A Thesis for the Degree of Master of Science

Genome-Wide Association between Histone H3 Lysine Methylation and DNA Methylation in an Intergeneric Hybrid, x*Brassicoraphanus*

속간교잡종 배무채에서 히스톤 H3 라이신 9 와 DNA 메틸레이션의 전장유전체연관

FEBRUARY, 2016

HYE YEON SOH

MAJOR IN INTERDISCIPLINARY PROGRAM IN
AGRICULTURAL GENOMICS
THE GRADUATE SCHOOL OF
SEOUL NATIONAL UNIVERSITY

Genome-Wide Association between Histone H3 Lysine Methylation and DNA Methylation in an Intergeneric Hybrid, x*Brassicoraphanus*

UNDER THE DIRECTION OF DR. JIN HOE HUH
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
SEOUL NATIONAL UNIVERSITY

BY
HYE YEON SOH

MAJOR IN
INTERDISCIPLINARY PROGRAM IN AGRICULTURAL GENOMICS
THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

FEBRUARY, 2016

APPROVED AS A QUALIFIED THESIS OF HYE YEON SOH
FOR THE DEGREE OF MASTER OF SCIENCE
BY THE COMMITTEE MEMBERS

CHAIRMAN _____
Yong-Hwan Lee, Ph.D.

VICE-CHAIRMAN _____
Jin Hoe Huh, Ph.D.

MEMBER _____
Doil Choi, Ph.D.

Genome-Wide Association between Histone H3 Lysine Methylation and DNA Methylation in an Intergeneric Hybrid, x*Brassicoraphanus*

HYE YEON SOH

INTERDISCIPLINARY PROGRAM IN

AGRICULTURAL GENOMICS

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

ABSTRACT

Epigenetic changes in allopolyploids play a pivotal role in reprogramming homeologous gene expression and genomic stability. Dimethylation of H3K9 (H3K9me2) marks heterochromatin and is crucial in controlling DNA methylation and silencing transposable elements in *Arabidopsis*. In this study, chromatin immunoprecipitation accompanied with sequencing method (ChIP-seq) was employed to profile the distribution of H3K9me2 in intergeneric hybrid, x*Brassicoraphanus*. Compared with methylome data, a high rate of association of H3K9me2 with DNA methylation was observed throughout the genome. In addition, occurrence and disappearance of H3K9me2 in x*Brassicoraphanus* after polyploidization is intensely correlated with that CHG methylation. Extinction of DNA methylation and H3K9me2 in the allopolyploid happened regardless of its

subgenomic location. However, the newly generated epigenetic states showed biased subgenomic proportion. These results confirm that epigenetic mechanisms are mutually correlated and suggest that effects of hybridization between two genera of genetic distance would trigger differential epigenetic modulation in x*Brassicoraphanus*.

Key Words**:** Epigenetic change, H3K9me2, x*Brassicoraphanus*, ChIP-seq, DNA methylation, allopolyploid

Student number: 2014-20725

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

H3K9me2	Histone H3 lysine 9 dimethylation
ChIP	Chromatin immunoprecipitation
H3K4me	Histone H3 lysine 4 methylation
H3K27me	Histone H3 lysine 27 methylation
FLC	Flowering Locus C
CCA1	CIRCADIAN CLOCK ASSOCIATED 1
LHY	LATE ELONGATED HYPOCOTYL
TOC1	TIMING OF CAB EXPRESSION1
GI	GIGANTEA
WGD	Whole Genome Duplication
PRMT	Protein Arginine Methyltransferase
HKMT	Histone Lysine Methyltransferase
Su(var)3-9	Suppressor of variegation 3-9
E(z)	Enhancer of zeste
TrxG	Trithorax group protein
ASH1	Absent, small, or homeotic discs 1
LSD1	Lysine-specific demethylase 1
JmjC	Jumonji C

SUVH	Su(var)3-9 homolog
KYP	KRYPTONITE
rDNA	ribosomal DNA
NGS	Next-Generation sequencing
TE	Transposable element
sRNA	Small RNA
DHM	Differential histone-marked region
DMR	Differentially methylated region
DEG	Differentially expressed gene
CMT	CHROMOMETHYLASE
DDM	DECREASE IN DNA METHYLATION
DRM	DOMAINS REARRANGED METHYLASE
TTS	Translation termination site
BS-seq	Bisulfite sequencing

INTRODUCTION

Polyploidization is one of the most important cues for eukaryotic evolution. Most of angiosperms and the majority of pteridophytes (~95 %) have experienced series of polyploidization (Soltis, 2014; Grant, 1981). Scrutinizing the grounds and outcomes of polyploidy is essential to understand the diversity of flowering plants and pteridophytes as well as other eukaryotic pedigrees. Numerous major crops such as oats, soybean, wheat, cotton, canola, and banana were generated through natural polyploidization, which allowed crop domestications and the advent of valuable traits (Chen and Yu, 2013). Genome-level investigations on these crops have recently been published and several genetic and epigenetic mechanisms related to genome hybridization have been gradually uncovered.

x*Brassicoraphanus* (2n=38) is a newly synthesized species from a cross between *Brassica rapa* and *Raphanus sativus* (Lee et al., 2002). Intergeneric hybridizations between these two species have been recurrently attempted since early 1800's and a genetically stable allopolyploid was developed through induced mutation with the microspore culture method (Lee et al., 2011). x*Brassicoraphanus* shows intermediate phenotypes between *B. rapa* and *R. sativus* in hypocotyl length, the leaf shape, and the structure of silique. One of its novel phenotypes is the white petal color while the flowers of *B. rapa* and *R. sativus* have yellow and purple color, respectively. x*Brassicoraphanus* is genetically stabilized within a short period of time as a rare case of intergeneric hybrid. The genome and epigenome of x*Brassicoraphanus* represent instant hybridization effects. The genetic distance

between the two parents would be beneficial to distinguish the parental origin of duplicate genes. Studying the unique allopolyploid, x*Brassicoraphanus* could provide comprehensive understanding of genome rearrangement occurring within an intergeneric hybrid and the developmental mechanism of neomorphism. Furthermore, speciation and evolutionary process derived from intergeneric crossing also could be explored.

Previous studies showed that histone modifications mediate gene transcription during growth and development of plants and animals by remodeling the chromatin structure (Berger, 2007; Zhang, 2008). In *Arabidopsis thaliana* active promoters contain several modifications, such as histone H3 and H4 acetylations and methylation of H3 lysine 4 (H3K4me 1/2/3) that are also known as euchromatic marks. By contrast, repressed gene expressions are associated with methylations of H3 lysine 9 (H3K9) and H3 lysine 27 (H3K27) that are known to mark heterochromatin (Jenuwein and Allis, 2001; Li et al., 2007). Out of these several modifications, di-methylation of H3K9 (H3K9me2) is dispersed across pericentromeric heterochromatin and the genome-wide location of H3K9me2 is known to be highly correlated with transposable elements and repeat regions (Lippman et al., 2004; Turck et al., 2007; Bernatavichute et al., 2008).

Epigenetic changes in newly hybridized interspecific allopolyploids are supposed to play a crucial role in genomic reorganization and reprogramming chromatin structures that give rise to gene expression changes (Osborn et al., 2003; Chen, 2007). In *Arabidopsis*, a synthetic allotetraploid shows different phenotypes including flowering time and morphological vigor compared to its parents *A. thaliana* (*Ler*) and *A. arenosa* (Wang et al., 2006; Ni et al., 2009). Biochemical and

genome-level studies revealed that changes of histone modifications in the promoters of *FLC* gene and the circadian clock genes (*CCA1*, *LHY*, *TOC1*, and *GI*) are associated with the altered expression in allopolyploids and hybrids (Wang et al., 2006; Ni et al., 2009).

In this study, chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) analysis was employed to profile H3K9me2 across the entire x*Brassicoraphanus* genome. With DNA methylation data, nucleotide-resolution histone modification (H3K9me2) profiling data were analyzed to understand general states of epigenome in x*Brassicoraphanus* and the epigenetic correlation between DNA methylation and H3K9me2 was revealed. Also by comparing genome-wide distributions of H3K9me2 between x*Brassicoraphanus* and its parents, *B. rapa* and *R. sativus*, epigenetic changes including DNA methylation and subsequent transcriptomic changes after polyploidization were explored. Generation of DNA methylation and H3K9me2 showed differences in terms of their subgenomic sites. Hybridization effects and genomic incompatibility between two progenitors are considered to be putative causes for this phenomenon. This research would provide insights into comprehensive epigenetic states occurred after genome hybridization and set up the basis for further studies of epigenetic mechanisms of polyploidization.

LITERATURE REVIEWS

1. Polyploidy

Polyploidy depicts an organism or a cell that possesses multiple sets of chromosomes. In the majority of cases, even numbered sets of chromosomes are included in polyploids. The research of eukaryotic genome sequence suggests that genomes of many species show marks of polyploidal ancestor (Paterson, 2005; Becak and Kobashi, 2004; Van de Peer and Meyer, 2005).

Polyploidization typically gives rise to biodiversity and instantaneous speciation, acting as a major driving force of macro-evolution (Levin, 1983; Levin, 2002; Mayrose et al., 2011). This process is a common phenomenon that can be easily found in plants and also among fish and amphibians (Leitch et al., 2008; Comai, 2005; Wendel, 2000; Song and Chen, 2015; Hegarty and Hiscock, 2008; Doyle et al., 2008; Soltis et al., 2014) Most of the flowering plants are believed to have experienced whole-genome duplications (WGDs) (Soltis et al., 2014). Two rounds of WGDs before the divergence of flowering plants are estimated to have occurred and consequently diversify genes that are crucial for the development of seeds and flowers, which gives clues for Charles Darwin's abominable mystery (Jiao et al., 2011; Van de Peer et al., 2009).

There are two types of polyploidy: autopolyploidy and allopolyploidy that are grouped by its origin and the arrangement of chromosomes (Stebbins, 1950; Song and Chen, 2015). Autopolyploids (AAAA in an autotetraploid) are derived from a single genome duplication of one individual or different plants within a species. On the other hand, allopolyploids are formed by hybridizing and doubling

of chromosomes between two different species. Therefore, two parental genomes exist in a nucleus and allotetraploids (AABB) have two of each pair of equivalent chromosomes from the parent species, or homeologous chromosomes that are also known as homeologs. Some polyploid plants such as sugarcane are hard to distinguish between them (Premachandran et al., 2011). Since intraspecific pairing is more frequent than interspecific mating, allopolyploidy was estimated to be less widespread than autopolyploidy (Stebbins, 1971). However, allopolyploids are more universal than autopolyploids in nature (Schatlowski and Kohler, 2012; Hegarty and Hiscock, 2008).

When a nucleus of allopolyploid contains two parental genomes, the expression levels of homeologous genes could display divergence from the arithmetic average, also known as additivity, of parental gene expression patterns. The putative causes of those nonadditive gene expression patterns are maternal-paternal influences, gene dosage balance, *cis*- and *trans*- regulations, and dysregulation of transposons (Yoo et al., 2014). As maternal organellar genomes are generally inherited in plants, the expression dominance toward the maternal parent was observed in many polyploids; cotton, coffee, *Spartina anglica*, and *Tragopogon* (Yoo et al., 2014). Secondly, the gene-balance hypothesis explains that the function of the whole complex is regulated by the stoichiometry among parts of macromolecular complexes (Birchler and Veitia, 2012). According this concept, unequally fractionated subgenomes can describe the expression dominance toward one homeologous gene. On the other hand, combined *cis*- and *trans*- regulatory systems from two divergent parents can modulate the expression of homeologs. Due to sequence similarities between two parents, trans-acting factors may

crosstalk over homeologous *cis*-elements in allopolyploids. Lastly, epigenetic changes during interspecific hybridization may alter the expressions of transposable elements, which leads to the modulation of homeologs in allopolyploids (McClintock, 1984; Parisod et al., 2009).

2. Histone modification

Histone proteins comprise the nucleosome, a fundamental unit of chromatin, and several modifications on histone proteins modulate chromatin structure through epigenetic mechanisms. Histone modification mainly occurs on amino (N) terminal tails that protrude from core histone proteins. Various posttranslational modifications can take place on histone residues by forming covalent bonds with different chemical groups, which include acetylation, methylation, phosphorylation, ubiquitination, sumoylation, glycosylation, and ADP-ribosylation (Liu et al., 2010). The histone code hypothesis proposes that the orchestration of histone modifications determines chromatin structures, which leads to the activation or repression of gene transcription. (Jenuwein et al., 2001)

Histone methylation, one of the most commonly studied forms, plays a crucial role in regulating diverse biological processes ranging from heterochromatin formation to transcriptional regulation. It can be mono- and di-methylated on arginine or mono-, di-, and tri- methylated on lysine residues by protein arginine methyltransferases (PRMTs) and histone lysine methyltransferases (HKMTs), respectively.

Lysine methylation of histone is found in both transcriptionally silenced and active chromatin regions, according to the location of lysine residues binding

to methyl groups and the degree of methylation. Conjugating a methyl group to lysine residues does not alter the net charge of the modified products, but the hydrophobicity increases. Also, specific proteins that recognize the methylated residues may be recruited to newly-made binding surfaces or intra- and intermolecular interactions may change their patterns.

First discovered in *Drosophila melanogaster* are SET domain proteins conserved in three proteins and named after them; Suppressor of variegation 3-9 (Su(var)3-9), Enhancer of zeste (E(z)), and trithorax group proteins (TrxG) (Dillon et al., 2005). SET domain proteins are known to catalyze the lysine methylation and there are 41 and 37 genes that encode SET domain proteins respectively in *Arabidopsis* and rice genomes (Gendler et al., 2008). SET domain proteins in plants are homologous to those in animals and yeast and grouped into four categories; SU(VAR)3–9 groups, E(Z) homologs, TrxG groups, and ASH1 (absent, small, or homeotic discs 1) groups (Baumbusch et al., 2001; Springer et al., 2003; Zhao et al., 2004). Even though the specificity and characteristics of these SET proteins in plants have not been deeply studied in every detail, several genetic researches showed that they may affect the same lysine residues or related pathways as in animals or yeast (Liu et al., 2010). On the other hand, the methyl groups on lysine can be removed by histone demethylases that perform vital roles in regulating homeostasis of histone methylation. The two types of demethylases work in distinct ways; lysine-specific demethylase1 (LSD1) that oxidizes amine group and Jumonji C (JmjC) domain-containing proteins that hydroxylate the methylated substrate both result in withdrawing methyl groups from methylated lysine residues (Shi et al., 2004; Tsukada et al., 2006).

In *Arabidopsis*, Lys 4 (K4), Lys 9 (K9), Lys 27 (K27), and Lys 36 (K36) of histone H3 are main targets for histone lysine methylation. In general, methylation of histone H3K9 and H3K27 is related to silenced regions, whereas methylation of H3K4 and H3K36 is related to active gene expression (Berger et al., 2007). Mono- and di- methylation of H3K9 is especially more abundant in chromocenters than in euchromatin regions and highly associated with transposons and related repeat regions (Lippman et al., 2004; Turck et al., 2007; Bernatavichute et al., 2008). In *Arabidopsis* genome, ten SU(VAR)3-9 homologues are expected to function (SUVH1-10; Baumbusch et al., 2001). Among them, KRYPTONITE/SUVH4 (KYP), SUVH5 and SUVH6 have been reported to be in charge of the propagation of H3K9me1 and H3K9me2 (Jackson et al., 2002; Ebbs et al., 2005; Ebbs et al., 2006; Jasencakova et al., 2003). *kyp suvh5 suvh6* triple mutants not only eliminate H3K9me2 levels, but also show the reduced level of DNA methylation and release the silent state of heterochromatin as transposons are reactivated (Ebbs et al., 2006). *kyp* mutants discharge the silencing state insignificantly and do not show any defects in morphology despite its role in repressing genes and transposable elements (Tariq and Paszkowski, 2004).

In *met1* plants that lack the function of maintenance of CG methylation, the distribution of H3K9me2 was found to be massively relocated across the genome (Mathieu et al., 2005; Soppe et al., 2002; Tariq et al., 2003), and the possibility that CG methylation manipulates the level of H3K9me2 was proposed (Mathieu et al., 2005; Soppe et al., 2002; Tariq et al., 2003; Tariq and Paszkowski, 2004). Moreover, KYP/SUVH4-mediated H3K9 methylation was reported to manage non-CG methylation (Jackson et al., 2002; Malagnac et al., 2002). Chromatin

immunoprecipitation (ChIP) analysis accompanied with high-resolution genome wide microarray (ChIP-chip) also showed the high correlation between H3K9me2 and CHG DNA methylation (Bernatavichute et al., 2008). Histone H3K9 methylation is essential to maintain non-CG methylation levels and they play a pivotal role in repressing transposons, ribosomal RNA genes (rDNA), and repeat sequences (Liu et al., 2010).

3. Sequencing technologies and ChIP-seq

After the DNA structure as double helix was discovered by Watson and Crick in 1953 (Watson and Crick, 1953), Frederick Sanger proposed a modern DNA sequencing method with chain-terminating inhibitors (Sanger et al. 1977). In this method, DNA polymerase synthesizes complementary DNA sequences using normal 2'-deoxynucleotides (dNTPs) and 2', 3'-dideoxynucleotides (ddNTPs) that terminate the reaction. Then DNA fragments with various sizes are separated using gel electrophoresis method and examined to uncover its sequence. After several modifications has been made, the automated Sanger sequencing has been spread out for two decades but limitations of this first-generation technology made it to be replaced by next-generation sequencing (NGS). These new technologies include steps of template preparation, sequencing and imaging, and data analysis (Metzker, 2010). Compared to the former method, NGS platforms have drastically lowered the cost and enabled higher throughput.

Since the genome of *Arabidopsis* was completed in 2000 (Initiative AG, 2000), a number of plant genomes have been sequenced with NGS technology and exceed more than 100 (Michael and Jackson, 2013). Even with the handy tools for

genome assembly, plant genome still remains to have many barriers that block complete analysis because of its tremendous genome size, repeat content, paralogy, and heterozygosity. Less than 200 bp-long reads from NGS technologies are too short to resolve complexity of plant genomes. While most mammalian genomes have been assigned into chromosome scale regions (Gnerre et al., 2011), most draft plant genomes comprise numerous contigs and scaffolds that shatter with embedded gaps. Among plant genomes, the *Arabidopsis* genome is unquestionably known to be the best organized one, but still scattered in 102 contigs and its total gap length is 185,644 bp (TAIR 10; Lamesch et al., 2012).

A third generation sequencing technology has emerged to cover the limitations of the former technologies and focuses on single molecule DNA. This new generation platform can be grouped into three categories; sequencing-by-synthesis technology that utilizes DNA polymerase to synthesize a single molecule of DNA, nanopore-sequencing technology that detects individual bases as DNA strand goes through the nanopore, and microscopy technology that directly makes image of DNA molecules. These single-molecule-based sequencing technologies have provided many advantages. First, they can process more throughput for much shorter time at the lower price. Also from a diminutive amount of starting material, they can produce longer reads that are helpful in assembling the de novo genome (Schadt et al., 2010).

To study the interactions between DNA and protein, Varshavsky and his colleagues first proposed the chromatin immunoprecipitation (ChIP) method in 1988 (Solomon et al., 1988). This experiment is basically composed of three steps; to cross-link between DNA and proteins with formaldehyde, to immunoprecipitate

the protein-DNA complex with specific antibody, and to reverse the cross-links and analyze recovered DNA. Since DNA and proteins are randomly cross-linked and non-specific interactions can occur during the immunoprecipitation step, ChIP-derived DNA is compared to a mock sample that is collected after the immunoprecipitation process done without antibody. For genome-wide analysis of ChIP-selected DNA in yeast, Ren and colleagues primarily compared those two samples by hybridizing to a genomic microarray (ChIP-chip) (Ren et al., 2000). With the advent of next-generation sequencing technologies, ChIP-chip have been substituted by ChIP followed by sequencing m (ChIP-seq). ChIP-seq provides many advantages including higher resolution, lower price per the same amount of data, less complicated process, and requirement for less input quantity and fewer replicates (Mardis, 2007).

MATERIALS AND METHODS

Plant materials and growth conditions

x*Brassicoraphanus*, *Brassica rapa* cv. Chiifu-401-42, and *Raphanus sativus* cv. WK10039 plants were grown on 1 X Murashige and Skoog medium (Duchefa, Haarlem, The Netherlands) in a growth chamber under 16 h of illumination for 14 days. The whole seedlings including shoots and roots were harvested together for the analysis of chromatin modifications. Plants grown for 14 days and for 3 months are shown in Figure 1.

Isolation and immunoprecipitation of chromatin

Chromatin Immunoprecipitation (ChIP) was performed using the seedlings (~ 3 g) of x*Brassicoraphanus*, *B. rapa*, and *R. sativus* respectively. Whole plant tissues were combined and cross-linked in 1% formaldehyde for 15 min using vacuum infiltration method and the cross-linking reaction was quenched with 0.125 M of glycine. Plants were washed three times with deionized water and blotted to remove water. Frozen tissues in liquid nitrogen were ground to fine powder and resuspended in 30 ml of extraction buffer 1 (0.4 M sucrose, 10 mM Tris-HCl pH 8.0, 5 mM beta-mercaptoethanol, 1 mM phenylmethylsulfonyl [PMSF], and complete protease inhibitor; Roche Diagnostics, Basel, Switzerland), then filtered through two layers of Miracloth (Merk Miilipore, Guyancourt, France) twice. The filtrate was centrifugated at 2,800 g at 4 ^{o}C for 25 min. The pellet was resuspended in 1 ml of extraction buffer 2 (0.25 M sucrose, 10 mM Tris-HCl p H 8.0, 10 mM $MgCl_2$, 1% Triton X-100, 5 mM beta-mercaptoethanol, 1 m

M PMSF, and complete protease inhibitor; Roche Diagnostics, Basel, Switzerland) and centrifugated at 12,000 g at 4 ^{o}C for 10 min. This procedure was performed twice for better nuclei isolation. The pellet was resuspended in 300 μl of extraction buffer 3 (1.7 M sucrose, 10 mM Tris-HCl pH 8.0, 2 mM $MgCl_2$, 0.15% Triton X-100, 5 mM beta-mercaptoethanol, 1 mM PMSF, and complete protease inhibitor; Roche Diagnostics, Basel, Switzerland) and layered on top of 1.2 ml of clean extraction buffer 3, then centrifugated at 16,000 g at 4 ^{o}C for 1 hr. The chromatin pellet was resuspended in 300 μl of nuclei lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS, and complete protease inhibitor; Roche Diagnostics, Basel, Switzerland) and sonicated with a Sonifier (Branson Sonifier, Danbury, USA) to shear DNA into 0.3 ~ 1 kb. The sonicated chromatin was centrifugated and the supernatant was diluted 10-fold with ChIP dilution buffer (1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.0, 167 mM NaCl). Seven ug of the antibody against histone H3 dimethyl Lys 9 (AB1220, Abcam, Cambridge, UK) was reacted respectively with 1.5 mg of protein G Dynabeads (Invitrogen, CA, USA) for 2 hr at 4 ^{o}C, then chromatin extract was incubated with antibody/magnetic bead mix at 4 ^{o}C overnight. After several washes, immunocomplex was eluted twice from the beads with 30ul of elution buffer (1% SDS, 0.1 M sodium bicarbonate) and reverse cross-linked in the presence of 200 mM NaCl at 65 ^{o}C overnight. To remove all proteins proteinase K was treated, and then DNA was purified by phenol/chloroform/isoamylalcohol extraction and precipitated with ethanol. The pellet was resuspended with 30 μl of tridistilled water. A small aliquot of untreated sonicated chromatin was reverse cross-linked and used as total input DNA control. ChIP experiments were conducted with two biological replicates samples to reduce

errors derived from experimental variations.

Computational analysis of sequencing data

ChIP-seq libraries were constructed as described in the Illumina ChIP sequencing kit (Illumina, CA, USA). DNA fragments with a range of ~500 bp were excised and amplified for cluster generation and sequencing. All DNA libraries were sequenced on a HiSeq2500 (50 single end reads) (Illumina, CA, USA). The sequencing reads were quality controlled with FastQC and aligned to the reference genome of x*Brassicoraphanus* by Bowtie software using default settings (Langmead et al., 2009). Reads that could be assigned to only one region of the genome were used for downstream analyses. To find significant peaks compared to control, MACS software (Zhang et al., 2008) was employed and common regions from two biological replicates were extracted with mergePeaks command from Homer software (Heinz et al., 2010). For comparative analysis between x*Brassicoraphanus* and its parent, *B. rapa* and *R. sativus*, all peaks were classified into three groups; common peaks found in both allopolyploid and its progenitors, xB-specific peaks and BrRs-specific peaks that only exist in x*Brassicoraphanus* or in *B. rapa* and *R. sativus.*

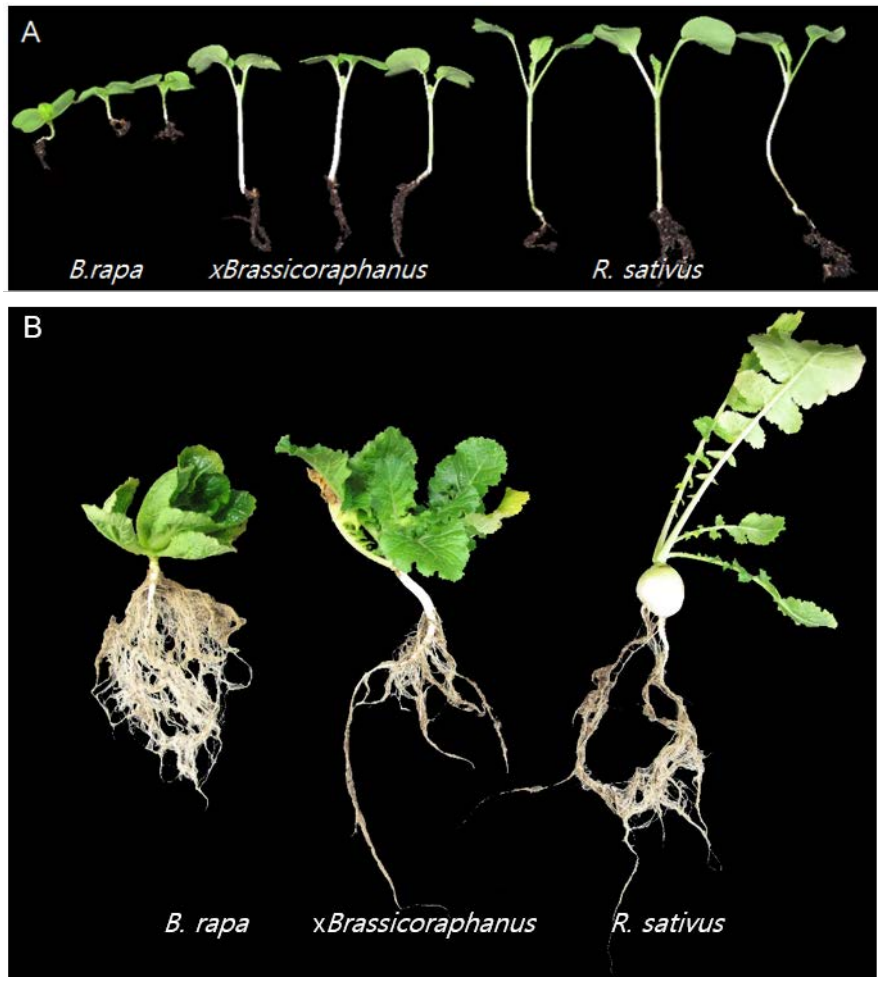

Figure 1. Phenotypes of *Brassica rapa*, x*Brassicoraphanus*, and *Raphanus sativus*. (A) Whole seedlings grown for 14 days that were used in this study and (B) whole plants grown for 3 months (Courtesy of Dr. Gibum Yi).

RESULTS

Genome-wide detection of H3K9me2-enriched regions

Cross-linking chromatin immunoprecipitation sequencing (X-ChIP-seq) methods employing formaldehyde fixation and sonication were utilized to get the genome-wide histone modification pattern data for x*Brassicoraphanus*, *B. rapa* cv. Chiifu, and *R. sativus* cv. WK10039. About 30 million reads were obtained from the ChIP-seq process for each sample. All reads obtained from three genera were mapped to the genome of x*Brassicoraphanus* and uniquely aligned reads (6~30%) were only used for the peak calling process (Table 1). By comparing with control, significantly H3K9me2-enriched regions (peaks) were detected for each replicate. The common peaks were merged from two replicates and regarded as representative H3K9me2-related regions for each plant. For x*Brassicoraphanus*, 12,068 peaks were detected with the length of 9,324,563 bp of sequenced genome (1.35%) (Table 2). And 11,986 and 11,589 peaks were allocated to *B. rapa* and *R. sativus* spanning 10,043,125 bp and 11,863,241 bp, respectively.

To investigate the epigenomic changes occurred after polyploidization, differential binding sites of histone modification (DHM; differential histone modification-marked region) between x*Brassicoraphanus* and its progenitors were identified. The peak file of x*Brassicoraphanus* and that of the parents were compared and the whole regions were classified into three types; the common peaks that exist in both x*Brassicoraphanus* and its parents, x*Brassicoraphanus*-specific peaks (xB DHMs) and *B. rapa* or *R. sativus*-specific peaks (BR DHMs) that are enriched in x*Brassicoraphanus* and in its parents respectively. A total of

4,173 peaks covering 8,208,133 bp are sorted into common peaks (Table 3). Table 3 shows that xB DHMs contain 2,424 regions spanning 1,974,707 bp and 8,978 peaks are identified as BR DHMs with the length of 3,755,736 bp.

Table 1. ChIP-seq counts and mapping results

Sample	# Total reads	# Unique mapped reads	(%)
xB.1_input	40,459,912	11,890,694	(30.15)
xB.1_H3K9me2	41,202,029	6,573,472	(16.38)
xB.2_input	39,754,351	11,784,202	(30.51)
xB.2_H3K9me2	36,160,104	2,337,146	(6.82)
Br.1_input	38,504,444	11,430,177	(30.47)
Br.1_H3K9me2	42,555,140	5,233,934	(12.84)
Br.2_input	24,251,434	3,451,534	(15.15)
Br.2_H3K9me2	39,172,728	4,893,917	(12.79)
Rs.1_input	39,559,995	8,316,826	(21.74)
Rs.1_H3K9me2	27,737,166	3,623,603	(13.39)
Rs.2_input	38,077,542	11,162,452	(29.76)
Rs.2_H3K9me2	40,393,014	4,843,834	(12.19)

Table 2. General information of H3K9me2 detection

Sample	Respective peak			Common peak		
	Total length	Average length	Total number	Total length	Average length	Total number
xB.1_H3K9me2	22,288,484	518.5	42,979	9,324,563	772.6	12,068
xB.2_H3K9me2	11,546,007	237.0	48,705			
Br.1_H3K9me2	17,180,834	370.9	46,313	10,043,125	837.9	11,986
Br.2_H3K9me2	11,863,260	550.3	21,558			
Rs.1_H3K9me2	10,316,041	608.8	16,944	11,863,241	1023.6	11,589
Rs.2_H3K9me2	16,422,659	464.4	35,357			

Table 3. Summary of differential histone modification binding sites and common sites

Type	Total length	Average length	Total number
Common peak	8,208,133	1,967	4,173
xB-specific peak (xB DHM)	1,974,707	814	2,424
BrRs-specific peak (BR DHM)	3,755,736	418	8,978

Characterizing the distribution of H3K9me2 level

For the downstream analyses related to H3K9me2 distributions in x*Brassicoraphanus*, both xB DHMs and the common peaks were utilized (Table 4). In the same way BR DHMs and the common peaks were included in H3K9me2 associated regions in the parents. Figure 2 describes the distribution of H3K9me2 level and fractional DNA methylation levels of x*Brassicoraphanus* and the parents along 19 pseudochromosomes (10 Ax chromosomes from A genome and 9 Rx chromosomes from R genome) of x*Brassicoraphanus* in 100 kb windows. H3K9me2 is highly enriched in gene-poor regions and tightly associated with DNA methylation (Figure 2). The locations of xB DHMs and BR DHMs are not restricted to the specific regions but are dispersed across the genome. Also, the distribution of histone modification was calculated relative to inter and intragenic regions in x*Brassicoraphanus* (Figure 3). It was found that H3K9me2 is predominantly wide-spread within intergenic regions (66.9%) followed by the coding regions (19.9%). The percentages of H3K9me2 in translation termination sites (TTS) and in the promoter regions are similar to each other (6.1% and 7.1%).

To understand detailed distributions of H3K9me2 in x*Brassicoraphanus*, 6,597 peaks were divided according to their subgenomic locations and the lengths of allocated peaks were merged. Figure 4A shows that less than half of the total length of H3K9me2-enriched regions (43.5%) are positioned in the *B. rapa* subgenome (AxAx) while the rest in the *R. sativus* subgenome (RxRx). When the whole lengths of identified H3K9me2 regions of *B. rapa* and *R. sativus* are compared to each other, the higher coverage of the peaks in *R. sativus* than that in *B. rapa* is observed (Figure 4B). To investigate the effect of subgenome dominance on

histone modification, the subgenomic locations of xB DHMs and BR DHMs were examined (Table 5). One third of xB DHMs occur in the Ax subgenome and the rest two-thirds are positioned in the Rx subgenome. BR DHMs, however, are divided equally to each subgenome. In the case of Common peaks, subgenomic distribution is similar to that of H3K9me2 peaks in x*Brassicoraphanus*. Disparity of xB DHMs in each subgenome would explain Rx-weighted H3K9me2 distribution in x*Brassicoraphanus*.

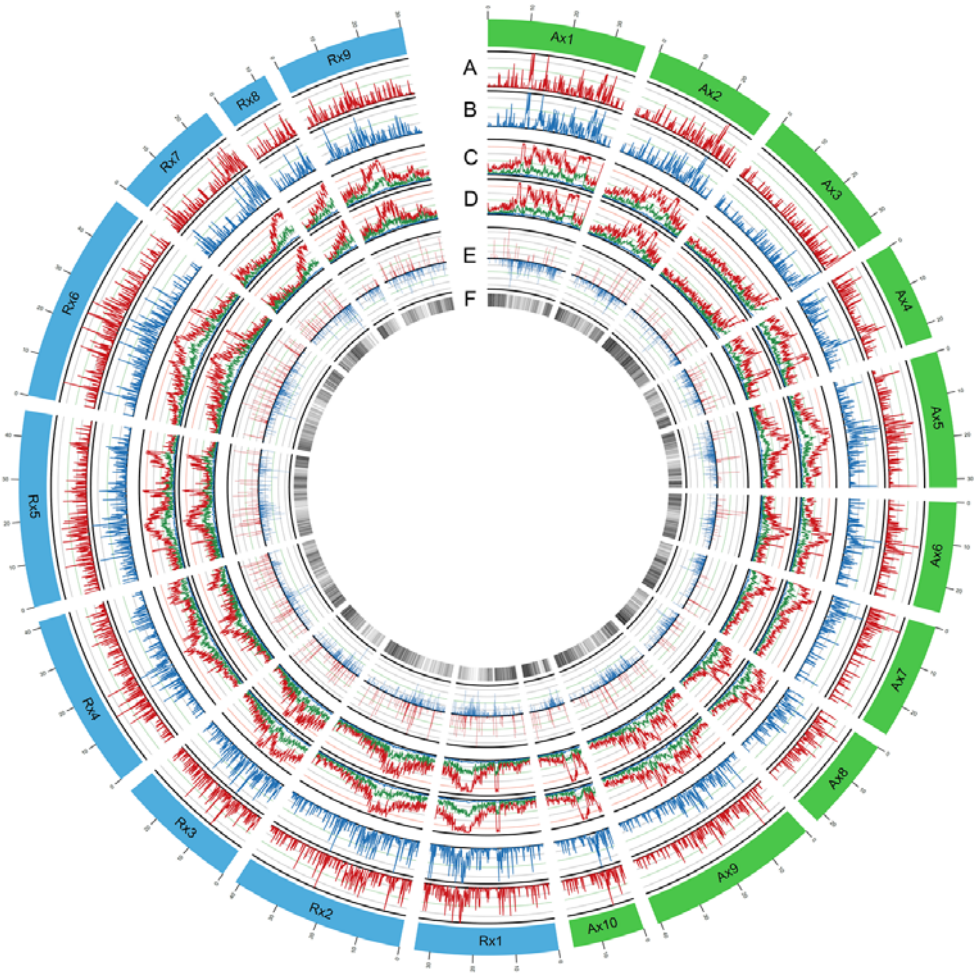

Figure 2. Genome-wide distribution of H3K9me2 and DNA methylation in x*Brassicoraphanus*. The x*Brassicoraphanus* genome is composed of two subgenomes from *B. rapa* and *R. sativus* including 10 Ax chromosomes and 9 Rx chromosomes. Tracks shown are **(A, B)** H3K9me2 densities of x*Brassicoraphanus* (A) and of its parents, *B. rapa* and *R. sativus* (B). **(C, D)** DNA methylation levels of x*Brassicoraphanus* (C) and of its parents (D). CG, CHG, CHH methylation levels are indicated as red, green, and blue lines, respectively. **(E)** Differential H3K9me2 binding sites. x*Brassicoraphanus*-specific site (red) and the parents-specific site (blue) **(F)** Gene density. The scales are in 100kb windows.

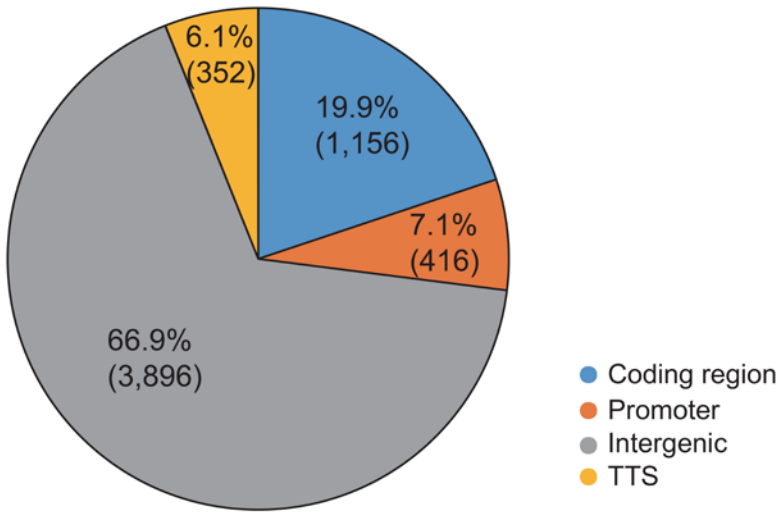

Figure 3. The distribution of H3K9me2 within different sites of the x*Brassicoraphanus* genome.

According to its location the whole peaks (6,597) were grouped into four regions; Coding region, promoter (-2 kb from the translation start site), intergenic, and translation termination site (TTS). When located in the region from -1 kb to 100 bp, the peaks are assigned to TSS. For TTS, a criterion is from -100 bp to 1 kb.

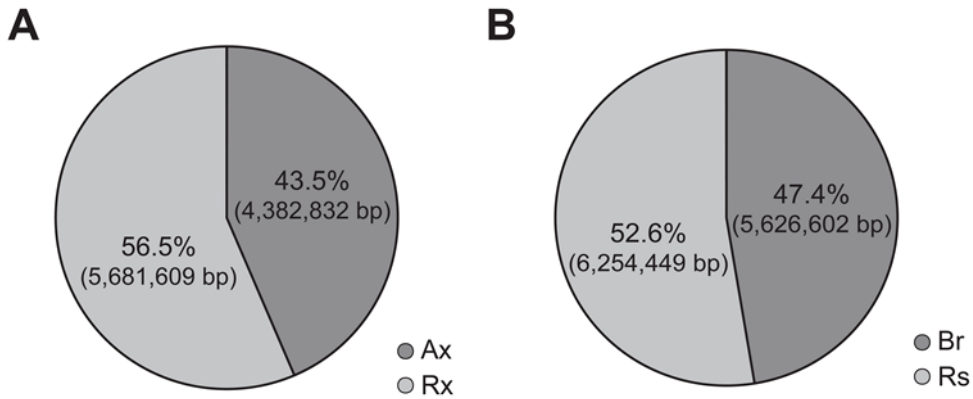

Figure 4. Distributions of H3K9me2 within subgenomic chromosomes of x*Brassicoraphanus* and in its parents. (A) The proportion of H3K9me2 levels in x*Brassicoraphanus* according to its subgenome. All peaks of H3K9me2 were divided based on their subgenomic locations and the total lengths were presented. Dark grey represents Ax subgenome and light grey represents Rx genome. **(B)** The proportion of H3K9me2 levels between *B. rapa* (dark grey) and *R. sativus* (light grey). Whole lengths of peaks of each genus were calculated.

Table 4. Summary of H3K9me peaks of x*Brassicoraphanus* and those of *B. rapa* and *R. sativus*

	Total length	Average Length	Total number
x*Brassicoraphanus* (xB)	10,182,840	1,544	6,597
B. rapa and *R. sativus* (BR)	11,963,869	910	13,151

Table 5. Subgenomic proportions of xB DHMs, BR DHMs, and common peaks in x*Brassicoraphanus*

	xB DHM		BR DHM		Common peak	
	Total length	%	Total length	%	Total length	%
Ax	669,483	33.90	1,913,253	50.94	3,713,349	45.24
Rx	1,252,848	63.44	1,825,688	48.61	4,428,761	53.96
Not assigned	52,376	2.65	16,795	0.45	66,023	0.80
Total	1,974,707	100	3,755,736	100	8,208,133	100

The correlation between H3K9me2 and DNA methylation

One of the well-known factors that induce H3K9me2 is DNA methylation. To investigate the correlation between H3K9me2 and DNA methylation, DNA methylome data from bisulfite sequencing (BS-seq) that display nucleotide-resolution data of quantitative methylation levels (analyzed and kindly provided by Hosub Shin) were compared with the ChIP-seq data in two different ways. First, the average levels of DNA methylation in different contexts, CG, CHG, and CHH, were calculated within H3K9me2 positive and negative regions in x*Brassicoraphanus* (Figure 5A). The DNA methylation levels of all three types within H3K9me2 positive regions are much higher than those within genome and the regions devoid of H3K9me2. Compared to a two-fold increase in average CG methylation level, non-CG methylation was three times more enriched within H3K9me2-related regions.

There are negligible differences of the average DNA methylation levels between genomic area and H3K9me2 negative zones. The general aspect of Figure 5A is connected to that in *Arabidopsis* (Bernatavichute et al., 2008), which verifies the quality of ChIP-seq results. Second, the ratio of methylated cytosines (>50% for CG, >20% for CHG, and >0% for CHH; Bernatavichute et al., 2008) relative to the whole number of cytosines was figured out according to the degree of H3K9me2 enrichment (Figure 5B). In the CG context, almost all cytosines (99.56%) are methylated in H3K9me2 positive areas while the half are methylated across the genome (47.29%) and H3K9me2-deficient regions (45.46%). A quarter of cytosines in the CHG context are methylated in x*Brassicoraphanus* but a significant portion of H3K9me positive regions (82.12%) corresponds to the

regions with DNA methylation (Figure 5B). Similarly, the percentage of CHH methylated cytosines increases from 17% to 52% in accordance with dimethylation of H3K9. In short, H3K9me2 and DNA methylation are mutually correlated within x*Brassicoraphanus* genome, where non-CG methylation is significantly correlated with the status of H3K9me2.

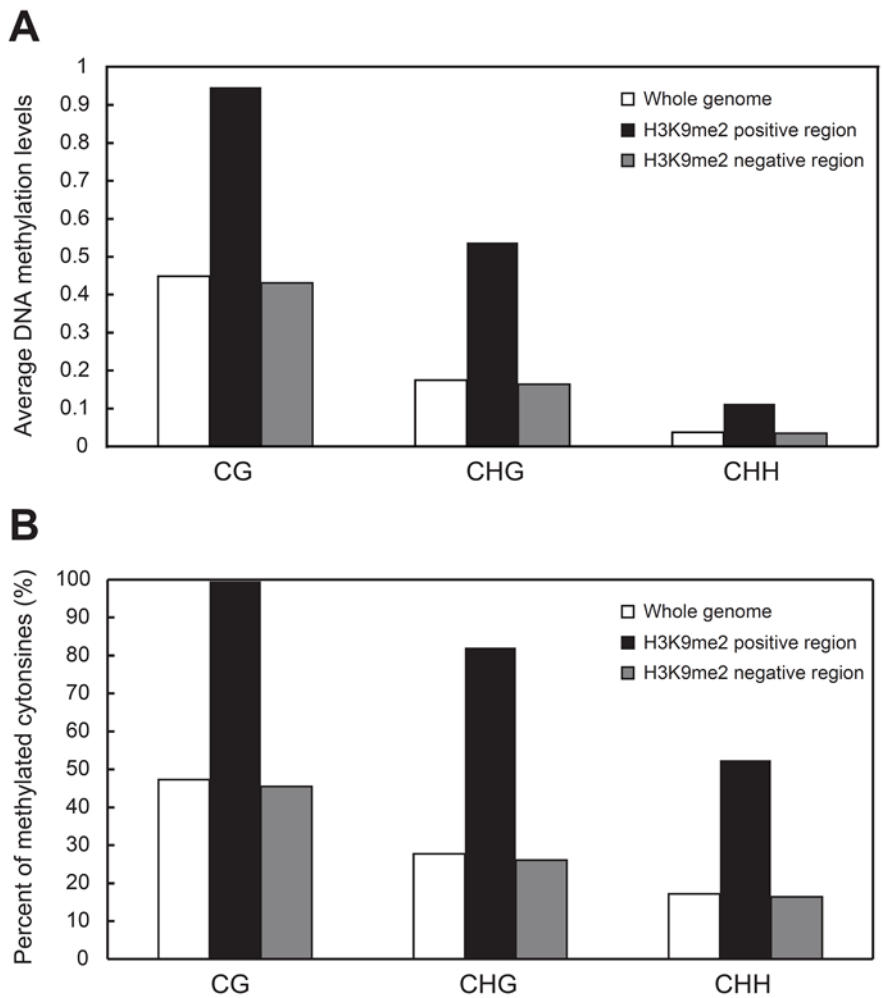

Figure 5. Correlation of H3K9 dimethylation with DNA methylation in x*Brassicoraphanus*.

(A) Average DNA methylation levels throughout the genome (white), in H3K9me2-related region (black), and in H3K9me2 negative region (grey). All methylated cytosines within each region were used to compute average methylation levels in CG, CHG, and CHH contexts. **(B)** The number of qualified methylcytosines (>50% for CG, >20% for CHG, and >0% for CHH; Bernatavichute et al., 2008) was presented in a ratio to total cytosines within each region according to DNA methylation contexts.

Coincidence of DHMs and differentially methylated regions (DMRs)

Since the maintenance mechanism of DNA methylation patterns with histone H3K9me2 has been discovered in previous studies (Stroud et al., 2014; Bernatavichute et al., 2008), the regions of xB DHMs and BR DHMs were associated with hyper DMRs and hypo DMRs that exhibit the areas where relatively higher levels of DNA methylation are found in x*Brassicoraphanus* than in the parents and vice versa. Out of 1,974,707 bps of xB DHMs, nearly 9% falls within hyper DMRs whereas a small fraction (1.47%) is overlapped with hypo DMRs (Figure 6). On the other hand, one fifth of BR DHMs that span 772,585 bps overlap with hypo DMRs, while only 6.1% of BR DHMs are found in the regions of hyper DMRs. To summarize, xB DHMs are more correlated with hyper DMRs and BR DHMs are more connected to hypo DMRs.

To test whether the occurrence and disappearance of H3K9me2 after polyploidization correlate with the different contexts of DNA methylation, the total length of DMR-associated DHMs was calculated according to the contexts; CG, CHG, and CHH. Table 6 displays that almost all (97.7%) of the xB DHMs that overlap with hyper-DMRs are highly correlated with CHG-methylated hyper DMRs. Interestingly, BR DHMs also have a high level of interrelation with hypo DMRs in the CHG context. These results indicate that the differentially CHG-methylated regions are correlated with the high proportion of both xB DHMs and BR DHMs.

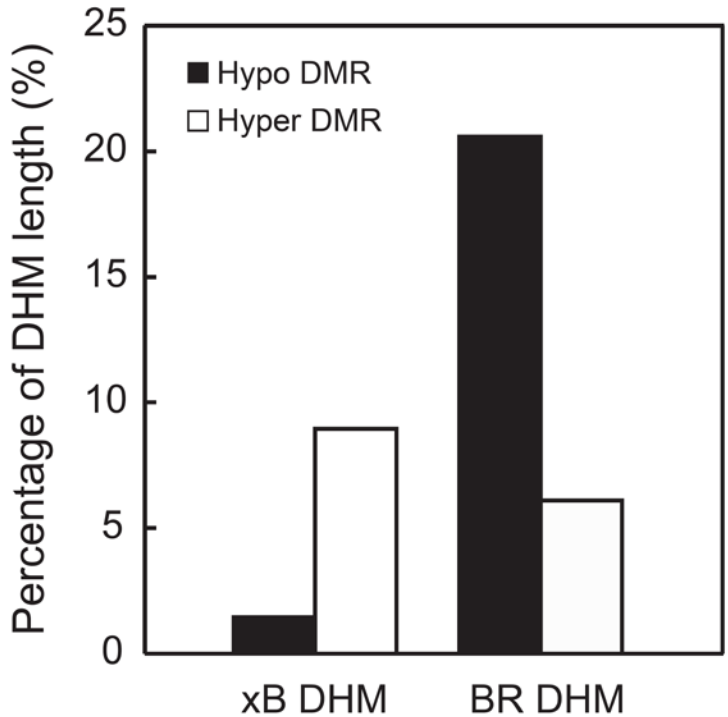

Figure 6. The synchronicity of DHM (Differential H3K9me2-marked region) and DMR (Differentially methylated region) in x*Brassicoraphanus*. The percentage of DHM that is associated with DMR. Each DHM was compared with hypo DMR and hyper DMR and DMR-associated DHMs were merged to calculate its proportion of length to total length of DHM.

Table 6. Summary of the association of DMRs with DHMs in x*Brassicoraphanus*

	Context	xB DHM			BR DHM		
		Total number	Total length	%	Total number	Total length	%
Hyper DMR	CG	90	107,774	61.0	21	13,655	6.0
	CHG	156	172,598	97.7	686	220,700	96.3
	CHH	67	114,926	65.1	20	12,229	5.3
Hypo DMR	CG	1	2,443	8.4	276	235,125	30.4
	CHG	18	20,194	69.8	862	731,787	94.7
	CHH	11	13,339	46.1	329	274,100	35.5

DHM and DMR coordinate gene expression changes

To test if generation and disappearance of H3K9me2 and DNA methylation affects gene expression patterns in x*Brassicoraphanus*, the regions including hyper-DMRs and xB DHMs were compared with differentially expressed genes (DEGs) between x*Brassicoraphanus* and the parents. The expression levels of associated DEGs in x*Brassicoraphanus* were generally found to decrease than in the parents (Figure 7A). For the regions associated with hypo DMRs and BR DHMs, the expression changes of associated DEGs in x*Brassicoraphanus* increase (Figure 7B).

To investigate the patterns of gene expression changes, DEGs found in the regions associated with DMRs and DHMs were classified into up-regulated DEGs (up DEGs), down-regulated DEGs (down DEGs). In Table 7, down DEGs (31.4%) are highly enriched in the regions of hyper DMRs and xB DHMs (Fisher's exact test, P = 9.82E-08). For the regions associated with hypo DMRs and BR DHMs, the proportion of up DEGs is much higher than observed in x*Brassicoraphanus* (Fisher's exact test, P = 0.0017). Taken together, hyper DMR and xB DHMs are often associated with the lower expression of genes in x*Brassicoraphanus* than its parents and hypo-DMR and BR DHMs with the higher expression in x*Brassicoraphanus* than in its parents.

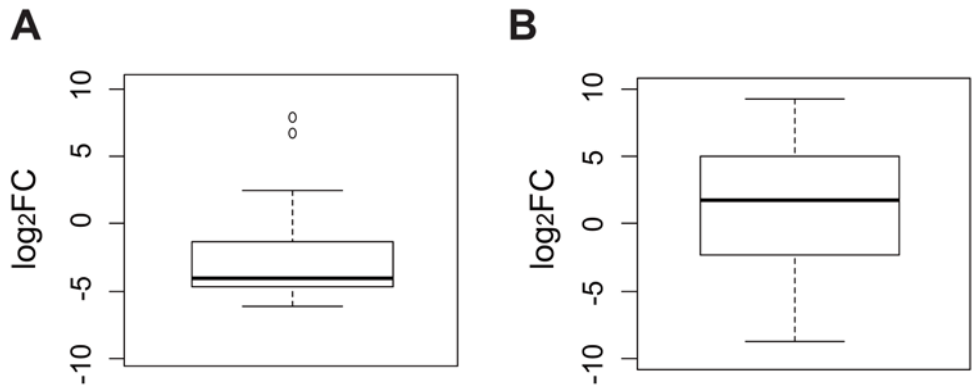

Figure 7. Expression changes of differentially expressed genes (DEGs) associated with DHM and DMR in x*Brassicoraphanus* compared to the parents. Fold changes (log_2) of DEG expression are displayed. Fold change (FC) = (expression in x*Brassicoraphanus*) / (expression in the parents) **(A)** DEGs associated with the regions including hyper DMR and xB DHM. **(B)** DEGs related to the regions of hypo DMR and BR DHM.

Table 7. Summary of the association of DEGs with DHMs and DMRs

	hyper DMR & xB DHM		hypo DMR & BR DHM		x*Brassicoraphanus*	
Up DEG	4	(7.8%)	46	(11.6%)	5,115	(7.2%)
Down DEG	16	(31.4%)	25	(6.3%)	4,676	(6.6%)
Non DEG	31	(60.8%)	325	(82.1%)	61,231	(86.2%)
Total	51	(100%)	396	(100%)	71,022	(100%)

Expression patterns of the enzymes modulating epigenetic modification

To investigate whether some factors that are known to modify epigenetic states of chromatin are engaged in epigenetic changes in x*Brassicoraphanus*, the genes encoding such enzymes were identified in x*Brassicoraphanus* and their expression patterns were comparatively analyzed with the RNA-seq data (analyzed and kindly provided by Hosub Shin). CHROMOMETHYLASE (CMT), DOMAINS REARRANGED METHYLASE (DRM), and DECREASE IN DNA METHYLATION (DDM) are supposed to mediate non-CG DNA methylation and five, eight, and two of their homologues were found respectively in x*Brassicoraphanus* for each family (Table 8). Also, 23 genes were found to be homologous to SUVH that transfers methyl groups to lysine of histones. Most of genes were expressed similarly between x*Brassicoraphanus* and its parents and four of them were confirmed as DEGs (Table 8). In the case of DRM homologs, even though they are classified as DEGs, the disparity of FPKM is not huge and their expression levels are very low. The expression levels of xB01275r and xB11965r that are homologs of SUVH slightly increase in x*Brassicoraphanus* (Table 8). However, the homologs of DRM2, CMT2, CMT3, DDM1, SUVH4, and SUVH5 that are supposed to play crucial roles in maintaining and initially establishing DNA methylation and H3K9me2 in *Arabidopsis* show no expression difference between x*Brassicoraphanus* and the parents (Table 8). Other cues but epigenetic modification enzymes are expected to trigger epigenetic changes in x*Brassicoraphanus* after polyploidization.

Table 8. Expression patterns of the genes related to epigenetic modifications

At ID	xB ID	xB FPKM	BR FPKM	DEG
CMT1	xB23077r	24.68	23.16	None
CMT1	xB66307b	16.00	23.26	None
CMT2	xB12574r	14.35	9.14	None
CMT2	xB25616b	12.79	16.60	None
CMT3	xB57468b	0.58	0.75	None
DDM1	xB47302b	9.10	15.70	None
DDM1	xB64642b	31.56	33.84	None
DRM1	xB03359b	0	0.69	Down
DRM1	xB71947b	771.06	1155.73	None
DRM2	xB10199b	17.71	22.81	None
DRM2	xB12043r	0.53	0.05	UP
DRM2	xB19415r	14.35	19.97	None
DRM2	xB58631b	0.49	1.63	None
DRM3	xB40032b	5.37	8.65	None
DRM3	xB41361r	7.28	7.18	None
SUVH1	xB01275r	13.11	5.68	UP
SUVH1	xB11666r	0	0.19	None
SUVH1	xB20043r	7.03	13.24	None
SUVH1	xB20044r	8.61	8.28	None
SUVH1	xB43996b	9.76	11.93	None
SUVH1	xB56005b	0.16	0.51	None
SUVH1	xB73419b	20.32	21.43	None
SUVH1	xB73420b	11.69	11.08	None
SUVH2	xB54470b	3.15	5.44	None
SUVH2	xB78524b	0.00	0.00	None
SUVH3	xB20911r	16.17	13.35	None
SUVH3	xB24889r	7.18	9.73	None
SUVH3	xB57856b	6.43	10.02	None
KYP/SUVH4	xB00911r	32.61	37.96	None
KYP/SUVH4	xB44396b	50.55	66.39	None
SUVH5	xB29881b	4.21	5.39	None
SUVH5	xB78679b	3.40	2.90	None
SUVH7	xB03235b	0	0	None
SUVH7	xB62149b	0	0	None
SUVH7	xB62150b	0	0	None
SUVH9	xB11965r	22.69	10.62	UP
SUVH9	xB22032r	0.98	2.27	None
SUVH9	xB35894b	24.13	27.27	None

DISCUSSION

During polyploidization epigenetic changes are known to be crucial for stabilizing the hybrid genome. To assess whether histone modifications are involved in mediating epigenomic changes in x*Brassicoraphanus,* the distribution of H3K9me2 across the genome was profiled by the ChIP-seq method in this study. Furthermore, the BS-seq data were combined to figure out the correlation between H3K9me2 and DNA methylation. Also, epigenomic differences between x*Brassicoraphanus,* an allotetraploid, and its parents, *B. rapa* and *R. sativus*, were investigated to comprehend the hybridization effects on gene expression changes. Due to the complexity of ChIP procedures, the mapping results of sequenced reads are not uniform. However, this weakness was complemented by using two biological repeats and utilizing overlapped regions for downstream analyses. Since the length of identified regions that are associated with H3K9me2 varies and is generally long, most results are based on the total length of peaks.

Consistent with previous studies in *Arabidopsis*, H3K9me2 and DNA methylation are mutually colocalized in the x*Brassicoraphanus* genome and their distributions are inversely proportional to the gene density. This tendency is derived from their high concentration at heterochromatin and repeat regions. The exceedingly enriched DNA methylation within H3K9me2 regions implies that general mechanisms modulating those epigenetic states and interdependency between them are universal. Furthermore, genome-wide dispersion of xB DHM and BR DHM suggests that genesis and vanishment of H3 Lys 9 methylation after genome hybridization occurs at random over the whole genome and is considerably

associated with that of CHG methylation. This correlation is coherent with the proposed self-reinforcing state of CHG methyltransferase and H3K9me2-related histone methyltransferase in *Arabidopsis* (Johnson et al., 2007). The association with DEGs indicates that when H3K9me2 and DNA methylation are generated in an allotetraploid, the expression of genes within that area declines in general. On the contrary, if they are dissipated, genes are expressed relatively in high level. Thus, polyploidization between *B. rapa* and *R. sativus* triggered epigenetic changes in genome level, which leads to transcriptional reprogramming.

When epigenetic changes in x*Brassicoraphanus* compared to the progenitors are examined, the occurrence rate of H3K9me2 in Rx subgenome was much higher than that in Ax subgenome (63.4% and 33.9%). Since *B. rapa* was used as a maternal progenitor, the genome of *R. sativus* in an allopolyploid would have experienced a different maternal environment. Responding to cytoplasm of *B. rapa*, chromatin from *R. sativus* would adjust by itself through genome-wide modification of H3K9me2 levels. However, the locations of newly or more established DNA methylation in x*Brassicoraphanus* specifically scatter over Ax subgenome (74.6%, length-based). In contrast to subgenomic differences in forming the new patterns of histone modification and DNA methylation, perishments of the epigenetic patterns equally happened regardless of subgenomic. These results suggest that after hybridization, epigenetic changes have occurred in subgenome-specific ways; more H3K9me2 was preferentially generated towards Rx subgenome but DNA methylation was generated towards Ax subgenome. These biased alterations may have been derived from the expression changes of epigenetic factors or transposons and sRNAs that are explosively activated through

polyploidization (Chen and Yu, 2013). As to assumed factors concerned in epigenomic changes in x*Brassicoraphanus*, there was no striking discrepancy of gene expressions between homeologs. Therefore, associations of sRNAs and transposable elements with DHM or DMR could be further explored to trace mechanisms having induced epigenome changes after intergeneric hybridization.

This study conducted a genome-level investigation of H3K9me2 in an intergeneric hybrid, x*Brassicoraphanus*, and its correlation with DNA methylation was explored. Furthermore, general epigenetic changes in x*Brassicoraphanus* relative to its parents were identified. To my best knowledge, this is the first report on epigenomic variations in a polyploid compared with the progenitors with ChIP-seq method. And this may help to understand the hybridization effects orchestrating chromatin structures more clearly. Furthermore, the study under the subject of intergeneric hybrid, which is rarely found in nature, could provide clues to how two genetically distinct genomes were combined in a nucleus and to the mechanisms permitting genome hybridization, which leads to explanation for evolutionary process in eukaryotes.

REFERENCES

Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I., Reuter, G., and Aalen, R.B. (2001) The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Research 29, 4319-4333.

Becak, M.L. and Kobashi, L.S. (2004) Evolution by polyploidy and gene regulation in Anura. Genetics and Molecular Research 3, 195-212.

Berger, S.L. (2007) The complex language of chromatin regulation during transcription. Nature 447, 407-412.

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. PLoS ONE 3, e3156.

Birchler, J.A. and Veitia, R.A. (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proceedings of the National Academy of Sciences of the United States of America 109, 14746-14753.

Chen, Z.J. (2007) Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids. Annual Review of Plant Biology 58, 377-406.

Chen, Z. J. and Yu, H. H. (2013) Genetic and epigenetic mechanisms for polyploidy and hybridity. *In* Chen, Z. J. and Birchler, J. A. (eds.) Polyploid

and Hybrid Genomics. 335-354. Wiley-Blackwell, New York.

Comai, L. (2005) The advantages and disadvantages of being polyploid. Nature Reviews Genetics 6, 836-846.

Conesa, A., Götz, S, García-Gómez, J.M., Terol, J., Talón, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674-3676.

Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005) The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biology 6, 227.

Doyle, J.J., Flagel, L.E., Paterson, A.H., Rapp, R.A., Soltis, D.E., Soltis, P.S., and Wendel, J.F. (2008) Evolutionary genetics of genome merger and doubling in plants. Annual Review of Genetics 42, 443-461.

Ebbs, M.L., Bartee, L., and Bender, J. (2005) H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. Molecular and Cellular Biology 25, 10507-10515.

Ebbs, M.L. and Bender, J. (2006) Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. Plant Cell 18, 1166-1176.

Gendler, K., Paulsen, T., Napoli, C. (2008) ChromDB: the chromatin database. Nucleic Acids Research 36, 298-302.

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., and Sykes, S. (2011) High-quality draft

assemblies of mammalian genomes from massively parallel sequence data. Proceedings of the National Academy of Sciences of the United States of America 108, 1513-1518.

Grant, V. (1981) Plant Speciation, Columbia University Press, New York.

Hegarty, M.J. and Hiscock, S.J. (2008) Genomic clues to the evolutionary success of polyploid plants. Current Biology 18, R435-R444.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime *cis*-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell 38, 576-589.

Initiative A.G. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796-815.

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, 556-560.

Jasencakova, Z., Soppe, W.J.J., Meister, A., Gernand, D., and Turner, B.M. (2003) Histone modifications in *Arabidopsis* - high methylation of H3 lysine 9 is dispensable for constitutive heterochromatin. Plant Journal 33, 471-480.

Jenuwein, T. and Allis, C.D. (2001) Translating the Histone Code. Science 293, 1074-1080.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S. et al. (2011) Ancestral

polyploidy in seed plants and angiosperms. Nature 473, 97-100.

Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I.R., Callis, J., and Jacobsen, S.E. (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. Current Biology 17, 379-384.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., and Garcia-Hernandez, M. (2012) The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research 40, 1202-1210.

Langmead B., Trapnell C., Pop M., and Salzberg S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10, R25.

Lee, S.S., Choi, W.J., and Woo, J.G. (2002) Development of a new vegetable crop in *Brassicoraphanus* by hybridization of *Brassica campestris* and *Raphanus sativus*. Journal of Korean Society for Horticultural Science 43, 693-698.

Lee, S.S., Lee, S.A., Yang, J., and Kim, J. (2011) Developing stable progenies of x*Brassicoraphanus*, an intergeneric allopolyploid between *Brassica rapa* and *Raphanus sativus*, through induced mutation using microspore culture. Theoretical and Applied Genetics 122, 885-891.

Leitch, A.R. and Leitch, I.J. (2008) Genomic plasticity and the diversity of polyploid plants. Science 320, 481-483.

Levin, D.A. (1983) Polyploidy and novelty in flowering plants. The American Naturalist 122, 1-25

Levin, D.A. (2002) The Role of Chromosomal Change in Plant Evolution. Oxford University Press, New York.

Li, B., Carey, M., and Workman, J.L. (2007) The Role of Chromatin during Transcription. Cell 128, 707-719.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430, 471-476.

Liu, C., Lu, F., Cui, X., and Cao, X. (2010) Histone Methylation in Higher Plants. Annual Review of Plant Biology 61, 395-420.

Malagnac, F., Bartee, L., and Bender, J. (2002) An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. The EMBO Journal 21, 6842-6852.

Mardis, E.R. (2007) ChIP-seq: welcome to the new frontier. Nature 4, 613-614.

Mathieu, O., Probst, A.V., and Paszkowski, J. (2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis*. The EMBO Journal 24, 2783-2791.

Mayrose, I., Zhan, S.H., Rothfels, C.J., Magnuson-Ford, K., Barker, M.S., Rieseberg, L.H., and Otto, S.P. (2011) Recently formed polyploid plants diversify at lower rates. Science 333, 1257.

McClintock, B. (1984) The significance of responses of the genome to challenge. Science 226, 792-801.

Metzker, M.L. (2010) Sequencing technologies — the next generation. Nature Reviews Genetics 11, 31-46.

Michael, T.P. and Jackson, S. (2013) The first 50 plant genomes. Plant Genome 6.

Ni, Z., Kim, E.D., Ha, M., Lackey, E., Liu, J., Zhang, Y., Sun, Q., and Chen, Z.J. (2009) Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. Nature 457, 327-331.

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. (2003) Understanding mechanisms of novel gene expression in polyploids. Trends in Genetics 19, 141-147.

Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M.A., and Ainouche, M. (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytologist 184, 1003-1015.

Paterson, A.H. (2005) Polyploidy, evolutionary opportunity, and crop adaptation. Genetica 123, 191-196.

Premachandran, M.N., Prathima, P.T., and Lekshmi, M. (2011) Sugarcane and polyploidy: a review. Journal of Sugarcane Research 1, 1-15.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T.L., Wilson, C.J., Bell, S.P., and Young, R.A. (2000) Genome-wide location and function of DNA binding proteins. Science 22, 2306-2309.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-

terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74, 5463-5467.

Schadt, E.E., Turner S., and Kasarskis, A. (2010) A window into third-generation sequencing. Human Molecular Genetics 19, 227-240.

Schatlowski, N. and Kohler, C. (2012) Tearing down barriers: understanding the molecular mechanisms of interploidy hybridizations. Journal of Experimental Botany 63, 6059-6067.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119, 941-953.

Solomon, M.J., Larsen, P.L., and Varshavsky, A. (1988) Mapping protein-DNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene. Cell 53, 937-947.

Soltis, D.E., Visger, C.J., and Soltis, P.S. (2014) The polyploidy revolution then. . . and now: stebbins revisited. American Journal of Botany 101, 1057-1078.

Song, Q. and Chen, Z.J. (2015) Epigenetic and developmental regulation in plant polyploids. Current Opinion in Plant Biology 24, 101-109.

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*. The EMBO Journal 21, 6549-6559.

Springer, N.M., Napoli, C.A., Selinger, D.A., Pandey, R., Cone, K.C., Chandler, V.L., Kaeppler, H.F., and Kaeppler, S.M. (2003) Comparative analysis of

SET domain proteins in maize and *Arabidopsis* reveals multiple duplications preceding the divergence of monocots and dicots. Plant Physiology 132, 907-925.

Stebbins, G.L. (1950) Variation and Evolution in Plants. Columbia University Press, New York.

Stebbins, G.L. (1971) Chromosomal Evolution in Higher Plants. Edward Arnold, London.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014) Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. Nature structural & molecular biology 21, 64-72.

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

Tariq, M., Saze, H., Probst, A.V., Lichota, J., Habu, Y., and Paszkowski, J. (2003) Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. Proceedings of the National Academy of Sciences of the United States of America 100, 8823-8827.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. Nature 439, 811-816.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007) *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genetics 3, e86.

Van de Peer, Y. and Meyer, A (2005) Large-scale gene and ancient genome duplications. *In* Gregory, T. R. (eds.) The Evolution of the Genome. 330-363. Elsevier, San Diego.

Van de Peer, Y., Maere, S., and Meyer, A. (2009) The evolutionary significance of ancient genome duplications. Nature Reviews Genetics 10, 725-732.

Wang, J., Tian, L., Lee, H.S., Wei, N.E., Jiang, H., Watson, B., Madlung, A., Osborn, T.C., Doerge, R.W., Comai, L. et al. (2006) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics 172, 507-517.

Watson, J.D. and Crick, F.H.C. (1953) Molecular structure of nucleic acids. Nature 171, 737-738.

Wendel, J.F. (2000) Genome evolution in polyploids. Plant Molecular Biology 42, 225-249.

Ye, J. , Fang, L. , Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, Y., Bolund, L., and Wang, J. (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Research 34, 293-297.

Yoo, M.J., Liu, X., Pires, J.C., Soltis, P.S., and Soltis, D.E. (2014) Nonadditive gene expression in polyploids. Annual Review of Genetics 48, 485-517.

Zhang, X. (2008) The Epigenetic Landscape of Plants. Science 320, 489-492.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. et al. (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biology 9, R137.

Zhao, Z. and Shen, W.H. (2004) Plants contain a high number of proteins showing sequence similarity to the animal SUV39H family of histone methyltransferases. Annals of the New York Academy of Sciences 1030, 661-669.

초 록

후성유전학적 변화는 배수화된 식물에서 동위 유전자의 발현을 조절하고 유전체의 안정성에 영향을 끼친다. 히스톤 변형 중 히스톤 단백질 H3의 아홉 번째 라이신에 두 개의 메틸기가 붙은 경우 (H3K9me2), 애기장대에서는 이질염색질에서 주로 찾아볼 수 있다. 또한 DNA 메틸레이션을 조절하며 트랜스포존의 발현을 억제한다고 알려져 있다. 염색질 면역침강법과 시퀀싱을 진행하여, 배추와 무의 속간교잡으로 탄생한 배무채에서 H3K9me2의 위치를 추적하였다. 메틸레이션 데이터와 비교하였을 때 배무채의 유전체에서 H3K9me2와 DNA 메틸레이션의 높은 상관관계가 발견되었다. 이러한 경향성은 배수화 현상이 일어나서 배무채가 생성되고 이 때 사라지거나 새롭게 생성된 H3K9me2에서도 발견할 수 있었고 주로 CHG 메틸레이션과 연관되어 있었다. 부모와 비교하였을 때 배무채에서 DNA 메틸레이션 레벨이 낮아진 지역과 H3K9me2이 없어진 지역은 배무채 전체 유전체 지역에서 고르게 분포한다. 반면 DNA 메틸레이션의 수준이 높아진 지역과 H3K9me2이 새로 생성된 지역은 배추, 혹은 무 하위유전체에 특이적으로 많이 존재한다. 이러한 결과는 후성유전학적 메커니즘이 서로 연관되어 있고 두 속이 교잡되면서 많은 영향을 받았다는 것을 보여준다.

주요어: 히스톤 변형, 염색질면역침강법, 배무채, 속간교잡종, DNA 메틸레이션, 후성유전학

학번: 2014-20725

A Thesis for the Degree of Master of Science

Genome-Wide Association between Histone H3 Lysine Methylation and DNA Methylation in an Intergeneric Hybrid, x*Brassicoraphanus*

속간교잡종 배무채에서 히스톤 H3 라이신 9 와 DNA 메틸레이션의 전장유전체연관

FEBRUARY, 2016

HYE YEON SOH

MAJOR IN INTERDISCIPLINARY PROGRAM IN
AGRICULTURAL GENOMICS
THE GRADUATE SCHOOL OF
SEOUL NATIONAL UNIVERSITY

Genome-Wide Association between Histone H3 Lysine Methylation and DNA Methylation in an Intergeneric Hybrid, x*Brassicoraphanus*

UNDER THE DIRECTION OF DR. JIN HOE HUH
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
SEOUL NATIONAL UNIVERSITY

BY
HYE YEON SOH

MAJOR IN
INTERDISCIPLINARY PROGRAM IN AGRICULTURAL GENOMICS
THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

FEBRUARY, 2016

APPROVED AS A QUALIFIED THESIS OF HYE YEON SOH
FOR THE DEGREE OF MASTER OF SCIENCE
BY THE COMMITTEE MEMBERS

CHAIRMAN _____
Yong-Hwan Lee, Ph.D.

VICE-CHAIRMAN _____
Jin Hoe Huh, Ph.D.

MEMBER _____
Doil Choi, Ph.D.

Genome-Wide Association between Histone H3 Lysine Methylation and DNA Methylation in an Intergeneric Hybrid, x*Brassicoraphanus*

HYE YEON SOH

INTERDISCIPLINARY PROGRAM IN

AGRICULTURAL GENOMICS

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

ABSTRACT

Epigenetic changes in allopolyploids play a pivotal role in reprogramming homeologous gene expression and genomic stability. Dimethylation of H3K9 (H3K9me2) marks heterochromatin and is crucial in controlling DNA methylation and silencing transposable elements in *Arabidopsis*. In this study, chromatin immunoprecipitation accompanied with sequencing method (ChIP-seq) was employed to profile the distribution of H3K9me2 in intergeneric hybrid, x*Brassicoraphanus*. Compared with methylome data, a high rate of association of H3K9me2 with DNA methylation was observed throughout the genome. In addition, occurrence and disappearance of H3K9me2 in x*Brassicoraphanus* after polyploidization is intensely correlated with that CHG methylation. Extinction of DNA methylation and H3K9me2 in the allopolyploid happened regardless of its

subgenomic location. However, the newly generated epigenetic states showed biased subgenomic proportion. These results confirm that epigenetic mechanisms are mutually correlated and suggest that effects of hybridization between two genera of genetic distance would trigger differential epigenetic modulation in x*Brassicoraphanus*.

Key Words**:** Epigenetic change, H3K9me2, x*Brassicoraphanus*, ChIP-seq, DNA methylation, allopolyploid

Student number: 2014-20725

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

H3K9me2	Histone H3 lysine 9 dimethylation
ChIP	Chromatin immunoprecipitation
H3K4me	Histone H3 lysine 4 methylation
H3K27me	Histone H3 lysine 27 methylation
FLC	Flowering Locus C
CCA1	CIRCADIAN CLOCK ASSOCIATED 1
LHY	LATE ELONGATED HYPOCOTYL
TOC1	TIMING OF CAB EXPRESSION1
GI	GIGANTEA
WGD	Whole Genome Duplication
PRMT	Protein Arginine Methyltransferase
HKMT	Histone Lysine Methyltransferase
Su(var)3-9	Suppressor of variegation 3-9
E(z)	Enhancer of zeste
TrxG	Trithorax group protein
ASH1	Absent, small, or homeotic discs 1
LSD1	Lysine-specific demethylase 1
JmjC	Jumonji C

SUVH	Su(var)3-9 homolog
KYP	KRYPTONITE
rDNA	ribosomal DNA
NGS	Next-Generation sequencing
TE	Transposable element
sRNA	Small RNA
DHM	Differential histone-marked region
DMR	Differentially methylated region
DEG	Differentially expressed gene
CMT	CHROMOMETHYLASE
DDM	DECREASE IN DNA METHYLATION
DRM	DOMAINS REARRANGED METHYLASE
TTS	Translation termination site
BS-seq	Bisulfite sequencing

INTRODUCTION

Polyploidization is one of the most important cues for eukaryotic evolution. Most of angiosperms and the majority of pteridophytes (~95 %) have experienced series of polyploidization (Soltis, 2014; Grant, 1981). Scrutinizing the grounds and outcomes of polyploidy is essential to understand the diversity of flowering plants and pteridophytes as well as other eukaryotic pedigrees. Numerous major crops such as oats, soybean, wheat, cotton, canola, and banana were generated through natural polyploidization, which allowed crop domestications and the advent of valuable traits (Chen and Yu, 2013). Genome-level investigations on these crops have recently been published and several genetic and epigenetic mechanisms related to genome hybridization have been gradually uncovered.

x*Brassicoraphanus* (2n=38) is a newly synthesized species from a cross between *Brassica rapa* and *Raphanus sativus* (Lee et al., 2002). Intergeneric hybridizations between these two species have been recurrently attempted since early 1800's and a genetically stable allopolyploid was developed through induced mutation with the microspore culture method (Lee et al., 2011). x*Brassicoraphanus* shows intermediate phenotypes between *B. rapa* and *R. sativus* in hypocotyl length, the leaf shape, and the structure of silique. One of its novel phenotypes is the white petal color while the flowers of *B. rapa* and *R. sativus* have yellow and purple color, respectively. x*Brassicoraphanus* is genetically stabilized within a short period of time as a rare case of intergeneric hybrid. The genome and epigenome of x*Brassicoraphanus* represent instant hybridization effects. The genetic distance

between the two parents would be beneficial to distinguish the parental origin of duplicate genes. Studying the unique allopolyploid, x*Brassicoraphanus* could provide comprehensive understanding of genome rearrangement occurring within an intergeneric hybrid and the developmental mechanism of neomorphism. Furthermore, speciation and evolutionary process derived from intergeneric crossing also could be explored.

Previous studies showed that histone modifications mediate gene transcription during growth and development of plants and animals by remodeling the chromatin structure (Berger, 2007; Zhang, 2008). In *Arabidopsis thaliana* active promoters contain several modifications, such as histone H3 and H4 acetylations and methylation of H3 lysine 4 (H3K4me 1/2/3) that are also known as euchromatic marks. By contrast, repressed gene expressions are associated with methylations of H3 lysine 9 (H3K9) and H3 lysine 27 (H3K27) that are known to mark heterochromatin (Jenuwein and Allis, 2001; Li et al., 2007). Out of these several modifications, di-methylation of H3K9 (H3K9me2) is dispersed across pericentromeric heterochromatin and the genome-wide location of H3K9me2 is known to be highly correlated with transposable elements and repeat regions (Lippman et al., 2004; Turck et al., 2007; Bernatavichute et al., 2008).

Epigenetic changes in newly hybridized interspecific allopolyploids are supposed to play a crucial role in genomic reorganization and reprogramming chromatin structures that give rise to gene expression changes (Osborn et al., 2003; Chen, 2007). In *Arabidopsis*, a synthetic allotetraploid shows different phenotypes including flowering time and morphological vigor compared to its parents *A. thaliana* (*Ler*) and *A. arenosa* (Wang et al., 2006; Ni et al., 2009). Biochemical and

genome-level studies revealed that changes of histone modifications in the promoters of *FLC* gene and the circadian clock genes (*CCA1*, *LHY*, *TOC1*, and *GI*) are associated with the altered expression in allopolyploids and hybrids (Wang et al., 2006; Ni et al., 2009).

In this study, chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) analysis was employed to profile H3K9me2 across the entire x*Brassicoraphanus* genome. With DNA methylation data, nucleotide-resolution histone modification (H3K9me2) profiling data were analyzed to understand general states of epigenome in x*Brassicoraphanus* and the epigenetic correlation between DNA methylation and H3K9me2 was revealed. Also by comparing genome-wide distributions of H3K9me2 between x*Brassicoraphanus* and its parents, *B. rapa* and *R. sativus*, epigenetic changes including DNA methylation and subsequent transcriptomic changes after polyploidization were explored. Generation of DNA methylation and H3K9me2 showed differences in terms of their subgenomic sites. Hybridization effects and genomic incompatibility between two progenitors are considered to be putative causes for this phenomenon. This research would provide insights into comprehensive epigenetic states occurred after genome hybridization and set up the basis for further studies of epigenetic mechanisms of polyploidization.

LITERATURE REVIEWS

1. Polyploidy

Polyploidy depicts an organism or a cell that possesses multiple sets of chromosomes. In the majority of cases, even numbered sets of chromosomes are included in polyploids. The research of eukaryotic genome sequence suggests that genomes of many species show marks of polyploidal ancestor (Paterson, 2005; Becak and Kobashi, 2004; Van de Peer and Meyer, 2005).

Polyploidization typically gives rise to biodiversity and instantaneous speciation, acting as a major driving force of macro-evolution (Levin, 1983; Levin, 2002; Mayrose et al., 2011). This process is a common phenomenon that can be easily found in plants and also among fish and amphibians (Leitch et al., 2008; Comai, 2005; Wendel, 2000; Song and Chen, 2015; Hegarty and Hiscock, 2008; Doyle et al., 2008; Soltis et al., 2014) Most of the flowering plants are believed to have experienced whole-genome duplications (WGDs) (Soltis et al., 2014). Two rounds of WGDs before the divergence of flowering plants are estimated to have occurred and consequently diversify genes that are crucial for the development of seeds and flowers, which gives clues for Charles Darwin's abominable mystery (Jiao et al., 2011; Van de Peer et al., 2009).

There are two types of polyploidy: autopolyploidy and allopolyploidy that are grouped by its origin and the arrangement of chromosomes (Stebbins, 1950; Song and Chen, 2015). Autopolyploids (AAAA in an autotetraploid) are derived from a single genome duplication of one individual or different plants within a species. On the other hand, allopolyploids are formed by hybridizing and doubling

of chromosomes between two different species. Therefore, two parental genomes exist in a nucleus and allotetraploids (AABB) have two of each pair of equivalent chromosomes from the parent species, or homeologous chromosomes that are also known as homeologs. Some polyploid plants such as sugarcane are hard to distinguish between them (Premachandran et al., 2011). Since intraspecific pairing is more frequent than interspecific mating, allopolyploidy was estimated to be less widespread than autopolyploidy (Stebbins, 1971). However, allopolyploids are more universal than autopolyploids in nature (Schatlowski and Kohler, 2012; Hegarty and Hiscock, 2008).

When a nucleus of allopolyploid contains two parental genomes, the expression levels of homeologous genes could display divergence from the arithmetic average, also known as additivity, of parental gene expression patterns. The putative causes of those nonadditive gene expression patterns are maternal-paternal influences, gene dosage balance, *cis*- and *trans*- regulations, and dysregulation of transposons (Yoo et al., 2014). As maternal organellar genomes are generally inherited in plants, the expression dominance toward the maternal parent was observed in many polyploids; cotton, coffee, *Spartina anglica*, and *Tragopogon* (Yoo et al., 2014). Secondly, the gene-balance hypothesis explains that the function of the whole complex is regulated by the stoichiometry among parts of macromolecular complexes (Birchler and Veitia, 2012). According this concept, unequally fractionated subgenomes can describe the expression dominance toward one homeologous gene. On the other hand, combined *cis*- and *trans*- regulatory systems from two divergent parents can modulate the expression of homeologs. Due to sequence similarities between two parents, trans-acting factors may

crosstalk over homeologous *cis*-elements in allopolyploids. Lastly, epigenetic changes during interspecific hybridization may alter the expressions of transposable elements, which leads to the modulation of homeologs in allopolyploids (McClintock, 1984; Parisod et al., 2009).

2. Histone modification

Histone proteins comprise the nucleosome, a fundamental unit of chromatin, and several modifications on histone proteins modulate chromatin structure through epigenetic mechanisms. Histone modification mainly occurs on amino (N) terminal tails that protrude from core histone proteins. Various posttranslational modifications can take place on histone residues by forming covalent bonds with different chemical groups, which include acetylation, methylation, phosphorylation, ubiquitination, sumoylation, glycosylation, and ADP-ribosylation (Liu et al., 2010). The histone code hypothesis proposes that the orchestration of histone modifications determines chromatin structures, which leads to the activation or repression of gene transcription. (Jenuwein et al., 2001)

Histone methylation, one of the most commonly studied forms, plays a crucial role in regulating diverse biological processes ranging from heterochromatin formation to transcriptional regulation. It can be mono- and di-methylated on arginine or mono-, di-, and tri- methylated on lysine residues by protein arginine methyltransferases (PRMTs) and histone lysine methyltransferases (HKMTs), respectively.

Lysine methylation of histone is found in both transcriptionally silenced and active chromatin regions, according to the location of lysine residues binding

to methyl groups and the degree of methylation. Conjugating a methyl group to lysine residues does not alter the net charge of the modified products, but the hydrophobicity increases. Also, specific proteins that recognize the methylated residues may be recruited to newly-made binding surfaces or intra- and intermolecular interactions may change their patterns.

First discovered in *Drosophila melanogaster* are SET domain proteins conserved in three proteins and named after them; Suppressor of variegation 3-9 (Su(var)3-9), Enhancer of zeste (E(z)), and trithorax group proteins (TrxG) (Dillon et al., 2005). SET domain proteins are known to catalyze the lysine methylation and there are 41 and 37 genes that encode SET domain proteins respectively in *Arabidopsis* and rice genomes (Gendler et al., 2008). SET domain proteins in plants are homologous to those in animals and yeast and grouped into four categories; SU(VAR)3–9 groups, E(Z) homologs, TrxG groups, and ASH1 (absent, small, or homeotic discs 1) groups (Baumbusch et al., 2001; Springer et al., 2003; Zhao et al., 2004). Even though the specificity and characteristics of these SET proteins in plants have not been deeply studied in every detail, several genetic researches showed that they may affect the same lysine residues or related pathways as in animals or yeast (Liu et al., 2010). On the other hand, the methyl groups on lysine can be removed by histone demethylases that perform vital roles in regulating homeostasis of histone methylation. The two types of demethylases work in distinct ways; lysine-specific demethylase1 (LSD1) that oxidizes amine group and Jumonji C (JmjC) domain-containing proteins that hydroxylate the methylated substrate both result in withdrawing methyl groups from methylated lysine residues (Shi et al., 2004; Tsukada et al., 2006).

In *Arabidopsis*, Lys 4 (K4), Lys 9 (K9), Lys 27 (K27), and Lys 36 (K36) of histone H3 are main targets for histone lysine methylation. In general, methylation of histone H3K9 and H3K27 is related to silenced regions, whereas methylation of H3K4 and H3K36 is related to active gene expression (Berger et al., 2007). Mono- and di- methylation of H3K9 is especially more abundant in chromocenters than in euchromatin regions and highly associated with transposons and related repeat regions (Lippman et al., 2004; Turck et al., 2007; Bernatavichute et al., 2008). In *Arabidopsis* genome, ten SU(VAR)3-9 homologues are expected to function (SUVH1-10; Baumbusch et al., 2001). Among them, KRYPTONITE/SUVH4 (KYP), SUVH5 and SUVH6 have been reported to be in charge of the propagation of H3K9me1 and H3K9me2 (Jackson et al., 2002; Ebbs et al., 2005; Ebbs et al., 2006; Jasencakova et al., 2003). *kyp suvh5 suvh6* triple mutants not only eliminate H3K9me2 levels, but also show the reduced level of DNA methylation and release the silent state of heterochromatin as transposons are reactivated (Ebbs et al., 2006). *kyp* mutants discharge the silencing state insignificantly and do not show any defects in morphology despite its role in repressing genes and transposable elements (Tariq and Paszkowski, 2004).

In *met1* plants that lack the function of maintenance of CG methylation, the distribution of H3K9me2 was found to be massively relocated across the genome (Mathieu et al., 2005; Soppe et al., 2002; Tariq et al., 2003), and the possibility that CG methylation manipulates the level of H3K9me2 was proposed (Mathieu et al., 2005; Soppe et al., 2002; Tariq et al., 2003; Tariq and Paszkowski, 2004). Moreover, KYP/SUVH4-mediated H3K9 methylation was reported to manage non-CG methylation (Jackson et al., 2002; Malagnac et al., 2002). Chromatin

immunoprecipitation (ChIP) analysis accompanied with high-resolution genome wide microarray (ChIP-chip) also showed the high correlation between H3K9me2 and CHG DNA methylation (Bernatavichute et al., 2008). Histone H3K9 methylation is essential to maintain non-CG methylation levels and they play a pivotal role in repressing transposons, ribosomal RNA genes (rDNA), and repeat sequences (Liu et al., 2010).

3. Sequencing technologies and ChIP-seq

After the DNA structure as double helix was discovered by Watson and Crick in 1953 (Watson and Crick, 1953), Frederick Sanger proposed a modern DNA sequencing method with chain-terminating inhibitors (Sanger et al. 1977). In this method, DNA polymerase synthesizes complementary DNA sequences using normal 2'-deoxynucleotides (dNTPs) and 2', 3'-dideoxynucleotides (ddNTPs) that terminate the reaction. Then DNA fragments with various sizes are separated using gel electrophoresis method and examined to uncover its sequence. After several modifications has been made, the automated Sanger sequencing has been spread out for two decades but limitations of this first-generation technology made it to be replaced by next-generation sequencing (NGS). These new technologies include steps of template preparation, sequencing and imaging, and data analysis (Metzker, 2010). Compared to the former method, NGS platforms have drastically lowered the cost and enabled higher throughput.

Since the genome of *Arabidopsis* was completed in 2000 (Initiative AG, 2000), a number of plant genomes have been sequenced with NGS technology and exceed more than 100 (Michael and Jackson, 2013). Even with the handy tools for

genome assembly, plant genome still remains to have many barriers that block complete analysis because of its tremendous genome size, repeat content, paralogy, and heterozygosity. Less than 200 bp-long reads from NGS technologies are too short to resolve complexity of plant genomes. While most mammalian genomes have been assigned into chromosome scale regions (Gnerre et al., 2011), most draft plant genomes comprise numerous contigs and scaffolds that shatter with embedded gaps. Among plant genomes, the *Arabidopsis* genome is unquestionably known to be the best organized one, but still scattered in 102 contigs and its total gap length is 185,644 bp (TAIR 10; Lamesch et al., 2012).

A third generation sequencing technology has emerged to cover the limitations of the former technologies and focuses on single molecule DNA. This new generation platform can be grouped into three categories; sequencing-by-synthesis technology that utilizes DNA polymerase to synthesize a single molecule of DNA, nanopore-sequencing technology that detects individual bases as DNA strand goes through the nanopore, and microscopy technology that directly makes image of DNA molecules. These single-molecule-based sequencing technologies have provided many advantages. First, they can process more throughput for much shorter time at the lower price. Also from a diminutive amount of starting material, they can produce longer reads that are helpful in assembling the de novo genome (Schadt et al., 2010).

To study the interactions between DNA and protein, Varshavsky and his colleagues first proposed the chromatin immunoprecipitation (ChIP) method in 1988 (Solomon et al., 1988). This experiment is basically composed of three steps; to cross-link between DNA and proteins with formaldehyde, to immunoprecipitate

the protein-DNA complex with specific antibody, and to reverse the cross-links and analyze recovered DNA. Since DNA and proteins are randomly cross-linked and non-specific interactions can occur during the immunoprecipitation step, ChIP-derived DNA is compared to a mock sample that is collected after the immunoprecipitation process done without antibody. For genome-wide analysis of ChIP-selected DNA in yeast, Ren and colleagues primarily compared those two samples by hybridizing to a genomic microarray (ChIP-chip) (Ren et al., 2000). With the advent of next-generation sequencing technologies, ChIP-chip have been substituted by ChIP followed by sequencing m (ChIP-seq). ChIP-seq provides many advantages including higher resolution, lower price per the same amount of data, less complicated process, and requirement for less input quantity and fewer replicates (Mardis, 2007).

MATERIALS AND METHODS

Plant materials and growth conditions

x*Brassicoraphanus*, *Brassica rapa* cv. Chiifu-401-42, and *Raphanus sativus* cv. WK10039 plants were grown on 1 X Murashige and Skoog medium (Duchefa, Haarlem, The Netherlands) in a growth chamber under 16 h of illumination for 14 days. The whole seedlings including shoots and roots were harvested together for the analysis of chromatin modifications. Plants grown for 14 days and for 3 months are shown in Figure 1.

Isolation and immunoprecipitation of chromatin

Chromatin Immunoprecipitation (ChIP) was performed using the seedlings (~ 3 g) of x*Brassicoraphanus*, *B. rapa*, and *R. sativus* respectively. Whole plant tissues were combined and cross-linked in 1% formaldehyde for 15 min using vacuum infiltration method and the cross-linking reaction was quenched with 0.125 M of glycine. Plants were washed three times with deionized water and blotted to remove water. Frozen tissues in liquid nitrogen were ground to fine powder and resuspended in 30 ml of extraction buffer 1 (0.4 M sucrose, 10 mM Tris-HCl pH 8.0, 5 mM beta-mercaptoethanol, 1 mM phenylmethylsulfonyl [PMSF], and complete protease inhibitor; Roche Diagnostics, Basel, Switzerland), then filtered through two layers of Miracloth (Merk Miilipore, Guyancourt, France) twice. The filtrate was centrifugated at 2,800 g at 4 ^{o}C for 25 min. The pellet was resuspended in 1 ml of extraction buffer 2 (0.25 M sucrose, 10 mM Tris-HCl p H 8.0, 10 mM $MgCl_2$, 1% Triton X-100, 5 mM beta-mercaptoethanol, 1 m

M PMSF, and complete protease inhibitor; Roche Diagnostics, Basel, Switzerland) and centrifugated at 12,000 g at 4 ^{o}C for 10 min. This procedure was performed twice for better nuclei isolation. The pellet was resuspended in 300 μl of extraction buffer 3 (1.7 M sucrose, 10 mM Tris-HCl pH 8.0, 2 mM $MgCl_2$, 0.15% Triton X-100, 5 mM beta-mercaptoethanol, 1 mM PMSF, and complete protease inhibitor; Roche Diagnostics, Basel, Switzerland) and layered on top of 1.2 ml of clean extraction buffer 3, then centrifugated at 16,000 g at 4 ^{o}C for 1 hr. The chromatin pellet was resuspended in 300 μl of nuclei lysis buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS, and complete protease inhibitor; Roche Diagnostics, Basel, Switzerland) and sonicated with a Sonifier (Branson Sonifier, Danbury, USA) to shear DNA into 0.3 ~ 1 kb. The sonicated chromatin was centrifugated and the supernatant was diluted 10-fold with ChIP dilution buffer (1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.0, 167 mM NaCl). Seven ug of the antibody against histone H3 dimethyl Lys 9 (AB1220, Abcam, Cambridge, UK) was reacted respectively with 1.5 mg of protein G Dynabeads (Invitrogen, CA, USA) for 2 hr at 4 ^{o}C, then chromatin extract was incubated with antibody/magnetic bead mix at 4 ^{o}C overnight. After several washes, immunocomplex was eluted twice from the beads with 30ul of elution buffer (1% SDS, 0.1 M sodium bicarbonate) and reverse cross-linked in the presence of 200 mM NaCl at 65 ^{o}C overnight. To remove all proteins proteinase K was treated, and then DNA was purified by phenol/chloroform/isoamylalcohol extraction and precipitated with ethanol. The pellet was resuspended with 30 μl of tridistilled water. A small aliquot of untreated sonicated chromatin was reverse cross-linked and used as total input DNA control. ChIP experiments were conducted with two biological replicates samples to reduce

errors derived from experimental variations.

Computational analysis of sequencing data

ChIP-seq libraries were constructed as described in the Illumina ChIP sequencing kit (Illumina, CA, USA). DNA fragments with a range of ~500 bp were excised and amplified for cluster generation and sequencing. All DNA libraries were sequenced on a HiSeq2500 (50 single end reads) (Illumina, CA, USA). The sequencing reads were quality controlled with FastQC and aligned to the reference genome of x*Brassicoraphanus* by Bowtie software using default settings (Langmead et al., 2009). Reads that could be assigned to only one region of the genome were used for downstream analyses. To find significant peaks compared to control, MACS software (Zhang et al., 2008) was employed and common regions from two biological replicates were extracted with mergePeaks command from Homer software (Heinz et al., 2010). For comparative analysis between x*Brassicoraphanus* and its parent, *B. rapa* and *R. sativus*, all peaks were classified into three groups; common peaks found in both allopolyploid and its progenitors, xB-specific peaks and BrRs-specific peaks that only exist in x*Brassicoraphanus* or in *B. rapa* and *R. sativus.*

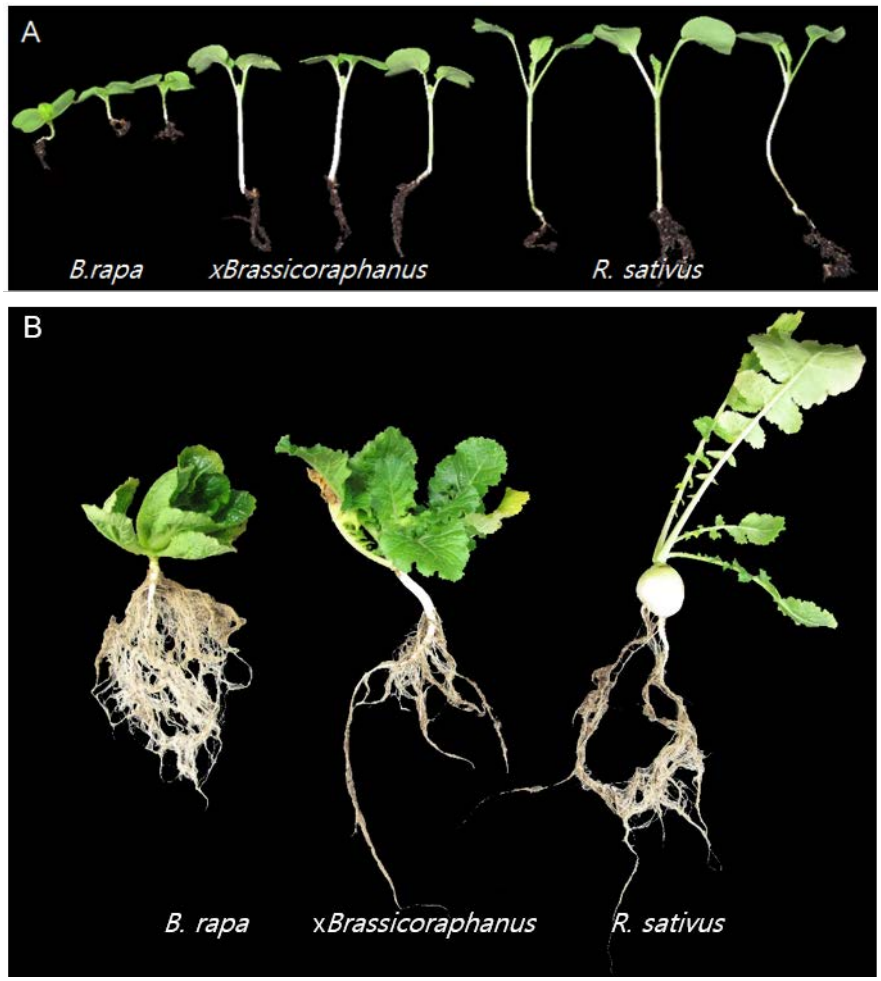

Figure 1. Phenotypes of *Brassica rapa*, x*Brassicoraphanus*, and *Raphanus sativus*. (A) Whole seedlings grown for 14 days that were used in this study and (B) whole plants grown for 3 months (Courtesy of Dr. Gibum Yi).

RESULTS

Genome-wide detection of H3K9me2-enriched regions

Cross-linking chromatin immunoprecipitation sequencing (X-ChIP-seq) methods employing formaldehyde fixation and sonication were utilized to get the genome-wide histone modification pattern data for x*Brassicoraphanus*, *B. rapa* cv. Chiifu, and *R. sativus* cv. WK10039. About 30 million reads were obtained from the ChIP-seq process for each sample. All reads obtained from three genera were mapped to the genome of x*Brassicoraphanus* and uniquely aligned reads (6~30%) were only used for the peak calling process (Table 1). By comparing with control, significantly H3K9me2-enriched regions (peaks) were detected for each replicate. The common peaks were merged from two replicates and regarded as representative H3K9me2-related regions for each plant. For x*Brassicoraphanus*, 12,068 peaks were detected with the length of 9,324,563 bp of sequenced genome (1.35%) (Table 2). And 11,986 and 11,589 peaks were allocated to *B. rapa* and *R. sativus* spanning 10,043,125 bp and 11,863,241 bp, respectively.

To investigate the epigenomic changes occurred after polyploidization, differential binding sites of histone modification (DHM; differential histone modification-marked region) between x*Brassicoraphanus* and its progenitors were identified. The peak file of x*Brassicoraphanus* and that of the parents were compared and the whole regions were classified into three types; the common peaks that exist in both x*Brassicoraphanus* and its parents, x*Brassicoraphanus*-specific peaks (xB DHMs) and *B. rapa* or *R. sativus*-specific peaks (BR DHMs) that are enriched in x*Brassicoraphanus* and in its parents respectively. A total of

4,173 peaks covering 8,208,133 bp are sorted into common peaks (Table 3). Table 3 shows that xB DHMs contain 2,424 regions spanning 1,974,707 bp and 8,978 peaks are identified as BR DHMs with the length of 3,755,736 bp.

Table 1. ChIP-seq counts and mapping results

Sample	# Total reads	# Unique mapped reads	(%)
xB.1_input	40,459,912	11,890,694	(30.15)
xB.1_H3K9me2	41,202,029	6,573,472	(16.38)
xB.2_input	39,754,351	11,784,202	(30.51)
xB.2_H3K9me2	36,160,104	2,337,146	(6.82)
Br.1_input	38,504,444	11,430,177	(30.47)
Br.1_H3K9me2	42,555,140	5,233,934	(12.84)
Br.2_input	24,251,434	3,451,534	(15.15)
Br.2_H3K9me2	39,172,728	4,893,917	(12.79)
Rs.1_input	39,559,995	8,316,826	(21.74)
Rs.1_H3K9me2	27,737,166	3,623,603	(13.39)
Rs.2_input	38,077,542	11,162,452	(29.76)
Rs.2_H3K9me2	40,393,014	4,843,834	(12.19)

Table 2. General information of H3K9me2 detection

Sample	Respective peak			Common peak		
	Total length	Average length	Total number	Total length	Average length	Total number
xB.1_H3K9me2	22,288,484	518.5	42,979	9,324,563	772.6	12,068
xB.2_H3K9me2	11,546,007	237.0	48,705			
Br.1_H3K9me2	17,180,834	370.9	46,313	10,043,125	837.9	11,986
Br.2_H3K9me2	11,863,260	550.3	21,558			
Rs.1_H3K9me2	10,316,041	608.8	16,944	11,863,241	1023.6	11,589
Rs.2_H3K9me2	16,422,659	464.4	35,357			

Table 3. Summary of differential histone modification binding sites and common sites

Type	Total length	Average length	Total number
Common peak	8,208,133	1,967	4,173
xB-specific peak (xB DHM)	1,974,707	814	2,424
BrRs-specific peak (BR DHM)	3,755,736	418	8,978

Characterizing the distribution of H3K9me2 level

For the downstream analyses related to H3K9me2 distributions in x*Brassicoraphanus*, both xB DHMs and the common peaks were utilized (Table 4). In the same way BR DHMs and the common peaks were included in H3K9me2 associated regions in the parents. Figure 2 describes the distribution of H3K9me2 level and fractional DNA methylation levels of x*Brassicoraphanus* and the parents along 19 pseudochromosomes (10 Ax chromosomes from A genome and 9 Rx chromosomes from R genome) of x*Brassicoraphanus* in 100 kb windows. H3K9me2 is highly enriched in gene-poor regions and tightly associated with DNA methylation (Figure 2). The locations of xB DHMs and BR DHMs are not restricted to the specific regions but are dispersed across the genome. Also, the distribution of histone modification was calculated relative to inter and intragenic regions in x*Brassicoraphanus* (Figure 3). It was found that H3K9me2 is predominantly wide-spread within intergenic regions (66.9%) followed by the coding regions (19.9%). The percentages of H3K9me2 in translation termination sites (TTS) and in the promoter regions are similar to each other (6.1% and 7.1%).

To understand detailed distributions of H3K9me2 in x*Brassicoraphanus*, 6,597 peaks were divided according to their subgenomic locations and the lengths of allocated peaks were merged. Figure 4A shows that less than half of the total length of H3K9me2-enriched regions (43.5%) are positioned in the *B. rapa* subgenome (AxAx) while the rest in the *R. sativus* subgenome (RxRx). When the whole lengths of identified H3K9me2 regions of *B. rapa* and *R. sativus* are compared to each other, the higher coverage of the peaks in *R. sativus* than that in *B. rapa* is observed (Figure 4B). To investigate the effect of subgenome dominance on

histone modification, the subgenomic locations of xB DHMs and BR DHMs were examined (Table 5). One third of xB DHMs occur in the Ax subgenome and the rest two-thirds are positioned in the Rx subgenome. BR DHMs, however, are divided equally to each subgenome. In the case of Common peaks, subgenomic distribution is similar to that of H3K9me2 peaks in x*Brassicoraphanus*. Disparity of xB DHMs in each subgenome would explain Rx-weighted H3K9me2 distribution in x*Brassicoraphanus*.

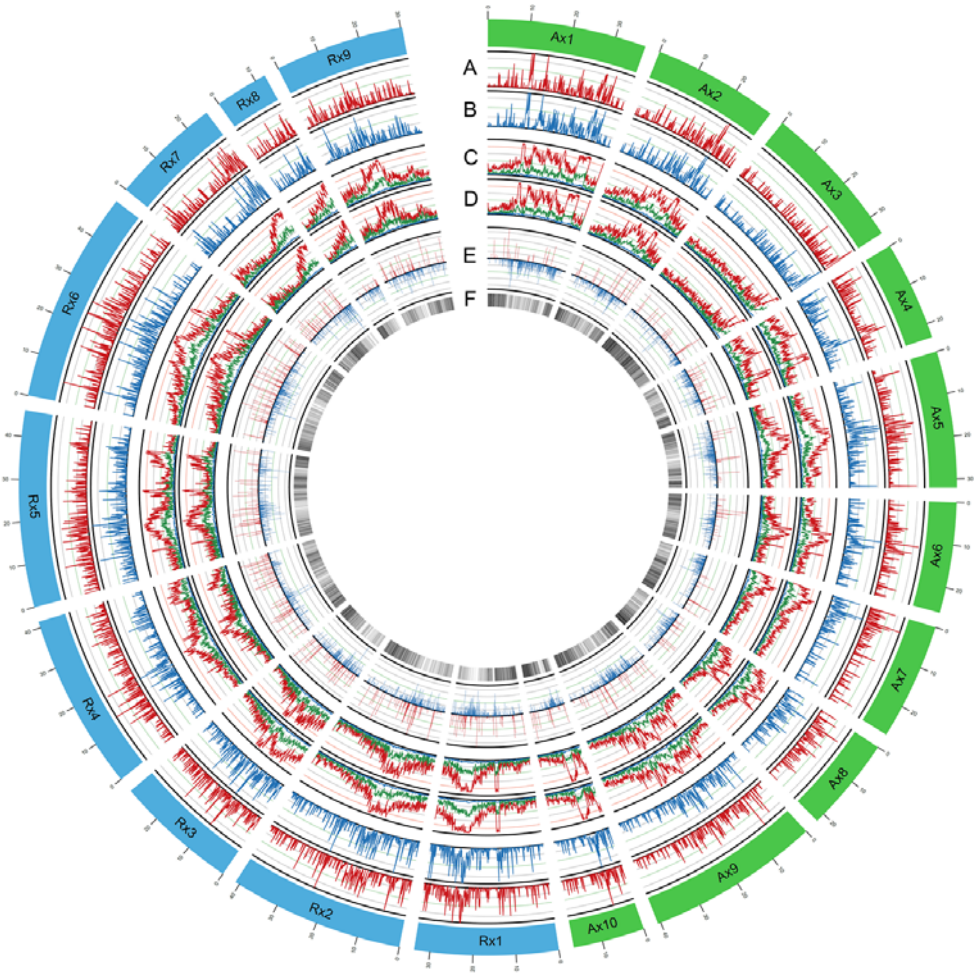

Figure 2. Genome-wide distribution of H3K9me2 and DNA methylation in x*Brassicoraphanus*. The x*Brassicoraphanus* genome is composed of two subgenomes from *B. rapa* and *R. sativus* including 10 Ax chromosomes and 9 Rx chromosomes. Tracks shown are **(A, B)** H3K9me2 densities of x*Brassicoraphanus* (A) and of its parents, *B. rapa* and *R. sativus* (B). **(C, D)** DNA methylation levels of x*Brassicoraphanus* (C) and of its parents (D). CG, CHG, CHH methylation levels are indicated as red, green, and blue lines, respectively. **(E)** Differential H3K9me2 binding sites. x*Brassicoraphanus*-specific site (red) and the parents-specific site (blue) **(F)** Gene density. The scales are in 100kb windows.

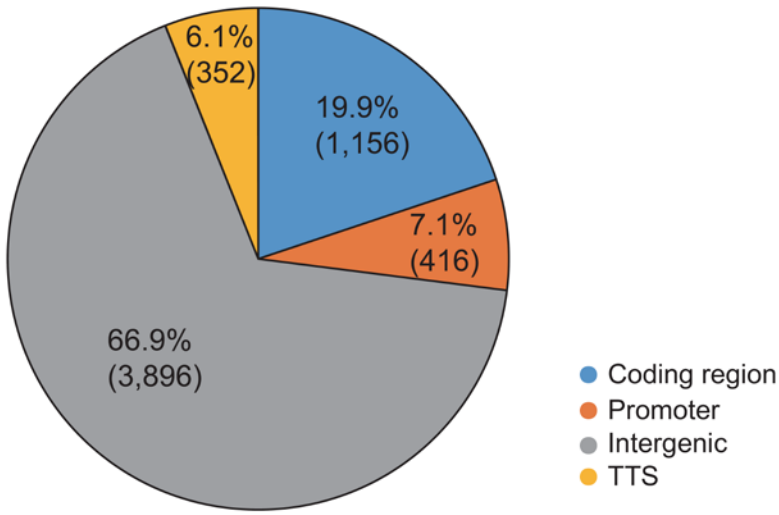

Figure 3. The distribution of H3K9me2 within different sites of the x*Brassicoraphanus* genome.

According to its location the whole peaks (6,597) were grouped into four regions; Coding region, promoter (-2 kb from the translation start site), intergenic, and translation termination site (TTS). When located in the region from -1 kb to 100 bp, the peaks are assigned to TSS. For TTS, a criterion is from -100 bp to 1 kb.

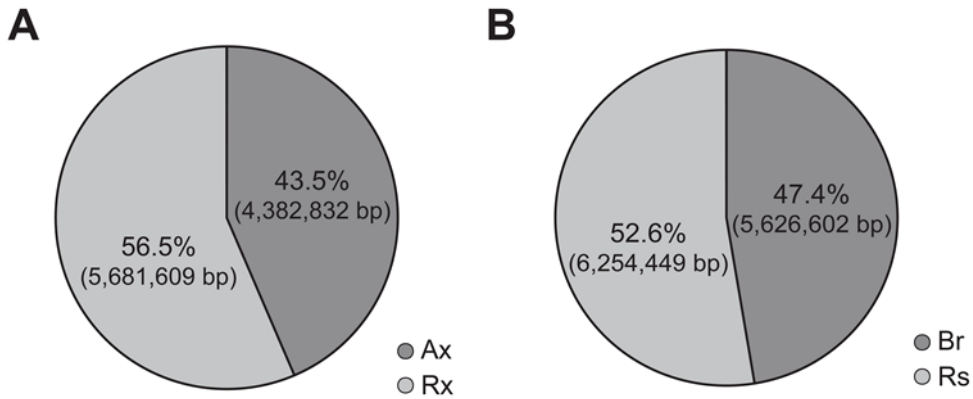

Figure 4. Distributions of H3K9me2 within subgenomic chromosomes of x*Brassicoraphanus* and in its parents. (A) The proportion of H3K9me2 levels in x*Brassicoraphanus* according to its subgenome. All peaks of H3K9me2 were divided based on their subgenomic locations and the total lengths were presented. Dark grey represents Ax subgenome and light grey represents Rx genome. **(B)** The proportion of H3K9me2 levels between *B. rapa* (dark grey) and *R. sativus* (light grey). Whole lengths of peaks of each genus were calculated.

Table 4. Summary of H3K9me peaks of x*Brassicoraphanus* and those of *B. rapa* and *R. sativus*

	Total length	Average Length	Total number
x*Brassicoraphanus* (xB)	10,182,840	1,544	6,597
B. rapa and *R. sativus* (BR)	11,963,869	910	13,151

Table 5. Subgenomic proportions of xB DHMs, BR DHMs, and common peaks in x*Brassicoraphanus*

	xB DHM		BR DHM		Common peak	
	Total length	%	Total length	%	Total length	%
Ax	669,483	33.90	1,913,253	50.94	3,713,349	45.24
Rx	1,252,848	63.44	1,825,688	48.61	4,428,761	53.96
Not assigned	52,376	2.65	16,795	0.45	66,023	0.80
Total	1,974,707	100	3,755,736	100	8,208,133	100

The correlation between H3K9me2 and DNA methylation

One of the well-known factors that induce H3K9me2 is DNA methylation. To investigate the correlation between H3K9me2 and DNA methylation, DNA methylome data from bisulfite sequencing (BS-seq) that display nucleotide-resolution data of quantitative methylation levels (analyzed and kindly provided by Hosub Shin) were compared with the ChIP-seq data in two different ways. First, the average levels of DNA methylation in different contexts, CG, CHG, and CHH, were calculated within H3K9me2 positive and negative regions in x*Brassicoraphanus* (Figure 5A). The DNA methylation levels of all three types within H3K9me2 positive regions are much higher than those within genome and the regions devoid of H3K9me2. Compared to a two-fold increase in average CG methylation level, non-CG methylation was three times more enriched within H3K9me2-related regions.

There are negligible differences of the average DNA methylation levels between genomic area and H3K9me2 negative zones. The general aspect of Figure 5A is connected to that in *Arabidopsis* (Bernatavichute et al., 2008), which verifies the quality of ChIP-seq results. Second, the ratio of methylated cytosines (>50% for CG, >20% for CHG, and >0% for CHH; Bernatavichute et al., 2008) relative to the whole number of cytosines was figured out according to the degree of H3K9me2 enrichment (Figure 5B). In the CG context, almost all cytosines (99.56%) are methylated in H3K9me2 positive areas while the half are methylated across the genome (47.29%) and H3K9me2-deficient regions (45.46%). A quarter of cytosines in the CHG context are methylated in x*Brassicoraphanus* but a significant portion of H3K9me positive regions (82.12%) corresponds to the

regions with DNA methylation (Figure 5B). Similarly, the percentage of CHH methylated cytosines increases from 17% to 52% in accordance with dimethylation of H3K9. In short, H3K9me2 and DNA methylation are mutually correlated within x*Brassicoraphanus* genome, where non-CG methylation is significantly correlated with the status of H3K9me2.

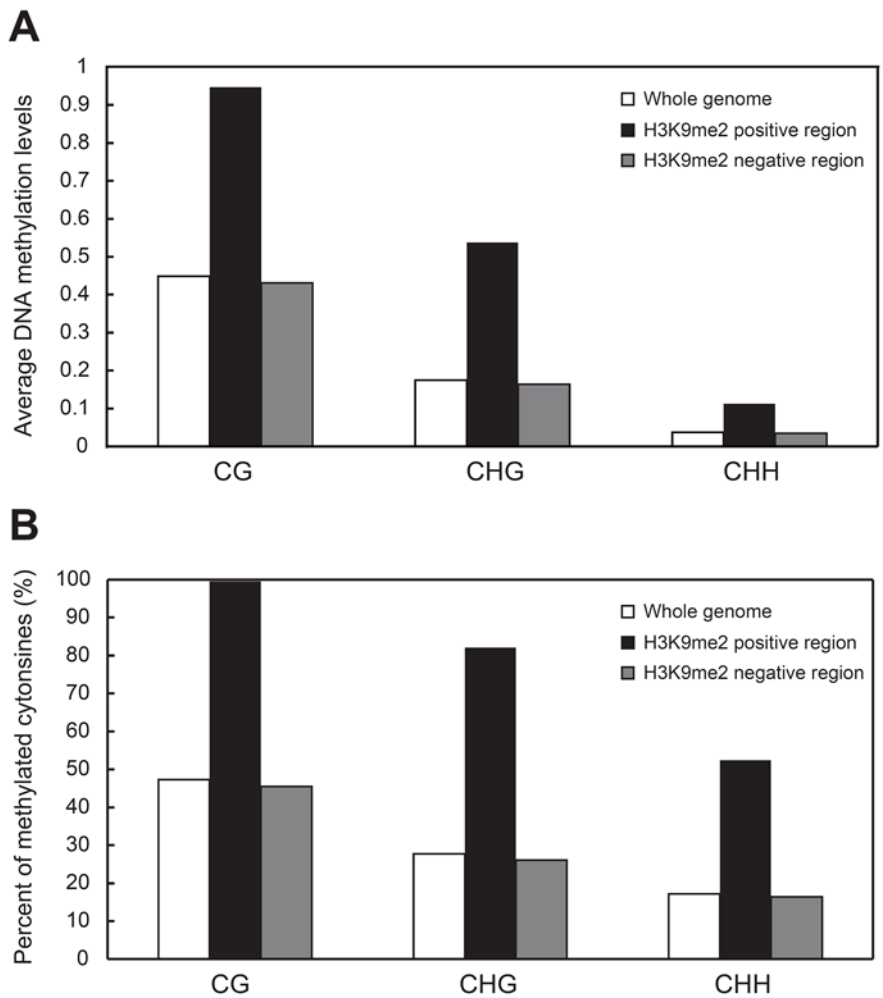

Figure 5. Correlation of H3K9 dimethylation with DNA methylation in x*Brassicoraphanus*.

(A) Average DNA methylation levels throughout the genome (white), in H3K9me2-related region (black), and in H3K9me2 negative region (grey). All methylated cytosines within each region were used to compute average methylation levels in CG, CHG, and CHH contexts. **(B)** The number of qualified methylcytosines (>50% for CG, >20% for CHG, and >0% for CHH; Bernatavichute et al., 2008) was presented in a ratio to total cytosines within each region according to DNA methylation contexts.

Coincidence of DHMs and differentially methylated regions (DMRs)

Since the maintenance mechanism of DNA methylation patterns with histone H3K9me2 has been discovered in previous studies (Stroud et al., 2014; Bernatavichute et al., 2008), the regions of xB DHMs and BR DHMs were associated with hyper DMRs and hypo DMRs that exhibit the areas where relatively higher levels of DNA methylation are found in x*Brassicoraphanus* than in the parents and vice versa. Out of 1,974,707 bps of xB DHMs, nearly 9% falls within hyper DMRs whereas a small fraction (1.47%) is overlapped with hypo DMRs (Figure 6). On the other hand, one fifth of BR DHMs that span 772,585 bps overlap with hypo DMRs, while only 6.1% of BR DHMs are found in the regions of hyper DMRs. To summarize, xB DHMs are more correlated with hyper DMRs and BR DHMs are more connected to hypo DMRs.

To test whether the occurrence and disappearance of H3K9me2 after polyploidization correlate with the different contexts of DNA methylation, the total length of DMR-associated DHMs was calculated according to the contexts; CG, CHG, and CHH. Table 6 displays that almost all (97.7%) of the xB DHMs that overlap with hyper-DMRs are highly correlated with CHG-methylated hyper DMRs. Interestingly, BR DHMs also have a high level of interrelation with hypo DMRs in the CHG context. These results indicate that the differentially CHG-methylated regions are correlated with the high proportion of both xB DHMs and BR DHMs.

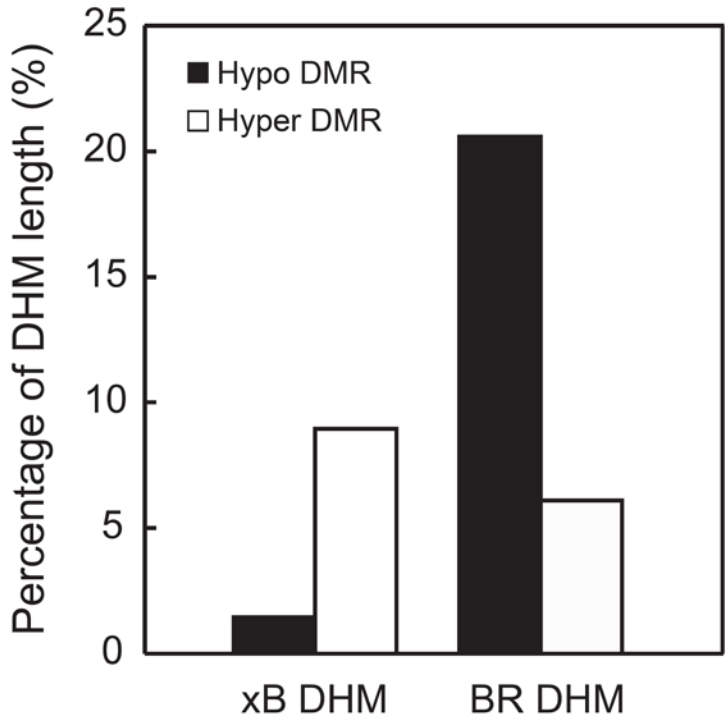

Figure 6. The synchronicity of DHM (Differential H3K9me2-marked region) and DMR (Differentially methylated region) in x*Brassicoraphanus*. The percentage of DHM that is associated with DMR. Each DHM was compared with hypo DMR and hyper DMR and DMR-associated DHMs were merged to calculate its proportion of length to total length of DHM.

Table 6. Summary of the association of DMRs with DHMs in x*Brassicoraphanus*

	Context	xB DHM			BR DHM		
		Total number	Total length	%	Total number	Total length	%
Hyper DMR	CG	90	107,774	61.0	21	13,655	6.0
	CHG	156	172,598	97.7	686	220,700	96.3
	CHH	67	114,926	65.1	20	12,229	5.3
Hypo DMR	CG	1	2,443	8.4	276	235,125	30.4
	CHG	18	20,194	69.8	862	731,787	94.7
	CHH	11	13,339	46.1	329	274,100	35.5

DHM and DMR coordinate gene expression changes

To test if generation and disappearance of H3K9me2 and DNA methylation affects gene expression patterns in x*Brassicoraphanus*, the regions including hyper-DMRs and xB DHMs were compared with differentially expressed genes (DEGs) between x*Brassicoraphanus* and the parents. The expression levels of associated DEGs in x*Brassicoraphanus* were generally found to decrease than in the parents (Figure 7A). For the regions associated with hypo DMRs and BR DHMs, the expression changes of associated DEGs in x*Brassicoraphanus* increase (Figure 7B).

To investigate the patterns of gene expression changes, DEGs found in the regions associated with DMRs and DHMs were classified into up-regulated DEGs (up DEGs), down-regulated DEGs (down DEGs). In Table 7, down DEGs (31.4%) are highly enriched in the regions of hyper DMRs and xB DHMs (Fisher's exact test, P = 9.82E-08). For the regions associated with hypo DMRs and BR DHMs, the proportion of up DEGs is much higher than observed in x*Brassicoraphanus* (Fisher's exact test, P = 0.0017). Taken together, hyper DMR and xB DHMs are often associated with the lower expression of genes in x*Brassicoraphanus* than its parents and hypo-DMR and BR DHMs with the higher expression in x*Brassicoraphanus* than in its parents.

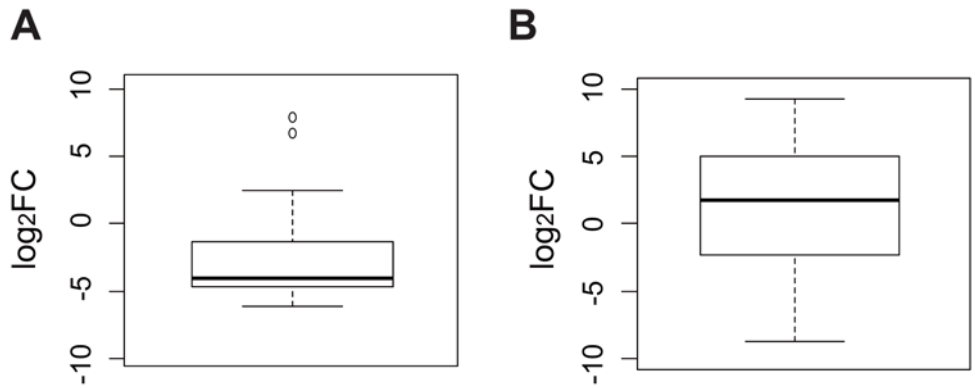

Figure 7. Expression changes of differentially expressed genes (DEGs) associated with DHM and DMR in x*Brassicoraphanus* compared to the parents. Fold changes (log_2) of DEG expression are displayed. Fold change (FC) = (expression in x*Brassicoraphanus*) / (expression in the parents) **(A)** DEGs associated with the regions including hyper DMR and xB DHM. **(B)** DEGs related to the regions of hypo DMR and BR DHM.

Table 7. Summary of the association of DEGs with DHMs and DMRs

	hyper DMR & xB DHM		hypo DMR & BR DHM		x*Brassicoraphanus*	
Up DEG	4	(7.8%)	46	(11.6%)	5,115	(7.2%)
Down DEG	16	(31.4%)	25	(6.3%)	4,676	(6.6%)
Non DEG	31	(60.8%)	325	(82.1%)	61,231	(86.2%)
Total	51	(100%)	396	(100%)	71,022	(100%)

Expression patterns of the enzymes modulating epigenetic modification

To investigate whether some factors that are known to modify epigenetic states of chromatin are engaged in epigenetic changes in x*Brassicoraphanus*, the genes encoding such enzymes were identified in x*Brassicoraphanus* and their expression patterns were comparatively analyzed with the RNA-seq data (analyzed and kindly provided by Hosub Shin). CHROMOMETHYLASE (CMT), DOMAINS REARRANGED METHYLASE (DRM), and DECREASE IN DNA METHYLATION (DDM) are supposed to mediate non-CG DNA methylation and five, eight, and two of their homologues were found respectively in x*Brassicoraphanus* for each family (Table 8). Also, 23 genes were found to be homologous to SUVH that transfers methyl groups to lysine of histones. Most of genes were expressed similarly between x*Brassicoraphanus* and its parents and four of them were confirmed as DEGs (Table 8). In the case of DRM homologs, even though they are classified as DEGs, the disparity of FPKM is not huge and their expression levels are very low. The expression levels of xB01275r and xB11965r that are homologs of SUVH slightly increase in x*Brassicoraphanus* (Table 8). However, the homologs of DRM2, CMT2, CMT3, DDM1, SUVH4, and SUVH5 that are supposed to play crucial roles in maintaining and initially establishing DNA methylation and H3K9me2 in *Arabidopsis* show no expression difference between x*Brassicoraphanus* and the parents (Table 8). Other cues but epigenetic modification enzymes are expected to trigger epigenetic changes in x*Brassicoraphanus* after polyploidization.

Table 8. Expression patterns of the genes related to epigenetic modifications

At ID	xB ID	xB FPKM	BR FPKM	DEG
CMT1	xB23077r	24.68	23.16	None
CMT1	xB66307b	16.00	23.26	None
CMT2	xB12574r	14.35	9.14	None
CMT2	xB25616b	12.79	16.60	None
CMT3	xB57468b	0.58	0.75	None
DDM1	xB47302b	9.10	15.70	None
DDM1	xB64642b	31.56	33.84	None
DRM1	xB03359b	0	0.69	Down
DRM1	xB71947b	771.06	1155.73	None
DRM2	xB10199b	17.71	22.81	None
DRM2	xB12043r	0.53	0.05	UP
DRM2	xB19415r	14.35	19.97	None
DRM2	xB58631b	0.49	1.63	None
DRM3	xB40032b	5.37	8.65	None
DRM3	xB41361r	7.28	7.18	None
SUVH1	xB01275r	13.11	5.68	UP
SUVH1	xB11666r	0	0.19	None
SUVH1	xB20043r	7.03	13.24	None
SUVH1	xB20044r	8.61	8.28	None
SUVH1	xB43996b	9.76	11.93	None
SUVH1	xB56005b	0.16	0.51	None
SUVH1	xB73419b	20.32	21.43	None
SUVH1	xB73420b	11.69	11.08	None
SUVH2	xB54470b	3.15	5.44	None
SUVH2	xB78524b	0.00	0.00	None
SUVH3	xB20911r	16.17	13.35	None
SUVH3	xB24889r	7.18	9.73	None
SUVH3	xB57856b	6.43	10.02	None
KYP/SUVH4	xB00911r	32.61	37.96	None
KYP/SUVH4	xB44396b	50.55	66.39	None
SUVH5	xB29881b	4.21	5.39	None
SUVH5	xB78679b	3.40	2.90	None
SUVH7	xB03235b	0	0	None
SUVH7	xB62149b	0	0	None
SUVH7	xB62150b	0	0	None
SUVH9	xB11965r	22.69	10.62	UP
SUVH9	xB22032r	0.98	2.27	None
SUVH9	xB35894b	24.13	27.27	None

DISCUSSION

During polyploidization epigenetic changes are known to be crucial for stabilizing the hybrid genome. To assess whether histone modifications are involved in mediating epigenomic changes in x*Brassicoraphanus,* the distribution of H3K9me2 across the genome was profiled by the ChIP-seq method in this study. Furthermore, the BS-seq data were combined to figure out the correlation between H3K9me2 and DNA methylation. Also, epigenomic differences between x*Brassicoraphanus,* an allotetraploid, and its parents, *B. rapa* and *R. sativus*, were investigated to comprehend the hybridization effects on gene expression changes. Due to the complexity of ChIP procedures, the mapping results of sequenced reads are not uniform. However, this weakness was complemented by using two biological repeats and utilizing overlapped regions for downstream analyses. Since the length of identified regions that are associated with H3K9me2 varies and is generally long, most results are based on the total length of peaks.

Consistent with previous studies in *Arabidopsis*, H3K9me2 and DNA methylation are mutually colocalized in the x*Brassicoraphanus* genome and their distributions are inversely proportional to the gene density. This tendency is derived from their high concentration at heterochromatin and repeat regions. The exceedingly enriched DNA methylation within H3K9me2 regions implies that general mechanisms modulating those epigenetic states and interdependency between them are universal. Furthermore, genome-wide dispersion of xB DHM and BR DHM suggests that genesis and vanishment of H3 Lys 9 methylation after genome hybridization occurs at random over the whole genome and is considerably

associated with that of CHG methylation. This correlation is coherent with the proposed self-reinforcing state of CHG methyltransferase and H3K9me2-related histone methyltransferase in *Arabidopsis* (Johnson et al., 2007). The association with DEGs indicates that when H3K9me2 and DNA methylation are generated in an allotetraploid, the expression of genes within that area declines in general. On the contrary, if they are dissipated, genes are expressed relatively in high level. Thus, polyploidization between *B. rapa* and *R. sativus* triggered epigenetic changes in genome level, which leads to transcriptional reprogramming.

When epigenetic changes in x*Brassicoraphanus* compared to the progenitors are examined, the occurrence rate of H3K9me2 in Rx subgenome was much higher than that in Ax subgenome (63.4% and 33.9%). Since *B. rapa* was used as a maternal progenitor, the genome of *R. sativus* in an allopolyploid would have experienced a different maternal environment. Responding to cytoplasm of *B. rapa*, chromatin from *R. sativus* would adjust by itself through genome-wide modification of H3K9me2 levels. However, the locations of newly or more established DNA methylation in x*Brassicoraphanus* specifically scatter over Ax subgenome (74.6%, length-based). In contrast to subgenomic differences in forming the new patterns of histone modification and DNA methylation, perishments of the epigenetic patterns equally happened regardless of subgenomic. These results suggest that after hybridization, epigenetic changes have occurred in subgenome-specific ways; more H3K9me2 was preferentially generated towards Rx subgenome but DNA methylation was generated towards Ax subgenome. These biased alterations may have been derived from the expression changes of epigenetic factors or transposons and sRNAs that are explosively activated through

polyploidization (Chen and Yu, 2013). As to assumed factors concerned in epigenomic changes in x*Brassicoraphanus*, there was no striking discrepancy of gene expressions between homeologs. Therefore, associations of sRNAs and transposable elements with DHM or DMR could be further explored to trace mechanisms having induced epigenome changes after intergeneric hybridization.

This study conducted a genome-level investigation of H3K9me2 in an intergeneric hybrid, x*Brassicoraphanus*, and its correlation with DNA methylation was explored. Furthermore, general epigenetic changes in x*Brassicoraphanus* relative to its parents were identified. To my best knowledge, this is the first report on epigenomic variations in a polyploid compared with the progenitors with ChIP-seq method. And this may help to understand the hybridization effects orchestrating chromatin structures more clearly. Furthermore, the study under the subject of intergeneric hybrid, which is rarely found in nature, could provide clues to how two genetically distinct genomes were combined in a nucleus and to the mechanisms permitting genome hybridization, which leads to explanation for evolutionary process in eukaryotes.

REFERENCES

Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I., Reuter, G., and Aalen, R.B. (2001) The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. Nucleic Acids Research 29, 4319-4333.

Becak, M.L. and Kobashi, L.S. (2004) Evolution by polyploidy and gene regulation in Anura. Genetics and Molecular Research 3, 195-212.

Berger, S.L. (2007) The complex language of chromatin regulation during transcription. Nature 447, 407-412.

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. PLoS ONE 3, e3156.

Birchler, J.A. and Veitia, R.A. (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proceedings of the National Academy of Sciences of the United States of America 109, 14746-14753.

Chen, Z.J. (2007) Genetic and Epigenetic Mechanisms for Gene Expression and Phenotypic Variation in Plant Polyploids. Annual Review of Plant Biology 58, 377-406.

Chen, Z. J. and Yu, H. H. (2013) Genetic and epigenetic mechanisms for polyploidy and hybridity. *In* Chen, Z. J. and Birchler, J. A. (eds.) Polyploid

and Hybrid Genomics. 335-354. Wiley-Blackwell, New York.

Comai, L. (2005) The advantages and disadvantages of being polyploid. Nature Reviews Genetics 6, 836-846.

Conesa, A., Götz, S, García-Gómez, J.M., Terol, J., Talón, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21, 3674-3676.

Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005) The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biology 6, 227.

Doyle, J.J., Flagel, L.E., Paterson, A.H., Rapp, R.A., Soltis, D.E., Soltis, P.S., and Wendel, J.F. (2008) Evolutionary genetics of genome merger and doubling in plants. Annual Review of Genetics 42, 443-461.

Ebbs, M.L., Bartee, L., and Bender, J. (2005) H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. Molecular and Cellular Biology 25, 10507-10515.

Ebbs, M.L. and Bender, J. (2006) Locus-specific control of DNA methylation by the *Arabidopsis* SUVH5 histone methyltransferase. Plant Cell 18, 1166-1176.

Gendler, K., Paulsen, T., Napoli, C. (2008) ChromDB: the chromatin database. Nucleic Acids Research 36, 298-302.

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., and Sykes, S. (2011) High-quality draft

assemblies of mammalian genomes from massively parallel sequence data. Proceedings of the National Academy of Sciences of the United States of America 108, 1513-1518.

Grant, V. (1981) Plant Speciation, Columbia University Press, New York.

Hegarty, M.J. and Hiscock, S.J. (2008) Genomic clues to the evolutionary success of polyploid plants. Current Biology 18, R435-R444.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime *cis*-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell 38, 576-589.

Initiative A.G. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796-815.

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, 556-560.

Jasencakova, Z., Soppe, W.J.J., Meister, A., Gernand, D., and Turner, B.M. (2003) Histone modifications in *Arabidopsis* - high methylation of H3 lysine 9 is dispensable for constitutive heterochromatin. Plant Journal 33, 471-480.

Jenuwein, T. and Allis, C.D. (2001) Translating the Histone Code. Science 293, 1074-1080.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S. et al. (2011) Ancestral

polyploidy in seed plants and angiosperms. Nature 473, 97-100.

Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I.R., Callis, J., and Jacobsen, S.E. (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. Current Biology 17, 379-384.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., and Garcia-Hernandez, M. (2012) The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Research 40, 1202-1210.

Langmead B., Trapnell C., Pop M., and Salzberg S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 10, R25.

Lee, S.S., Choi, W.J., and Woo, J.G. (2002) Development of a new vegetable crop in *Brassicoraphanus* by hybridization of *Brassica campestris* and *Raphanus sativus*. Journal of Korean Society for Horticultural Science 43, 693-698.

Lee, S.S., Lee, S.A., Yang, J., and Kim, J. (2011) Developing stable progenies of x*Brassicoraphanus*, an intergeneric allopolyploid between *Brassica rapa* and *Raphanus sativus*, through induced mutation using microspore culture. Theoretical and Applied Genetics 122, 885-891.

Leitch, A.R. and Leitch, I.J. (2008) Genomic plasticity and the diversity of polyploid plants. Science 320, 481-483.

Levin, D.A. (1983) Polyploidy and novelty in flowering plants. The American Naturalist 122, 1-25

Levin, D.A. (2002) The Role of Chromosomal Change in Plant Evolution. Oxford University Press, New York.

Li, B., Carey, M., and Workman, J.L. (2007) The Role of Chromatin during Transcription. Cell 128, 707-719.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430, 471-476.

Liu, C., Lu, F., Cui, X., and Cao, X. (2010) Histone Methylation in Higher Plants. Annual Review of Plant Biology 61, 395-420.

Malagnac, F., Bartee, L., and Bender, J. (2002) An *Arabidopsis* SET domain protein required for maintenance but not establishment of DNA methylation. The EMBO Journal 21, 6842-6852.

Mardis, E.R. (2007) ChIP-seq: welcome to the new frontier. Nature 4, 613-614.

Mathieu, O., Probst, A.V., and Paszkowski, J. (2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in *Arabidopsis*. The EMBO Journal 24, 2783-2791.

Mayrose, I., Zhan, S.H., Rothfels, C.J., Magnuson-Ford, K., Barker, M.S., Rieseberg, L.H., and Otto, S.P. (2011) Recently formed polyploid plants diversify at lower rates. Science 333, 1257.

McClintock, B. (1984) The significance of responses of the genome to challenge. Science 226, 792-801.

Metzker, M.L. (2010) Sequencing technologies — the next generation. Nature Reviews Genetics 11, 31-46.

Michael, T.P. and Jackson, S. (2013) The first 50 plant genomes. Plant Genome 6.

Ni, Z., Kim, E.D., Ha, M., Lackey, E., Liu, J., Zhang, Y., Sun, Q., and Chen, Z.J. (2009) Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. Nature 457, 327-331.

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. (2003) Understanding mechanisms of novel gene expression in polyploids. Trends in Genetics 19, 141-147.

Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M.A., and Ainouche, M. (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytologist 184, 1003-1015.

Paterson, A.H. (2005) Polyploidy, evolutionary opportunity, and crop adaptation. Genetica 123, 191-196.

Premachandran, M.N., Prathima, P.T., and Lekshmi, M. (2011) Sugarcane and polyploidy: a review. Journal of Sugarcane Research 1, 1-15.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T.L., Wilson, C.J., Bell, S.P., and Young, R.A. (2000) Genome-wide location and function of DNA binding proteins. Science 22, 2306-2309.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-

terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74, 5463-5467.

Schadt, E.E., Turner S., and Kasarskis, A. (2010) A window into third-generation sequencing. Human Molecular Genetics 19, 227-240.

Schatlowski, N. and Kohler, C. (2012) Tearing down barriers: understanding the molecular mechanisms of interploidy hybridizations. Journal of Experimental Botany 63, 6059-6067.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell 119, 941-953.

Solomon, M.J., Larsen, P.L., and Varshavsky, A. (1988) Mapping protein-DNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene. Cell 53, 937-947.

Soltis, D.E., Visger, C.J., and Soltis, P.S. (2014) The polyploidy revolution then. . . and now: stebbins revisited. American Journal of Botany 101, 1057-1078.

Song, Q. and Chen, Z.J. (2015) Epigenetic and developmental regulation in plant polyploids. Current Opinion in Plant Biology 24, 101-109.

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*. The EMBO Journal 21, 6549-6559.

Springer, N.M., Napoli, C.A., Selinger, D.A., Pandey, R., Cone, K.C., Chandler, V.L., Kaeppler, H.F., and Kaeppler, S.M. (2003) Comparative analysis of

SET domain proteins in maize and *Arabidopsis* reveals multiple duplications preceding the divergence of monocots and dicots. Plant Physiology 132, 907-925.

Stebbins, G.L. (1950) Variation and Evolution in Plants. Columbia University Press, New York.

Stebbins, G.L. (1971) Chromosomal Evolution in Higher Plants. Edward Arnold, London.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014) Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. Nature structural & molecular biology 21, 64-72.

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

Tariq, M., Saze, H., Probst, A.V., Lichota, J., Habu, Y., and Paszkowski, J. (2003) Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. Proceedings of the National Academy of Sciences of the United States of America 100, 8823-8827.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. Nature 439, 811-816.

Turck, F., Roudier, F., Farrona, S., Martin-Magniette, M.L., Guillaume, E., Buisine, N., Gagnot, S., Martienssen, R.A., Coupland, G., and Colot, V. (2007) *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. PLoS Genetics 3, e86.

Van de Peer, Y. and Meyer, A (2005) Large-scale gene and ancient genome duplications. *In* Gregory, T. R. (eds.) The Evolution of the Genome. 330-363. Elsevier, San Diego.

Van de Peer, Y., Maere, S., and Meyer, A. (2009) The evolutionary significance of ancient genome duplications. Nature Reviews Genetics 10, 725-732.

Wang, J., Tian, L., Lee, H.S., Wei, N.E., Jiang, H., Watson, B., Madlung, A., Osborn, T.C., Doerge, R.W., Comai, L. et al. (2006) Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. Genetics 172, 507-517.

Watson, J.D. and Crick, F.H.C. (1953) Molecular structure of nucleic acids. Nature 171, 737-738.

Wendel, J.F. (2000) Genome evolution in polyploids. Plant Molecular Biology 42, 225-249.

Ye, J. , Fang, L. , Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, Y., Bolund, L., and Wang, J. (2006) WEGO: a web tool for plotting GO annotations. Nucleic Acids Research 34, 293-297.

Yoo, M.J., Liu, X., Pires, J.C., Soltis, P.S., and Soltis, D.E. (2014) Nonadditive gene expression in polyploids. Annual Review of Genetics 48, 485-517.

Zhang, X. (2008) The Epigenetic Landscape of Plants. Science 320, 489-492.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. et al. (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biology 9, R137.

Zhao, Z. and Shen, W.H. (2004) Plants contain a high number of proteins showing sequence similarity to the animal SUV39H family of histone methyltransferases. Annals of the New York Academy of Sciences 1030, 661-669.

초 록

후성유전학적 변화는 배수화된 식물에서 동위 유전자의 발현을 조절하고 유전체의 안정성에 영향을 끼친다. 히스톤 변형 중 히스톤 단백질 H3의 아홉 번째 라이신에 두 개의 메틸기가 붙은 경우 (H3K9me2), 애기장대에서는 이질염색질에서 주로 찾아볼 수 있다. 또한 DNA 메틸레이션을 조절하며 트랜스포존의 발현을 억제한다고 알려져 있다. 염색질 면역침강법과 시퀀싱을 진행하여, 배추와 무의 속간교잡으로 탄생한 배무채에서 H3K9me2의 위치를 추적하였다. 메틸레이션 데이터와 비교하였을 때 배무채의 유전체에서 H3K9me2와 DNA 메틸레이션의 높은 상관관계가 발견되었다. 이러한 경향성은 배수화 현상이 일어나서 배무채가 생성되고 이 때 사라지거나 새롭게 생성된 H3K9me2에서도 발견할 수 있었고 주로 CHG 메틸레이션과 연관되어 있었다. 부모와 비교하였을 때 배무채에서 DNA 메틸레이션 레벨이 낮아진 지역과 H3K9me2이 없어진 지역은 배무채 전체 유전체 지역에서 고르게 분포한다. 반면 DNA 메틸레이션의 수준이 높아진 지역과 H3K9me2이 새로 생성된 지역은 배추, 혹은 무 하위유전체에 특이적으로 많이 존재한다. 이러한 결과는 후성유전학적 메커니즘이 서로 연관되어 있고 두 속이 교잡되면서 많은 영향을 받았다는 것을 보여준다.

주요어: 히스톤 변형, 염색질면역침강법, 배무채, 속간교잡종, DNA 메틸레이션, 후성유전학

학번: 2014-20725