-176 lines (Figure 2.9, red arrows), which might be the transcripts from sub-set of centromere repeats, serving as the original trigger of dsRNA and siRNA formation.

To study the particular regions and DNA strands from which the centromere-associated siRNAs were derived, two 40-mer DNA oligos were synthezised for hybridization to *arenosa*-specific centromere antisense strand (Figure 2.10A). Northern blots showed that both oligos induced siRNAs and centromere-like transcripts (Figure 2.10B and C), thus no conclusion could be drawn based on this experiment.

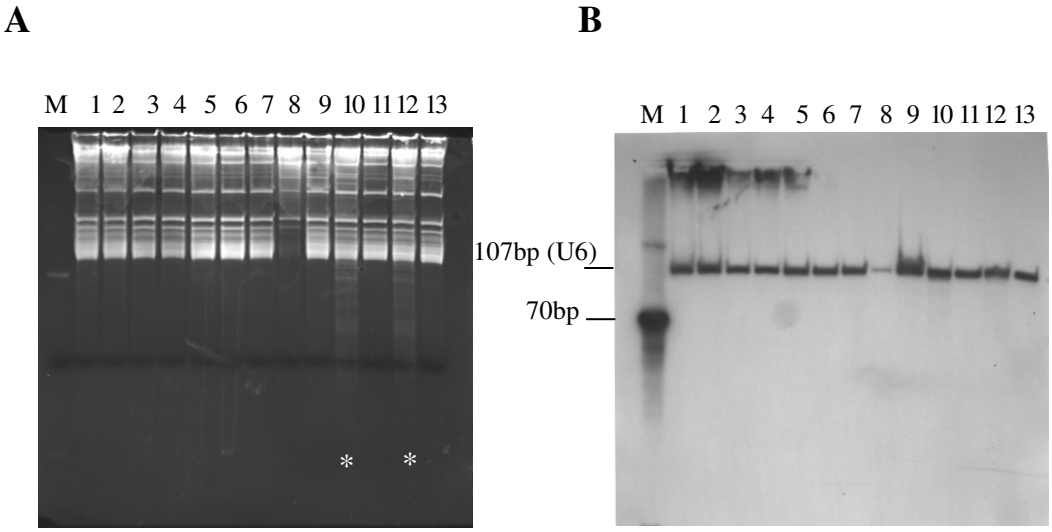


Figure 2.7 Northern blot for siRNA detection.

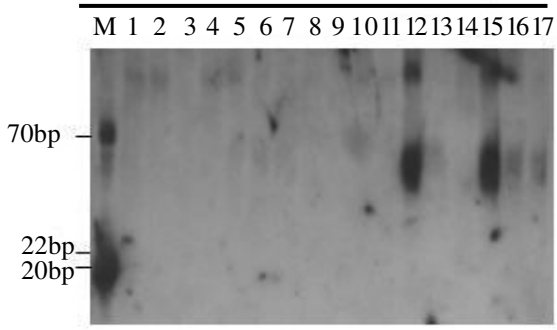
A) Ethidium bromide staining shows total RNA loading control. Asterisks show the lanes of *met1* lines which have higher accumulation of total siRNAs. **(B)** shows the membrane blotted with oligo of *Arabidopsis* U6 as hybridization and size control.
Lanes: 1-*At*2, 2-*At*4, 3-*Aa*, 4-*As*9502, 5-alloF1, 6-alloF2, 7-alloF3-*rdrp*-2 (RNAi knock down *At* RNA DEPENDENT RNA POLYMERASE2 in alloF3, 8-9502-*rdrp*-1 (RNAi knock-down in natural *As*9502), 9-9502-*rdrp*-2 (different transgenic line from 8), 10-*met1*-1, 11- *met1*-12, 12-*ddm1*-176 (RNAi know down *At* DDM1 in natural *As*9502), 13-*ddm1*-3 (different transgenic line from 12)

Figure 2.8 Centromere-associated siRNA accumulates preferentially from *thaliana*-specific centromeres.

Small RNA blots were hybridized with *thaliana*-specific centromere in **(A),** and *arenosa*-specific centromere in **(B)**. **(C)** is the comparison of signal strength analyzed by Fuji PhosphorImager. Marker shows similar exposure strength in **(A)** and **(B)**, and emission of siRNA signals were compared to the marker standard. **(D)** shows the relative size of centromere-derived siRNA. Arrows show the hybridization signal position of Mir173, which is a microRNA of 22nt in length, serving as size marker here.
Lanes: 1-*At*2, 2-*At*4, 3-*Aa*, 4-*As*9502, 5-alloF1, 6-alloF2, 7-alloF3, 8-alloF3-*rdrp*-2 (RNAi knock down *At* RNA DEPENDENT RNA POLYMERASE2 in alloF3, 9-9502-*rdrp*-1 (RNAi knock-down in natural *As*9502), 10-9502-*rdrp*-2 (different transgenic line from 8), 11-10-9502-*rdrp*-4, 12- *met1*-1, 13- *met1*-12, 14- *met1*-12, 15-*ddm1*-176 (RNAi know down *At* DDM1 in natural *As*9502), 16-*ddm1*-3 (different transgenic line from 12), 17-*ddm1*-3.

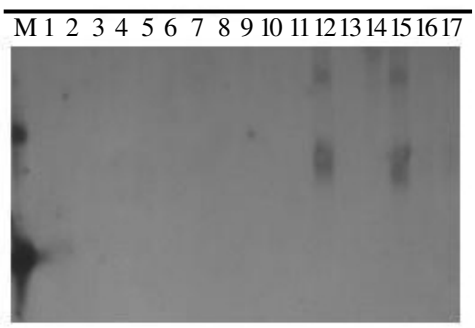
A

***At*-centromere**



B

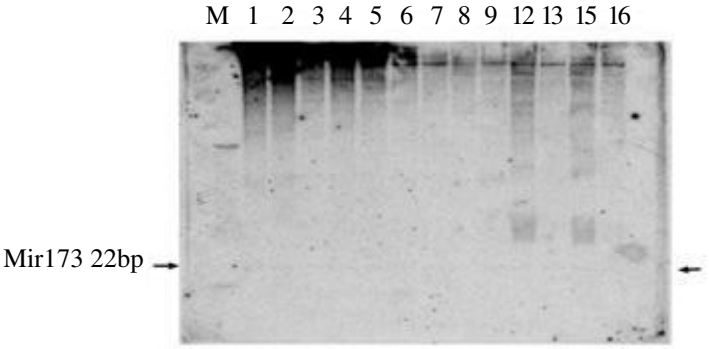
***Aa*-entromere**



C

Cen	lane	Signal strength (total emmision/marker)	Signal intensity (emission/area)
At	*met1-1*	45%	18%
	ddm1-176	45%	16%
Aa	*met1-1*	3%	8%
	ddm1-176	3%	9%

D



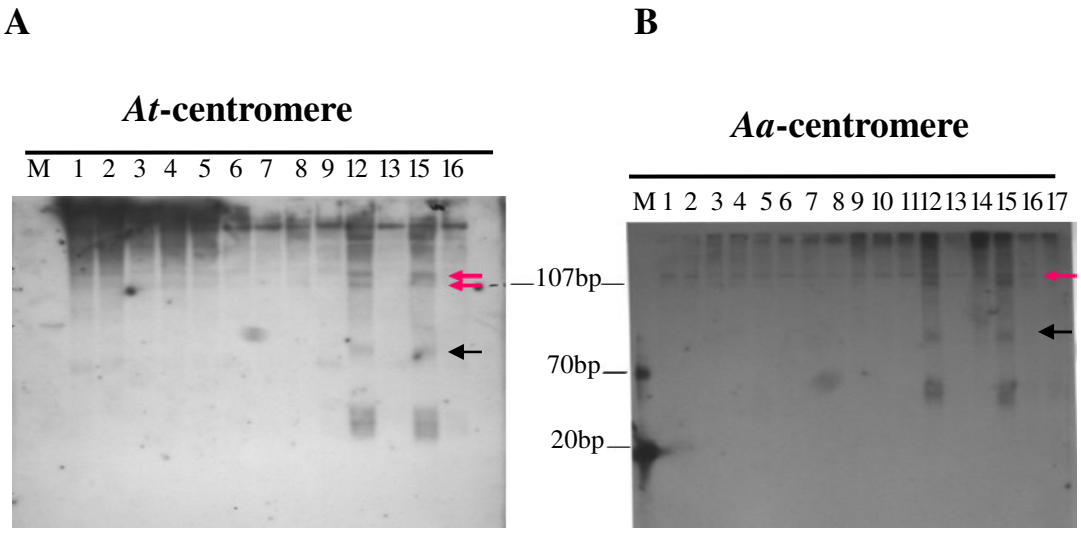
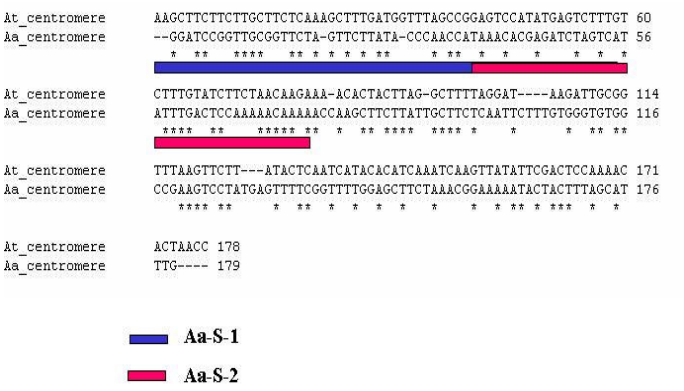


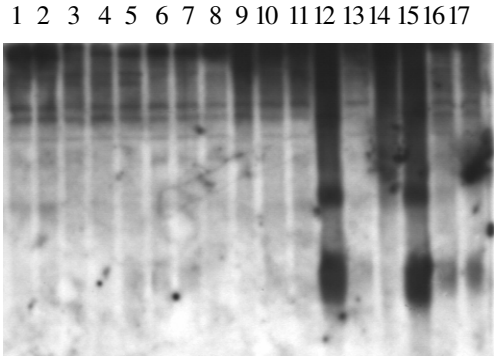
Figure 2.9 Intermediates from siRNAs processing and transcripts derived from centromeres.

In **(A)** and **(B)**, black arrows show the possible intermediates for siRNA processing (about 70bp), and red arrows show the possible transcripts derived from centromeres because of reduction of methylation.
Lanes: 1-*At*2, 2-*At*4, 3-*Aa*, 4-*As*9502, 5-alloF1, 6-alloF2, 7-alloF3, 8-alloF3-*rdrp*-2 (RNAi knock down *At* RNA DEPENDENT RNA POLYMERASE2 in alloF3, 9-9502-*rdrp*-1 (RNAi knock-down in natural *As*9502), 10-9502-*rdrp*-2 (different transgenic line from 8), 11-10-9502-*rdrp*-4, 12- *met1*-1, 13- *met1*-12, 14- *met1*-12, 15-*ddm1*-176 (RNAi know down *At* DDM1 in natural *As*9502), 16-*ddm1*-3 (different transgenic line from 12), 17-*ddm1*-3.

A



B



C

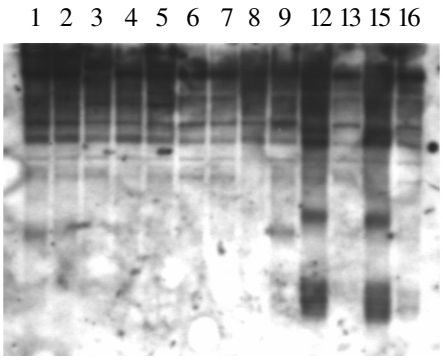


Figure 2.10 Origins of centromere-derived siRNAs.

Two 40-mer DNA oligos designed to detect siRNA origins are shown in **(A)**. They are designed to hybridize to the first 80bp region of the antisense strand of *arenosa*-specific centromeres. **(B)** and **(C)** are Northern blots hybridized with Aa-S-1 and Aa-S-2 oligos, respectively. Signals for siRNAs and centromere-like transcripts can be detected in both blots.
Lanes: 1-*At*2, 2-*At*4, 3-*Aa*, 4-*As*9502, 5-alloF1, 6-alloF2, 7-alloF3, 8-alloF3-*rdrp*-2 (RNAi knock down *At* RNA DEPENDENT RNA POLYMERASE2 in alloF3, 9-9502-*rdrp*-1 (RNAi knock-down in natural *As*9502), 10-9502-*rdrp*-2 (different transgenic line from 8), 11-10-9502-*rdrp*-4, 12- *met1*-1, 13- *met1*-12, 14- *met1*-12, 15-*ddm1*-176 (RNAi know down *At* DDM1 in natural *As*9502), 16-*ddm1*-3 (different transgenic line from 12), 17-*ddm1*-3.

DISCUSSION

In *met1*-RNAi *A. suecica* plants, endogenous *MET1* expression is reduced to less than 5% of that in WT plants, which leads to pleiotropic changes in phenotype and development. The increased severity of phenotypic effects in *met1* plants with successive generations suggests that the disorder of DNA methylation may disturb meiosis or gametogenesis by perturbing the normal structure and function of centromere. The inability to establish essential imprinting patterns because of lack of DNA methylation also affects the embryo development (Berger, 2004; Grini et al., 2002; Xiao et al., 2003).

Southern analysis detected an overall reduction of CG methylation in the *SCP* gene, two En/Spm transposons and centromeric repeats of *met1* lines, which correlated to their increased transcriptional levels, implying the impact of *MET1* on CG methylation and gene silencing in plants. In addition, an increase of CG methylation was found at the 5' transcriptional region of a *RD22* family gene, corresponding to the suppression of this gene in *met1* plants. Transpositional reactivations were detected for both En/Spm transposons examined in these experiments, in contrast to the incapability of transposition for *CACTA* transposon in spite of increase of transcripts in *Arabidopsis* mutants that express antisense *MET1* gene (Lippman et al, 2003). This may be due to the fact that the transposon genes were selected from microarray gene list that displayed highest transcriptional up-regulation. Thus, as just one of the major epigenetic regulatory systems, DNA methylation probably works in a comprehensive way and provides partial shaping force for gene expression (Kishimoto et al., 2001).

Differential demethylation of *thaliana*- and *arenosa*- specific centromeres in *met1 suecica* implied a different sensitivity of progenitor genomes to hypo-methylation. The detection of modest demethylation of *arenosa*-specific centromere in WT *A. suecica* 9502 implicated a possible chromatin structure change upon polyploidization. Resynthesized allotetraploid lines could be included to investigate whether the loosening of *arenosa*-originated centromere structure is an instant effect or a evolutionary result in polyploidization.

Northern analyses revealed that the different levels of centromere demethylation might be associated with siRNAs preferential accumulation from one genome (*A. thaliana*), which correlated with the siRNA directed DNA methylation for transcriptional silencing of heterochromatic region (Lippman and Martienssen, 2004). However the origins of the centromere-derived siRNAs were not identifiable in this experiment by the blotting method, so there is little information about how the putative dsRNA precursors of the siRNAs arise from heterochromatic regions. One possibility is that RdRp may convert a rare transcript from heterochromatic DNA into a double-stranded form which serves as the precursor of siRNA formation (Hamilton et al., 2002). Bidirectional transcription from heterochromatin has been detected in *met1* and *ddm1* mutants by tilling microarray in *Arabidopsis* (Lippman et al., 2004), arguing that transcription from sense as well as antisense strand may be another source of dsRNA. Moreover, it was also proposed that readthrough transcription of inverted repeats could account for the occurrence of dsRNA from heterochromatin (Sijen and Plasterk, 2003). More blots may be needed to determine the particular origins and DNA strands the siRNAs are derived

from.

MATERIALS AND METHODS

Plant Materials

All plants were grown in vermiculite mixed with 10% soil in a growth chamber with growth conditions of 22°/18° (day/night) and 16 hr of illumination per day.

A. thaliana diploid ecotype Lansberg *erecta* (Ler) was used throughout the experiments. Autotetraploid *A. thaliana* was obtained through colchicine treatment of Ler (Comai et al., 2000a) (accession no. CS3900 in the Arabidopsis Biological Resource Center (ABRC), $2n = 4x = 20$). Autotetraploid *A. arenosa* (accession no. 3901, $2n = 4x = 32$) and natural *A. suecica* (9502) (2n=4x=26) were used. Rosette leaves were collected at ~16-leaf stage for all the plants. Leaf tissues were collected from individual plant for *met1* transgenic lines, and pooled for other plants for DNA and RNA extraction.

DNA and RNA Analysis

For all the plant materials, RNA and DNA were isolated separately to ensure the best quality. RNA was extracted using TRIzol reagent (Invitrogen Life Technologies), following the manufacture instruction. Genomic DNA was isolated following lab protocol described below. Tissues frozen in liquid nitrogen was grounded to powder, and then

mixed with three volumes (wt/vol) of extraction buffer [200 mM Tris·HCl, (pH7.5), 250 mM NaCl, 25 mM EDTA (pH 8.0), and 0.5%SDS)]. After centrifuge at 3,500 rpm in 4°C for 15 min, the supernatant was extracted twice with phenol/chloroform/isoamyl alcohol and precipitated with 2x volumes of 95% ethanol plus 1/10 volume of 3 M NaOAC. Pellet was washed with 70% ethanol follow by a RNaseA treatment, and dissolved in ddH_2O for standard southern. DNA probes were obtained from PCR followed by purification (Qiagen PCR purification kit) and were prepared by a random priming method (Amersham) with radioactive labeling (α-P^{32} ATP). Hybridization was performed following the method of Church and Gilbert (Church and Gilbert, 1984).

Total RNA (25μg) was subjected to electrophoresis in an polyacrylamide gel (15%, polyacrylamide: bisacrylamide = 19:0.5) and blotted onto Hybond-N+ membrane (Amersham Pharmacia) for small RNA Northern blot. 5' end labeling was performed. 30-50ul total labeling mixture contained 1ul of T4 polynecleotide kinase, 3-5ul of 10x kinase buffer, 3-5ul of γ-P^{32} ATP (3000 Ci/ul), 1ul of gene-specific primer (50uM), and ddH_2O and the incubation was kept at 37° for at least 1 hr or room temperature over night. Hybridization was performed following the method of Church and Gilbert (Church and Gilbert, 1984). Signals were detected after at least over night exposure to Kodak X-ray films or analyzed by PhosphorImager BAS1800II system (Fuji, Japan).

Primers and PCR

For At1g44070: 5'-TCCATGCTTGAGCGAAAAAGA-3' (forward primer) and 5'-CATCCCCGACATCGTAATCAA-3' (reverse primer).

For At4g08010: 5'-AGAGATTGTGCATGGGCCTTT-3' (forward primer) and 5'-GCATCTCCGACATCGTAATCG-3' (reverse primer).

For *A. thaliana* specific centromere: 5'-AAGCTTCTTCTTGCTTCTCAAA-3' (forward primer) and 5'-GGTTAGTGTTTTGGAGTCGAA-3' (reverse primer).

For *A. arenosa* specific centromere: 5'-GGATCCGGTTGCGGTTCTAGTT-3' (forward primer) and 5'-CAAATGCTAAAGTAGTATTTT-3' (reverse primer).

Oligo sequence for U6:

GATAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCCTGCGCAAGGATGACACGCATAAATCGAGAA.

DNA oligo sequence for Mir173 blot: GTGATTTCTCTCTGCAAGCGAA.

Aa-S-1 sequence: GGATCCGGTTGCGGTTCTAGTTCTTATACCCAACCATAAA.

Aa-S-2 sequence: CACGAGATCTAGTCATATTTGACTCCAAAAACAAAAACCA.

The PCR reaction included 1 cycle of 94° for 2 min, 30 cycles of 94° for 30 sec, 52° for 30 sec, and 72° for 1 min.

CHAPTER III

DNA METHYLATION PLAYING AS A FACTOR IN *Arabidopsis* POLYPLOIDIZATION

INTRODUCTION

To explore the genome-wide gene expression changes in polyploid, amplified fragment length polymorphism (AFLP) cDNA display analysis is often performed (Vos et al., 1995; Mei et al., 2004). However the need of investigating large-scale gene expression profiles led to a alternative method which is more informative and enables higher through-put. Microarray technology entails features of Southern and Northern blotting. To conduct microarray analysis, fragmented DNA is attached to a solid surface and then probed with known DNA or RNA samples to compare the genetic diversity between two or more interested treatments (DNA or RNA samples) (Fodor 1991; Hughes 2001; Singh-Gasson, 1999). Oligo DNA microarray analysis is used in the Chen Lab, and currently relies on microarray featuring 26,000 unique 70-mer oligos designed within 1,000bp of the 3' end of characterized genes based on publicly available *Arabidopsis thaliana* sequence to represent the whole genome of *A. thaliana* (Operon/Qiagen, http://oligos.qiagen.com/arrays/oligosets_arabidopsis.php). The oligos are spotted on

poly-L-lysine- or SuperAmine-coated glass slides, as described in previous paper (Wang et al, 2005). After hybridization and signal capturing, data are normalized and analyzed with a linear model which was described before (Lee et al., 2004).

As described in Wang et al.'s paper (2004), microarray experiments have been conducted to individually compare the gene expression of two independently-derived resynthesized *A. suecica* allotetraploid lines, allo733 and allo738 (the fourth generation of synthetic allotetraploid), with the mid-value of the additional expression from equal-amount mixture of two progenitors, *A. thaliana* tetraploid and *A. arenosa* (mid-parent). More than 1,300 genes were identified to have non-additive expression (Wang et al., 2005), among which >65% were genes that were suppressed in the allotetraploid. Moreover, >94% of the repressed genes have higher expression in *A. thaliana* progenitor rather than in *A. arenosa*, suggesting an imperfect genomic dominance during interspecific hybridization.

To study the genome-wide changes in gene expression alterations in *met1* transgenic plants, microarray experiments were also performed to compare *met1*-RNAi *A. suecica* with wild type *A. suecica* (WT) (Wang et al., data unpublished). Using GeneSpring software (Silicon Genetics) for data analysis (p=0.05, >= 2-fold change), about 200 genes were identified as significantly differentially expressed genes. By comparative analysis, we could identify methylation-affected and allopolyploidization-related genes, followed by the study of progenitor origins for the reactivated homoeologous genes.

RESULTS

Transposons Are Over-represented Among the Genes That Are Affected in *met1*-RNAi *A. suecica* Plants

Based on microarray data in *met1*-RNAi *A. suecica*, up- and down-regulated genes each represent 50% of the total number of affected genes (Wang et al, data unpublished), indicating that there are some other mechanisms regulating gene expression besides DNA methylation. To study the content of the affected genes in hypo-methylation background, genes were subjected to categorization into 15 functional categories modified from the published *Arabidopsis* functional categories from *Arabidopsis* Information Resource (TAIR) GO annotation (http://www.arabidopsis.org/tools/bulk/go/index.jsp). In Figure 3.1, the representative percentage for each functional category was calculated by dividing the percentage of genes in this category identified in particular microarray experiment with the standard percentage in this category in WT *Arabidopsis*. For example, 3.64% of affected genes in *met1*-1 are transposable elements, whereas 5.51% of the total number of genes in *WT Arabidopsis* entire genome are transposable elements, thus the percentage of the category of transposable elements in *met1*vsWT microarray will be 100% x 3.64% / 5.51%=66% (Figure 3.1). By comparing the category percentages of three microarray data sets (*met1* vs WT, allo733 vs mid-parent and allo738 vs mid-parent), we got an overall understanding of the structures of gene populations from microarray experiments. The two parallel comparisons, allo733 vs mid-parent and allo738

vs mid-parent do not overlap with each other 100% because of the individual variances. However, they show overall the same trend that categories of cell rescue and defense, energy synthesis, metabolism, and hormone regulation are over represented (>100% of the standard percentage, Figure 3.1). It argues that the polyploidization process can entail or cause genomic shock to the plant genome. Altered and novel gene expressions need to be established to modify and regulate metabolism and growth to adapt to the increase dosage of genome. It also suggests that hormone regulation may play an important role in re-establishing this harmony in polyploid. For the genes detected in *met1* vs WT microarray, no category is over-represented compared to the standard percentage (no larger than 100% of the standard), indicating that DNA methylation directly regulates the expressions of only some of the genes in the genome. It is noteworthy that transposons are the only category that is relatively over-represented in *met1* vs WT compared to allo vs mid-parents experiments, indicating that DNA hypo-methylation mainly affects transposons and heterochromatic regions. However, when considered relative to genome composition, the outcomes could have somewhat different implications. The standard percentage of transposons in the whole genome is higher than what is affected in *met1*, indicating that not all transposons are affected in *met1* hypo-methylation background.

To study the chromosomal location and distribution of the affected genes in *met1* background, *in silico* methods were used to chromosomally map the ~200 genes identified in *met1* vs WT microarray experiments onto the five chromosomes of *Arabidopsis*.

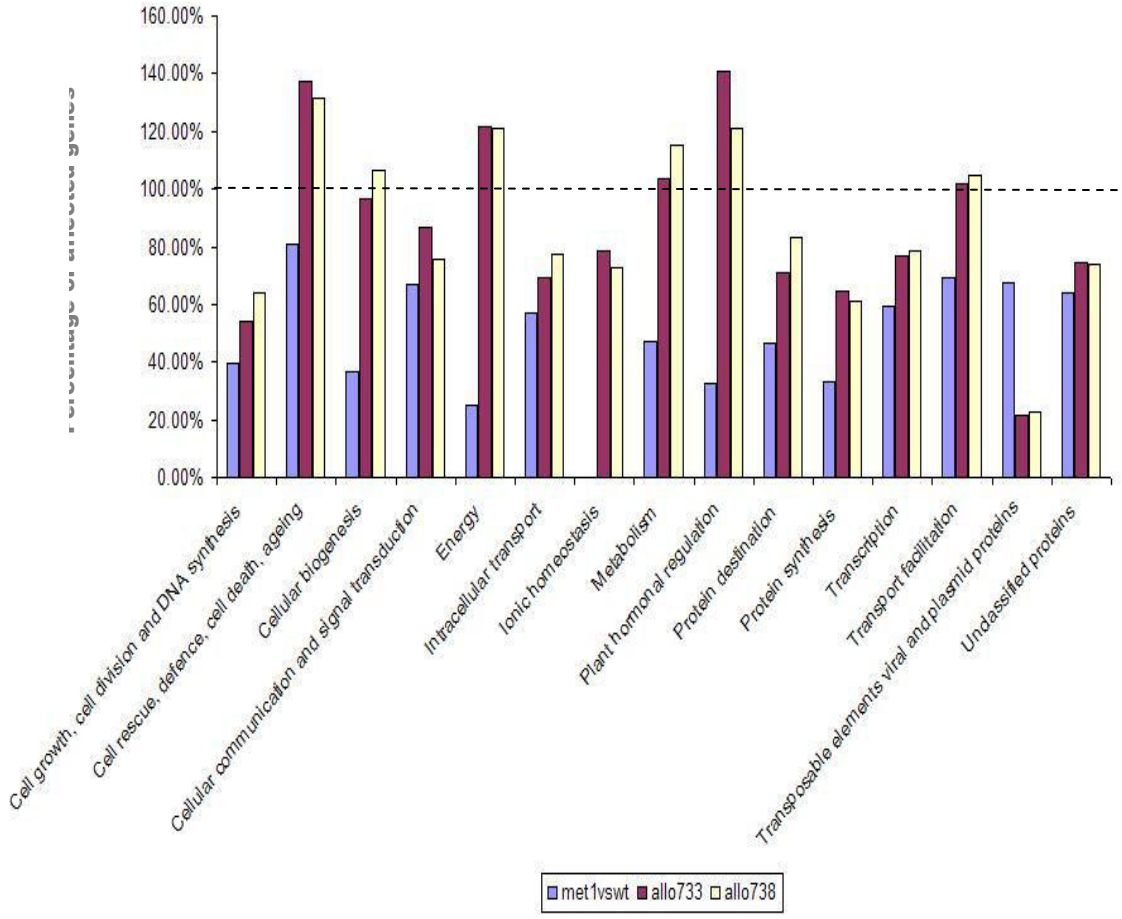


Figure 3.1 Differential effects of gene expression by methylation or polyploidization by gene category.

Categorized genes from three separate microarray data sets are shown here. The y-axis denotes the percentage of genes affected relative to gene content 100% x (percentage of affected genes / percentage of recognized genes in *A.t.*). Broken line stands for the standard of each functional category represented in entire genome of WT *Arabidopsis* diploid.
Legend: *met1* vs wt: microarray experiment comparing gene expression of *met1*-1 with wild type *A. suecica*9502.
allo733: microarray experiment comparing the gene expression of synthetic allotetraploid line allo733 with ½ of the equal-amount mixture of *A. thaliana* tetraploid and *A. arenosa* (mid-parent).
allo738: microarray of comparing allo738 and mid-parent.

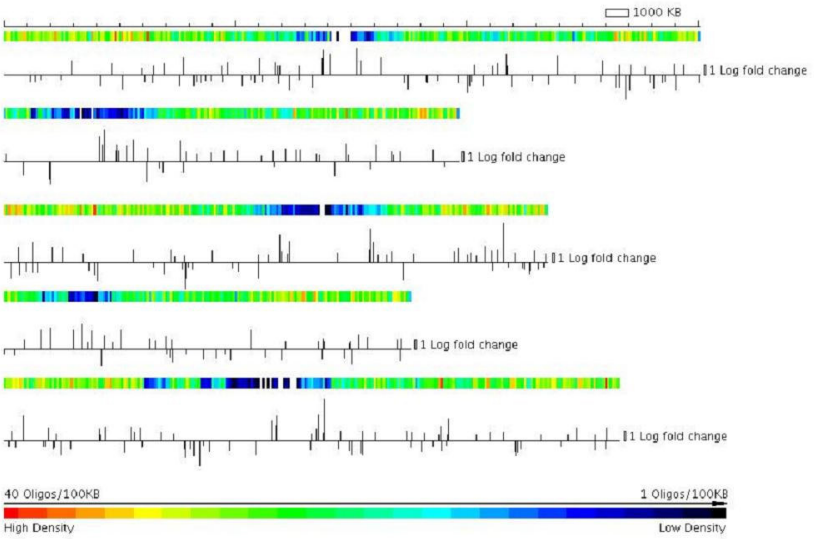


Figure 3.2 Chromosomal distribution of methylation-related genes.

The color bars stand for the five chromosome pairs in diploid *Arabidopsis thaliana* (2n=2x=10, n=x=5). Lines sticking upwards and downwards represent the up- and down-regulated genes identified by *met1* vs WT microarray experiment. Length of each vertical line represents the up- or down-regulation level of the affected gene. A nearly even distribution of genes can be observed, as well as major up-regulations condensed in centromeres and heterochromatins, shown as dark blue regions in the bar.

The positions of the hypo-methylation-related genes are mostly evenly distributed along the chromosome arms, indicating that DNA methylation has a genome-wide and sporadic impact on the genes located in euchromatic regions (Figure 3.2). Up-regulated genes are clustered towards centromeric regions and other heterochromatic regions, in contrast to the alternative distribution of up- and down-regulated genes along chromosome arms, correlating with our assumption that heterochromatic regions are primarily subjected to repression by DNA methylation.

Determine the Methylation-affected and Allopolyploidization-related Genes

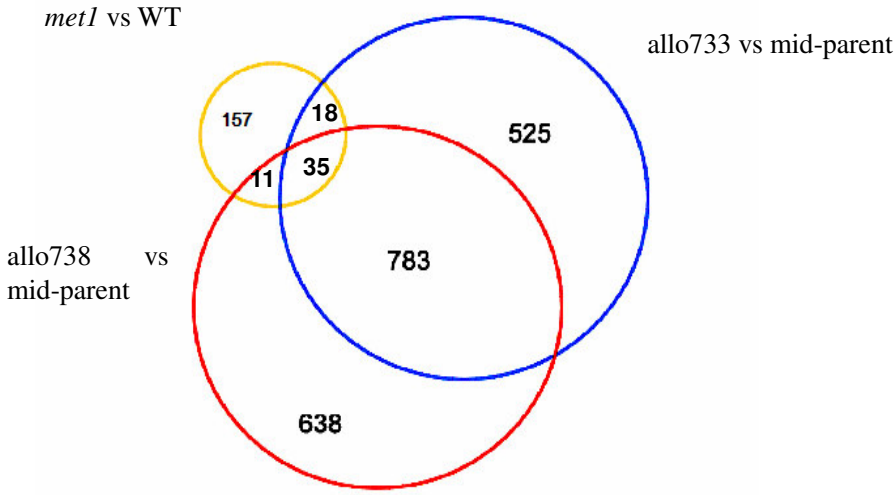
For the ~1,3000 genes that were identified in allo lines vs mid-parent microarray experiments, different regulatory mechanisms might exist and work together to cause the alterations in gene expressions in allotetraploid. Here we focused on studying the role that DNA methylation plays in polyploidization. By combining the three sets of microarray data, we were able to define a list of the methylation-affected allopolyploidization-related genes. In other words, the data set of met1 vs WT ∩ allo733 vs mid-parent ∩ allo738 vs mid-parent is shared by all three sets of microarray experiments, and the genes in this shared data set are the ones that are associated with allopolyploidization and affected by DNA methylation level change as well. There are total 74 genes containing in this shared set (Table 3.1). As mentioned above, individual variances existed between allo733 and allo738, so not every observation or differential gene expression could be overlapped between these two allotetraploid lines. For instance, At2g14580 is

down-regulated in allo733 by 0.57947-fold, but up-regulated in allo738 by 3.8471-fold. Because the experiments included four replications for each allo vs mid-parents comparison, and the resulting statistics indicated that such huge divergence between allo733 and allo738 is unlikely observed due to error. Indeed, some of the 74 genes we picked were counted twice, because they showed opposite directions of expression changes between allo733 and allo738. Results are summarized in the Venn diagram in Figure 3.3A, and the full gene list is given in Table 3.1. Of the 74 affected genes, up- and down-regulation in *met1* affected 40 genes and 34 genes, respectively. In contrast, the numbers of the up- and down-regulation genes are 21 and 53, respectively in allotetraploids. The shared set shows the same trend of gene expression changes as the trend shown in separate microarray data sets, which is that *met1* has almost same amount of up- and down-regulation compared to WT and the allotetraploid lines have most of genes down-regulated compared to mid-parent. For the genes with changes in the same direction, the average value is shown, whereas for the genes with diverged regulation trends, the amounts (folds) changes in allo733 and allo738 are both shown with "/" between them in the Table 3.2.

Based on the different up- and down-regulation trends of the experiments of allo and *met1*, the 74 genes could be grouped into four sub-groups, including down-regulated in both allo and *met1*, up-regulated in both allo and *met1*, up-regulated in allo but down-regulated in *met1*, and down-regulated in allo but up-regulated in *met1*. As described before, most of the genes are repressed in synthetic allotetraploids (Wang et al., 2005). To investigate whether DNA methylation is responsible for the suppression

of those genes, a study was focused on the last sub-group of genes mentioned above. As shown in Table 3.1, this sub-group is the largest portion in the four sub-groups, which contains almost half of the genes defined in the shared set. It implies that genes are re-pressed in the allotetraploids due to increased DNA methylation during polyploidization, but are reactivated when the DNA methylation is reduced in *met1* defect background.

A



B

Objective(microarray experiment)	Up-regulated	Down-regulated	Total
met1 (*met1* vs WT)	111	110	221
met1 (*met1*∩allo733∩allo738)	40	34	74
allo (*met1*∩allo733∩allo738)	21	53	74

Figure 3.3 Summary of the comparison analysis.

(A) Venn diagram comparing the three sets of microarray data. Numbers refer to the gene number identified in each microarray analysis. (B) Brief summary of the number of genes that are up- and down- regulated in *met1* vs WT and the overlap of genes affected by the treatments

Table 3.1 List of the genes in the shared set.

List of all the genes grouped into four sub-groups that are affected in both *met1* and allotetraploids. Oligo ID, locus ID, functional description, and fold changes for those genes in allotetraploids and *met1* are listed in the table.

Gene	Oligo_ID	Locus	Description	Average or both fold chage in allo733 and all738 (allo vs mid-parents)	Average fold change in *met1*(*met1* vs wt)
Downregulated in both allo and met1 microarray					
1	A001720_01	At1g72290	trypsin and protease inhibitor family protein / Kunitz family protein similar to water-soluble chlorophyll protein [Raphanus sativus var. niger] GI:16945735, BnD22 drought induced protein [Brassica napus] GI:17813; contains Pfam profile PF00197: Trypsin a	0.41559	0.189040625
2	A001732_01	At1g54650	expressed protein similar to Actin-binding protein ABP140 (Swiss-Prot:Q08641) [Saccharomyces cerevisiae]	0.56803	0.323654844
3	A006524_01	At2g37870	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam protease inhibitor/seed storage/LTP family domain PF00234	0.68885	0.480669419
4	A007535_01	At2g15960	expressed protein	0.488135	0.51756479
5	A007921_01	At2g14580	pathogenesis-related protein, putative	0.57947/3.847	0.258815271
6	A010805_01	At3g03870	expressed protein predicted using genefinder	0.7321/1.1538	0.630275329
7	A011810_01	At3g04720	hevein-like protein (HEL) identical to SP|P43082 Hevein-like protein precursor {Arabidopsis thaliana}; similar to SP|P09762 Wound-induced protein WIN2 precursor {Solanum tuberosum}; contains Pfam profile PF00187: Chitin recognition protein	0.459265	0.532601116
8	A012747_01	At3g15450	expressed protein similar to auxin down-regulated protein ARG10	0.630265	0.683958782
9	A014791_01	At4g27450	expressed protein similar to auxin down-regulated protein ARG10 [Vigna radiata] GI:2970051, wali7 (aluminum-induced protein) [Triticum aestivum] GI:451193	0.275685	0.491164479
10	A016636_01	At5g55450	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam protease inhibitor/seed storage/LTP family domain PF00234	0.87006/ 1.2701	0.508477471

Table 3.1 (continued)

Gene	Oligo_ID	Locus	Description	Average fold chage in allo (allo vs mid-parents)	Average fold change in *met1*(*met1* vs wt)
11	A018006_01	At5g24780	vegetative storage protein 1 (VSP1) identical to SP|O49195 Vegetative storage protein 1 precursor {Arabidopsis thaliana}; contains Pfam profile PF03767: HAD superfamily (subfamily IIIB) phosphatase	0.065851	0.175802131
12	A019450_01	At4g25490	DRE-binding protein (DREB1BCRT/CRE-binding factor 1 (CBF1) / transcriptional	0.74369/1.2881	0.542082589
13	A021053_01	At5g24770	vegetative storage protein 2 (VSP2) identical to SP|O82122 Vegetative storage protein 2 precursor {Arabidopsis thaliana}; contains Pfam profile PF03767: HAD superfamily (subfamily IIIB) phosphatase	0.0970295	0.196367648
14	A021371_01	AP001306		0.542725	0.262740185
15	A022007_01	At2g05510	glycine-rich protein	0.80196	0.232570815
16	A022124_01	At2g14610	pathogenesis-related protein 1 (PR-1) identical to GB:M90508 SP|P33154	0.21271/1.3808	0.329183356
17	A024214_01	At4g08320	tetratricopeptide repeat (TPR)-containing protein glutamine-rich tetratricopeptide repeat (TPR) containing protein (SGT) - Rattus norvegicus,PID:e1285298 (SP|O70593); contains Pfam profile PF00515 TPR Domain	0.809105	0.372610641
18	A025813_01	At1g57990	purine permease-related low similarity to purine permease	0.56552	1.816748717
19	A025929_01	At3g49120	peroxidase, putative identical to peroxidase [Arabidopsis thaliana] gi|405611|emb|CAA50677	0.51177	0.532177736
Upregulated in both allo and met1 microarray					
1	A012940_01	At4g10060	expressed protein contains Pfam domain PF04685: Protein of unknown function, DUF608	1.3457	2.004034673
2	A020559_01	At3g60530	zinc finger (GATA type) family protein	1.37585	0.652054673
3	A021656_01	At2g10880	Arabidopsis thaliana chromosome II section 57 of 255 of the complete sequence. Sequence from clones F16G22, F15K19	1.38185	3.46490441
4	A021767_01	At3g03830	auxin-responsive protein, putative similar to auxin-inducible SAUR (Small Auxin Up RNAs) (GI:3043536) [Raphanus sativus]	1.22075	0.661109007
5	A023935_01	At1g60030	xanthine/uracil permease family protein contains Pfam profile: PF00860 permease family	1.97095	2.468128384
6	A008551_01	At2g11090	Arabidopsis thaliana expressed protein , gypsy-like retrotransposon family has a	1.2404/0.72391	10.46914267

Table 3.1 (continued)

Gene	Oligo_ID	Locus	Description	Average fold chage in allo (allo vs mid-parents)	Average fold change in met1(met1 vs wt)
Upregulated in allo but downregulated in met1 microarray					
1	A005592_01	At5g25610	dehydration-responsive protein (RD22) identical to SP|Q08298 Dehydration-responsive protein RD22 precursor {Arabidopsis thaliana}	1.84595	0.503011183
2	A006813_01	AC006220		4.01475	0.34351552
3	A007362_01	At2g33830	dormancy/auxin associated family protein contains Pfam profile: PF05564 dormancy/auxin associated protein	1.56545	0.406764325
4	A007921_01	At2g14580	pathogenesis-related protein, putative	0.57947/3.847	0.258815271
5	A010805_01	At3g03870	expressed protein predicted using genefinder	0.7321/1.1538	0.630275329
6	A013100_01	At4g13030	expressed protein	1.23885	0.509544554
7	A014556_01	At4g04350	leucyl-tRNA synthetase, putative / leucine--tRNA ligase, putative similar to SP|P36430 Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS) {Bacillus subtilis}; contains Pfam profile PF00133: tRNA synthetases class I (I, L, M and V)	1.483	0.54828603
8	A016040_01	At5g55590	pectinesterase family protein contains Pfam profile: PF01095 pectinesterase	1.62435	0.404876263
9	A016636_01	At5g55450	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam protease inhibitor/seed storage/LTP family domain PF00234	0.87006/ 1.2701	0.508477471
10	A019450_01	At4g25490	DRE-binding protein (DREB1BCRT/CRE-binding factor 1 (CBF1) / transcriptional	0.74369/1.2881	0.542082589
11	A021217_01	At1g75830	plant defensin-fusion protein	2.3872	0.533938647
12	A021449_01	AB017065		1.53695	0.479153713
13	A022124_01	At2g14610	pathogenesis-related protein 1 (PR-1) identical to GB:M90508 SP|P33154	0.21271/1.3808	0.329183356
14	A024434_01	At1g60610	expressed protein	1.37745	0.560608489
15	A024532_01	At1g63000	Arabidopsis thaliana Unknown protein (F16P17.17) mRNA, complete cds	1.32115	0.587091298
downregulated in allo but upregulated in met1 microarray					
1	A002960_01	At1g72950	disease resistance protein (TIR-NBS class),putative domain signature TIR-NBS exists,	0.35984	1.633648427
2	A005691_01	At5g54145	Arabidopsis thaliana unknown protein mRNA, complete cds	0.35714	2.102644023
3	A006013_01	At3g01190	peroxidase 27 (PER27) (P27) (PRXR7) identical to SP|Q43735 Peroxidase 27 precursor (EC 1.11.1.7) (Atperox P27) (PRXR7) (ATP12a) {Arabidopsis thaliana}	0.629695	1.960219265
4	A006343_01	At2g38470	WRKY family transcription factor contains Pfam profile: PF03106 WRKY DNA -binding domain;	0.45568	3.10561654

Table 3.1 (continued)

Gene	Oligo_ID	Locus	Description	Average fold chage in allo (allo vs mid-parents)	Average fold change in met1(met1 vs wt)
5	A006418_01	At2g38720	microtubule associated protein (MAP65/ASE1) family protein low similarity to myosin [Schistosoma japonicum] GI:3941320; contains Pfam profile PF03999: Microtubule associated protein (MAP65/ASE1 family)	0.5234	2.035847082
6	A008551_01	At2g11090	Arabidopsis thaliana expressed protein , gypsy-like retrotransposon family has a	1.2404/0.72391	10.46914267
7	A009328_01	At3g56880	VQ motif-containing protein contains PF05678: VQ motif	0.50475	2.538497247
8	A009954_01	At3g50970	dehydrin xero2 (XERO2) / low-temperature-induced protein LTI30 (LTI30) identical to dehydrin Xero 2 (Low-temperature-induced protein LTI30) [Arabidopsis thaliana] SWISS-PROT:P42758	0.39832	1.702355173
9	A010341_01	At3g44260	Arabidopsis thaliana putative CCR4-associated factor 1	0.16402	4.245897747
10	A012770_01	At3g55980	zinc finger (CCCH-type) family protein contains Pfam domain, PF00642: Zinc finger C-x8-C-x5-C-x3-H type (and similar) and Pfam domain, PF00023: Ankyrin repeat	0.37791	3.65779051
11	A017902_01	At5g42050	expressed protein similar to gda-1 [Pisum sativum] GI:2765418	0.572255	2.548161018
12	A017938_01	At5g06320	harpin-induced family protein / HIN1 family	0.46462	2.706765416
13	A017969_01	At5g54490	calcium-binding EF-hand protein, putative similar to EF-hand Ca2+-binding protein CCD1 [Triticum aestivum] GI:9255753; contains INTERPRO:IPR002048 calcium-binding EF-hand domain	0.36535	1.814499145
14	A018090_01	At5g22250	CCR4-NOT transcription complex protein	0.186365	1.826066352
15	A018673_01	At5g01920	protein kinase family protein contains eukaryotic protein kinase domain, INTERPRO:IPR000719	0.36431	2.309265464
16	A019308_01	At5g11740	arabinogalactan-protein (AGP15) identical to gi|10880507|gb|AAG24283	0.569225	2.107049867
17	A019620_01	At3g60280	thaliana blue copper-binding protein III mRNA, complete cds	0.492675	2.03291641
18	A020065_01	At5g62680	proton-dependent oligopeptide transport (POT) family protein contains Pfam profile: PF00854 POT family	0.564965	3.238166792
19	A020113_01	AV567323		0.60846	2.324670522
20	A020214_01	At3g47960	proton-dependent oligopeptide transport (POT) family protein contains Pfam profile: PF00854 POT family	0.63115	2.039642359
21	A020390_01	At2g22500	Arabidopsis thaliana putative mitochondrial dicarboxylate carrier protein (F14M13.10/At2g22500) mRNA, complete cds	0.443375	2.563411339

Table 3.1 (continued)

	Oligo_ID	LOCUS	DESCRIPTION	average fold chage in allo (allo vs mid-parents)	average fold change in met1(met1 vs wt)
22	A020420_01	At1g55450	embryo-abundant protein-related similar to embryo-abundant protein GI:1350531 from [Picea glauca]	0.422245	2.056413001
23	A020461_01	At5g20230	plastocyanin-like domain-containing protein	0.0709615	1.91501183
24	A020482_01	At2g23810	senescence-associated family protein similar to senescence-associated protein 5 [Hemerocallis hybrid cultivar] gi|3551954|gb|AAC34855	0.51554	2.362349421
25	A020492_01	At1g18740	expressed protein	0.45589	2.098447143
26	A023602_01	AC007063		0.312025	5.215152173
27	A023938_01	AC004512		0.740775	6.306730322
28	A024108_01	At4g38550	expressed protein	0.62598	2.343735647
29	A024298_01	At4g30270	MERI-5 protein (MERI-5) (MERI5B) / endo-xyloglucan transferase / xyloglucan endo-1,4-beta-D-glucanase (SEN4) identical to endo-xyloglucan transferase gi:944810, SP|P24806 MERI-5 protein precursor (Endo-xyloglucan transferase) (Xyloglucan endo-1,4-beta-D-g	0.377605	2.436086272
30	A024510_01	At3g05490	rapid alkalinization factor (RALF) family protein similar to RALF precursor [Nicotiana tabacum] GI:16566316	0.614405	1.611279109
31	A024557_01	At1g27730	zinc finger (C2H2 type) family protein (ZAT10) / salt-tolerance zinc finger protein (STZ) identical to salt-tolerance zinc finger protein GB:CAA64820 GI:1565227 from [Arabidopsis thaliana]; contains Pfam domain, PF00096: Zinc finger, C2H2 type; identical	0.47249	2.901337099
32	A024989_01	At4g04570	protein kinase family protein contains Pfam domain PF00069: Protein kinase domain	0.733885	1.844338303
33	A025268_01	AC000106		0.49354	2.012372745
34	A025813_01	At1g57990	purine permease-related low similarity to purine permease	0.56552	1.816748717

The Reactivated Genes in *met1*-RNAi *A. suecica* Lines May Be Predominately Derived from the *A. thaliana* Genome

Although microarray technology is handy for large-scale study of differential gene expression profile, it can not distinguish the hybridization from different homoeologous expressions. Thus, in order to determine the progenitor origin of the differentially expressed homoeologous genes in *A. suecica*, single strand conformation polymorphism (SSCP) analysis was performed for the selected genes from Table 3.1 which are marked as shaded in the table. *A. thaliana* and *A. arenosa* share high sequence similarity, especially in the gene coding region, in allotetraploid *A. suecica*, *A. thaliana* homoeologs and *A. arenosa* homoeologs will very likely have the same size of RT-PCR amplicons and will not be separable on regular agarose or polyacrylamide gel electrophoresis. SSCP is a technique that separates DNA or RNA based on not only their fragment sizes but also their conformations, in other words, the sequence content. Homoeologous genes from *A. thaliana* and *A. arenosa* may develop a few nucleotides difference during evolution and the sequence polymorphisms will cause different migration rates in electrophoresis even though the amplicons are same in length.

Eleven genes selected from Table 3.1 were subjected to RT-PCR and SSCP analysis followed by silver staining, adapted from the online protocol from Hiroyuki Takeda's lab (National Institute of Genetics in Japan), to verify the microarray result as well as study the origin of progenitor for either up- or down-regulated homoeologs in synthetic allotetraploids and *met1*-RNAi *A. suecica* plants. In addition, two genes lo-

cated in heterochromatic region, At5g36180 (tested with Southern blots in Chapter II) and At1g36470 (a transposon-related gene), were selected from *met1* vs WT microarray data set, to study the origin or progenitor for the reactivated transcripts. Primers for RT-PCR were designed based on *A. thaliana* sequence exclusively from the exon region of the genes of interest to ensure the amplification of both *A. thaliana* and *A. arenosa*.

Figures 3.4-3.7 show the gel pictures of RT-PCR and SSCP results for the genes. Most of RT-PCR results correlate with microarray data, confirming the reliability of microarray analysis. The expression changes and progenitor origins for the reactivations or suppressions of homoeologs detected by SSCP are summarized in Table 3.2. In the 11 genes that have been examined, major reactivations in *met1* seem to derive from *A. thaliana* homoeologs (5 out of 7 genes) and repressions are derived mainly from *A. arenosa* homoeologs (all 4 genes), suggesting an unequal sensitivity of progenitor genomes in response to DNA methylation regulation in allopolyploid. Moreover, for the two genes only significant in met1 vs WT microarray analysis, At5g36180 and At1g36470, major reactivations also seem to be derived from *A. thaliana* genome (Figure 3.8). Novel expressions can also be observed for these two genes because of relieving of dense DNA methylation in *met1*-RNAi *A. suecica*.

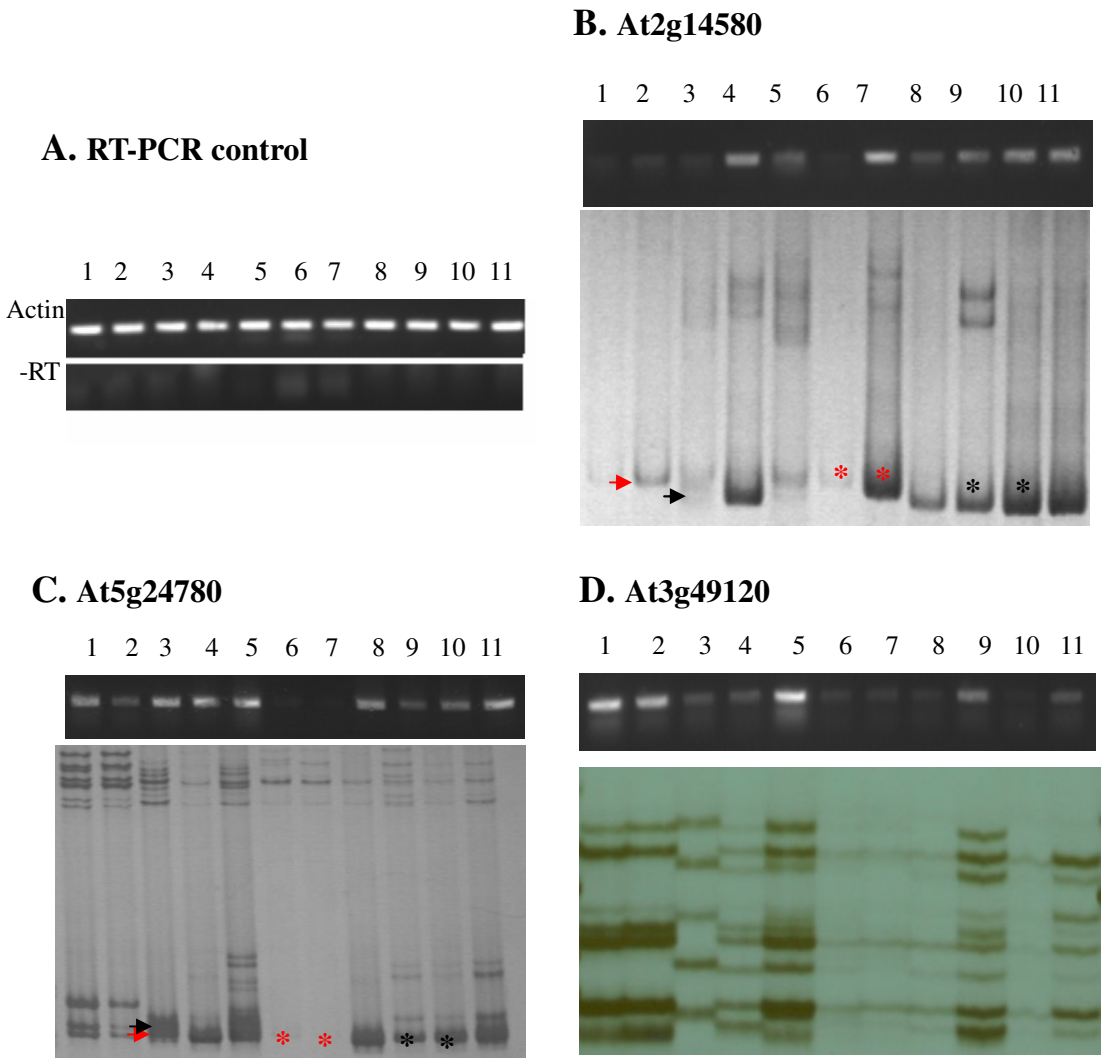


Figure 3.4 RT-PCR and SSCP for the genes that are down-regulated in allo733 and/or allo738, and *met1*.

(A) Actin control and –RT control for the RNA samples. **(B) - (D)** RT PCR results and SSCP gel for At2g14580, At5g24780, At3g49120 respectively. At2g14580 is actually down-regulated in allo733 and up-regulated in allo738. Asterisks show the reactivations and arrows show the progenitor origins if only one homoeolog is expressed. If both homoeologs are expressed, no symbol will be shown. Red symbols are for synthetic allo lines and black symbols are for *met1* line. If overlapped, two arrows symbols will be given side by side. **(D)** Expression of At3g49120 in met1 doesn't correlate with microarray result.
Lanes: 1-*At*2, 2-*At*4, 3-*Aa*, 4-*As*9502, 5-alloF1, 6-allo733, 7-allo738, 8-9502-*RdRP*, 9-9502-*met1*-1, 10-9502-*met1*-12, 11-9502-*ddm1*-176.

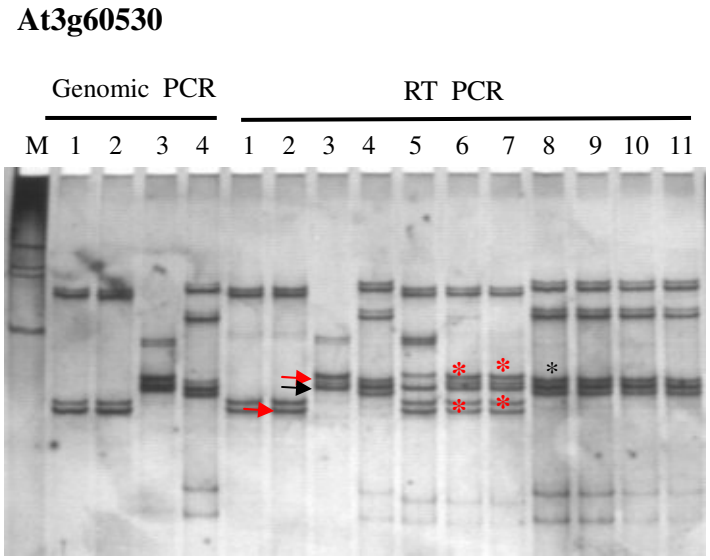


Figure 3.5 SSCP for At3g60530 which is up-regulated in allo733, allo738, and *met1*.

This gene is located about 1Mb to telomere heterochromatin. Genomic PCR lanes are shown by side as control.
Lanes: 1-*At*2, 2-*At*4, 3-*Aa*, 4-*As*9502, 5-alloF1, 6-allo733, 7-allo738, 8-9502-*RdRP*, 9-9502-*met1*-1, 10-9502-*met1*-12, 11-9502-*ddm1*-176.

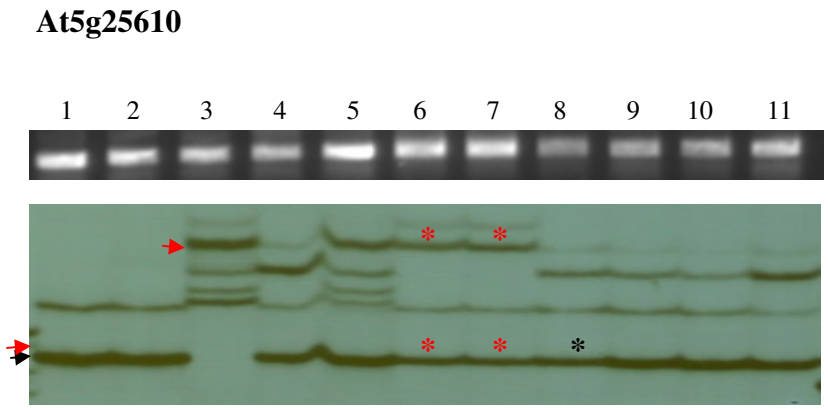


Figure 3.6 RT-PCR and SSCP for At5g25610 which is up-regulated in allo733 and allo738, but down-regulated in *met1*.

Lanes: 1-*At*2, 2-*At*4, 3-*Aa*, 4-*As*9502, 5-alloF1, 6-allo733, 7-allo738, 8-9502-*RdRP*, 9-9502-*met1*-1, 10-9502-*met1*-12, 11-9502-*ddm1*-176.

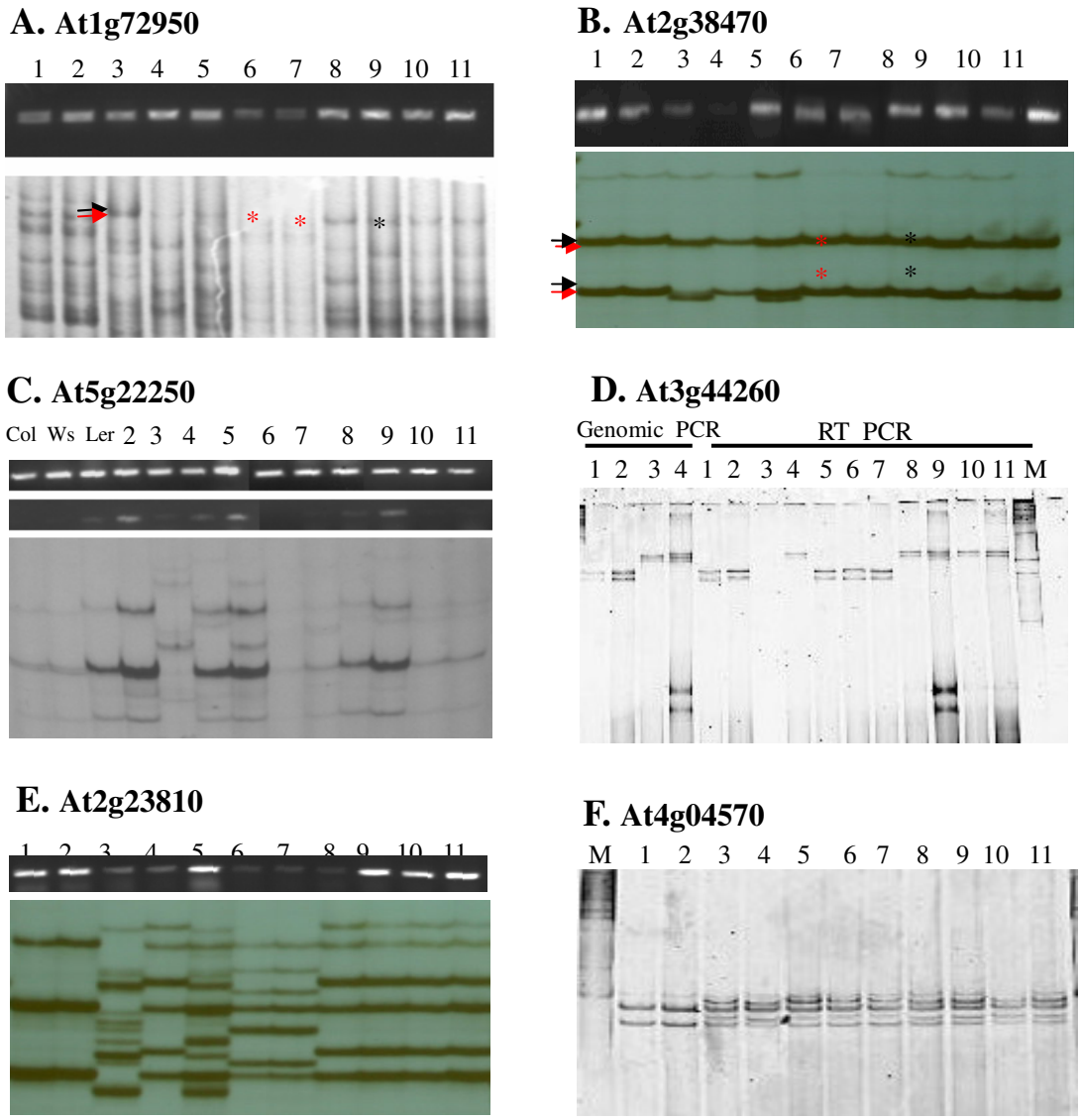


Figure 3.7 RT-PCR and SSCP for the genes that are down-regulated in allo733 and allo738, but up-regulated in *met1*.

(A) - (F) Gel pictures for the selected genes, with the locus IDs shows on top of the individual pictures. **(F)** At4g04570 is a gene with around 2Mb distance to centromere heterochromatin.
Lanes: Col-At2 Columbia ecotype, Ws-At2 Ws ecotype, Ler-At2 Lansberg ecotype, 1-*At*2, 2-*At*4, 3-*Aa*, 4-*As*9502, 5-alloF1, 6-allo733, 7-allo738, 8-9502-*RdRP*, 9-9502-*met1*-1, 10-9502-*met1*-12, 11-9502-*ddm1*-176.

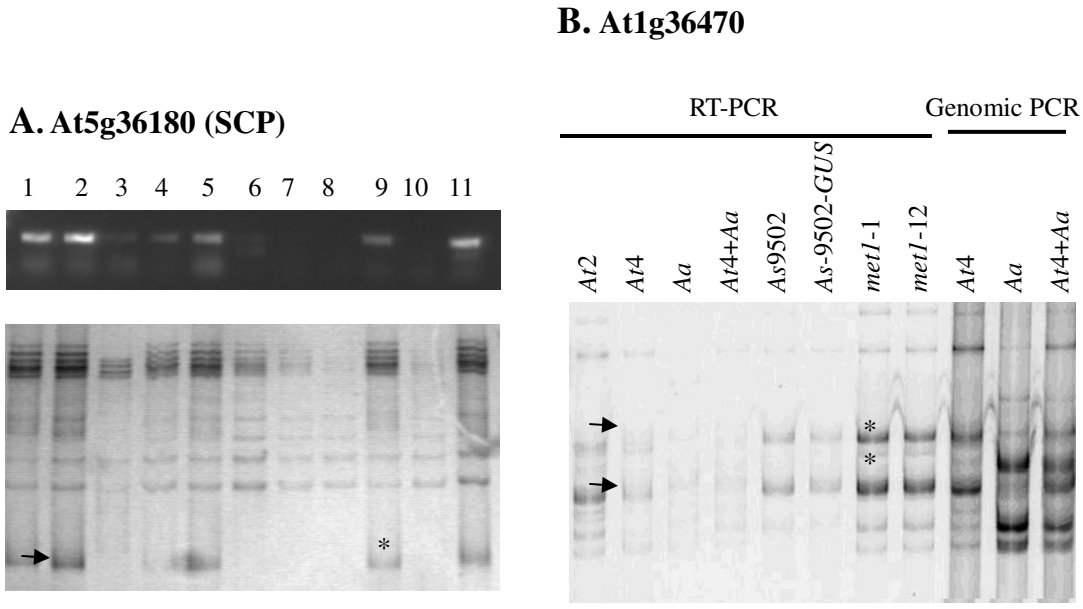


Figure 3.8 RT-PCR and SSCP for two genes only significant in *met1* microarray experiment

(A) Gel pictures for At5g36180, which is the *SCP* gene analyzed in Chapter I. **(B)** SSCP picture for At1g36470, which is a Tnp2/En/Spm CACTA-element located close to centromeric region.
Lanes: 1-*At*2, 2-*At*4, 3-*Aa*, 4-*As*9502, 5-alloF1, 6-allo733, 7-allo738, 8-9502-*RdRP*, 9-9502-*met1*-1, 10-9502-*met1*-12, 11-9502-*ddm1*-176. 9502-GUS is the As9502 plant transpired with GUS reporter as transformation control for 9502-met1. *At*4 + *Aa* is the equal amount mix of *At*4 and *Aa* as template to show no amplification bias here.

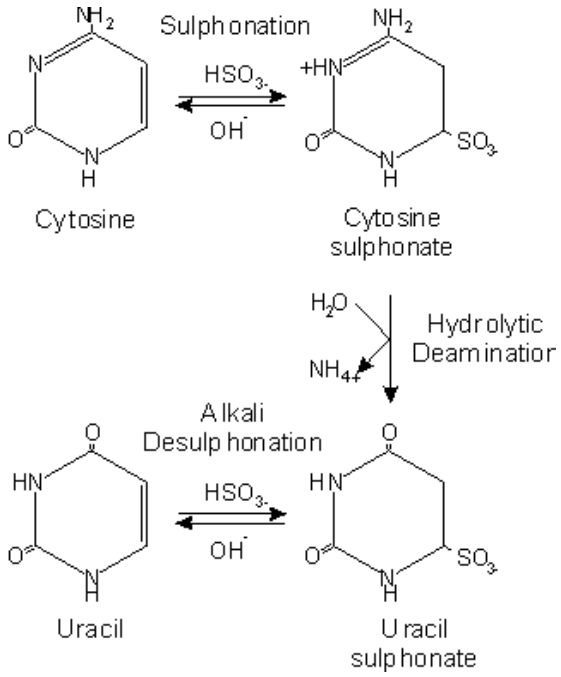
Table 3.2 Summary of SSCP analysis.

Sub-group	Locus	Description	Genome-specific expression change based on SSCP result
Downregulated in both allo and met1 microarray(19)	At2g14580	pathogenesis-related protein, putative	*Aa* homoeolog is expressed, up-regulated in allo738, and down-regulated in allo733 and *met1*-1
	At5g24780	vegetative storage protein 1 (VSP1) identical to SP|O49195 Vegetative storage protein 1 precursor {Arabidopsis thaliana}; contains Pfam profile PF03767: HAD superfamily (subfamily IIIB) phosphatase	*At* and *Aa* homoeologs are both expressed in *As*9502; *Aa* homoeolog is expressed and down-regulated in allos and down-regulated in *met1*-1
	At3g49120	peroxidase, putative identical to peroxidase [Arabidopsis thaliana] gi|405611|emb|CAA50677	*At* and *Aa* homoeologs are both expressed, down-regulated in allos, and up-regulated in *met*1-1
Upregulated in both allo and met1 microarray (6)	At3g60530	zinc finger (GATA type) family protein	*At* and *Aa* homoeologs are both expressed; *Aa* homoeolog is major affected; this gene is also up-regulated in 9502-*RdRp* line
Upregulated in allo but downregulated in met1 microarray (15)	At5g25610	dehydration-responsive protein (RD22) identical to SP|Q08298 Dehydration-responsive protein RD22 precursor {Arabidopsis thaliana}	*At* and *Aa* homoeologs are both expressed; No dramatic changes are detected in allos; *Aa* homoeolog is downregulated in *met1*-1
downregulated in allo but upregulated in met1 microarray (34)	At1g72950	disease resistance protein (TIR-NBS class),putative domain signature TIR-NBS exists,	*Aa* homoeolog is expressed, down-regulated in allos and up-regulated in *met1*-1
	At2g38470	WRKY family transcription factor contains Pfam profile: PF03106 WRKY DNA -binding domain;	*At* homoeolog is expressed, down-regulated in allos and up-regulated in *met1*-1
	At5g22250	CCR4-NOT transcription complex protein	*At* homoeolog is expressed, down-regulated in allos and up-regulated in *met1*-1
	At3g44260	Arabidopsis thaliana putative CCR4-associated factor 1	*At* homoeolog is expressed and down-regulated in allo lines*; Aa homoeolog* is expressed and up-regulated in *met1 lines.*
	At2g23810	senescence-associated family protein similar to senescence-associated protein 5 [Hemerocallis hybrid cultivar] gi|3551954|gb|AAC34855	*At* and *Aa* homoeologs both expressed; *At* homoeolog is down-regulated in allos and up-regulated in *met1*-1
	At4g04570	protein kinase family protein contains Pfam domain PF00069: Protein kinase domain	*At* and *Aa* homoeologs are both expressed, down-regulated in allos and up-regulated in met1-1
Only detected in met1 microarray	At5g36180	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MAB16, serine carboxypeptidase (SCP)	*At* and *Aa* homoeologs are both expressed; *At* homoeolog is up-regulated in *met1*-1
	At1g36470	Tnp2/En/Spm CACTA-element transposon	*At* and *Aa* homoeologs are both expressed; major reactivations are derived from *At*

Up-regulation of *A. thaliana* Homoeologs May Be Associated with Decease of DNA Methylation at *A. thaliana* Specific Promoters

Although we assume that in *met1*-RNAi *A. suecica* lines the up-regulation of a certain gene should correlate to DNA hypo-methylation of this gene, or at least at promoter region, evidence would still be needed to prove a direct relation between them. Hence, bisulfite sequencing was performed on the promoter regions of At2g23810 and At2g38470, following the protocol published online (www.protocol-online.org). This method allows a precise analysis of methylation in a certain region by converting all unmethylated cytosines into uracils, while methylated cytosines remain unchanged, thus the methylated and unmethylated cytosines can be distinguished after sequencing (Hayatsu et al., 1970, Figure 3.9). Because we didn't have leave tissue or genomic DNA from allo733 or allo738 lines, allo701 genomic DNA was used in these experiments, which is another individually-evolved line of resynthesized allotetraploid, in the fourth selfing generation. 500bp upstream of the transcriptional start sites for genes of interest were chosen to be the target regions for promoter analysis. Since *A. thaliana* and *A. arenosa* would be more diverged at promoter regions, PCR primers were designed only to detect the methylation status of *thaliana*-specific promoters.

A



B

After bisulfite treatment:

C → T

$^{m}C \rightarrow C$

Figure 3.9 Bisulfite treatment.

(A) is the diagram showing the procedures of chemical changes of cytosine upon bisulfite treatment. (B) shows after bisulfite treatment, unmethylated cytosines will turn into thymines, while methylated cytosines remain unchanged.

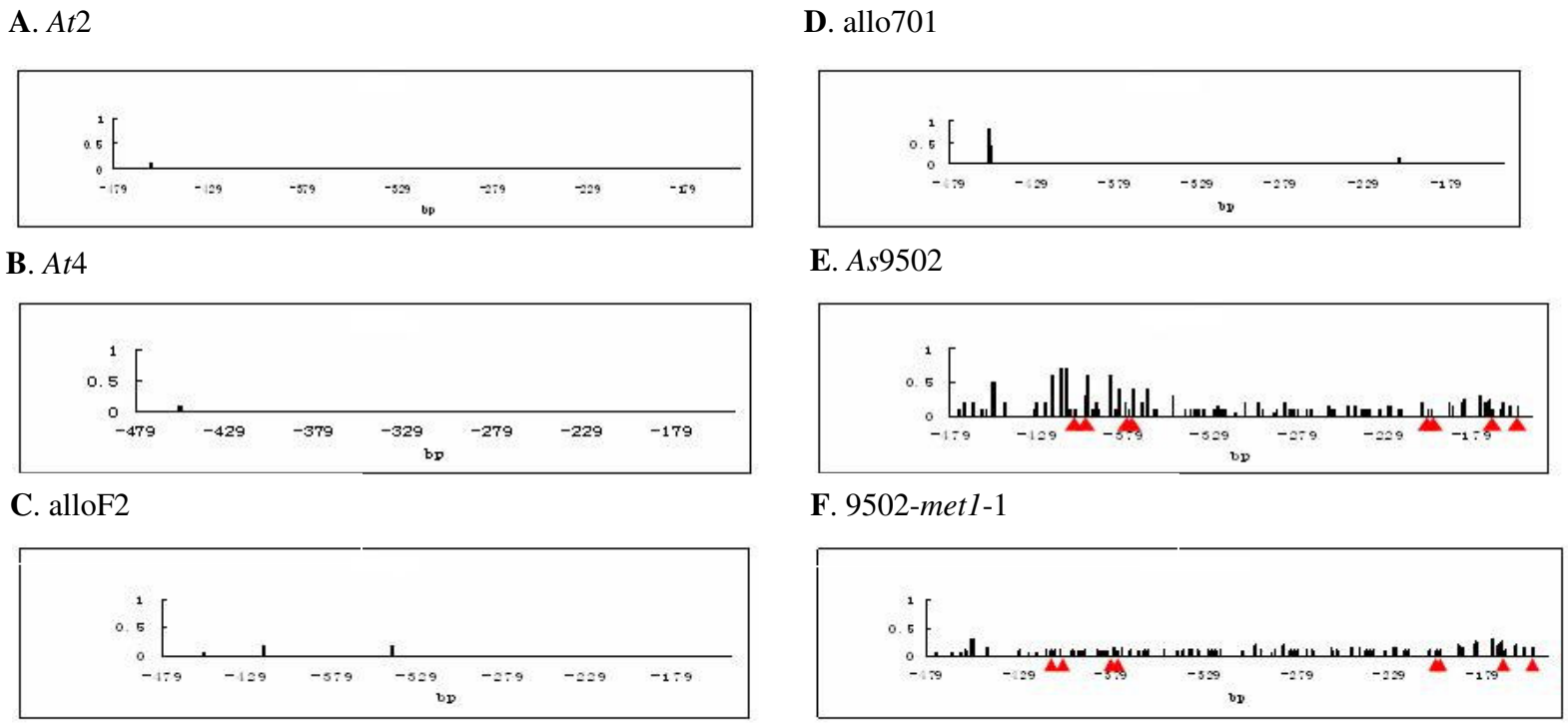


Figure 3.10 Methylation at *thaliana*-specific promoter for At2g23810 in different plants.

Based on bisulfite sequencing analysis results, **(A)-(F)** show the methylation status of the *A. thaliana*-specific promoter for gene At2g23810 in different species. Red arrow heads show CG methylation sites. Others are non-CG sites

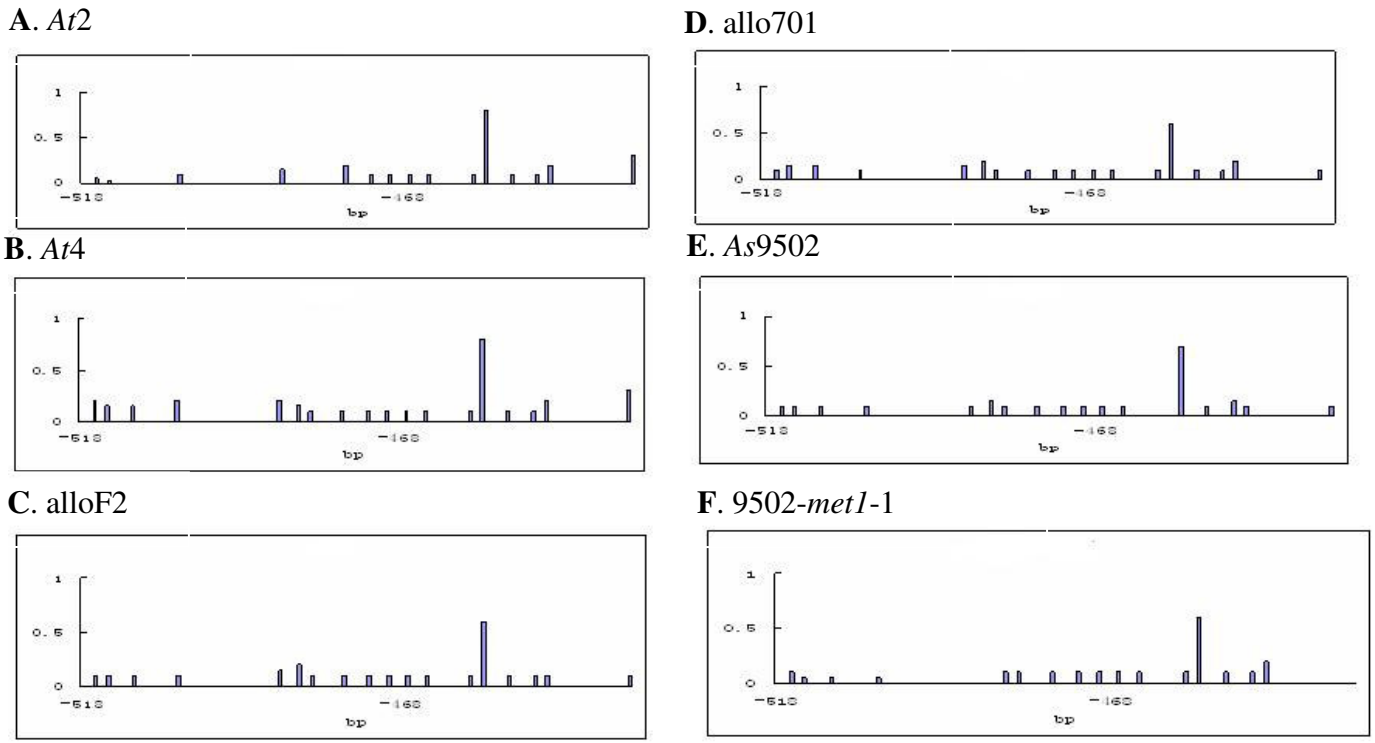


Figure 3.11 Methylation at *thaliana*-specific promoter for At2g38470 in different plants.

(A)-(F) show the methylation status of the *A. thaliana*-specific promoter for gene At2g38470 in different species

After bisulfite reaction, nested PCR products were directly sequenced without cloning. It is also suggested in the protocol that direct sequencing of the PCR amplicon is the best method of determining the degree of methylation in the population of amplicons. The percentage methylation was measure by the ratio of sequencing signal peak height of Cs to Ts at the same site. Methylation at every C site in the promoter regions is shown in Figure 3.10 and 3.11 for gene At2g23810 and At2g38470 respectively. Based on information from Figure 3.10, we can tell that promoter for At2g23810 is completely lack in methylation in *A. thaliana* diploid and tetraploid, *A. arenosa*, as well as allo701, which corresponds to the high expression of *thaliana* homoeolog in these species. In natural *A. suecica* 9502, much increased methylation can be observed in the whole promoter region at both CG and non-CG sites, especially at -300bp~-450bp region, whereas in *met1*-RNAi *A. suecica*, an overall of decrease of methylation occurs in promoter region, at both CG and non-CG sites. It corresponds with low level of *thaliana*-derived transcription in *A. suecica*9502 and its reactivation in *met1*-RNAi line. Hence the up-regulation of *thaliana* homoeolog is associated with decease of DNA methylation at *thaliana*-specific promoter. For gene At2g38470, no obvious changes of methylation status can be observed, due to the distance of designed fragment from the actual promoter region (Figure 3.11).

DISCUSSION

By mapping the genes identified by met1 vs WT microarray experiment, we understand that DNA methylation has a genome-wide and balanced effect on gene expression for euchromatic gene, as well as a condensed major suppressing effect on centromeric and other heterochromatic genes.

Comparative analysis was done here to combine the data from three microarray experiments together, in order to identify the DNA methylation-affected and polyploidization-associated genes. Microarray experiments and statistic analysis provide advantages in searching for the interesting genes and define the target ones in a huge gene pool. Many different mechanisms may play and interact with each other in the complicated process of polyploidization and evolution, among which DNA methylation is the one we are focus on. It is noticeable that, the up- and down-regulated genes each count about half of the total affected genes in *met1*-RNAi *A. suecica* compared to wild type plants, on the contrast of dominance of repression (>65% of total) presented in synthetic allotetraploid compared to the mid-parent (Wang et al, 2005). SSCP analysis provides a feasible way to study the expression from homoeologous alleles in allopolyploid. By using SSCP we are able to identify that most of the reactivations in *met1*-RNAi line are derived from *A. thaliana* parent. It hints that in polyploidization process, methylation may unequally influence the genomes from two progenitors. In this study, *A. thaliana* genome is more suppressed by DNA methylation in the allotetraploid, and in *met1*-RNAi plants, because of relieving of DNA methylation, many genes from *A. thaliana*

get reactivated. Moreover, bisulfite sequencing analysis reveals that the reactivations of *A. thaliana*-specific transcripts may associate with demethylation of *thaliana*-specific promoter for the genes in *met1*-RNAi *A. suecica*. More bisulfite sequencing data are needed for a more convincing conclusion. On the other hand, *A. arenosa* specific promoters can be cloned by Genome Walking technology and bisulfite sequencing for *arenosa*-specific promoters can be done to add strength to the conclusion.

MATERIALS AND METHODS

SSCP and Silver Staining

Protocol is modified from the online protocol from Hiroyuki Takeda's lab (National Institute of Genetics in Japan). In brief, 12% polyacrylamide gel is made containing 0.5XTBE, 0.5% of 10% APS, 0.1% of TEMED and ddH_2O. Allow the gel to be solidified and run the gel in 0.5x TBE for appropriate time for good separation. We use 200V for 10 hr running for a 20x20cm gel.

Fix the polyacrylamide gel with 10% acetic acid treatment for 30min, followed by washing of ddH_2O for 3 times, 5min each time. Soak the gel in 1% (w/v) $AgNo_3$ staining buffer, containing 0.15% of formaldehyde, shake for 30min. Rinse the gel with ddH_2O for exactly 10sec. Develop the staining by soaking the gel in 6% (w/v) cold Na_2co_3 buffer with 0.15% of formaldehyde and 0.24% of $Na_2S_2O_3$ and shake. Bands will be visible after around 10min. Don't over stain. Stop the staining reaction by adding the

acetic acid.

Bisulfite Sequencing

Exactly following the online protocol available at:

http://www.methods.info/Methods/DNA_methylation/Bisulphite_sequencing.html

RT-PCR and Primers

RT-PCR was conducted exactly following the user instruction of Invitrogen SuperScript III One-Step RT-PCR kit.

Primers for SSCP:

For Actin: 5'- GCACACTGGTGTCATGGTTGGT-3' (forward primer) and 5'-AACGGCCTGAATGGCAACATAC -3' (reverse primer).

For At2g14580: 5'- TGAGTGTAGTGACCGCAAACTC -3' (forward primer) and 5'-GTTCCCTTGAAGGCTCAAGACA-3' (reverse primer).

For At5g24780**:** 5'- AGGGAATACTAGAGAGGAGAGT -3' (forward primer) and 5'-GAACACCATCTTTGGGAACGAA -3' (reverse primer).

For At2g38470**:** 5'- GACTCGTCCAAACAATCAAGCT-3' (forward primer) and 5'-GATCCAAACTGGCATTGTACAC-3' (reverse primer).

For At3g44260**:** 5'- CCTCATCCTCTGAAACGCATGC -3' (forward primer) and 5'-CGGCGTGATTTTCAAATCCGAT -3' (reverse primer).

For At3g49120**:** 5'- TGCCTCCAGCTTTGACCCGTT-3' (forward primer) and 5'-AGCGATACGAGGGTCCGACCT-3' (reverse primer).

For At5g25610**:** 5'-ACCAAACACTCCCATTCCCAA -3' (forward primer) and 5'-ACCGCCTTTACCGACTCCTAC -3' (reverse primer).

For At1g72950**:** 5'- GACACTCGCCGCAGCTTCATC -3' (forward primer) and 5'-CAGCGACTTCTCCGATCTGTC -3' (reverse primer).

For At2g23810**:** 5'- GCTGGTGAAGCTATTGAAGGA -3' (forward primer) and 5'-TGTGTTCCCGTTGTGTTCTTG -3' (reverse primer).

For At5g36180**:** 5'- CCGTTGTGCGAAACTGAAACTC-3' (forward primer) and 5'-TGACCATCCATGGCCTCCAGTC -3' (reverse primer).

Primers for bisulfite sequencing

For At2g23810**:** 5'- GGAAGTTAAAAAGTTATTTTTTTAAATGAA -3' (forward primer), 5'- AACTTAACCAAATTCCACCAACTAA -3' (reverse primer), 5'-GAAGTTAAAAAGTTATTTTTTTAAATGAAT-3' (nested forward primer) and 5'-AATTCCACCAACTAAAATTAAAATC-3' (nested reverse primer).

For At2g38470**:** 5'- AGATTTTGTGGTTTTGATTTTTAAA -3' (forward primer), 5'-AAATAATCTTATATTCAATTAATCTATTCA -3' (reverse primer), 5'-ATTGGTTAGATTTTGTGGTTTTGAT-3' (nested forward primer) and 5'-AATCTATTCATAATAAAATTAAATCTCCTA -3' (nested reverse primer).

CHAPTER IV

SUMMARY AND DISCUSSION

Polyploidization is a prominent phenomenon among plants and it favors plants by increasing their adaptabilities to different circumstances. There are several mechanisms involved in the gene regulation in polyploids and polyploidization. Here we focus on the impact of DNA methylation on chromatin structure and gene expression in synthetic allotetraploid and natural line *A. suecica*, by studying *met1*-RNAi *A. suecica* transgenic plants. Transcriptional reactivations were found to be coupled with decrease of DNA methylation for some genes in *met1* plants, for instance the *SCP* gene. Two transposons, At4g08010 and At1g44070, showed transpositional reactivations as well as transcriptional reactivations. Intriguingly, we observe preferential demethylation occurring in *arenosa*-specific centromeres in *met1*-RNAi *A. suecica*, which is associated with centromere-derived siRNA accumulation. These observations suggest that DNA methylation serves as a chromatin remodeler to alter gene expression by changing the DNA methylation status of the region in which they are located.

Chromosomal mapping of the genes which have altered expression profiles in *met1*-RNAi *A. suecica* based on microarray analysis reveals that DNA methylation has a genome-wide and sporadic influence on euchromatic regions and a major repressive effect on heterochromatic regions. Transposons and other genes located in the heterochromatic regions undergo a huge increase of transcription as the result of hypo-methylation.

Comparative analysis between three different microarray data sets identified a gene list of methylation-affected and polyploidization-related genes. Further study reveals that the reactivations of those genes in *met1* plants may mainly derive from *A. thaliana* genome due to decrease of DNA methylation in *thaliana*-specific promoters. This supports the hypothesis that *thaliana* genome is mostly repressed in the allotetraploid by methylation and other mechanisms, although we don't know why *thaliana* genome is more susceptible to DNA methylation changes rather than *arenosa* genome. But this conclusion is almost opposite of the previous observation of preferential demethylation for *arenosa*-specific centromeres, in other words, *arenosa* genome may be more sensitive to DNA methylation changes at least in centromere region. Thus, genes located in euchromatic regions and heterochromatic regions may respond differently upon the shaping force of DNA methylation in polyploidization process.

REFERENCES

Allfrey, V.G., Faulkner, R., and Mirsky, A.E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc Natl Acad Sci U S A **51,** 786-794.

Antequera, F., Boyes, J., and Bird, A. (1990). High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. Cell **62,** 503-514.

Aufsatz, W., Mette, M.F., Matzke, A.J., and Matzke, M. (2004). The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides. Plant Mol. Biol. **54**, 793-804.

Bartee, L., Malagnac, F., and Bender, J. (2001). Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes Dev. **15**, 1753-1758.

Bennett, M.D., and Leitch, I.J. (2005). Nuclear DNA amounts in angiosperms: progress, problems and prospects. Ann. Bot. (Lond.) **95**, 45-90.

Berger, F. (2004). Plant sciences. Imprinting--a green variation. Science **5657**, 483-485.

Bernstein, E., Denli, A.M., and Hannon, G.J. (2001). The rest is silence. RNA **11**, 1509-1521.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. **16**, 6-21.

Cao, X., Aufsatz, W., Zilberman, D., Mette, M.F., Huang, M.S., Matzke, M., and Jacobsen, S.E. (2003). Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. Curr. Biol. **13**, 2212-2217.

Cao, X., and Jacobsen, S.E. (2002). Role of the arabidopsis DRM methyltransferases in

de novo DNA methylation and gene silencing. Curr Biol. **12**, 1138-1144.

Cao, X., Springer, N.M., Muszynski, M.G., Phillips, R.L., Kaeppler. S., and Jacobsen. S.E. (2000). Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. Proc Natl. Acad. Sci. USA **97**, 4979-4984.

Chan, S.W., Henderson, I.R., and Jacobsen, S.E. (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat. Rev. Genet. **6**, 351-360.

Chen, T., and Li, E. (2004). Structure and function of eukaryotic DNA methyltransferases. Curr. Top. Dev. Biol. **60**, 55-89.

Chen, Z.J., and Pikaard, C.S. (1997). Transcriptional analysis of nucleolar dominance in polyploid plants: biased expression/silencing of progenitor rRNA genes is developmentally regulated in Brassica. Proc. Natl. Acad. Sci. USA **94**, 3442-3447.

Colot, V., and Rossignol, J.L. (1999). Eukaryotic DNA methylation as an evolutionary device. Bioessays. **21**, 402-411.

Comai, L. (2000b). Genetic and epigenetic interactions in allopolyploid plants. Plant Mol. Biol. **43**, 387-399.

Comai, L., Tyagi, A.P., Winter, K., Holmes-Davis, R., Reynolds, S.H., Stevens, Y., and Byers, B. (2000a). Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. Plant Cell **12**, 1551-1568.

Denti, M.A., Boutla, A., Tsagris, M., and Tabler, M. (2004). Short interfering RNAs specific for potato spindle tuber viroid are found in the cytoplasm but not in the nucleus. Plant J. **5**, 762-769.

Eden, S., Hashimshony, T., Keshet, I., Cedar, H., and Thorne, A.W. (1998). DNA methylation models histone acetylation. Nature **394,** 842.

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by

21- and 22-nucleotide RNAs. Genes Dev**. 2**, 188-200.
Feldman, M., and Levy, A.A. (2005). Allopolyploidy--a shaping force in the evolution of wheat genomes. Cytogenet Genome Res. **1-3**, 250-8.

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

Gendrel, A.V., Lippman, Z., Yordan, C., Colot, V., and Martienssen, R.A. (2002). Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. Science **5588**, 1871-1873.

Grini, P.E., Jurgens, G., and Hulskamp, M. (2002). Embryo and endosperm development is disrupted in the female gametophytic capulet mutants of Arabidopsis. Genetics **4**, 1911-1925.

Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell **1**, 23-34.

Haaf, T. (1995). The effects of 5-azacytidine and 5-azadeoxycytidine on chromosome structure and function: implications for methylation-associated cellular processes. Pharmacol Ther. **1**, 19-46.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science **5441**, 950-952.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J. **21:** 4671–4679

Hassig, C.A., and Schreiber, S.L. (1997). Nuclear histone acetylases and deacetylases and transcriptional regulation: HATs off to HDACs. Curr. Opin. Chem Biol. **3**, 300-308.

Hayatsu, H., Wataya, Y., Kai, K., and Iida, S. (1970). Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. Biochemistry **9**, 2858-2865.

Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science **5531**, 834-838.

Imamura, T., Ohgane, J., Ito, S., Ogawa, T., Hattori, N., Tanaka, S., and Shiota, K. (2001). CpG island of rat sphingosine kinase-1 gene: tissue-dependent DNA methylation status and multiple alternative first exons. Genomics **76**, 117-125.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science **293,** 1074-1080.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature **416**, 556-60.

Johnson, L., Cao, X., and Jacobsen, S. (2002). Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Curr. Biol. **12**, 1360-1367.

Jones, L., Ratcliff, F., and Baulcombe, D.C. (2001). RNA-directed transcriptional gene silencing in plants can be inherited independently of the RNA trigger and requires Met1 for maintenance. Curr. Biol**. 11**, 747-757.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet **19,** 187-191.

Jones, P.L., Wade, P.A., and Wolffe, A.P. (2001). Purification of the MeCP2/Histone Deacetylase Complex from Xenopus laevis. Methods Mol. Biol. **181**, 297-307.

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.

Kashkush, K., Feldman, M., and Levy, A.A. (2002). Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics **160**, 1651-1659.

Kashkush, K., Feldman, M., and Levy, A.A. (2003). Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat. Genet. **33**, 102-106.

Kato, M., Miura, A., Bender, J., Jacobsen, S.E., and Kakutani, T. Role of CG and non-CG methylation in immobilization of transposons in Arabidopsis. Curr. Biol. **13**, 421-426.

Kessler, C., and Manta, V. (1990). Specificity of restriction endonucleases and DNA modification methyltransferases a review (Edition 3). Gene **92**, 1-248.

Ko, Y.G., Nishino, K., Hattori, N., Arai, Y., Tanaka, S., and Shiota, K. (2005). Stage-by-stage change in DNA methylation status of Dnmt1 locus during mouse early development. J. Biol. Chem. **280,** 9627-9634.

Koch, M., Haubold, B., Mitchell-Olds, T. (2001). Molecular systematics of the Brassicaceae: evidence from coding plastidic matK and nuclear Chs sequences. Am. J. Bot. **88**, 534-544.

Lee, H.S., and Chen, Z.J. (2001). Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. Proc. Natl. Acad. Sci. USA **98:** 6753–6758

Lee, H.S., Wang, J., Tian, L., Jiang, H., Black, M.A., Madlung, A., Watson, B., Lukens, L., Pires, J.C., Wang, J.J., Comai, L., Osborn, T.C., Doerge, R.W. and Chen, Z.J. (2004). Sensitivity of 70-mer oligonucleotides and cDNAs for microarray analysis of gene expression in *Arabidopsis* and *Brasssica*. Plant Biotechnol. J. **2**, 45–57

Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. Science **5543**, 862-864.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell **69**, 915-926.

Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., and Jacobsen, S.E. (2001). Requirement of CHROMOMETHYLASE3 for mainte-

nance of CpXpG methylation. Science 292, 2077-2080.
Lippman, Z. and Martienssen, R. (2004). The role of RNA interference in heterochromatic silencing. Nature **431**, 364-370.

Lippman, Z., Gendrel, AV., Black, M., Vaughn, M.W, Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., Carrington, J.C., Doerge, R.W., Colot, V., Martienssen, R. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature **430**, 471–476.

Liu, B., Vega, J.M., and Feldman, M. (1998). Rapid genomic changes in newly synthesized amphiploids of Triticum and Aegilops. II. Changes in low-copy coding DNA sequences. Genome **41**, 535-542.

Madlung, A., Masuelli, R.W., Watson, B., Reynolds, S.H., Davison, J., and Comai, L. (2002). Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic Arabidopsis allotetraploids. Plant Physiol. **129**, 733-746.

Malagnac, F., Bartee, L., and Bender, J. (2002). An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. EMBO J. **24**, 6842-6852.

Mann, M. B., and Smith, H. O. (1977). Specificity of *Hpa II* and *Hae III* DNA methylases. Nucleic Acids Res. **4,** 4211–4221.

Martienssen, R.A., and Colot, V. (2001). DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science **5532**, 1070-1074.

McClintock, B. (1965). The control of gene action in maize. Brookhaven Symp Biol **18**, 162–184.

Mette, M.F., Aufsatz, W., van der Winden, J., Matzke, M.A., and Matzke, A.J. (2000). Transcriptional silencing and promoter methylation triggered by double-stranded RNA. EMBO J. **19**, 5194-201.

Mishra, K.K., and Handa, A.K. (1998). Post-transcriptional silencing of pectin methylesterase gene in transgenic tomato fruits results from impaired pre-mRNA processing. Plant J. **14**, 583–592.

Morel, J.B., Mourrain, P., Beclin, C., and Vaucheret, H. (2000). DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. Curr. Biol. **10**, 1591-1594.

Nagaki, K., Talbert, P.B., Zhong, C.X., Dawe, R.K., Henikoff, S., and Jiang, J. (2003). Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres. Genetics **3**, 1221-1225.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature **393,** 386-389.

Osborn, T.C., Butrulle, D.V., Sharpe, A.G., Pickering, K.J., Parkin, I.A., Parker, J.S., and Lydiate, D.J. (2003b). Detection and effects of a homeologous reciprocal transposition in Brassica napus. Genetics **165**, 1569-1577.

Osborn, T.C., Pires, J.C., Birchler, J.A., Auger, D.L., Chen, Z.J., Lee, H.S., Comai, L., Madlung, A., Doerge, R.W., Colot, V., and Martienssen, R.A. (2003a). Understanding mechanisms of novel gene expression in polyploids. Trends Genet. **3**, 141-147.

Otto, S.P. (2000). Detecting the form of selection from DNA sequence data. Trends Genet. **12**, 526-9.

Otto, S.P., and Whitton, J. (2000). Polyploid incidence and evolution. Annu. Rev. Genet. **34**, 401-437.

Pikaard, C.S. (2000a). The epigenetics of nucleolar dominance. Trends Genet. **16**, 495-500.

Pikaard C.S. (2000b). Nucleolar dominance: uniparental gene silencing on a multi-megabase scale in genetic hybrids. Plant Mol. Biol. **43**, 163-177.
Probst, A.V., Fransz, P.F., Paszkowski, J., and Scheid, O.M. (2003). Two means of transcriptional reactivation within heterochromatin. Plant J. **4**, 743-9.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B.D., Sun, Z.W., Schmid, M., Opravil,

S., Mechtler, K., Ponting, C.P., Allis, C.D., and Jenuwein, T. (2000). Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature **6796**, 593-599.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature **6772**, 901-906.

Ruiz-Garcia, L., Cervera, M.T., Martinez-Zapater, J.M. (2005). DNA methylation increases throughout Arabidopsis development. Planta [Epub ahead of print]

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

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J. **5**, 1121-1131.

Selker, EU., Freitag, M., Kothe, G.O,, Margolin, B.S., Rountree, M.R., Allis, C.D., and Tamaru, H. (2002). Induction and maintenance of nonsymmetrical DNA methylation in Neurospora. Proc. Natl. Acad. Sci. USA **99, Suppl 4,** 16485-16490.

Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., Levy, A.A. (2001). Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell **13**, 1749-1759.

Shiota, K., Kogo, Y., Ohgane, J., Imamura, T., Urano, A., Nishino, K., Tanaka, S., and Hattori, N. (2002). Epigenetic marks by DNA methylation specific to stem, germ and somatic cells in mice. Genes. Cells **7**, 961-969.

Sijen, T., Vijn, I., Rebocho, A., van Blokland, R., Roelofs, D., Mol, J.N., and Kooter, J.M. (2001). Transcriptional and posttranscriptional gene silencing are mechanistically related. Curr. Biol. **11,** 436-440.

Sijen, T. and Plasterk, R.H. (2003). Transposon silencing in the *Caenorhabditis*

elegans germ line by natural RNAi. Nature **426**, 310–314

Singer, T., Yordan, C., and Martienssen, R.A. (2001). Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). Genes. Dev. **15**, 591-602.

Sleutels, F., and Barlow, D.P. (2002). The origins of genomic imprinting in mammals. Adv. Genet. **46**, 119-163.

Soltis, P.S., and Soltis D.E. (2000). The role of genetic and genomic attributes in the success of polyploids. Proc. Natl. Acad. Sci. USA **97**, 7051-7057.

Song, K., Lu, P., Tang, K., and Osborn, T.C. (1995). Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc. Natl. Acad. Sci. USA **92**, 7719-7723.

Song, J., Ugai, H., Kanazawa, I., Sun, K., and Yokoyama, K.K. (2001). Independent repression of a GC-rich housekeeping gene by Sp1 and MAZ involves the same cis-elements. J. Biol. Chem. **276**, 19897-19904.

Soppe, W.J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M.S., Jacobsen, S.E., Schubert, I., and Fransz, P.F. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in Arabidopsis. EMBO J**. 23**, 6549-6559.

Spencer, V.A., and Davie, J.R. (1999). Role of covalent modifications of histones in regulating gene expression. Gene **240,** 1-12.

Steward, N., Ito, M., Yamaguchi, Y., Koizumi, N., and Sano, H. (2002) Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. J. Biol. Chem**. 277**, 37741-37746.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature **403,** 41-45.

Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. Genes Dev. **5**, 599-606.

Tariq, M., and Paszkowski, J. (2004). DNA and histone methylation in plants. Trends. Genet. **20**, 244-251.

Tompa, R., McCallum, C.M., Delrow, J., Henikoff, J.G., van Steensel, B., and Henikoff, S. (2002). Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. Curr. Biol. **12**, 65-68.

Van Houdt, H., Ingelbrecht, I., Van Montagu, M., and Depicker, A. (1997). Post-transcriptional silencing of a neomycin phosphotransferase II transgene correlates with the accumulation of unproductive RNAs and with increased cytosine methylation of 3' flanking regions. Plant J. **12,** 379-392.

Wada, Y., Miyamoto, K., Kusano, T., and Sano, H. (2004) Association between up-regulation of stress-responsive genes and hypomethylation of genomic DNA in tobacco plants. Mol. Genet. Genomics **271**, 658-666.

Wang, J., Tian, L., Madlung, A., Lee, H.S, Chen, M., Lee, J.J., Watson, B., Kagochi, T., Comai, L., and Chen, Z.J. (2004). Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. Genetics **167**, 1961-1973.

Wang, J., Lee, J.J., Tian, L., Lee, H.S., Chen, M., Rao, S., Wei, E.N., Doerge, R.W., Comai, L., and Chen, Z.J. (2005) Methods for genome-wide analysis of gene expression changes in polyploids. Methods Enzymol. **395**, 570-596.

Wassenegger, M., and Pelissier, T. (1998). A model for RNA-mediated gene silencing in higher plants. Plant Mol. Biol. **2,** 349-362.

Workman, J.L., and Kingston, R.E. (1998). Alteration of nucleosome structure as a mechanism of transcriptional regulation. Annu. Rev. Biochem. **67**, 545-579.

Xiao, W., Gehring, M., Choi, Y, Margossian. L, Pu. H., Harada, J.J., Goldberg, R.B., Pennell, R.I., and Fischer, R.L. (2003). Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. Dev. Cell **6**,891-901.

Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman,

D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and functional diversification of small RNA pathways in plants. PLoS Biol**. 5,** E104.

Yoshida, M., Horinouchi, S., and Beppu, T. (1995). Trichostatin A and trapoxin: novel chemical probes for the role of histone acetylation in chromatin structure and function. Bioessays **17,** 423-430.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell **1**, 25-33.

Zhong, C.X,, Marshall, J.B., Topp, C., Mroczek, R., Kato, A., Nagaki, K., Birchler, J.A., Jiang, J., and Dawe, R.K. (2002). Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell **11**, 2825-2836.

VITA

Name: Meng Chen

Address: Heep Center, Room 610, Department of Soil and Crop Sciences,
Texas A&M University 2474, College Station, TX 77843-2474, USA

Education: B.S., 2002, Biotechnology, Beijing University, PR China
M.S., 2005, Genetics, Texas A&M University, USA

Email Address: chenmeng@neo.tamu.edu

Publications:

Wang, J., Lee, H.-S., Tian, L., Chen, M., Lee, J.J., Madlung, A., Comai, L., and Chen, Z.J. (2004). Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. Genetics 167: 1961-1973.

Chen, Z. J., Wang, J., Tian, L., Lee, H.-S., Wang, J.J., Chen, M., Lee, J.J., Josefsson, C., Madlung, A., Watson, B, Pires, J.C., Lippman, Z., Vaughn, M., Colot, V., Doerge, R.W., Martienssen, R.A., Comai, L., and Osborn, T.C. (2004). The development of an Arabidopsis model system for genome-wide analysis of polyploidy effects. Biol. J. Linn. Soc. 82:689-700.

Tian, L., Wang, J., Fong, M.K., Chen, M., Cao, H., Gelvin, S.B., and Chen, Z.J. (2003).l Genetic control of developmental changes induced by disruption of *Arabidopsis* histone deacetylase 1 (*AtHD*) expression. Genetics 165:399-409