A DISSERTATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Stabilization of hybrid genome and reconstruction of transcriptome network in x*Brassicoraphanus*

속간 잡종 배무채의 유전체 안정화 및 전사체 조절 네트워크 재정립

FEBRUARY 2020

HOSUB SHIN

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY

DEPARTMENT OF PLANT SCIENCE

COLLEGE OF AGRICULTURE AND LIFE SCIENCES

THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

Stabilization of hybrid genome and reconstruction of transcriptome network in x*Brassicoraphanus*

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

BY
HOSUB SHIN

MAJOR IN HORTICULTURAL SCIENCE AND BIOTECHNOLOGY
DEPARTMENT OF PLANT SCIENCE

FEBRUARY 2020

APPROVED AS A QUALIFIED DISSERTATION OF HOSUB SHIN
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY THE COMMITTEE MEMBERS

CHAIRMAN _____
Doil Choi, Ph.D.

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

MEMBER _____
Byoung-Cheorl Kang, Ph.D.

MEMBER _____
Tae-Jin Yang, Ph.D.

MEMBER _____
Soo-Sung Lee, Ph.D.

Stabilization of hybrid genome and reconstruction of transcriptome network in x*Brassicoraphanus*

HOSUB SHIN

Department of Plant Science, Seoul National University

ABSTRACT

Hybridization and polyploidization have facilitated the evolution of many plant species while resulting in speciation and production of novel characteristics to increase fitness in new environments. These phenomena have occurred in various organisms, especially in plants, and the evolution of complex and various sizes of plant genomes has been attributed to polyploidization events. Although increased genomic content and large changes of gene expression levels in polyploid genome provide various advantages for environmental adaptation, most early generated polyploid

plants suffer from genomic instability that results in sterility and inviability. x*Brassicoraphanus* is an intergeneric allopolyploid between Chinese cabbage and radish, which is a rare case of genetic merging two different genus species. Unlike most neoallopolyploid plants, x*Brassicoraphanus* is fertile and genetically stable, but little is known about the stabilization of hybrid genome between extremely divergent species and the molecular mechanism of transcriptional and epigenetic changes. In this study, I investigated the genomic, transcriptomic and epigenomic changes in a new intergeneric allotetraploid species x*Brassicoraphanus*. For the genomic analysis, *de novo* assembly of x*Brassicoraphanus* genome was performed. Complete set of both parental chromosomes without apparent genome structure changes was observed and hypermethylation of transposable elements by small RNA in *trans* were proposed. In addition, genome-wide transcriptional changes relative to parental expression are dramatically observed, and the majority of the duplicated genes are adjusted to similar levels due to the high similarity of *cis*-elements and sharing common transcription factors. This study demonstrates that a certain level of parental genome divergence is helpful to suppress genomic shocks in the early generation of polyploidy and compatibility of regulatory elements would

contribute to the massive reconstruction of transcription control network following after transcriptome shock.

Keywords:
x*Brassicoraphanus*, allopolyploidization, genome stability, genome shock, transcriptome shock, epigenetic regulation

Student number: 2015-30373

CONTENTS

CHAPTER 2. Divergence of *cis*- and *trans*-regulatory elements drives reconstruction of transcriptome network in x*Brassicoraphanus*

LIST OF TABLES

CHAPTER 1

CHAPTER 2

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

LIST OF ABBREVIATION

Bo	*Brassica oleracea*
Br	*Brassica rapa*
DEGs	Differentially expressed genes
DMC	Differentially expressed cytosines
DMR	Differentially expressed regions
ELD	Expression level dominance
FISH	Fluorescence in situ hybridization
FPKM	Fragments per kilobase of transcript per million mapped reads
GO	Gene ontology
LTR	Long terminal repeat
miRNA	microRNA
RdDM	RNA-directed DNA methylation
Rs	*Raphanus sativus*
siRNA	small interfering RNA
TEs	Transposable elements
WGD	whole-genome duplication
x*B*	x*Brassicoraphanus*

GENERAL INTRODUCTION

Polyploidization is the increase in genome size accompanied by the multiplication of sets of chromosomes. This phenomenon can be caused by the formation of unreduced gametes by abnormal cell division, hybridization between different species followed by chromosome doubling, and fusion of unreduced gametes between different species (Comai, 2005). Since Kihara & Ono first described polyploidy (Kihara and Ono, 1926), polyploidization has been generally classified into two categories: autopolyploidy and allopolyploidy. Autopolyploidy results from the multiplication of the chromosome set within a single species by either the union of two unreduced gametes or somatic chromosome doubling. In contrast, allopolyploidy is generated from hybridization followed by chromosome doubling between two different species (Ramsey and Schemske, 1998).

Although genomic instability leading to sterility and inviability is frequently detected in the early generation of polyploid individuals with two or more genomes present in a nucleus (Chen et al., 2018; Xiong et al., 2011; Zhang et al., 2013), polyploids are expected to exhibit evolutionary advantages for adaptation and survival through morphological changes.

Hybridization between divergent genomes can instantly trigger phenotypic changes via massive alterations in gene expression (Yoo et al., 2014), and increased genomic content could play a role in producing novel phenotypes via sub-/neofunctionalization of duplicated gene pairs derived from parental species (Cheng et al., 2018).

Prevalence of polyploidization

Polyploidization is a general phenomenon found in species from bacteria to eukaryotes. Prokaryotes are thought to possess a single copy of a circular chromosome like *Escherichia coli*, but several species, such as *Azotobacter vinelandii* (Maldonado et al., 1994), *Pseudomonas putida* (Pecoraro et al., 2011), *Desulfovibrio gigas* (Postgate et al., 1984), and *Halobacterium salinarum* (Breuert et al., 2006), have high numbers of genomic copies, confirming the existence of polyploidy in prokaryotes. In eukaryotes, many polyploid species have been reported. Several strains of *Saccharomyces cerevisiae* and *Candida albicans* exist as unicellular, polyploid yeast (Hubbard et al., 1985; Pomper et al., 1954). During the evolution of invertebrates, ancient polyploidization events were identified in the genomes of chelicerates (Nossa et al., 2014; Schwager et al., 2017), mollusks (Hallinan and Lindberg, 2011), and various insects (Li et al., 2018).

Polyploidy has also been frequently observed in amphibians (Schmid et al., 2015), fish (Mable et al., 2011; Zhou and Gui, 2017), but is rarely observed in other vertebrate species such as birds and mammals (Otto, 2007).

Polyploidization is thought to play an important role in the evolution of plants. Evidence of polyploidy in green algae, which is the evolutionary predecessor of land plants, has been reported. Polyploidy in various chlorophyte and streptophyte green algae was estimated by measuring their nuclear DNA content (Kapraun, 2007), and cytological analysis showed multiple chromosomes in various species of green algae, suggesting the existence of polyploidy (Casanova, 2015). In nonvascular land plants, the genomes of *Marchantia polymopha* (liverworts) and *Physcomitrella patens* (mosses) show evidence of whole-genome duplications (WGDs), indicating that polyploidization events occurred in these early land plants (Bowman et al., 2017; Lang et al., 2018). While polyploidization is relatively rare in gymnosperms, genome sequencing of conifer clades showed that polyploidization events did occur during early conifer evolution (Li et al., 2015). A few natural polyploids, such as the tetraploids *Juniperus chinesis* and *Fitzroya cupressoides*, and the hexaploid *Sequoia sepervirens* (Ahuja, 2005), have been reported, and various levels of polyploidy have been found in *Ginkgo biloba* (Smarda et al., 2018).

Polyploidy is predominantly observed in angiosperms and is thought to be important for the diversification of flowering plants (Soltis et al., 2009). High levels of unreduced gametes are found in natural populations of angiosperms (Ramsey and Schemske, 1998, 2002), and previous studies have shown that polyploidization events in 70% of flowering plants have been observed based on counting chromosome numbers (Masterson, 1994). In addition, many domesticated crop species, including oilseed rape (Chalhoub et al., 2014), mustard (Yang et al., 2016), wheat (International Wheat Genome Sequencing, 2014), cotton (Zhang et al., 2015), and strawberry (Edger et al., 2019), have undergone polyploidization events. Whole-genome sequencing analyses have revealed that many diploid plants possess redundant genomic fragments (Cheng et al., 2018; Jiao et al., 2011; Ruprecht et al., 2017). This observation suggests that all flowering plants are paleopolyploidy, which is the result of ancient WGDs, and are thought to have undergone at least one polyploidization event throughout plant lineages.

Advantages of polyploidization for evolution

Once two genomes have merged in a nucleus, the newly formed polyploid appears to have a potential survival advantage. Gene redundancy

acquired immediately after polyploidization is an undeniable advantage (Comai, 2005). Owing to decreased frequencies of incidence of recessive homozygotes, the effects of recessive alleles are masked by dominant alleles and the possibility of the occurrence of deleterious effects by mutations would be reduced. In addition, gene redundancy provides genetic materials for the diversification of gene functions. Sub-/neofunctionalization events—in which a gene acquires a new function—are more likely to occur in polyploids than in diploids, which have a lower frequency of segmental duplications (Adams and Wendel, 2005; Blanc and Wolfe, 2004; Cheng et al., 2018). Natural or synthetic polyploids usually show massive altered expression patterns as nonadditive expression, in which the expression level of the progeny is not equal to the average expression level of the parents (Yoo et al., 2014). Such genome-wide rewiring of the transcriptome may result from *cis*- or *trans*-regulatory divergence between parental genomes (McManus et al., 2010; Shi et al., 2012; Tirosh et al., 2009), as well as from epigenetic changes (Shen et al., 2017; Song and Chen, 2015; Song et al., 2017), producing novel characteristics to increase ecological competitiveness relative to the parental diploid species (Chen, 2007; Miller et al., 2015; Yoo et al., 2014). Increased genetic variation and novel phenotypes provided by polyploidization may contribute to increased

tolerance of environmental changes and possible ecological adaptation (Hahn et al., 2012). In an *in vitro* evolution analysis of yeast, polyploidy was able to accelerate evolutionary adaptation via a high rate of advantageous mutations (Selmecki et al., 2015). Several studies of salinity tolerance showed that tetraploid have a relatively high salt tolerance than diploid in *Arabidopsis*, citrus and sugar beet (Chao et al., 2013; Khalid et al., 2020; Wu et al., 2019), and other studies revealed that increased tolerances for salt and drought stress were attributed to polyploidization-induced expression of genes from hormone response pathways in rice and citrus (Ruiz et al., 2016; Yang et al., 2014). A study of the *Cardamine* showed that allotetraploid plants have general tolerance for both dry and wet conditions, unlike their parental species, which possess a special tolerance for dry or wet conditions, suggesting that polyploid plants may obtain ecological niches from both parental species (Shimizu-Inatsugi et al., 2017).

Environmental changes and stresses can induce polyploidization. The formation of unreduced gametes is an important process in polyploidy, and environmental stresses, such as heat, cold, nutrient limitation, and wounding, can increase the levels of unreduced gametes by abnormal cell division during meiosis (Sora et al., 2016; Wang et al., 2017). As described above, various phenotypic characteristics and increased genetic variation

during polyploidization may contribute to the enlargement of ecological niches and adaptation to unstable environments. This attribute might help organisms survive harsh environmental conditions. For example, evidence of numerous polyploidization events in the lineages of plants and animals were found to have occurred during the Permian-Triassic and Cretaceous-Paleogene extinction events, suggesting that polyploidy has the potential to confer ecological tolerance or fitness for survival during dramatic changes in the environment, particularly via advantageous changes in gene expression and increases in sets of genes (Fawcett et al., 2009; Van de Peer et al., 2017; Vanneste et al., 2014).

Genomic instability in newly synthesized polyploid plants

There are many spontaneous hybrids and polyploids in nature; however, strong postzygotic hybridization barriers exist to prevent gene flow between species, which lead to outbreeding depression, reduced viability, and infertility (Abbott et al., 2013; Todesco et al., 2016). Several mechanisms, such as abnormal endosperm development, self-incompatibility, and genetic and epigenetic changes, have been proposed to explain such postzygotic barriers (Dion-Cote and Barbash, 2017). Among them, “genome shock” was first suggested by Barbara McClintock to

describe genomic changes, such as large-scale chromosome reorganization and transcriptional activation of transposable elements, such that multiple copies of the genome are suddenly generated in the nucleus. This phenomenon is one of the detrimental effects resulting from hybridization or polyploidization.

Aneuploidy and massive chromosome rearrangements have been proposed as hybridization barriers that promote meiotic failure and inviability of hybrids. Many synthetic allopolyploid plants, such as rapeseed, tobacco, and wheat, have gone through gene loss, chromosome mispairing, transposon activation, altered methylation, and rearrangements that lead to aneuploidy over several generations (Xiong et al., 2011; Zhang et al., 2013; Chen et al., 2018). Chromosome rearrangement is a well-known factor inducing genomic instability in the *Brassica* family (Song et al., 1995), especially *Brassica napus* (AACC; 2n = 38), which is a natural allopolyploid resulting from the hybridization of its ancestors [*Brassica rapa* (AA; $2n = 2x = 20$) and *Brassica oleracea* (CC; $2n = 2x = 18$)]. In cytological studies of newly synthetic *B. napus*, gene conversion, homoeologous recombination, chromosomal breakage, fragment loss, and aneuploidy were frequently reported (Chalhoub et al., 2014; Song et al., 1995; Xiong et al., 2011).

Expression patterns of duplicated genes in polyploids

In newly formed allopolyploids, two divergent parental genomes reside in a single nucleus, and the expression level of duplicated genes could be adjusted to a new nucleus environment. Recent studies of allopolyploids showed that massive transcriptional changes were induced in allopolyploids by the merge of parental genome (Yoo et al., 2013; Combes et al., 2015; Wu et al., 2016; Ye et al., 2016; Zhang et al., 2016; Wu et al., 2018; Zhang et al., 2018). This genome-wide transcriptional changes relative to parental expression has been termed “transcriptome shock” (Hegarty et al., 2006; Buggs et al., 2011). In the studies of interspecific hybrids and allopolyploids, expression patterns have been showed comparing with the average expression level of parental gene expressions. These expression patterns were represented with several terms, such as “additivity” and “non-additivity”. Additivity can be considered as the expression conservation of parental genes and non-additivity, as a major feature of transcriptome shock, was referred to as altered expression in which the expression level of the hybrid is not equal to the average of their parental expression levels (Yoo et al., 2014). These non-additive expressions often contribute to ecological competitiveness, adaptation, and evolutionary plasticity relative to their

parent species (Chen, 2007; Abbott et al., 2013). The pattern of non-additive expressions is classified into transgressive up- and down-regulation, and expression level dominance (ELD), which indicate that the total expression of homeologous pairs changed to be similar to the expression of the single gene in one of the parents, and ELD has been the most remarkable phenomenon in almost allopolyploid plants (Rapp et al., 2009; Chelaifa et al., 2010; Bardil et al., 2011; Grover et al., 2012; Yoo et al., 2013; Wu et al., 2016; Zhang et al., 2016).

Transcriptional regulation is generally dependent on *cis*- or *trans*-regulatory elements (Rockman and Kruglyak, 2006; Williams et al., 2007). Since the *trans*-acting factors can interact both *cis*-elements of duplicated genes derived from both parental genomes in a single nucleus of polyploids, gene expression changes in polyploidy could be regulated by differences in *cis* and *trans*-elements between parental genomes (Tirosh et al., 2009; McManus et al., 2010; He et al., 2012; Shi et al., 2012). For example, if *cis*-element of duplicated genes are identical, shared transcription factor can bind both of duplicated genes and expressions of duplicated gene were changed to be similar transcription levels. Several studies in hybrid species of yeast, *Drosophila*, *Arabidopsis*, maize, and rice showed that such effects of *cis*- and *trans*-divergence is important to transcription regulation in

polyploidy, and proposed that genetic distance between parental genomes might be related with expression changes in polyploidy (Springer and Stupar, 2007; Tirosh et al., 2009; McManus et al., 2010; Shi et al., 2012; Wu et al., 2016).

Diploidization of a polyploid genome

After WGDs or hybridization between different species, followed by chromosome doubling, polyploid plants typically undergo genomic reconfigurations over a long period of time, leading to the downsizing of the genome. This accumulation of genomic changes eventually leads to diploid-like genomic structures. Several processes lead to downsizing of the genome in polyploids. Nonhomologous recombination between repeat sequences is associated with the loss of DNA in polyploid genomes (Devos et al., 2002). Chromosome size or number can be reduced by translocations between nonhomologous and homoeologous chromosomes (Mandakova et al., 2016; Mandakova et al., 2010; Mandakova et al., 2017b). Newly synthesized polyploids often suffer from chromosomal instability, causing aneuploidy over several generations. The entire loss of a chromosome could be associated with the downsizing of polyploid genomes (Mestiri et al., 2010; Wu et al., 2018; Xiong et al., 2011). Chromosomal fractionation also occurs

during the process of diploidization (Freeling et al., 2015). Duplicated genes or regulatory elements can be fractionated, generating single-copy genes (De Smet et al., 2013); this fractionation process can be induced predominantly in one parental subgenome, which is called biased fractionation (Thomas et al., 2006). This phenomenon has been found in many polyploids, including maize (Schnable et al., 2011), Chinese cabbage (Cheng et al., 2012), and cotton (Renny-Byfield et al., 2015). Depending on the degree of diploidization, the polyploid genome can be divided into neopolyploid, mesopolyploid, and paleopolyploid types, based on their structural changes (Hohmann et al., 2015). Neopolyploids are recently formed polyploids, such as oilseed rape and monkeyflower, which exhibit chromosomal evidence of polyploidy, with easily distinguishable parental subgenomes (Chalhoub et al., 2014; Edger et al., 2017). A mesopolyploid genome is more fractionized than a neopolyploid genome due to the process of diploidization. Mesopolyploid genomes often show biased subgenome fractionation (Liu et al., 2014; Mandakova et al., 2017a; Wang et al., 2011). Paleopolyploid genomes are typically highly diploidized genomes showing diploid-like meiosis. Newly formed diploid-like genomes may acquire novel characteristics during the diploidization process, resulting in phenotypic diversity during speciation. After then, rapid genomic changes in diploid

species might be induced by the formation of polyploids for evolutionary adaptations. These repeated cycles of polyploidization and diploidization may provide genomic diversification during the evolution of plant lineages.

U's triangle and intergeneric allotetraploid x*Brassicoraphanus*

Three Brassica species, *B. napus* (AACC; 2n = 4x = 38), *B. juncea* (AABB; 2n = 4x = 36), and *B. carinata* (BBCC; 2n = 4x = 34), were naturally emerged by polyploidization among three diploid species, *B. rapa* (AA; 2n = 2x = 20), *B. nigra* (BB; 2n = 2x = 16), *B. oleracea* (CC; 2n=2x=18). These evolutionary relationship among *Brassica* species was described by the "U's triangle" (U, 1935), and this model suggests that the synthesis of species is one of the driving forces for plant evolution. Hybridization can arise between two extremely divergent species belonging to a different genus. From 1826 by Sageret (Oost, 1984), the intergeneric hybrids between *Brassica* and *Raphanus* have been continuously reported (Karpechenko, 1928; Mcnaughton, 1973; Dolstra, 1982; Lee et al., 2011). Even with many attempts to generate intergeneric hybrids between *Brassica* and *Raphanus* only a few successful examples were reported. The recently developed x*Brassicoraphanus* (AARR; 2n = 4x = 38), an intergeneric

hybrid between *B. rapa* (AA; 2n = 2x = 20) and *R. sativus* (RR; 2n = 2x = 18), was generated by embryo rescue after crossing and following microspore culture with induced mutation (Lee et al., 2011). It has been self-fertilized over 10 generations resulting in stable fertility (Lee et al., 2011). As a rare case of intergeneric allopolyploids, x*Brassicoraphanus* is a promising material to study extreme polyploidization events and provides a model to address questions.

This study focused on the intergeneric allopolyploidization effects on genomic, epigenomic and transcriptomic changes using bioinformatics analysis. The thesis work address the following two topic:

Chapter 1: Parental divergence allows hybrid genome stabilization in x*Brassicoraphanus*

Chapter 2: Divergence of *cis*- and *trans*-regulatory elements drives reconstruction of transcriptome network in x*Brassicoraphanus*

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., *et al.* (2013). Hybridization and speciation. J. Evol. Biol. **26,** 229-246.

Adams, K.L., and Wendel, J.F. (2005). Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. **8,** 135-141.

Ahuja, M.R. (2005). Polyploidy in gymnosperms: Revisited. Silvae Genet. **54,** 59-69.

Bardil, A., de Almeida, J.D., Combes, M.C., Lashermes, P., and Bertrand, B. (2011). Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. New Phytol. **192,** 760-774.

Blanc, G., and Wolfe, K.H. (2004). Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell **16,** 1679-1691.

Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S.Q., Ishizaki, K. *et al.* (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. Cell **171,** 287.

Breuert, S., Allers, T., Spohn, G., Soppa, J. (2006). Regulated polyploidy in halophilic archaea. Plos One **1,** e92.

Buggs, R.J.A., Zhang, L.J., Miles, N., Tate, J.A., Gao, L., Wei, W. *et al.* (2011). Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. Curr. Biol. **21,** 551-556.

Casanova, M.T. (2015). Chromosome numbers in Australian charophytes (Characeae, Charophyceae). Phycologia **54,** 149-160.

Chalhoub, B., Denoeud, F., Liu, S.Y., Parkin, I.A.P., Tang, H.B., Wang, X.Y. *et al.* (2014). Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. Science **345,** 950-953.

Chao, D.Y., Dilkes, B., Luo, H.B., Douglas, A., Yakubova, E. *et al.* (2013). Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. Science **341,** 658-659.

Chelaifa, H., Monnier, A., and Ainouche, M. (2010). Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina x townsendii* and *Spartina anglica* (Poaceae). New Phytol. **186,** 161-174.

Chen, S., Ren, F., Zhang, L., Liu, Y., Chen, X., Li, Y. *et al.* (2018). Unstable allotetraploid tobacco genome due to frequent homeologous recombination, segmental deletion, and chromosome loss. Mol. Plant **11,** 914-927.

Chen, Z.J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu. Rev. Plant Biol. **58,** 377-406.

Cheng, F., Wu, J., Cai, X., Liang, J.L., Freeling, M., and Wang, X.W. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. Nat. Plants **4,** 258-268.

Cheng, F., Wu, J., Fang, L., Sun, S.L., Liu, B., Lin, K., Bonnema, G. *et al.* (2012). Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. Plos One **7**.

Comai, L. (2005). The advantages and disadvantages of being polyploid. Nat. Rev. Genet. **6,** 836-846.

Combes, M.C., Hueber, Y., Dereeper, A., Rialle, S., Herrera, J.C., and Lashermes, P. (2015). Regulatory divergence between parental

alleles determines gene expression patterns in hybrids. Genome Biol. Evol. **7,** 1110-1121.

De Smet, R., Adams, K.L., Vandepoele, K., Van Montagu, M.C., Maere, S., and Van de Peer, Y. (2013). Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc. Natl. Acad. Sci. U.S.A. **110,** 2898-2903.

Devos, K.M., Brown, J.K.M., and Bennetzen, J.L. (2002). Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res. **12,** 1075-1079.

Dion-Cote, A.M., and Barbash, D.A. (2017). Beyond speciation genes: an overview of genome stability in evolution and speciation. Curr. Opin. Genet. Dev. **47,** 17-23.

Dolstra, O. (1982). Synthesis and fertility of *Brassicoraphanus* and ways of transferring *Raphanus* characters to *Brassica* (Pudoc).

Edger, P.P., Smith, R., McKain, M.R., Cooley, A.M., Vallejo-Marin, M., Yuan, Y., *et al.* (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. Plant Cell **29,** 2150-2167.

Edger, P.P., Poorten, T., VanBuren, R., Hardigan, M.A., Colle, M., McKain, M.R., *et al.* (2019). Origin and evolution of the octoploid strawberry genome. Nat. Genet. **51,** 541.

Fawcett, J.A., Maere, S., and Van De Peer, Y. (2009). Plants with double genomes might have had a better chance to survive the cretaceous–tertiary extinction event. Proc. Natl. Acad. Sci. U.S.A. **106,** 5737-5742.

Freeling, M., Scanlon, M.J., and Fowler, J.E. (2015). Fractionation and

subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. Curr. Opin. Genet. Dev. **35,** 110-118.

Gaeta, R.T., Pires, J.C., Iniguez-Luy, F., Leon, E., and Osborn, T.C. (2007). Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell **19,** 3403-3417.

Grover, C.E., Gallagher, J.P., Szadkowski, E.P., Yoo, M.J., Flagel, L.E., and Wendel, J.F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol. **196,** 966-971.

Hahn, M.A., van Kleunen, M., and Muller-Scharer, H. (2012). Increased phenotypic plasticity to climate may have boosted the invasion success of polyploid *Centaurea stoebe*. Plos One **7,** e50248.

Hallinan, N.M., and Lindberg, D.R. (2011). Comparative analysis of chromosome counts infers three paleopolyploidies in the *Mollusca*. Genome Biol. Evol. **3,** 1150-1163.

He, F., Zhang, X., Hu, J.Y., Turck, F., Dong, X., Goebel, U., Borevitz, J., *et al.* (2012). Genome-wide analysis of *cis*-regulatory divergence between species in the *Arabidopsis* genus. Mol. Biol. Evol. **29,** 3385-3395.

Hegarty, M.J., Barker, G.L., Wilson, I.D., Abbott, R.J., Edwards, K.J., and Hiscock, S.J. (2006). Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. Curr. Biol. **16,** 1652-1659.

Hohmann, N., Wolf, E.M., Lysak, M.A., and Koch, M.A. (2015). A time-calibrated road map of Brassicaceae species radiation and evolutionary history. Plant Cell **27,** 2770-2784.

Hubbard, M.J., Poulter, R.T., Sullivan, P.A., and Shepherd, M.G. (1985). Characterization of a tetraploid derivative of *Candida albicans* ATCC 10261. J. Bacteriol. **161,** 781-783.

International Wheat Genome Sequencing, C. (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science **345,** 1251788.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., *et al.* (2011). Ancestral polyploidy in seed plants and angiosperms. Nature **473,** 97-100.

Kapraun, D.F. (2007). Nuclear DNA content estimates in green algal lineages: chlorophyta and streptophyta. Ann. Bot. **99,** 677-701.

Karpechenko, G.D. (1928). Polyploid hybrids of *Raphanus sativus* L. x*Brassica oleracea* L. Mol. Gen. Genet. **48,** 1-85.

Khalid, M.F., Hussain, S., Anjum, M.A., Ahmad, S., Ali, M.A., Ejaz, S., *et al.* (2020). Better salinity tolerance in tetraploid vs diploid volkamer lemon seedlings is associated with robust antioxidant and osmotic adjustment mechanisms. J. Plant Physiol. **244,** 153071.

Kihara, H., and Ono, T. (1926). Chromosomenzahlen und systematische gruppierung der Rumex-arten. Cell Tissue Res. **4,** 475-481.

Lang, D., Ullrich, K.K., Murat, F., Fuchs, J., Jenkins, J., Haas, F.B., *et al.* (2018). The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. Plant J. **93,** 515-533.

Lee, S.S., Lee, S.A., Yang, J., and Kim, J. (2011). Developing stable progenies of *xBrassicoraphanus*, an intergeneric allopolyploid between *Brassica rapa* and *Raphanus sativus*, through induced mutation using microspore culture. Theor. Appl. Genet. **122,** 885-891.

Li, Z., Baniaga, A.E., Sessa, E.B., Scascitelli, M., Graham, S.W., Rieseberg, L.H., *et al.* (2015). Early genome duplications in conifers and other seed plants. Sci Adv **1,** e1501084.

Li, Z., Tiley, G.P., Galuska, S.R., Reardon, C.R., Kidder, T.I., Rundell, R.J., and Barker, M.S. (2018). Multiple large-scale gene and genome duplications during the evolution of hexapods. Proc. Natl. Acad. Sci. U.S.A. **115,** 4713-4718.

Liu, S.Y., Liu, Y.M., Yang, X.H., Tong, C.B., Edwards, D., Parkin, I.A.P., *et al.* (2014). The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat. Commun. **5,** 3930.

Mable, B.K., Alexandrou, M.A., and Taylor, M.I. (2011). Genome duplication in amphibians and fish: an extended synthesis. J. Zool. **284,** 151-182.

Maldonado, R., Jimenez, J., and Casadesus, J. (1994). Changes of ploidy during the *Azotobacter vinelandii* growth cycle. J. Bacteriol. **176,** 3911-3919.

Mandakova, T., Gloss, A.D., Whiteman, N.K., and Lysak, M.A. (2016). How diploidization turned a tetraploid into a pseudotriploid. Am. J. Bot. **103,** 1187-1196.

Mandakova, T., Li, Z., Barker, M.S., and Lysak, M.A. (2017a). Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. Plant J. **91,** 3-21.

Mandakova, T., Joly, S., Krzywinski, M., Mummenhoff, K., and Lysak, M.A. (2010). Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. Plant Cell **22,** 2277-2290.

Mandakova, T., Pouch, M., Harmanova, K., Zhan, S.H., Mayrose, I., and Lysak, M.A. (2017b). Multispeed genome diploidization and diversification after an ancient allopolyploidization. Mol. Ecol. **26,** 6445-6462.

Masterson, J. (1994). Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science **264,** 421-424.

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

McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R., and Wittkopp, P.J. (2010). Regulatory divergence in *Drosophila* revealed by mRNA-seq. Genome Res. **20,** 816-825.

Mcnaughton, I.H. (1973). Synthesis and Sterility of *Raphanobrassica*. Euphytica **22,** 70-88.

Mestiri, I., Chague, V., Tanguy, A.M., Huneau, C., Huteau, V., Belcram, H., *et al.* (2010). Newly synthesized wheat allohexaploids display progenitor-dependent meiotic stability and aneuploidy but structural genomic additivity. New Phytol. **186,** 86-101.

Miller, M., Song, Q.X., Shi, X.L., Juenger, T.E., and Chen, Z.J. (2015). Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of *Arabidopsis*. Nat. Commun. **6,** 7456.

Nossa, C.W., Havlak, P., Yue, J.X., Lv, J., Vincent, K.Y., Brockmann, H.J., *et al.* (2014). Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. Gigascience **3,** 9.

Oost, E. (1984). x*Brassicoraphanus* Sageret or x*Raphanobrassica* Karpechenko? Cruciferae Newsletter **9,** 11-12.

Otto, S.P. (2007). The evolutionary consequences of polyploidy. Cell **131,** 452-462.

Pecoraro, V., Zerulla, K., Lange, C., and Soppa, J. (2011). Quantification of ploidy in proteobacteria revealed the existence of monoploid, (mero-)oligoploid and polyploid species. Plos One **6,** e16392.

Pomper, S., Daniels, K.M., and McKee, D.W. (1954). Genetic analysis of polyploid yeast. Genetics **39,** 343-355.

Postgate, J.R., Kent, H.M., Robson, R.L., and Chesshyre, J.A. (1984). The genomes of *Desulfovibrio gigas* and *D. vulgaris*. J. Gen. Microbiol. **130,** 1597-1601.

Ramsey, J., and Schemske, D.W. (1998). Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu. Rev. Ecol. Syst. **29,** 467-501.

Ramsey, J., and Schemske, D.W. (2002). Neopolyploidy in flowering plants. Annu. Rev. Ecol. Syst. **33,** 589-639.

Rapp, R.A., Udall, J.A., and Wendel, J.F. (2009). Genomic expression dominance in allopolyploids. BMC Biol. **7,** 18.

Renny-Byfield, S., Gong, L., Gallagher, J.P., and Wendel, J.F. (2015). Persistence of subgenomes in paleopolyploid cotton after 60 mya of evolution. Mol. Biol. Evol. **32,** 1063-1071.

Rockman, M.V., and Kruglyak, L. (2006). Genetics of global gene expression. Nat. Rev. Genet. **7,** 862-872.

Ruiz, M., Quinones, A., Martinez-Cuenca, M.R., Aleza, P., Morillon, R., Navarro, L., *et al.* (2016). Tetraploidy enhances the ability to exclude chloride from leaves in carrizo citrange seedlings. J. Plant Physiol. **205,** 1-10.

Ruprecht, C., Lohaus, R., Vanneste, K., Mutwil, M., Nikoloski, Z., Van

de Peer, Y., *et al.* (2017). Revisiting ancestral polyploidy in plants. Sci Adv **3,** e1603195.

Schmid, M., Evans, B.J., and Bogart, J.P. (2015). Polyploidy in Amphibia. Cytogenet. Genome Res. **145,** 315-330.

Schnable, J.C., Springer, N.M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc. Natl. Acad. Sci. U.S.A. **108,** 4069-4074.

Schwager, E.E., Sharma, P.P., Clarke, T., Leite, D.J., Wierschin, T., Pechmann, M., *et al.* (2017). The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biol. **15,** 62.

Selmecki, A.M., Maruvka, Y.E., Richmond, P.A., Guillet, M., Shoresh, N., Sorenson, A.L., *et al.* (2015). Polyploidy can drive rapid adaptation in yeast. Nature **519,** 349-352.

Shen, Y., Sun, S., Hua, S., Shen, E., Ye, C.Y., Cai, D., *et al.* (2017). Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid *Brassica napus* uncovers key roles for small RNAs. Plant J. **91,** 874-893.

Shi, X.L., Ng, D.W.K., Zhang, C.Q., Comai, L., Ye, W.X., and Chen, Z.J. (2012). *Cis*- and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. Nat. Commun. **3,** 950.

Shimizu-Inatsugi, R., Terada, A., Hirose, K., Kudoh, H., Sese, J., and Shimizu, K.K. (2017). Plant adaptive radiation mediated by polyploid plasticity in transcriptomes. Mol. Ecol. **26,** 193-207.

Smarda, P., Horova, L., Knapek, O., Dieck, H., Dieck, M., Razna, K., *et al.* (2018). Multiple haploids, triploids, and tetraploids found in

modern-day "living fossil" *Ginkgo biloba*. Hortic Res **5,** 55.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C.F., *et al.* (2009). Polyploidy and angiosperm diversification. Am. J. Bot. **96,** 336-348.

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

Song, Q., and Chen, Z.J. (2015). Epigenetic and developmental regulation in plant polyploids. Curr. Opin. Plant Biol. **24,** 101-109.

Song, Q., Zhang, T., Stelly, D.M., and Chen, Z.J. (2017). Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biol **18,** 99.

Sora, D., Kron, P., and Husband, B.C. (2016). Genetic and environmental determinants of unreduced gamete production in *Brassica napus*, *Sinapis arvensis* and their hybrids. Heredity **117,** 440-448.

Springer, N.M., and Stupar, R.M. (2007). Allelic variation and heterosis in maize: How do two halves make more than a whole? Genome Res. **17,** 264-275.

Thomas, B.C., Pedersen, B., and Freeling, M. (2006). Following tetraploidy in an *Arabidopsis* ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Res. **16,** 934-946.

Tirosh, I., Reikhav, S., Levy, A.A., and Barkai, N. (2009). A yeast hybrid provides insight into the evolution of gene expression regulation. Science **324,** 659-662.

Todesco, M., Pascual, M.A., Owens, G.L., Ostevik, K.L., Moyers, B.T.,

Hubner, S., *et al.* (2016). Hybridization and extinction. Evol. Appl. **9,** 892-908.

U, N. (1935). Genome-analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. J. Jpn. Genet. **7,** 784-794.

Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. Nat. Rev. Genet. **18,** 411.

Vanneste, K., Maere, S., and Van de Peer, Y. (2014). Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. Philos. Trans. R. Soc. Lond. B Biol. Sci. **369,** 1648.

Wang, J., Li, D., Shang, F., and Kang, X. (2017). High temperature-induced production of unreduced pollen and its cytological effects in *Populus*. Sci. Rep. **7,** 5281.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., *et al.* (2011). The genome of the mesopolyploid crop species *Brassica rapa*. Nat. Genet. **43,** 1035-1039.

Williams, R.B., Chan, E.K., Cowley, M.J., and Little, P.F. (2007). The influence of genetic variation on gene expression. Genome Res. **17,** 1707-1716.

Wu, G.-Q., Lin, L.-Y., Jiao, Q., and Li, S.-J. (2019). Tetraploid exhibits more tolerant to salinity than diploid in sugar beet (*Beta vulgaris L.*). Acta Physiol. Plant **41,** 52.

Wu, J., Lin, L., Xu, M., Chen, P., Liu, D., Sun, Q., *et al.* (2018a). Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. BMC Genomics **19,** 586.

Wu, Y., Sun, Y., Wang, X.T., Lin, X.Y., Sun, S., Shen, K., *et al.* (2016).

Transcriptome shock in an interspecific F1 triploid hybrid of *Oryza* revealed by RNA sequencing. J. Integr. Plant Biol. **58,** 150-164.

Wu, Y., Sun, Y., Sun, S., Li, G., Wang, J., Wang, B., *et al.* (2018b). Aneuploidization under segmental allotetraploidy in rice and its phenotypic manifestation. Theor. Appl. Genet. **131,** 1273-1285.

Xiong, Z., Gaeta, R.T., and Pires, J.C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc. Natl. Acad. Sci. U.S.A. **108,** 7908-7913.

Yang, J., Liu, D., Wang, X., Ji, C., Cheng, F., Liu, B., *et al.* (2016). The genome sequence of allopolyploid *Brassica juncea* and analysis of differential homoeolog gene expression influencing selection. Nat. Genet. **48,** 1225-1232.

Yang, P.M., Huang, Q.C., Qin, G.Y., Zhao, S.P., and Zhou, J.G. (2014). Different drought-stress responses in photosynthesis and reactive oxygen metabolism between autotetraploid and diploid rice. Photosynthetica **52,** 193-202.

Ye, B., Wang, R., and Wang, J. (2016). Correlation analysis of the mRNA and miRNA expression profiles in the nascent synthetic allotetraploid *Raphanobrassica*. Sci. Rep. **6,** 37416.

Yoo, M.J., Szadkowski, E., and Wendel, J.F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity **110,** 171-180.

Yoo, M.J., Liu, X., Pires, J.C., Soltis, P.S., and Soltis, D.E. (2014). Nonadditive gene expression in polyploids. Annu. Rev. Genet. **48,** 485-517.

Zhang, H., Gou, X., Zhang, A., Wang, X., Zhao, N., Dong, Y., *et al.*

(2016). Transcriptome shock invokes disruption of parental expression-conserved genes in tetraploid wheat. Sci. Rep. **6,** 26363.

Zhang, H., Bian, Y., Gou, X., Zhu, B., Xu, C., Qi, B., *et al.* (2013). Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. Proc. Natl. Acad. Sci. U.S.A. **110,** 3447-3452.

Zhang, M., Liu, X.K., Fan, W., Yan, D.F., Zhong, N.S., Gao, J.Y., *et al.* (2018). Transcriptome analysis reveals hybridization-induced genome shock in an interspecific F1 hybrid from *Camellia*. Genome **61,** 477-485.

Zhang, T.Z., Hu, Y., Jiang, W.K., Fang, L., Guan, X.Y., Chen, J.D., *et al.* (2015). Sequencing of allotetraploid cotton (*Gossypium hirsutum L.* acc. TM-1) provides a resource for fiber improvement. Nat. Biotechnol. **33,** 531-U252.

Zhou, L., and Gui, J. (2017). Natural and artificial polyploids in aquaculture. Aquacult. Fish. **2,** 103-111.

CHAPTER 1

Parental divergence allows hybrid genome stabilization in x*Brassicoraphanus*

ABSTRACT

Hybridization and subsequent genome duplication have facilitated the evolution of many plant species while producing novel characteristics to increase fitness to the new environment. x*Brassicoraphanus* is an intergeneric allotetraploid synthesized from a cross between Chinese cabbage (*Brassica rapa*) and radish (*Raphanus sativus*). Unlike most neoallopolyploid plants that often suffer from sterility and aneuploidy, x*Brassicoraphanus* is fertile and genetically stable, with many characteristics displayed as mixed phenotypes. Here, I showed that x*Brassicoraphanus* retains complete sets of parental chromosomes without apparent chromosome rearrangement, and burst of transposable elements. Besides, DNA methylation levels of genic body regions were maintained but some retrotransposons in x*Brassicoraphanus* were hypermethylated and transcriptionally silenced. This study suggests that a certain level of parental genome divergence is helpful to suppress genome shuffling, and the silencing of transposable elements by hypermethylation promotes the hybrid genome stabilization while ameliorating genomic shocks.

INTRODUCTION

Polyploidy is the condition of cells or organisms possessing multiple sets of chromosomes. These phenomena have occurred in various organisms, especially in plants, and the evolution of complex and various sizes of plant genomes has been attributed to polyploidization events (Van de Peer et al., 2017; Cheng et al., 2018).Although hybridization and allopolyploidization is one of the driving forces of the plant evolution, providing increasing phenotypic variability, ecological competitiveness and evolutionary plasticity for survival in nature (Wendel, 2000; Chen, 2007; Soltis and Soltis, 2009; Abbott et al., 2013; Van de Peer et al., 2017; Cheng et al., 2018), strong hybridization barriers generally prevent gene flows between different species in plants (Abbott et al., 2013). Several mechanisms have been proposed to explain the postzygotic barriers resulting from genome incompatibility between distantly related species (Lafon-Placette and Kohler, 2015; Dion-Cote and Barbash, 2017). Among them, interspecific hybrids and allopolyploids often experience genomic changes including gene loss, transposon activation, and chromosome

rearrangement, as well as epigenetic changes including altered DNA methylation. A “genome shock”, which elicits restructuring of the genome through changes of chromosomal organization or mobilization of transposable elements (TEs) upon hybridization, is proposed as one of the critical causes of genome destabilization upon hybridization (McClintock, 1984; Kashkush et al., 2002; Parisod et al., 2010; Petit et al., 2010; Diez et al., 2014; Jackson, 2017). For example, gene conversion, homoeologous recombination and aneuploidy are frequently detected in the recently formed allopolyploids (Song et al., 1995; Gaeta et al., 2007; Xiong et al., 2011; Zhang et al., 2013; Chalhoub et al., 2014; Chen et al., 2018), and the increase of TE activity through epigenetic changes such as DNA methylation and siRNAs were found in various interspecific hybrids or allopolyploids (Ha et al., 2009; He et al., 2013; Shen et al., 2017; Jiao et al., 2018). Moreover, small RNAs have been reported to affect genome stability by activating or silencing TEs as a buffer in allotetraploids (Ha et al., 2009; Barber et al., 2012).

The Brassicaceae family contains a variety of agronomically important crop species such as broccoli, cabbage, cauliflower, oilseed rape, radish and turnip, in addition to a model plant Arabidopsis. In particular, the genus *Brassica* is well known for hybridization between different species

within the same genus (interspecific hybridization). For instance, three diploid species including *Brassica rapa* (AA), *B. nigra* (BB), and *B. oleracea* (CC) can be hybridized to each other generating allotetraploid species *B. napus* (AACC), *B. juncea* (AABB) and *B. carinata* (BBCC), which is epitomized by the model of 'Triangle of U' (U, 1935). Hybridization between species in the Brassicaceae family is not restricted to interspecific hybridization. Since 1826 by Sageret (Oost, 1984), intergeneric hybrids between *Brassica* and *Raphanus* have been sporadically reported (Karpechenko, 1928; Mcnaughton, 1973; Dolstra, 1982). However, because of their genetic instability and poor fertility, very few hybrids have survived for many generations. The recently developed x*Brassicoraphanus koranhort* cv. BB1 (x*B*; AARR; 2n = 4x = 38, also known as Baemoochae) is an intergeneric allotetraploid between *B. rapa* (AA; 2n = 2x = 20) and *Raphanus sativus* (RR; 2n = 2x = 18) (Lee et al., 2011). Unlike other intergeneric hybrids in the Brassicaceae family, x*B* is self-fertile and genetically stable while displaying phenotypic uniformity in successive generations (more than thirty generations). Genetic and phenotypic stability of x*B* is very exceptional considering that many allopolyploids often display a high degree of genome instability and sterility issues, indicating that the

hybridization barrier was overcome immediately after the two genomes have merged.

I hypothesized that allopolyploidization events have somewhat ameliorated deleterious genome shock phenomena, and thereby overcome an intrinsic hybridization barrier between distantly related species. I here report the genome structure and epigenomic changes in a new allopolyploid species x*B*, while discussing the possible mechanisms by which two divergent genomes can successfully merge into a novel species.

MATERIALS AND METHODS

Plant materials

x*Brassicoraphanus* cv. BB1, *Brassica rapa* L. cv. Chiifu-401-42 (*Br*), and *Raphanus sativus* L. cv. WK10039 (*Rs*) were grown on 1x Murashige and Skoog (MS) medium (Duchefa) in a growth chamber under 16 hr of fluorescent light at 20 ± 10 μmol m^{-2} s^{-1}, 22°C for 14 days. The seedlings including shoots and roots were harvested together for whole genome-seq, RNA-seq, bisulfite (BS)-seq, chromatin immunoprecipitation (ChIP)-seq and small RNA-seq.

Genome sequencing, assembly and genome size estimation

Paired-end and mate-pair sequencing libraries with insert sizes of 200 bp, 400 bp, 3 kb, 8 kb 5 kb, 10 kb and 15 kb were constructed using KAPA library prep kit (Roche) and Illumina Mate Pair Library kit (Illumina) following the manufacturer's instructions (Table 1-1). The libraries were sequenced on an Illumina HiSeq 2000 platform. Prokaryotic sequences, duplicated reads, low quality reads and low frequency reads were filtered

out (Table 1-1). The preprocessed sequences were assembled using SOAPdenovo2 (Luo et al., 2015) with the best *k*-mer values for each library. To increase the length of scaffolds, serial scaffolding processes were carried out using SOAPdenovo2 (Luo et al., 2015) and SSPACE (Boetzer et al., 2011). Gaps in the scaffolds were reduced further using SOAPdenovo Gapcloser (Luo et al., 2015) and Platanus (Kajitani et al., 2014) (Table 1-2). The genome size of x*B* was estimated by flow cytometry analysis (FACSCalibur, BD Biosciences) as previously described (Huang et al., 2013). Genome data were visulizaed with Circos (Krzywinski et al., 2009).

Chloroplast genome assembly

The chloroplast genome was *de novo* assembled from the 1x coverage of whole-genome sequencing reads. The chloroplast genome was annotated with Dual Organellar GenoMe Annotator DOGMA (Wyman et al., 2004) and manually curated. The chloroplast genome was visualized using OrganellarGenomeDRAW (Lohse et al., 2013).

Assignment of scaffolds to A and R subgenomes

Whole-genome sequencing reads of *Rs* and *Br* from Brassica Database (BRAD) were mapped to the x*B* scaffolds using Bowtie

(Langmead et al., 2009). The number of mapped reads was counted and the scaffolds were assigned to A and R subgenomes of x*B*, based on a comparison of the number of parental reads (A subgenome of x*B*: >99% ratio of mapped reads from *Br*; R subgenome of x*B*: >99% ratio of mapped reads from *Rs*). Next, assigned x*B* scaffolds were anchored to the reference chromosomes of *Br* and *Rs* to build x*B* pseudo-chromosomes.

Gene and TE annotations and repeat analysis

Gene annotations of x*B* and *Rs* were performed following the previous annotation pipeline with minor modifications (Kim et al., 2014). Briefly, the annotation pipeline consisted of repeat masking, mapping of different protein sequence sets and mapping of RNA-seq reads. Independent *ab initio* predictions were performed with AUGUSTUS (Stanke et al., 2008). The EVidenceModeler (Haas et al., 2008) software combines *ab initio* gene predictions with protein and transcript alignments into weighted consensus gene structures. Gene annotation of *Br* was downloaded from Ensembl plant (ftp://ftp.ensemblgenomes.org/pub/plants/release-31/gff3/brassica_rapa/) and additional 1,700 genes were annotated using Exonerate (Slater and Birney, 2005). Functional annotation was performed through BLASTP against SwissProt and Plant RefSeq database. TE-related repeat sequences

were predicted by RepeatModeler (Smit and Hubley, 2008) and Repeatmasker (Smit et al., 2015). The proportion of repeat sequences was analyzed using dnaPipeTE (Goubert et al., 2015) with 1x whole genome sequencing reads. Whole genome sequencing reads from 8 cultivars (*R. sativus* cv. Wonkyo10039, *R. sativus* cv. Chungwoon, *R. sativus* cv. Kb68-1, *R. sativus* cv. Wonyeon25, *B. rapa* cv. Chiifu-401-42, *B. rapa* cv. CM218, *B. rapa* cv. Hagam and *B. rapa* cv. CR291M-64) were used for estimating repeat proportions of *Br* and *Rs*.

Fluorescence *in situ* hybridization (FISH) analysis

The sequences of 5S rDNA, 45S rDNA, *Rs*Cent1, *Rs*Cent2, *Br*Cent1, *Br*Cent2, *Rs*STRa, *Rs*STRb, *Br*STRa, *Br*STRb and telomere were used as probes (Table 1-3). The probes were labeled by nick translation with different fluorochromes. Root mitotic chromosome spreads and FISH procedures were performed according to the previous method (Waminal and Kim, 2012). For directly labeled probes, slides were immediately used for FISH after fixation with 4% paraformaldehyde, without subsequent pepsin and RNase pretreatment. Images were captured with an Olympus BX53 fluorescence microscope equipped with a Leica DFC365 FS CCD camera and processed using Cytovision ver. 7.2 (Leica Microsystems).

Identification of orthologous and homoeologous gene pairs

To identify orthologous gene pairs between parental genomes (*Br* vs. *Rs*), the reciprocal best BLAST hit was performed with >80% of identity and >80% of coverage. Syntenic regions were defined as contiguous regions containing at least five homologous gene pairs in *Br* and *Rs* genomes, and the pairs in the syntenic regions were determined as orthologous gene pairs. Homoeologous gene pairs between the progenitor genomes (A subgenome of x*B* vs. R subgenome of x*B*) were determined following the same standard.

RNA-seq analysis

Total RNA was extracted with RNeasy plant kit (Qiagen) as the manufacture's protocol. The DNase treated RNA samples, including two replicates for each of seedling, leaf, hypocotyl and flower, and one replicate for root of x*B* and its parents, were used for constructing RNA-seq libraries (Zhong et al., 2011). RNA sequencing was performed through Illumina HiSeq 2000 platform. The obtained raw reads were filtered using FASTX-Toolkit and low quality reads (Q < 20) were removed. The filtered reads were mapped on *Br* genome (https://plants.ensembl.org/Brassica_rapa), *Rs* genome (http://radish-genome.org/), and x*B* genome using Tophat (Trapnell

et al., 2009) with default parameters. The mapped read counts were calculated using HTSeq (Anders et al., 2015). Statistical tests of DEGs were performed using EdgeR (Robinson et al., 2010) with false discovery rate (FDR) < 0.05 and fragments per kilobase of transcript per million mapped reads (FPKM) $\log_2$ fold change > 1.

BS-seq analysis

Genomic DNA (5 μg) was used to construct BS-seq library with the KAPA Library kit (Roche) and EpiTect Bisulfite Kits (Qiagen) according to manufacturer's instructions. The libraries were sequenced using the Illumina HiSeq 2000. Raw reads were filtered using FASTX-Toolkit and low quality reads (Q < 20) were removed. Reads were mapped on x*B* genome using BISON (Ryan and Ehninger, 2014), with the parameters "--very-sensitive --score-min 'L,-0.6,-0.6'". Only cytosine sites with 4x coverage read depths were accepted for the subsequent analysis. Differentially methylated cytosines (DMCs) and regions (DMRs) were identified as described previously (Kim et al., 2019). In brief, DMCs were identified using Fisher's exact test (P < 0.05) between levels of methylation in x*B* and parental genomes. DMRs were identified based on the regions with a length ≥ 200 bp, ≥ 5 DMCs, and the mean methylation difference ≥ 0.3 for CG, ≥ 0.15 for

CHG, and ≥ 0.1 for CHH. For metagene plot of DNA methylation in gene bodies and repeat, regions of 2 kb upstream, downstream and body were divided into 50 bp windows and methylation levels were calculated for each window. Methylation data were visualized with the Integrated Genome Browser (Freese et al., 2016).

ChIP-seq analysis

ChIP was performed following the published protocol (Lee et al., 2007). Chromatin was immunoprecipitating with antibody against histone H3K9me2 (Abcam). ChIP-seq libraries were constructed as described in the Illumina ChIP sequencing kit (Illumina). DNA fragments with about 600 bp were excised and amplified for cluster generation and sequencing. All DNA libraries were sequenced on a HiSeq2500 (Illumina) with single end reads. The sequencing reads were quality controlled with FASTX-Toolkit and aligned to x*B* genome using Bowtie (Langmead et al., 2009) with parameters "-best -m1". H3K9me2-enriched regions were defined using SICER (Zang et al., 2009) (window size = 500, gap size = 600, FDR = 0.01) and overlapping regions between two biological replicates were identified using the MergePeaks module of the Homer software (Heinz et al., 2010).

Small RNA-seq analysis

The small RNA libraries were constructed using Illumina TruSeq Small RNA sample Prep kit (Illumina). The libraries were sequenced using the HiSeq 2000 (Illumina). The adaptor sequences were trimmed using cutadapt (Martin, 2011) with parameters "-g TACAGTCCGACGATC -a TGGAATTCTCGGGTGCCAAGG -m 18 -M 30". Low quality sequences were removed using FASTX-Toolkit with parameters "-q 20 -p 100". The quality-trimmed read sequences ranged from 18 to 30-nt were mapped to x*B* genome using Bowtie (Langmead et al., 2009) with parameters "-best -v 0". Mapped reads were classified into the ribosomal RNA, small nucleolar RNA, small nuclear RNA, signal recognition particle RNA, and transfer RNA using Rfam database version 12.1 (Nawrocki et al., 2014). Prediction of microRNA (miRNA) was performed with the miRDeep-P (Yang and Li, 2011) and ShortStack (Axtell, 2013b), and secondary structure was predicted using RNAfold. Candidate miRNAs were annotated by alignment to miRBase database version 21 (Kozomara and Griffiths-Jones, 2013). The target gene prediction of miRNA was perfomed by psRNATarget (Dai and Zhao, 2011). Differentially expressed miRNAs between xBrassicoraphanus and diploid parents were calculated with EdgeR (Robinson et al., 2010) with adjusted p-value < 0.05. Expression changes of miRNAs between

xBrassicoraphanus and parents were classified in to 12 possible pattern as previously reported (Rapp et al., 2009).

Northern blot analysis

Total RNA (10 μg) was electrophoresed in 1% formaldehyde denaturing gel and transferred onto a Hybond N+ membrane (GE Healthcare). The *BrGypsy*, *BrCopia* and *Actin* probes were amplified by PCR, and randomly labeled with [α-32P]dCTP (Perkin Elmer) using a Klenow fragment (3′ → 5′ exo−) (New England Biolabs). Hybridization was performed at 65°C overnight in the pre-hybridization solution containing 6X saline-sodium citrate buffer, 5X Denhardt's reagent, 1% sodium dodecyl sulfate. After hybridization, the membrane was exposed to an X-ray film (Fujifilm). Primer sequences are provided in Table 1-4.

Quantitative real-time polymerase chain reaction (qRT-PCR) analysis

cDNA synthesis was performed using the QuantiTect Reverse Transcription Kit (Qiagen) with the RNA samples which were used for transcriptome analysis. qRT-PCR was performed with a Rotor Gene Q real time PCR system (Qiagen) and QuantiFast SYBR Green PCR mater mix

(Qiagen) with the manufacture's protocol. The primer sequences for qRT-PCR were provided in Table 1-5.

Table 1-1. Summary of genomic reads from the x*Brassicoraphanus*.

Insert size	Raw data[a]	Step 1[b]	Step 2[c]	Step 3[d]	Filtered data
200 bp	39.1 Gb	38.7 Gb	33.7 Gb	26.5 Gb	26.3 Gb
400 bp	42.5 Gb	42.2 Gb	37.1 Gb	32.9 Gb	31.8 Gb
3 kbp	24.6 Gb	24.4 Gb	17.8 Gb	11.2 Gb	6.2 Gb
8 kbp	22.7 Gb	22.4 Gb	12.1 Gb	5.3 Gb	3.5 Gb
10 kbp	32.7 Gb	32.3 Gb	27.8 Gb	15.3 Gb	12.7 Gb
15 kbp	33.6 Gb	33.0 Gb	7.1 Gb	4.6 Gb	2.9 Gb
Total	195.0 Gb	193.0 Gb	135.5 Gb	95.7 Gb	83.4 Gb

[a]Original raw data.

[b]After bacterial genome removal.

[c]Amount of data after removing duplicated reads in Step 1.

[d]Amount of data after trimming low quality sequences in Step 2.

[e]Remained data in Step 3 after error correction.

Table 1-2. Statistics of x*Brassicoraphanus* genome assembly.

Step	Software	Library type	N50 (bp)	N90 (bp)	Total number	Total length (Mb)
Initial contig	SOAPdenovo2	PE	779	81	3,962,377	919.6
Scaffold 1	SOAPdenovo2	PE	17,985	1,608	167,211	586.1
Scaffold 2	SOAPdenovo2	MP	1,676,141	68,531	86,920	703.5
Scaffold 3	SSPACE	PE	1,769,221	72,908	61,802	702.3
Scaffold 4	SSPACE	MP	4,489,785	149,179	20,293	697.3
Scaffold 5	Gapcloser + Platanus	PE+MP	4,497,128	165,817	20,293	692.9
Final	adapter trimming		4,479,746	166,698	20,299	692.8

Table 1-3. Primers and oligo nucleotides used for FISH probes.

	Forward	Reverse	Label
5S rDNA	GATCCCATCAGAACTCC	GGTGCTTTAGTGCTGGTAT	FITC
45S rDNA	CACACCGCCCGTCGCTCCTACCGA	ACTCGCCGTTACTAGGGGAA	Coumarin
*Rs*Cent1	GGAGTTGTGAGTAAGAAGATATCCC	CTTAGATCGTGGTTCATCCTAGT	Coumarin
*Rs*Cent2	AGATATCCCACCTTCTATCCAGA	CTTAGATCGTGGTTCATCTTGGT	Cy5
*Br*Cent1	ACCAGGATGAATCACTTTGTAAG	TGGAACGACGAAGAAGCTGTGCTA	Texas red
*Br*Cent2	ACCAGGATGAATCGCGATGTACA	TAGAATGACGAAGAAGTTGTCATA	FITC
*Rs*STRa	TGCTAATTCCACTAATCTAGGCAGT	AGAGAAACGAAGACAACAGAGGT	FITC
*Rs*STRb	TGCTAATTCCACTAATCTAGGCTGT	AGAGTAACGGAGACTACAAAGGT	Texas red
*Br*STRa	GATCCTCCCCTTACATATTATGAATG	CAACTATGCAATAAAACAAAATTTTC	Texas red
*Br*STRc	CCCCTTACATATTATGAAGTTC	GTCGTATATTTTTAAATTTTCTC	FITC
Telomere	Oligoprobe: TTTAGGGTTTAGGGTTTAGGGTTAGGGTTTAGG		5’-Cy3 modification

Table 1-4. Primers used for northern blot probes.

	Sequence (5' to 3')
BrGypsy Forward	GCAAAGGCTTAATGGGATATC
BrGypsy Reverse	CATCATTCAGAGGTCAGGTC
BrCopia Forward	GTTACTTGATCAACATAACTCC
BrCopia Reverse	CCTTGTCTGATGTTAAGTCC
Actin Forward	CCGTATGAGCAAAGAGATCACA
Actin Reverse	TCATACCTGCTGGAATGTGC

Table 1-5. Primers for qRT-PCR.

	Sequence (5' to 3')
p1-1 Forward	TGTTGAGATAAGCTTGAAGATTAAGG
p1-1 Reverse	ATCTAGAATCAAAGATTGAAAGTATCTC
p1-2 Forward	CTCTTCTCTTCCAGTCTATACCG
p1-2 Reverse	TCTTCAAATTTTGTGTTGTTGCG
p2-1 Forward	GTGGGTGTTTAAGCTCAAAAGG
p2-1 Reverse	GGAGATCATCTTCTACCTTCTTTCG
p2-2 Forward	TGTTGGATTAAGCTTGAAGAGC
p2-2 Reverse	AACTCAAAGATTATAAGAATGCCTCTC
p3 Forward	CATCACCTGGATGTTAAAACCG
p3 Reverse	CTTGCCAAGATCACTCATCTC
p4-1 Forward	TTCGGGCTTTGGGTTGGG
p4-1 Reverse	GCCACCTTATTCTCACTGCC
p4-2 Forward	TTGGGGGATAAGGCACGTC
p4-2 Reverse	TAGAATGTGCTCAGCAAGCTC
p5-1 Forward	CTTGCATATTTGCATTGTTTTATCC
p5-1 Reverse	AGTGCTCGTCCAAAGACC
p5-2 Forward	AAGCTATTCCCTTTCTCTTTGG
p5-2 Reverse	CAAGCTCTCCTTAGAAAAGCC
p6-1 Forward	CCATTCATCCATGAGCATTTTCATC
p6-1 Reverse	GAGGTGTAAATGTGTGTCTCCAG
p6-2 Forward	GAGAAGTCTCTACCCAGTGACTC
p6-2 Reverse	TAGAATGTGCTCAGCAAGCTC
Actin Forward	GATCATGAAGTGTGACGTGG
Actin Reverse	TGCCTCATCATACTCAGCC

RESULTS

Phenotypes of x*Brassicoraphanus* intermediate between *B. rapa* and *R. sativus*

The overall morphology of x*B* is intermediate between the parent traits (Figure 1-1a). Briefly, x*B* leaf has characteristics of both the simple leaf shape of *Br* and lobed shape of *Rs* (Figure 1-1b). The x*B* hypocotyl length is roughly the average of those of *Br* and *Rs*, due mainly to difference in cell length rather than number (Figure 1-1). The long hypocotyl of *Rs* is tuberized to form a large bulbous taproot, whereas x*B* does not undergo significant root enlargement like that of *Rs*, but it does have a thicker root than *Br* (Figure 1-1a).

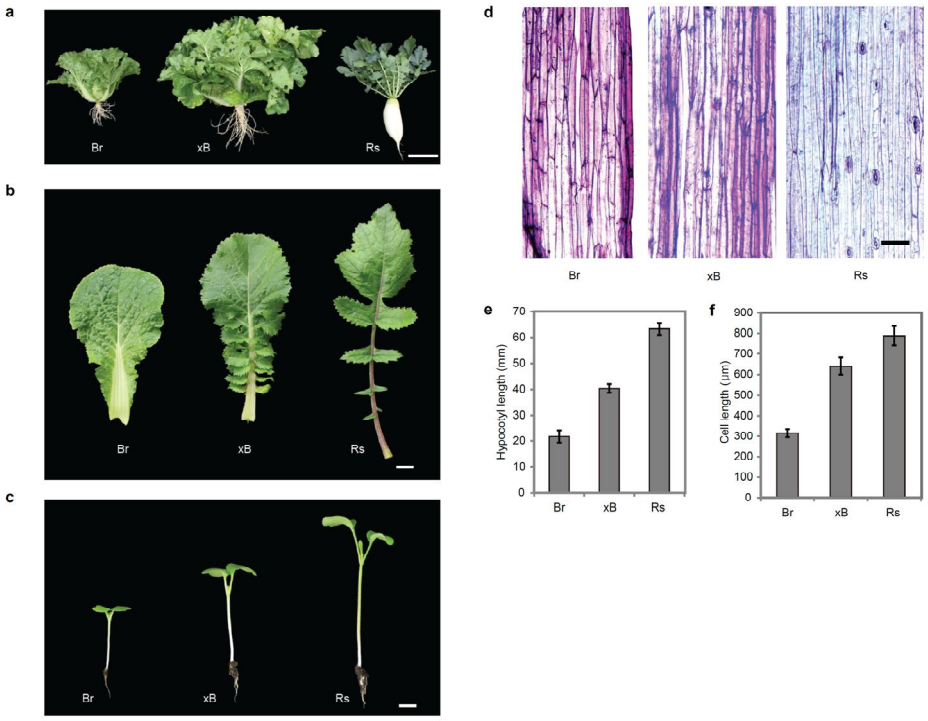


Figure 1-1. Phenotypes of x*Brassicoraphanus* between *B. rapa* and *R. sativus*.

(a) Photographs of whole plants grown in the field 40 days after planting, bar = 10 cm. (b) Mid-sized leaves from 100-day-old plants grown in a growth room, bar = 1 cm. (c) The 14-day-old seedlings, bar = 1 cm. (d) Microscopic images of hypocotyl epidermis cells. The center part of each hypocotyl was peeled and stained with toluidine blue-O, bar = 100 μm. (e) Average hypocotyl length from 10 seedlings with error bars indicating standard deviation. (f) Hypocotyl epidermis cell length was measured using microscopic images of three seedlings each.

Genomic feature of x*Brassicoraphanus*

x*B* is a fertile and genetically stable intergeneric allotetraploid synthesized from a cross between *Br* and *Rs*. x*B* genome was *de novo* assembled using 195.0 Gb of Illumina shotgun reads. A total of 68,454 contigs and 20,294 scaffolds were assembled with 28.5 Kb and 4.5 Mb of N50, respectively. The assembled 692.9 Mb sequence covered 69~71% of the x*B* genome (Table 1-6), the size of which was estimated to be 998.3 Mb with flow cytometry analysis using *Br* genome as a control (Table 1-7), close to the sum of *Br* (485 Mb) (Wang et al., 2011) and *Rs* (510 Mb) genomes (Jeong et al., 2016). The genome size of xB genome was underestimated using k-mer analysis (Figure 1-2). x*B* genome contains 87,861 annotated genes with 39.91% repeat regions, among which *Copia* and *Gypsy* classes of long terminal repeat (LTR) were predominant (3.15% of *Copia* and 8.56% of *Gypsy* in x*B* genome) (Table 1-8). The assembled chloroplast genome of x*B* (153,482 bp) has 113 genes including 79 protein genes, 4 ribosomal RNA genes, and 30 tRNA genes (Figure 1-3a and Table 1-9). The chloroplast x*B* genome was 99.9% identical to that of *Br*, indicating its maternal inheritance (Figure 1-3b). From whole x*B* genome (692.9 Mb), 333.5 Mb and 343.5 Mb of x*B* scaffolds were assigned to the *Br* and *Rs* reference genomes (Wang et al., 2011; Jeong et al., 2016)

(referred to as A_{Br} and R_{Rs} hereafter), respectively (Table 1-6), comprising two subgenomes of x*B* (referred to as A_{xB} and R_{xB} hereafter). In addition, the majority of A_{xB} and R_{xB} scaffolds were assigned to A_{Br} and R_{Rs} pseudo-chromosomes, respectively (279.8 Mb for A_{Br} and 301.8 Mb for R_{Rs}) (Table 1-6 and Figure 1-4). As shown in Figure 1-5, a complete set of chromosomes originated from A_{Br} and R_{Rs} was identified in x*B* genome. Genes are dispersed throughout every chromosome while repeat elements are distributed in the gene-poor regions (Figure 1-5). Differentially expressed genes (DEGs) whose expressions are up- or down-regulated relative to the progenitors are evenly distributed (Figure 1-5). DNA methylation is predominant in repeat-enriched regions (Figure 1-5), where DNA methylation is abundant at all CG, CHG and CHH (H = A, T or C) contexts. Differentially methylated regions (DMRs) refer to the regions where DNA methylation levels in x*B* are significantly different (absolute difference > 0.3 for CG, > 0.15 for CHG and > 0.1 for CHH) from those of *Br* and *Rs*. Approximately 75.2% of H3K9me2 repressive histone marks are enriched in repeat regions, whereas the rest are dispersed in both genic and intergenic regions (Figure 1-5). Small RNAs (18-30 nt) are distributed throughout the entire x*B* genome, and many of them are significantly associated with DNA methylation (Figure 1-5). Cytological observation

revealed a total of 19 chromosome pairs present in x*B* without apparent chromosomal rearrangements such as translocations and deletions (Figure 1-6). Previous studies reported that many synthetic allopolyploid plants such as rapeseed (Xiong et al., 2011), tobacco (Chen et al., 2018) and wheat (Zhang et al., 2013) go through massive chromosome reconstruction leading to transgressive gain or loss of chromosomes and/or aneuploidy over generations. However, my data indicate that both A_{Br} and R_{Rs} genomes can merge and reside in the same nucleus of x*B* without aberrations in chromosome configuration.

Table 1-6. Summary of the x*Brassicoraphanus* genome assembly statistics

Assembly information	Contig	Scaffold
Total length / Number	652.44 Mb / 68,454 ea	692.83 Mb / 20,299 ea
Average / Median	9.53 kb / 2.40 kb	34.13 kb / 901 bp
Max / Min length	190.62 kb / 200 bp	16.46 Mb / 213 bp
N50	28,581 bp (6,854th)	4,479,746 bp (49th)
N90	5,982 bp (24,969th)	166,698 bp (284th)
GC contents	35.75%	33.68%

Scaffold assignment	Total number	Assigned to A genome	Assigned to R genome	Unassigned
No. of scaffolds	20,299	7,790	7,364	5,145
Cumulative size (bp)	692,831,961	333,554,805	343,544,771	13,732,385
(% of total assembly)	(100%)	(48.43%)	(49.59%)	(1.98%)
No. of chromosomes assigned	213	129	84	
Size of chromosome assigned (bp)	581,691,615	279,795,674	301,895,941	
(% of total assembly)	(83.96%)	(83.38%)	(87.87%)	

Species	Protein-coding loci	Total CDS length (bp)	Average CDS length (bp)
x*Brassicoraphanus*	87,861	106.896.611	1,216
B. rapa	42,601	49,456,892	1,172
R. sativus	52,326	67,790,376	1,295

Table 1-7. Genome size estimation by flow cytometry for x*Brassicoraphanus*

		Peak value	CV (%)[a]	Genome size (Mb)
1st	x*B*/*Br*	197.90/98.29	7.07/8.20	976.5
2nd	x*B*/*Br*	190.69/92.29	4.38/5.01	1,002.1
Average				998.3

[a] Coefficient of variation

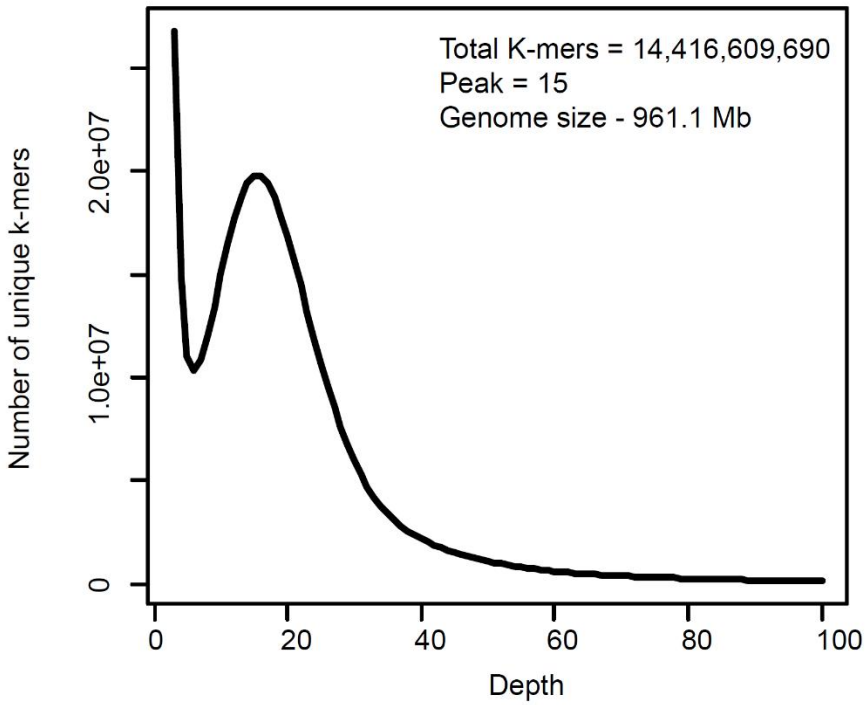


Figure 1-2. Genome size estimation using K-mer analysis.

The peak position was 15. Genome size was estimated as 961.1 Mb (K-mer count = 14,416,609,690 and Peak = 15)

Table 1-8. Annotation of repeat sequences in x*Brassicoraphanus* genome.

	x*Brassicoraphanus*		A_{xB} subgenome		R_{xB} subgenome	
Without N gaps	652.8 Mb		315.1 Mb		325.8 Mb	
Assembly size	692.8 Mb		333.0 Mb		342.5 Mb	
Type of TE	Length (bp)	% [a]	Length (bp)	% [a]	Length (bp)	% [a]
DNA elements	47,156,933	7.22	21,490,747	6.82	25,131,125	7.71
LINE elements	18,601,941	2.85	8,596,780	2.73	9,791,504	3.00
SINE elements	3,064,487	0.47	1,720,357	0.55	1,319,758	0.40
LTR/Others	1,286,854	0.20	542,628	0.17	713,288	0.22
LTR/*Gypsy*	55,864,303	8.56	27,947,271	8.87	26,337,315	8.08
LTR/*Copia*	20,590,492	3.15	7,525,233	2.39	12,587,084	3.86
rRNAs	1,048,741	0.16	828,446	0.26	40,448	0.01
Simple repeats	8,337,003	1.28	4,355,817	1.38	3,827,299	1.17
Others	9,722,660	1.49	3,041,544	0.97	5,896,261	1.81
Unknown	90,173,914	13.81	47,944,426	15.21	40,284,717	12.36
Total	255,847,328	39.19	123,993,819	39.34	125,928,799	38.64

[a] Repeat element / genome size (without N gaps)

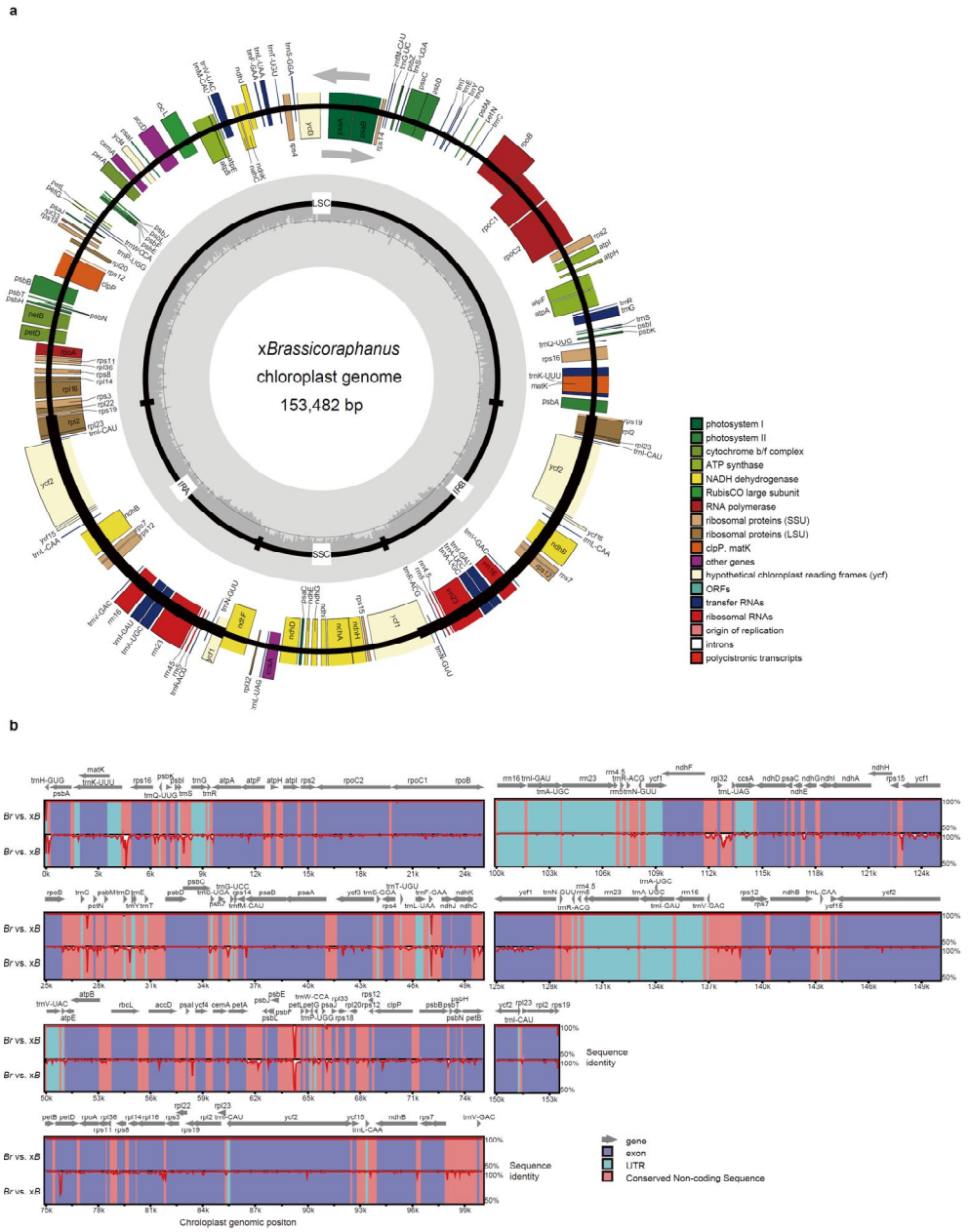

a
xBrassicoraphanus
chloroplast genome
153,482 bp
LSC
IRA
IRB
SSC
photosystem I
photosystem II
cytochrome b/f complex
ATP synthase
NADH dehydrogenase
RubisCO large subunit
RNA polymerase
ribosomal proteins (SSU)
ribosomal proteins (LSU)
clpP, matK
other genes
hypothetical chloroplast reading frames (ycf)
ORFs
transfer RNAs
ribosomal RNAs
origin of replication
introns
polycistronic transcripts
b
Br vs. xB
Sequence identity
gene
exon
UTR
Conserved Non-coding Sequence
Chroloplast genomic positon

Figure 1-3. Chloroplast genome of x*Brassicoraphanus*.

(a) The chloroplast genome of x*B* contains 113 genes including 79 protein-coding genes, 4 ribosomal RNA genes, and 30 transfer RNA genes. Genes were colored based on their functional groups. A pair of inverted repeats (IRA and IRB), small single-copy (SSC) and large single-copy (LSC) are represented in the inner circle. The grey bars in the circle indicate the GC contents. (b) Sequence alignment of the *Br*, *Rs* and x*B* chloroplast genomes. Grey arrows indicate the position and direction of each gene. Exon and untranslated region (UTR) and conserved non-coding sequences are represented in blue, red and green, respectively. Red lines show sequence identity between *Br* and x*B* (upper), and *Rs* and x*B* (lower).

Table 1-9. Chloroplast genome annotations of x*Brassicoraphanus* and its parental genomes

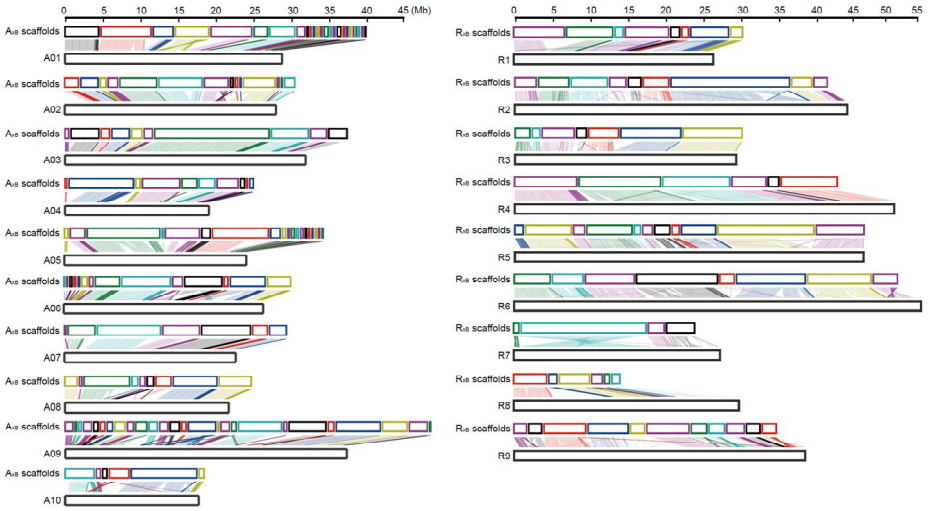
	Structural annotation				Number of genes			
	Size (bp)	LSC[a] (bp)	SSC[b] (bp)	IR[c] (bp)	Coding gene	rRNA	tRNA	Total Gene
x*B*	153,482	83,281	17,775	26,213	79	4	30	113
B. rapa	153,482	83,281	17,775	26,213	79	4	30	113
R. sativus	153,368	83,171	17,765	26,215	79	4	30	113

[a] Large single copy regions

[b] Small single copy regions

[c] Inverted regions

a


0 5 10 15 20 25 30 35 40 45 (Mb)
0 5 10 15 20 25 30 35 40 45 50 55
Axb scaffolds
A01
A02
A03
A04
A05
A06
A07
A08
A09
A10
Rxb scaffolds
R1
R2
R3
R4
R5
R6
R7
R8
R9

b

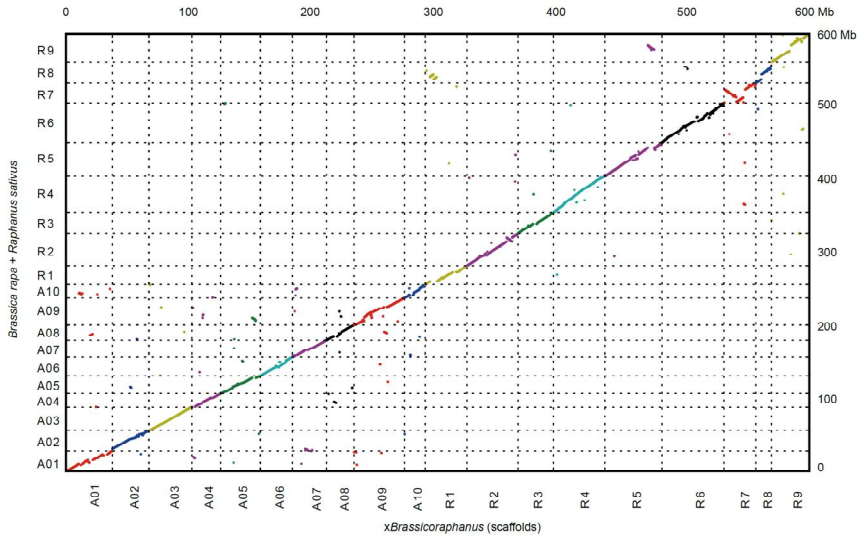

0 100 200 300 400 500 600 Mb
600 Mb
500
400
300
200
100
0
R9
R8
R7
R6
R5
R4
R3
R2
R1
A10
A09
A08
A07
A06
A05
A04
A03
A02
A01
Brassica rapa + Raphanus sativus
A01 A02 A03 A04 A05 A06 A07 A08 A09 A10 R1 R2 R3 R4 R5 R6 R7 R8 R9
xBrassicoraphanus (scaffolds)

Figure 1-4. Comparison of x*Brassicoraphanus* genome with its parental genomes.

(a) Assignment of x*B* scaffolds to its parental chromosomes. Gene pairs between x*B* and its parents are linked with colored lines. (b) Dot plot of x*B* (horizontal) and its parental genomes (vertical). The positions of gene pairs are plotted with colored dots. Horizontal and vertical axes indicate the length of chromosomes and scaffolds, respectively.

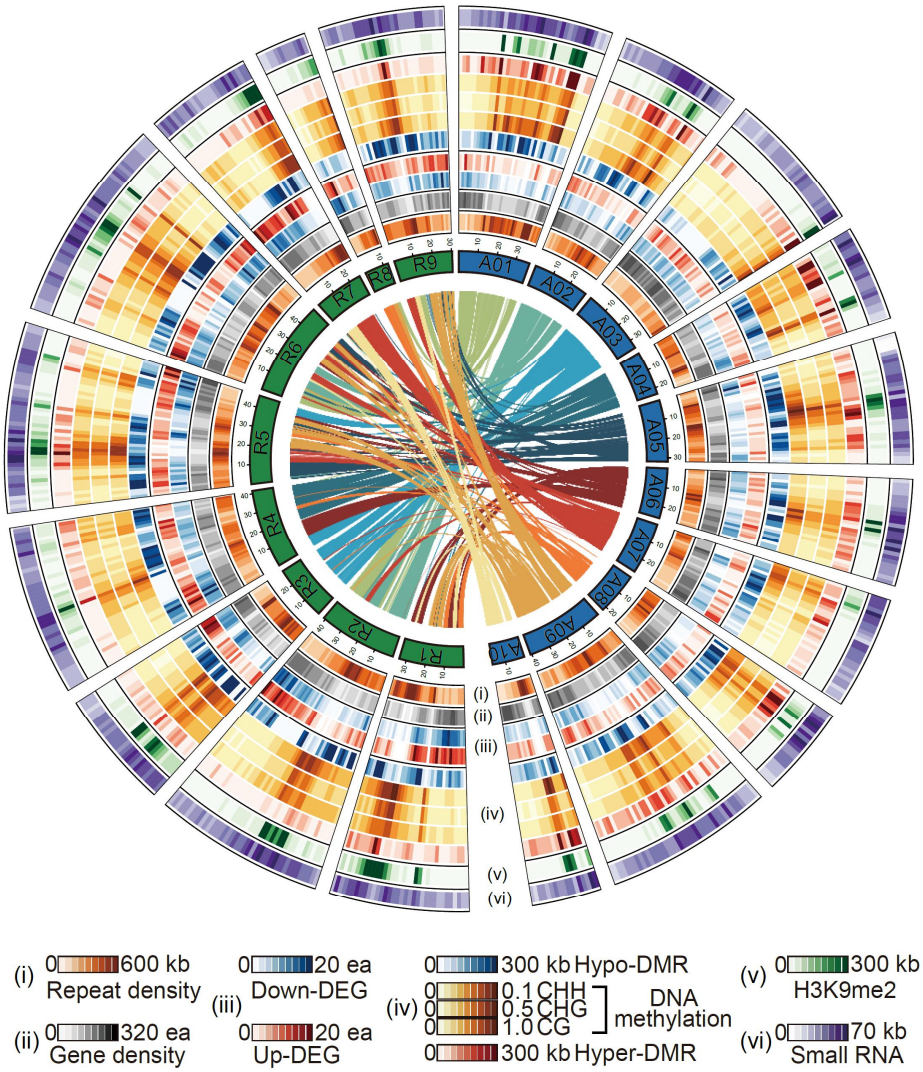


Figure 1-5. The genome of the x*Brassicoraphanus*.

The x*B* genome comprises 10 A_{xB} and 9 R_{xB} chromosomes. The data tracks represent (i) repeat density; (ii) gene density; (iii) DEGs between x*B* and its progenitor seedlings; (iv) CG, CHG, and CHH methylation levels and DMRs; (v) H3K9me2 repressive histone mark; and (vi) small RNAs. The

pseudo-chromosomes were divided into 1Mb windows. Lines in the inner circle represent syntenic relationships between A_{xB} and R_{xB}.

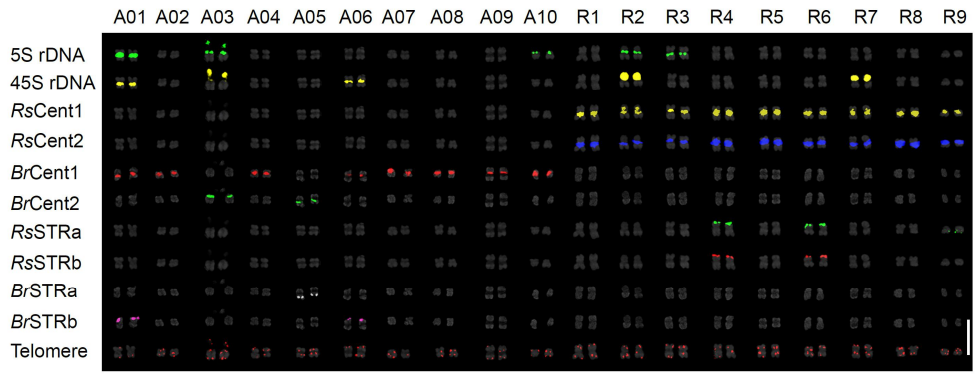


Figure 1-6. Conservation of parental chromosome in x*Brassicorphanus*. Multicolor Fluorescence *in situ* Hybridization (FISH) karyograms of x*B* with specific FISH probes for 5S rDNA, 45S rDNA, centromeric tandem repeats (Cent), short tandem repeats (STR) and telomere repeats.

Maintenance of TEs composition in x*Brassicoraphanus*

TEs play significant roles of genome reorganization and expansion, providing new genetic source for functional changes (Contreras et al., 2015; Kim et al., 2017) but reactivation of silenced TEs could cause genome instability in the hybrids (McClintock, 1984; Chenais et al., 2012). The portions of TEs sequences in WGS reads of *Br*, *Rs* and x*B* were investigated to estimate the abundance of TEs in x*B* and the progenitor genomes. Repeat sequences of *Copia* and *Gypsy* element were estimated as 6.97% and 10.35% in WGS reads of x*B*, respectively, showing more amount of TEs were detected than that of *de novo* annotation from assembled genomes (3.15% of *Copia* and 8.56% of *Gypsy* in x*B* genome) (Figure 1-7). Proportion of *Copia* and *Gypsy* sequences were observed as 6.02% and 9.39% in *Br*, and 9.80% and 8.64% in *Rs*, respectively. Interestingly, average proportions of *Copia* and *Gypsy* sequences in *Br* and *Rs* WGS reads were similar to those of x*B* (Figure 1-7), indicating that TE proportion was not dramatically changed after allopolyploidization. Recent studies showed that transcriptional activation or increase of mobility of TEs were observed in various interspecific crosses (Kashkush et al., 2003; Ungerer et al., 2009; Piednoel et al., 2015), but this results suggest that suppression of TEs mobility might contribute to acquiring stability of intergeneric hybrid genome.

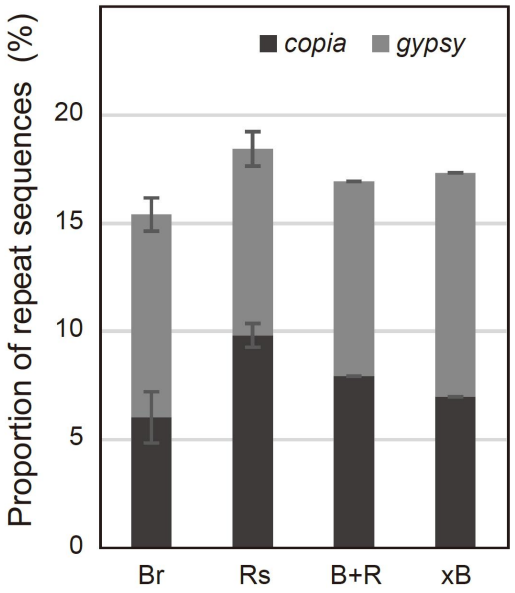


Figure 1-7. Comparison of repeat sequences in x*Brassicoraphanus* genome with its parental genomes.

Proportions of repeat sequences were estimated using WGS reads of *Br*, *Rs*, in silico *Br*+*Rs*, and x*B*. *Copia* and *Gypsy* were colored by dark grey and light grey, respectively. Repeat abundance of in silico *Br*+*Rs* is similar to that of x*B*. Error bars represent standard deviation.

Epigenetic changes in x*Brassicoraphanus*

Various neo-synthetic hybrids often go through epigenetic alterations such as DNA methylation, histone modification and small RNAs (Greaves et al., 2015), and these factors have important roles in the transcriptional regulation of plant development and genome stability of allopolyploidization (Gibney and Nolan, 2010; Song and Chen, 2015). To evaluate the effects of epigenetic changes on intergeneric hybridization, BS-seq, ChIP-seq and small RNA-seq were performed in *Br*, *Rs*, and x*B*.

In methylome analyses, DNA methylation is predominant in repeat-enriched regions (Figure 1-8a). A total of 1.4% and 1.2% of cytosines in x*B* were identified as hypermethylated and hypomethylated cytosines compared to progenitors, respectively (Figure 1-8b), and 15.9 Mb of x*B* genome (2.2% of the genome size) were defined as DMRs including 8.3 Mb hypo-DMRs and 7.6 Mb hyper-DMRs (Figure 1-8c). A Similar amount of hypo-DMRs were observed in A_{xB} and R_{xB} subgenomes, but Hyper-DMRs were preferentially induced in A_{xB} subgenome (Figure 1-8c). Most hyper- and hypo-DMRs were found to include CHG methylation changes (Figure 1-8d), and approximately 60.16% of hyper-DMRs are enriched in repeat regions, whereas hypo-DMRs are mainly observed in intergenic regions (Figure 1-8e).

From ChIP-seq analyses, a total of 58.2 Mb regions in x*B* was identified as H3K9me2 marks located regions, and a high level of DNA methylation was detected in the H3K9me2 positive regions (Figure 1-9a). The x*B* genome appeared to inherit approximately 93.4% of H3K9me2 marks (25.9 out of 27.6 Mb in *Br* and 25.5 out of 27.3 Mb in *Rs*) from the progenitors, and the A_{xB} and R_{xB} gained 2.5 Mb and 4.2 Mb of H3K9me2 regions after allopolyploidization, respectively (Figure 1-9b). Approximately 75.8% of the H3K9me2 regions in x*B* were located in repeat regions, but H3K9me loss regions were more observed in genic regions than repeat regions (Figure 1-9d).

For analyses of small RNA-seq data, small RNAs reads were classified into ribosomal RNA, small nucleolar RNA, small nuclear RNA, signal recognition particle RNA, transfer RNA and microRNA (miRNA). The other small RNA reads mapped on the genome were predicted as small interfering RNA (siRNA). The proportion of siRNAs in small RNA-seq reads was higher in x*B* (32.5%) than either of the progenitors (26.1% in *Br* and 22.8% in *Rs*), indicating the up-regulation of siRNA expression levels in x*B* (Figure 1-10a). Within a range of 20-25nt small RNAs, miRNAs were predominantly enriched in 21-nt RNAs in *Br*, *Rs* and x*B*. However, profiles of siRNAs were dissimilar between progenitor species. In *Br*, the 24-nt

RNAs were the most highly enriched (51.05% of 24-nt RNAs and 2.91% of 21-nt RNAs), whereas similar frequencies between 21- and 24-nt RNAs were observed in *Rs* (30.88% of 24-nt RNAs and 24.29% of 21-nt RNAs) (Figure 1-10b). In x*B*, frequencies of 21- and 24-nt siRNAs were similar to the averages between those of *Br* and *Rs* (40.41% of 24-nt RNAs and 9.54% of 21-nt RNAs) (Figure 1-10b). These results indicated that small RNAs were inherited in an additive manner after allopolyploidization in x*B*.

For analyses of miRNAs, 207 miRNAs were annotated, and 2,631 of miRNA target genes were predicted in x*B*. Expression changes of miRNAs were identified with statistical test, and 64 and 82 of differentially expressed miRNAs were found in A_{xB} and R_{xB}, respectively. Expression patterns of miRNAs in x*B* were classified according to previously defined criteria (Rapp et al., 2009; Grover et al., 2012), and about one-third of miRNA (88 miRNAs, 32.5%) showed additive expression pattern with no parental bias and few miRNAs (10 miRNAs, 4.8%) displayed patterns of transgressive up and down (Figure 1-11a). A total of 61 and 53 genes were predicted to be targeted by the transgressive up- and down-regulated miRNAs, respectively. In A_{xB} subgenome, expression levels of genes targeted by transgressive up- and down-regulated miRNAs were not changed, but slightly decreased and increased in R_{xB} subgenome,

respectively (Figure 1-11b). These results suggest that miRNA expression was also additively inherited, and it might play little roles for gene expression changes in x*B*.

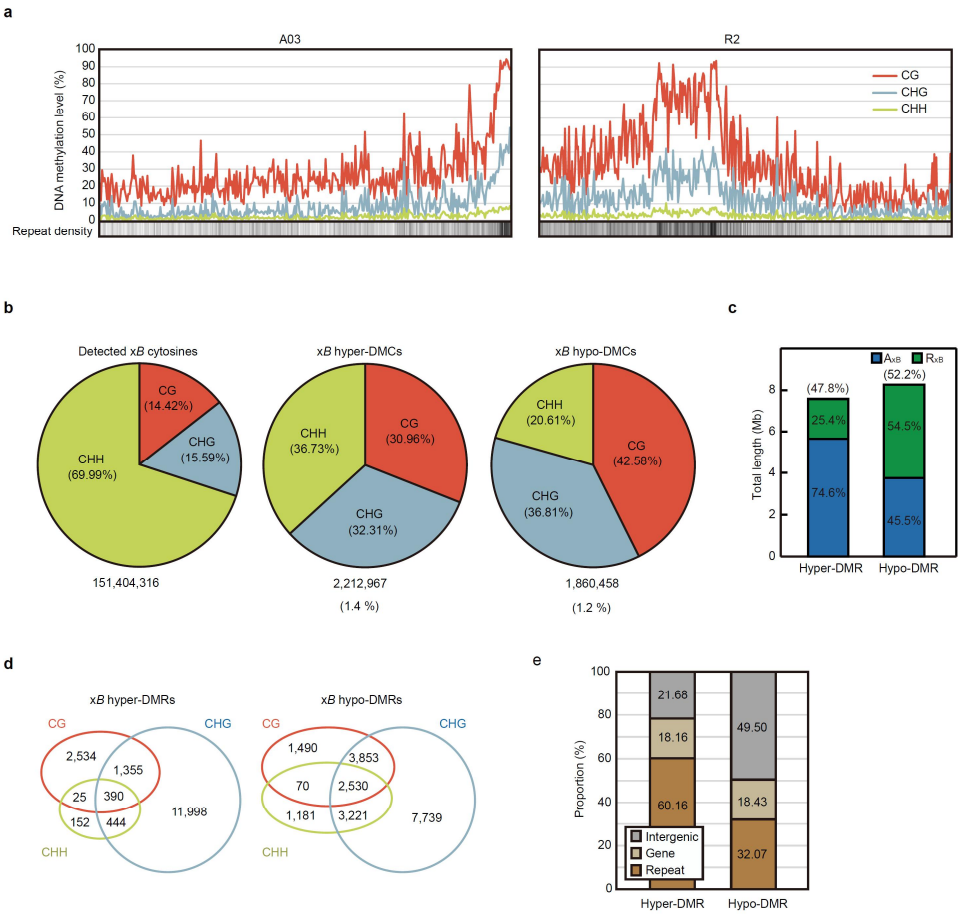


Figure 1-8. Genome-wide DNA methylation in x*Brassicoraphanus*.

(a) Examples of DNA methylation distribution in A_{xB} and R_{xB} pseudo-chromosomes. Red, blue and green lines represent CG, CHG and CHH methylation levels in a bin (100 kb), respectively. Repeat density was displayed in a heatmap. (b) Fractions of differentially methylated cytosines (DMCs) between x*B* and its parental genomes. The numbers and proportions of hyper- and hypo-DMCs are presented at the bottom of the chart. (c) Total lengths of hyper- and hypo-DMRs in x*B*. The ratios of

hyper- and hypo-DMRs are represented in parentheses. (d) Overlaps of the x*B* hyper- and hypo-DMRs in each sequence context. (e) Proportions of intergenic, genic and repeat regions in hyper- and hypo-DMRs.

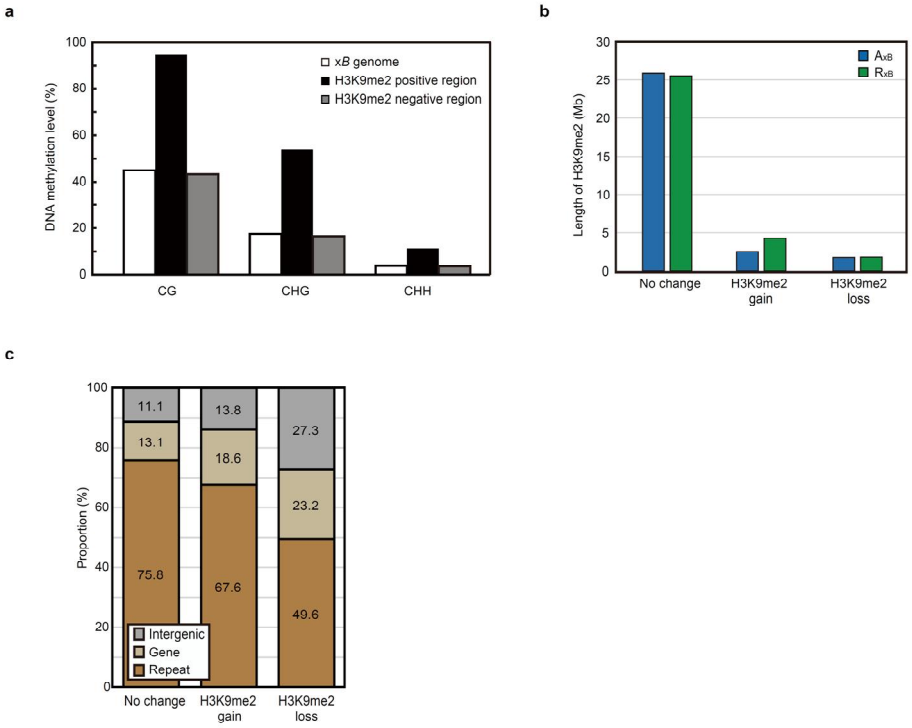


Figure 1-9. H3K9me2 modification of x*Brassicoraphanus*.

(a) Average DNA methylation levels of genome (white), H3K9me2 positive regions (black) and H3K9me2 negative regions (grey) in x*B*. (b) Differentially modified H3K9me2 regions between x*B* and its parents. The regions were classified into three types; shared regions (no change), x*B*-unique regions (H3K9me2 gain), and *Br*- or *Rs*-unique regions (H3K9me2 loss). Blue and green bars indicate A_{xB} and R_{xB} subgenomes, respectively. (c) DNA methylation levels of H3K9me2 gain and loss regions in x*B*. Black and grey bars represent average DNA methylation levels of x*B* and its parents, respectively. (d) Proportions of intergenic, genic and TE regions in the three types of H3K9me2 regions in x*B*.

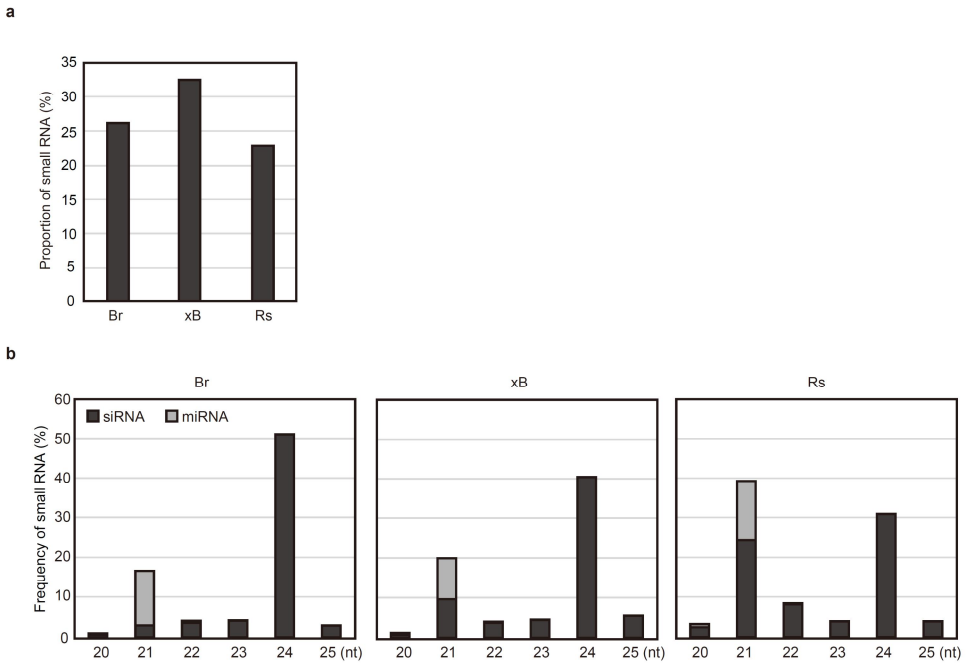


Figure 1-10. Small RNA analysis of x*Brassicoraphanus*.

(a) Proportion of siRNAs in small RNA libraries of x*B* (32.9%), *Br* (26.9%), and *Rs* (22.0%). (b) Size distribution of small RNAs in *Br*, x*B*, and *Rs*. Grey and black bars indicate proportions of miRNAs and other small RNAs, respectively. 21-nt and 24-nt siRNAs were abundantly observed in all species with distinct distribution (51.05% in *Br*, 40.41% in x*B*, and 30.88% in *Rs* for 24-nt siRNA; 2.91% in *Br*, 9.54% in x*B*, and 24.29% in *Rs* for 21-nt siRNA; 13.69% in *Br*, 10.44% in x*B*, and 14.95% in *Rs* for 21-nt miRNA).

a

Additivity		A expression level dominance		R expression level dominance		Transgressive up-regulation			Transgressive down-regulation			No change	Total
I	XII	IX	VI	II	XI	VIII	VI	V	VII	III	X		
47 (22.7%)	41 (19.8%)	9 (4.3%)	11 (5.3%)	4 (1.9%)	5 (2.4%)	1 (0.5%)	5 (2.4%)	0 (0.0%)	3 (1.4%)	1 (0.5%)	0 (0.0%)	80 (38.6%)	207

b

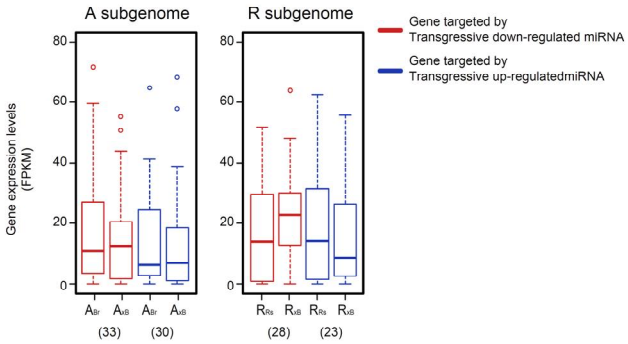


Figure 1-11. miRNA analysis of x*Brassicoraphanus*.

(a) Expression patterns of miRNA in x*B*. ♀, *B. rapa*; ♂, *R. sativus*; P, x*Brassicoraphanus*. (c) Gene expression levels targeted by transgressive up- and down-regulated miRNA. The number of genes targeted by transgressive regulated miRNA were represented in parentheses.

Correlation between epigenetic factors in x*Brassicoraphanus*

During allopolyploidization of x*B*, DNA methylation levels were decreased in H3K9me2 loss regions and increased in H3K9me2 gain regions in all sequence contexts (Figure 1-12a). Most hyper-DMRs overlapped H3K9me2-gain regions, whereas hypo-DMRs more overlapped H3K9me2-deficient regions (Figure 1-12b), indicating a positive correlation between DNA methylation and repressive H3K9me2 marks. In addition, higher levels of 24-nt RNAs were associated with hyper-DMRs in x*B* relative to the parents, whereas lower 24-nt RNA levels with hypo-DMRs, showing a positive correlation between expression level of 24-nt RNAs and DNA methylations (Figure 1-12c). In x*B*, Most of DMRs and histone gain or loss regions were mainly located in repeat regions rather than genic regions (Figures 1-8e and 1-9c), and only small number of genes was involved with DNA methylation or histone modifications, albeit negative correlations between expression and epigenetic changes (Figure 1-12d). These data indicate that epigenetic alterations are associated with each other, but massive gene expression change was not mainly associated with epigenetic alterations in x*B*.

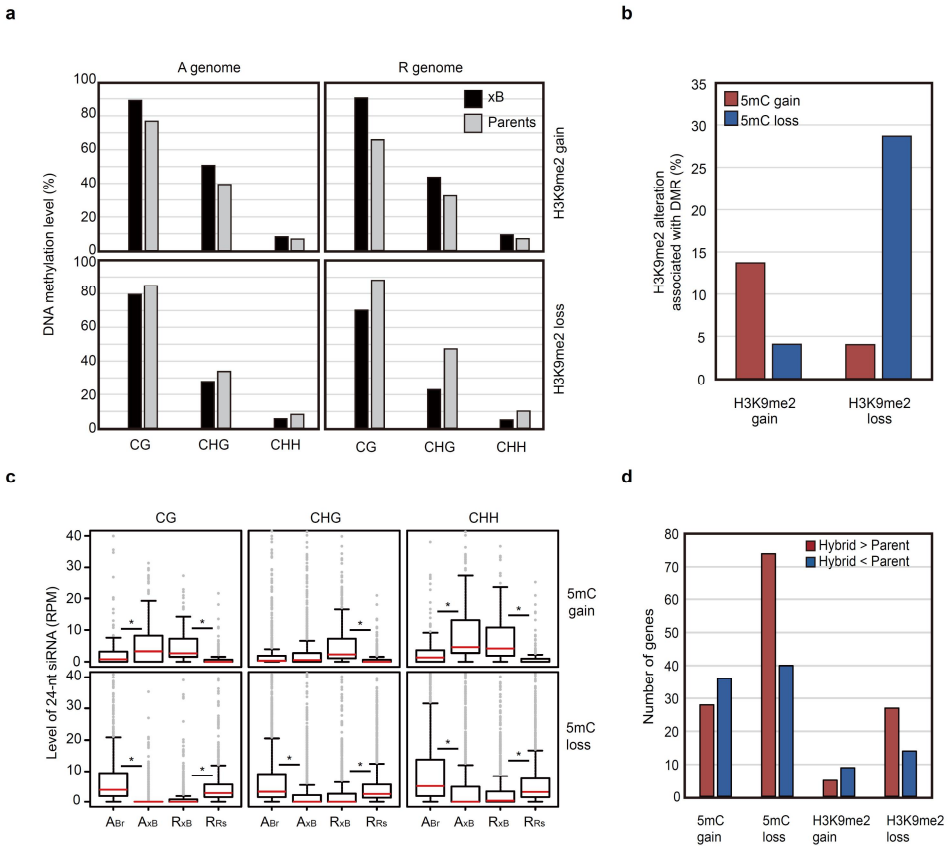


Figure 1-12. Association analysis of epigenetic changes.

(a) DNA methylation levels of H3K9me2 gain and loss regions in x*B*. Black and grey bars represent average DNA methylation levels of x*B* and its parents, respectively. (b) Relationships between DMR and H3K9me2 alteration. Red and blue bars indicate the proportion of H3K9me2 alteration regions overlapping by at least 1 bp with 5mC gain and loss regions. (c) Expression levels of 24-nt RNAs at CG, CHG and CHH DMRs in x*B* subgenome (A_{xB} and R_{xB}) and its parental genomes (A_{Br} and R_{Rs}). The Expression level of 24-nt RNAs was calculated as reads per million (RPM) (two-tailed Student's *t*-test, $^{*}P < 5.0e\text{-}5$). (d) Number of differentially

expressed genes related to DNA methylation and histone modification changes. Up- and down-regulated genes in the hybrid are colored in red and blue, respectively.

TE-specific hypermethylation in x*Brassicoraphanus*

The TE mobility can trigger genome instability by gene disruption and chromosome rearrangement, as first suggested by Barbara McClintock (McClintock, 1984; Chenais et al., 2012). The establishment of repressive chromatin modifications is required for the silencing of transcriptional activity of TEs. In plant, DNA methylation in all sequence contexts is abundant in TEs, and RNA-directed DNA methylation (RdDM) functions the initiation, maintenance and reinforcement of TE silencing (Law and Jacobsen, 2010; Kim and Zilberman, 2014). In x*B* and its progenitors, overall profiles of the methylation levels were observed in gene body and TE regions for each sequence context. Patterns of CG, CHG, and CHH methylation in x*B* were similar to those of *Arabidopsis* (Cokus et al., 2008), showing that DNA methylation levels are high in gene body, decrease at 5' and 3' shores, and increase again beyond the translation start and termination sites (Figure 1-13). Notably, the progenitor A_{Br} and R_{Rs} genomes have unique CG methylation patterns, in which A_{Br} has higher methylation levels than R_{Rs}, and this methylation asymmetry is inherited to hybrid subgenomes A_{xB} and R_{xB} (Figure 1-13). This pattern of gene body DNA methylation was represented in all sequence contexts, indicating genic regions of x*B* underwent only minor DNA methylation changes (Figure 1-

13). TEs are heavily methylated at all sequence contexts (92.6% in A_{Br} and 90.0% in R_{Rs} for CG; 35.0% in A_{Br} and 43.5% in R_{Rs} for CHG; and 7.5% in A_{Br} and 8.2% in R_{Rs} for CHH) (Figure 1-13). Interestingly, TE CHG methylation at A_{Br} is less than that of R_{Rs} but both A_{xB} and R_{xB} have similar CHG methylation levels in x*B*, showing that difference of CHG methylation levels between the parental genomes was lost (Figure 1-13). Moreover, I examined distributions of TEs on x*B* subgenomes and found that DNA transposons are widespread throughout the chromosome with little association with DMRs but LTR retrotransposons have the clusters highly associated with an increase in CHG methylation (Figure 1-14). I presumed that small RNAs produced from repetitive sequences may trigger DNA methylation via *trans*-acting mechanisms by RdDM (Law and Jacobsen, 2010; Wendel et al., 2016). In the analysis of small RNAs, about 11.8-12.9% of 24-nt RNAs from *Br* and *Rs* have a pairwise sequence identity and may share the same targets across the genomes (Figure 1-15a). Indeed, 10.4% of 24-nt RNAs from x*B* have indistinguishable origins with such potentials (Figure 1-15b). This strongly suggests that in the x*B* hybrid genome R_{xB}-originated siRNAs induce gain of CHG methylation at TEs on A_{xB} via RdDM in *trans*. I also found that transcription levels of TE elements which have higher CHG methylation levels in A_{xB} than A_{Br} were lower in x*B* than

the average value of parental expressions (Figure 1-16a-f). In addition, highly abundant *Copia* and *Gypsy* elements are moderately expressed in *Br* but silenced in x*B* seedlings (Figure 1-16f). Taken together, these findings suggest the possibility that RdDM-mediated DNA methylation induces TE silencing, which in turn stabilizes the x*B* hybrid genome.

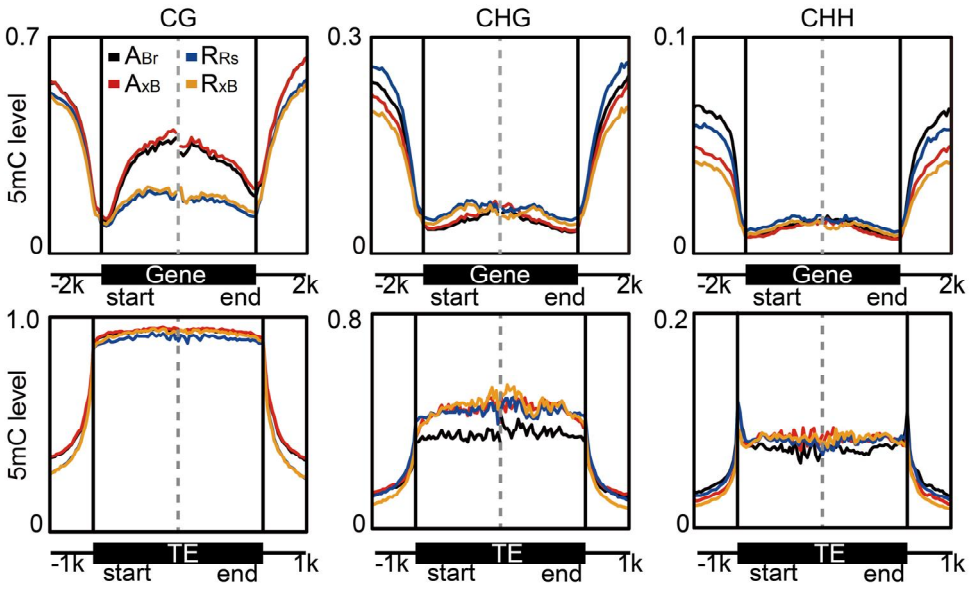


Figure 1-13. Meta plot of DNA methylation in x*Brassicoraphanus*.

Distribution of DNA methylation at gene body and TE regions in x*B* subgenomes (A_{xB} and R_{xB}) and its parental genomes (A_{Br} and R_{Rs}).

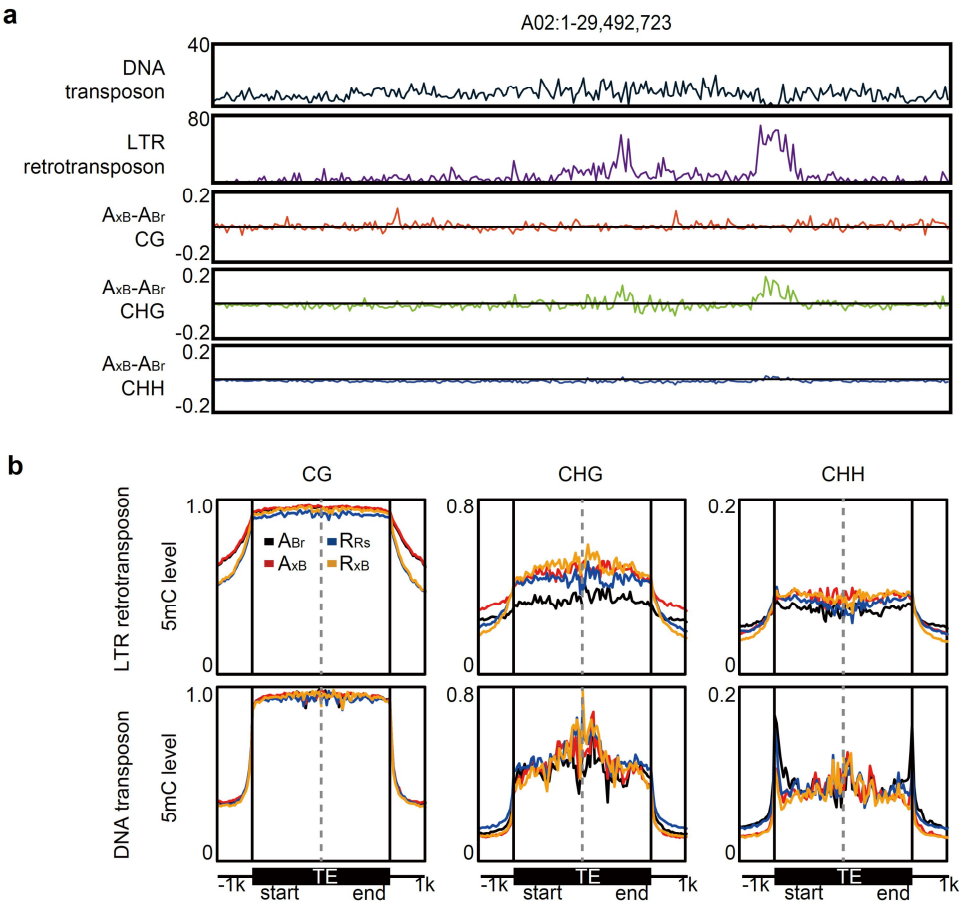


Figure 1-14. Hypermethylation at LTR retrotransposon in x*Brassicoraphanus*.

(a) Distributions of DNA transposons, LTR retrotransposons and DNA methylation difference between A_{Br} and A_{xB} across chromosome A02. The navy blue and purple lines are the base coverage rates of DNA transposons and LTR retrotransposons in a 100 kb bin, respectively. The differences of DNA methylation levels between A_{xB} and A_{Br} at CG, CHG and CHH contexts are colored by red, green and blue in a 100 kb bin, respectively. (b) Distribution of DNA methylation at LTR retrotransposon and DNA

transposon in x*B* subgenomes (A_{xB} and R_{xB}) and its parental genomes (A_{Br} and R_{Rs}).

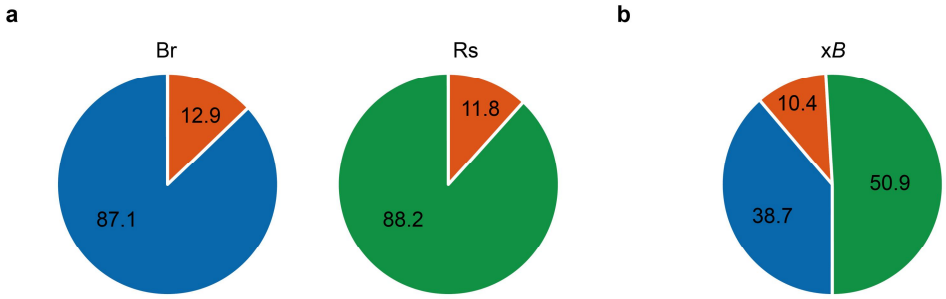


Figure 1-15. Proportion of 24-nt RNAs sharing targets across the genomes.

(a) Proportions of 24-nt RNAs from *Br* and *Rs* sharing targets across the genomes. The proportions of 24-nt RNAs mapped to both of *Br* and *Rs* genomes are represented in red, and the 24-nt RNAs uniquely mapped to *Br* and *Rs* genomes are shown in blue and green, respectively. (b) Proportion of 24-nt RNAs of x*B* sharing targets across the A_{xB} and R_{xB} subgenomes. The proportion of 24-nt RNAs mapped to both of A_{xB} and R_{xB} subgenomes is represented in red, and the 24-nt RNAs uniquely mapped to A_{xB} and R_{xB} subgenomes are shown in blue and green, respectively.

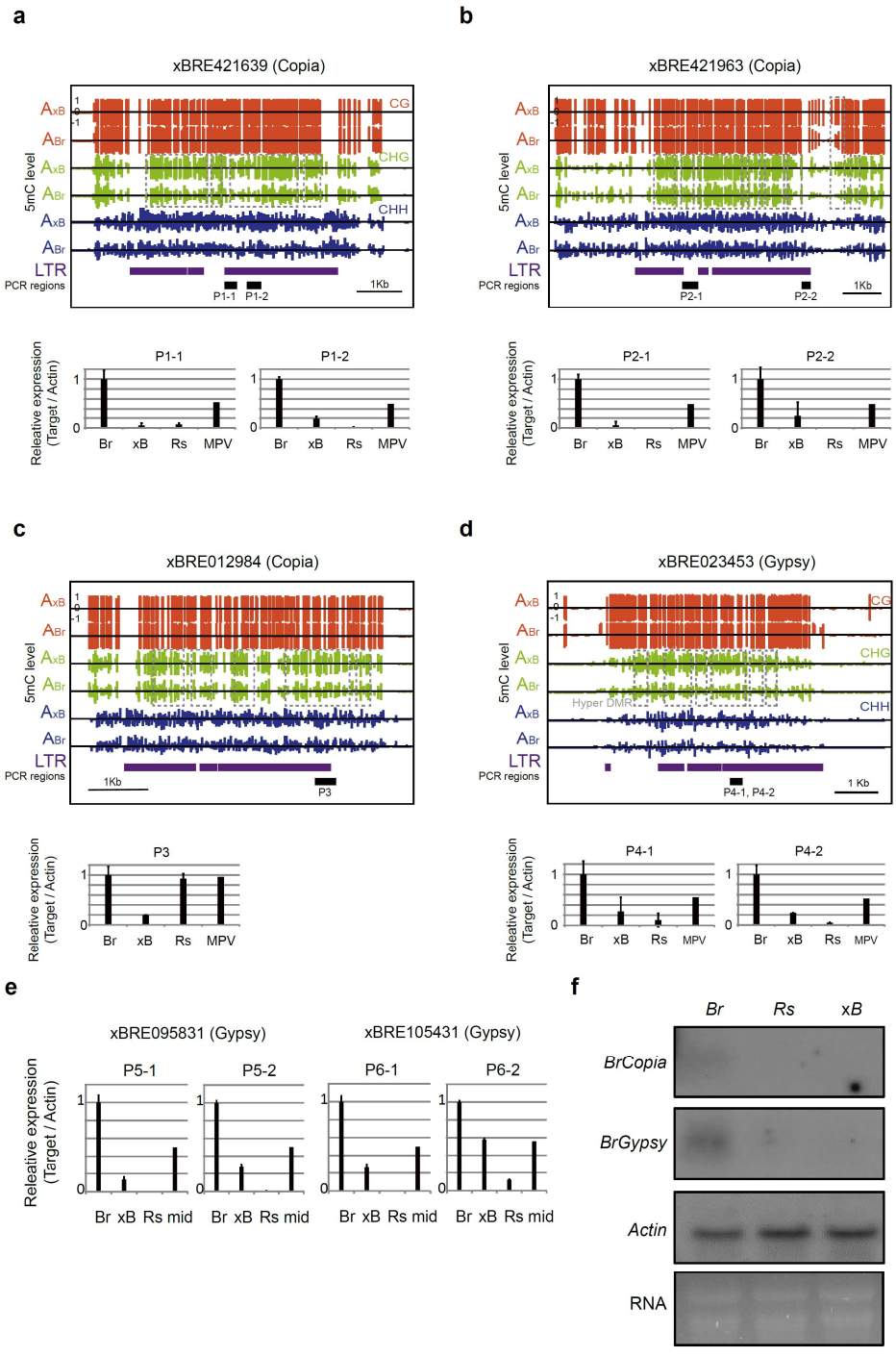

a
xBRE421639 (Copia)
AxB
ABr
CG
CHG
CHH
5mC level
LTR
PCR regions
P1-1
P1-2
1Kb
Releative expression (Target / Actin)
Br
xB
Rs
MPV
b
xBRE421963 (Copia)
P2-1
P2-2
c
xBRE012984 (Copia)
P3
d
xBRE023453 (Gypsy)
Hyper DMR
P4-1, P4-2
1 Kb
P4-1
P4-2
e
xBRE095831 (Gypsy)
xBRE105431 (Gypsy)
P5-1
P5-2
P6-1
P6-2
Br xB Rs mid
f
Br
Rs
xB
BrCopia
BrGypsy
Actin
RNA

Figure 1-16. DNA methylation distribution and expression level of hypermethylated LTR in A_{xB} and A_{Br}.

(a-d) Examples of methylation distribution at hypermethylated *Copia* and *Gypsy* class LTR in A_{xB} and A_{Br}. Genomic views at hypermethylated LTRs were represented in the upper panel. Positive and negative bars represent the methylation level of single cytosine on the Watson (+1) and Crick (-1) strands, respectively. PCR-amplified regions and hypermethylated regions were represented with black and grey boxes, respectively. Relative expressions of hypermethylated LTRs in A_{xB} and A_{Br} were displayed in the lower panel. (e) qPCR analysis of multi-copy LTRs in A_{xB} and A_{Br}. Relative expression levels are averages of two replicates with error bars representing standard deviation. (f) Northern blot for *BrCopia* and *BrGypsy*. *Actin* was used as a loading control.

DISCUSSION

Maintenance of chromosome stability in intergeneric allotetraploid

The *Brassica* genus includes various diploid and polyploid species that have economic importance for using oilseed or vegetable crops. The evolutionary relationship among *Brassica* species has been studied with three diploid species, *Br* (AA; 2n = 2x = 20), *B. nigra* (BB; 2n = 2x = 16), *B. oleracea* (CC; 2n = 2x = 18), and three allopolyploid species, *B. napus* (AACC; 2n = 4x = 38), *B. juncea* (AABB; 2n = 4x = 36) and *B. carinata* (BBCC; 2n = 4x = 34), which was described by the 'U's triangle (U, 1935). This model proposed that the genome of allotetraploid species consists of two genomes derived from each diploid, suggesting the birth of allopolyploid by natural hybridization between two species. Although natural allopolyploidization is one of the driving forces of speciation, hybridization barriers serve as a mechanism to prevent gene flow between related species (Abbott et al., 2013). In particular, the post-zygotic hybridization barrier that occurs after fertilization is often manifested as hybrid inviability or sterility (Dion-Cote and Barbash, 2017). Hybrid sterility is generally associated with a failure in meiosis. In resynthesized

allopolyploids formed between closely related species, aneuploidy and/or massive chromosome rearrangements are frequently observed and generally have abnormal chromosome behavior during meiosis inducing a detrimental effect on chromosome stability (Comai, 2005). The chromosomal instability such as homeologous recombination or aneuploidy is presumably caused by the collinearity and significant homology between the less divergent parental chromosomes. For instance, most phenotypic variations and aneuploidy in resynthesized *B. napus* lines are caused by homoeologous chromosomal rearrangements (Song et al., 1995; Gaeta et al., 2007; Xiong et al., 2011; Grandont et al., 2014). Among chromosomes of *Br* and *B. oleracea*, A1/C1 and A2/C2 chromosomes shared highly conserved syntenic regions each other (Parkin et al., 2005), and homoeologous pairings between A and C chromosome were observed frequently in A1/C1 and A2/C2 (Szadkowski et al., 2010; Szadkowski et al., 2011; Xiong et al., 2011), implying a relationship between syntenic similarity of chromosome and frequency of homoeologous recombination. In the study of hybrid genomes from the cross of three *Brassica* diploids, homoeologous pairings were more frequently observed in allohaploid *B. napus* (AC) compared to *B. juncea* (AB) and *B. carianta* (BC), showing that more divergent genome has a lower frequency of homoeologous pairing (Cui et al., 2012). In previous

cytological study of chromosome interactions among A, C and R genomes, allosyndesis trivalents A-C-C were more frequently observed than the other allosyndesis trivalents (C-C-R, A-R-R and A-A-R) in hybrid plants (ARCC) between *B. napus* (AACC) and *Raphanobrassica* (CCRR, 2N = 4x, 36) which is intergeneric allotetraploid between *Rs* and *B. oleracea*. These studies suggest that close genetic relationships between genomes in hybrid could promote homoeologous interactions (Zhan et al., 2017). My data showed *Br* and *Rs* chromosomes share little homology and collinearilty (Figure 1-17), and x*B* genome has full compliments of both *Br* and *Rs* chromosomes without genome reconfiguration (Figure 1-4 and 1-6). In addition, and it was reported that synapsis formation was not detected at meiotic prophase I of allodiploid x*B* (AR), indicating homoeologous interactions between A and R chromosomes are prevented during meiosis (Park, 2020). Taken together, these findings suggest that a merger of distantly divergent genomes may avoid such harmful homoeologous interactions while minimizing nonhomologous exchanges that often lead to aneuploidy and/or chromosome reshuffling, presumably due to parental genome divergence.

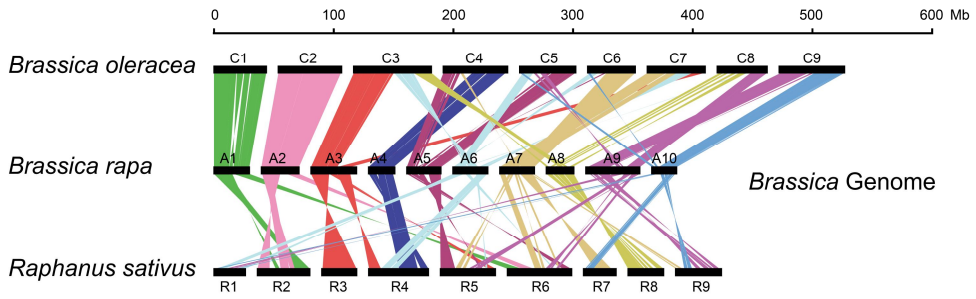


Figure 1-17. Schematic representation of collinearity regions.

Chromosomal synteny among in *Br* (AA), *B. oleracea* (CC), and *Rs* (RR). Each pair of orthologous genes was connected with a line. The genomes of *Br* and *B. oleracea* have similar chromosomal syntenic regions. Especially, highly syntenic similarities were observed in A1/C1 (green), and A2/C2 (pink) chromosomes. The genome of *Rs* was dissimilar with those of the other species.

Hypermethylation and silencing of TEs in x*Brassicoraphanus*

The term “genome shock” first suggested by Barbara McClintock, referred to as genomic changes including gene loss and large scale of chromosome rearrangement when the interspecific hybridization or allopolyploidization occurs (McClintock, 1984). Because mobility of TEs could induce genome instability by gene disruption and chromosome rearrangement (Parisod et al., 2010; Chenais et al., 2012), transcription of TEs were generally silenced through establishing repressive chromatin modifications. In plant genomes, transcription of TEs is repressed by high levels of DNA methylation within the three contexts of CG, CHG, and CHH sequences and a high level of histone mark H3K9me2 (Law and Jacobsen, 2010; Kim and Zilberman, 2014). In addition, transcriptional silencing could be established by RdDM, which functions the initiation, reinforcement and maintenance of DNA methylation in TEs (Law and Jacobsen, 2010; Kim and Zilberman, 2014; Fultz and Slotkin, 2017). In other words, DNA methylation, histone modifications, and small RNAs have an important role in genome stabilization of allopolyploidy (Gibney and Nolan, 2010; Song and Chen, 2015). Coincident with study of McClintock, several studies suggest that interspecific hybridization could increase TE insertion (Kashkush et al., 2003; Ungerer et al., 2009; Piednoel

et al., 2015), which can be resulted from activation of TE transcription through epigenetic changes such as DNA methylation or siRNAs (He et al., 2013; Shen et al., 2017; Jiao et al., 2018; Zhang et al., 2018).

In x*B*, DNA methylation changes were predominantly generated in TE regions than genic regions. Difference of DNA methylation patterns in gene body between *Br* and *Rs* was maintained after allopolyploidization, whereas pre-existed difference of CHG methylation level in TE were reduced, showing hypermethylation of retrotransposons and down-regulation of their expression levels. As a possible explanation for hypermethylation at LTRs, TE-derived siRNAs from R_{xB} subgenome may function in *trans* to increase DNA methylation on A_{xB} subgenome (Figure 1-18). Previous study reported that the activity of incoming homologous TEs could be silenced by identity-based silencing, which endogenous 24-nt siRNA produced from previously silenced TE functions in *trans* for reinforcement of DNA methylation on the homologous TE (Fultz and Slotkin, 2017). I found that 10.4 % of 24-nt siRNA in x*B* have possibilities to function *in trans*. This observation suggests that sequence similarity of TEs and crossability of siRNA between A_{xB} and R_{xB} may induce the reinforcement of DNA methylation of homologous TE. The concept of suppression TEs in *trans* during allopolyploidization were also suggested

with heterochromatic siRNAs (hc-siRNAs) which generated from repeat regions and associated with preventing invasion of transposon like "guardians of the genome" (Axtell, 2013a; Wendel et al., 2016). Although *trans*-acting activity of hc-siRNA has not been tested, hc-siRNA may function in *trans* via sequence homology, and hc-siRNAs derived from the subgenomes which has similar sequence homology might suppress TEs in *trans*. Since a burst of TEs in recent hybrid and allopolyploid plants could lead the genomic destabilization, suppression mechanisms of activated TEs are required for maintenance of genome stability after a few generations of allopolyploids. As a result, hypermethylation by *trans*-acting TE-derived siRNAs might be one of the mechanisms for silencing TEs in early generations for long term survival of allopolyploid plants.

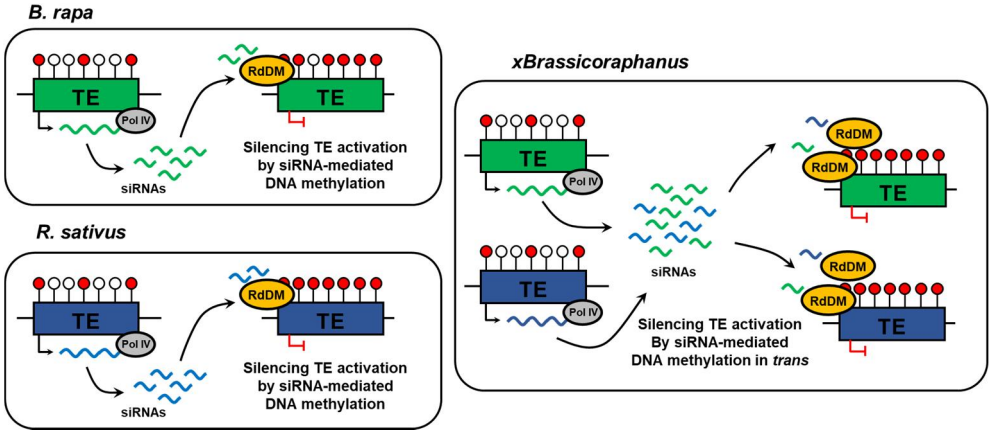


Figure 1-18. Hypermethylation possibly by siRNAs in *trans*.

siRNAs are generated from TEs by Pol IV. Transcription of TEs can be suppressed by RdDM (left panels). After allopolyploidization, siRNA derived from sequence-similar TEs of two parental subgenomes might function in *trans* and both TEs might be hypermethylated by RdDM.

Model of 'triangle of U' expanded to the intergeneric level

It is believed that the more distantly related the species, the stronger the hybridization barrier. On the contrary to this assumption, a certain extent of genome divergence could promote hybridization between distant species. This also proposes that interspecific/intergeneric hybridization may occur more frequently in nature than I have thought, and the model of 'triangle of U' (U, 1935) can be further expanded to the intergeneric level. Once established, polyploid plants undergo gradual but substantial genome reconstruction including differential deletion or retention of duplicated genes and biased genome fragmentation (Cheng et al., 2018). This eventually leads to a decrease in both chromosome number and genome size, with most of the polyploid property lost. Thus, evolution of land plants is likely to comprise the recurrent cycles of hybridization, diversification and diploidization (Wendel, 2015; Soltis et al., 2016). Furthermore, understanding the highly dynamic and flexible process of hybridization and polyploidization should provide a clue to Charles Darwin's 'abominable mystery' (Darwin, 1903) questioning the great diversification and expansion of angiosperms within a short geological time.

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., *et al.* (2013). Hybridization and speciation. J. Evol. Biol. **26,** 229-246.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics **31,** 166-169.

Axtell, M.J. (2013a). Classification and comparison of small RNAs from plants. Annu. Rev. Plant Biol. **64,** 137-159.

Axtell, M.J. (2013b). ShortStack: comprehensive annotation and quantification of small RNA genes. RNA **19,** 740-751.

Barber, W.T., Zhang, W., Win, H., Varala, K.K., Dorweiler, J.E., Hudson, M.E., *et al.* (2012). Repeat associated small RNAs vary among parents and following hybridization in maize. Proc. Natl. Acad. Sci. U.S.A. **109,** 10444-10449.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27,** 578-579.

Chalhoub, B., Denoeud, F., Liu, S.Y., Parkin, I.A.P., Tang, H.B., Wang, X.Y., *et al.* (2014). Early allopolyploid evolution in the post-neolithic *Brassica napus* oilseed genome. Science **345,** 950-953.

Chen, S., Ren, F., Zhang, L., Liu, Y., Chen, X., Li, Y., *et al.* (2018). Unstable allotetraploid tobacco genome due to frequent homeologous recombination, segmental deletion, and chromosome loss. Mol. Plant **11,** 914-927.

Chen, Z.J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu. Rev. Plant Biol. **58,** 377-406.

Chenais, B., Caruso, A., Hiard, S., and Casse, N. (2012). The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene **509,** 7-15.

Cheng, F., Wu, J., Cai, X., Liang, J.L., Freeling, M., and Wang, X.W. (2018). Gene retention, fractionation and subgenome differences in polyploid plants. Nat. Plants **4,** 258-268.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., *et al.* (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. Nature **452,** 215-219.

Comai, L. (2005). The advantages and disadvantages of being polyploid. Nat. Rev. Genet. **6,** 836-846.

Contreras, B., Vives, C., Castells, R., and Casacuberta, J.M. (2015). The impact of transposable elements in the evolution of plant genomes: from selfish elements to key players. In Evolutionary biology: Biodiversification from genotype to phenotype (Springer), pp. 93-105.

Cui, C., Ge, X., Gautam, M., Kang, L., and Li, Z. (2012). Cytoplasmic and genomic effects on meiotic pairing in *Brassica* hybrids and allotetraploids from pair crosses of three cultivated diploids. Genetics **191,** 725-738.

Dai, X., and Zhao, P.X. (2011). psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. **39,** W155-159.

Darwin, C. (1903). More letters of Charles Darwin: a record of his work in a series of hitherto unpublished letters. (D. Appleton).

Diez, C.M., Roessler, K., and Gaut, B.S. (2014). Epigenetics and plant genome evolution. Curr. Opin. Plant Biol. **18,** 1-8.

Dion-Cote, A.M., and Barbash, D.A. (2017). Beyond speciation genes: an overview of genome stability in evolution and speciation. Curr. Opin. Genet. Dev. **47,** 17-23.

Dolstra, O. (1982). Synthesis and fertility of *Brassicoraphanus* and ways of transferring *Raphanus* characters to *Brassica* (Pudoc).

Freese, N.H., Norris, D.C., and Loraine, A.E. (2016). Integrated genome browser: visual analytics platform for genomics. Bioinformatics **32,** 2089-2095.

Fultz, D., and Slotkin, R.K. (2017). Exogenous transposable elements circumvent identity-based silencing, permitting the dissection of expression-dependent silencing. Plant Cell **29,** 360-376.

Gaeta, R.T., Pires, J.C., Iniguez-Luy, F., Leon, E., and Osborn, T.C. (2007). Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell **19,** 3403-3417.

Gibney, E.R., and Nolan, C.M. (2010). Epigenetics and gene expression. Heredity **105,** 4-13.

Goubert, C., Modolo, L., Vieira, C., ValienteMoro, C., Mavingui, P., and Boulesteix, M. (2015). *De novo* assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). Genome Biol. Evol. **7,** 1192-1205.

Grandont, L., Cunado, N., Coriton, O., Huteau, V., Eber, F., Chevre, A.M., *et al.* (2014). Homoeologous chromosome sorting and

progression of meiotic recombination in *Brassica napus*: ploidy does matter! Plant Cell **26,** 1448-1463.

Greaves, I.K., Gonzalez-Bayon, R., Wang, L., Zhu, A., Liu, P.C., Groszmann, M., *et al.* (2015). Epigenetic changes in hybrids. Plant Physiol. **168,** 1197-1205.

Grover, C.E., Gallagher, J.P., Szadkowski, E.P., Yoo, M.J., Flagel, L.E., and Wendel, J.F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol. **196,** 966-971.

Ha, M., Lu, J., Tian, L., Ramachandran, V., Kasschau, K.D., Chapman, E.J., *et al.* (2009). Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. Proc. Natl. Acad. Sci. U.S.A. **106,** 17835-17840.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., *et al.* (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. **9,** R7.

He, G., Chen, B., Wang, X., Li, X., Li, J., He, H., *et al.* (2013). Conservation and divergence of transcriptomic and epigenomic variation in maize hybrids. Genome Biol. **14,** R57.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., *et al.* (2010). Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Mol. Cell **38,** 576-589.

Huang, H., Tong, Y., Zhang, Q.J., and Gao, L.Z. (2013). Genome size variation among and within *Camellia* species by using flow cytometric analysis. Plos One **8,** e64981.

Jackson, S.A. (2017). Epigenomics: dissecting hybridization and polyploidization. Genome Biol. **18,** 117.

Jeong, Y.M., Kim, N., Ahn, B.O., Oh, M., Chung, W.H., Chung, H., *et al.* (2016). Elucidating the triplicated ancestral genome structure of radish based on chromosome-level comparison with the *Brassica* genomes. Theor. Appl. Genet. **129,** 1357-1372.

Jiao, W., Yuan, J., Jiang, S., Liu, Y., Wang, L., Liu, M., *et al.* (2018). Asymmetrical changes of gene expression, small RNAs and chromatin in two resynthesized wheat allotetraploids. Plant J. **93,** 828-842.

Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., *et al.* (2014). Efficient *de novo* assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. **24,** 1384-1395.

Karpechenko, G.D. (1928). Polyploid hybrids of *Raphanus sativus* L. x*Brassica oleracea* L. Mol. Gen. Genet. **48,** 1-85.

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

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

Kim, J.S., Lim, J.Y., Shin, H., Kim, B.G., Yoo, S.D., Kim, W.T., *et al.* (2019). ROS1-dependent DNA demethylation is required for ABA-inducible *NIC3* expression. Plant Physiol. **179,** 1810-1821.

Kim, M.Y., and Zilberman, D. (2014). DNA methylation as a system of plant genomic immunity. Trends Plant Sci. **19,** 320-326.

Kim, S., Park, J., Yeom, S.I., Kim, Y.M., Seo, E., Kim, K.T., *et al.* (2017). New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biol. **18,** 210.

Kim, S., Park, M., Yeom, S.I., Kim, Y.M., Lee, J.M., Lee, H.A., *et al.* (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat. Genet. **46,** 270.

Kozomara, A., and Griffiths-Jones, S. (2013). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. **42,** D68-D73.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., *et al.* (2009). Circos: an information aesthetic for comparative genomics. Genome Res. **19,** 1639-1645.

Lafon-Placette, C., and Kohler, C. (2015). Epigenetic mechanisms of postzygotic reproductive isolation in plants. Curr. Opin. Plant Biol. **23,** 39-44.

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 Biol. **10,** R25.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. **11,** 204-220.

Lee, J., He, K., Stolc, V., Lee, H., Figueroa, P., Gao, Y., *et al.* (2007). Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. Plant Cell **19,** 731-749.

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. Theor. Appl. Genet. **122,** 885-891.

Lohse, M., Drechsel, O., Kahlau, S., and Bock, R. (2013). OrganellarGenomeDRAW-a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. **41,** W575-W581.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., *et al.* (2015). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. GigaScience **4,** s13742-13015-10069-13742.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. **17,** 10-12.

Masterson, J. (1994). Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. Science **264,** 421-424.

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

Mcnaughton, I.H. (1973). Synthesis and Sterility of *Raphanobrassica*. Euphytica **22,** 70-88.

Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., *et al.* (2014). Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. **43,** D130-D137.

Oost, E. (1984). x*Brassicoraphanus* Sageret or x*Raphanobrassica* Karpechenko? Cruciferae Newsletter. **9,** 11-12.

Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., *et al.* (2010). Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol. **186,** 37-45.

Park, H.R. (2020). Cytogenetic analysis for genome divergence and compatibility in the *Brassicaceae* family (Seoul National University).

Parkin, I.A., Gulden, S.M., Sharpe, A.G., Lukens, L., Trick, M., Osborn, T.C., *et al.* (2005). Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics **171,** 765-781.

Petit, M., Guidat, C., Daniel, J., Denis, E., Montoriol, E., Bui, Q.T., *et al.* (2010). Mobilization of retrotransposons in synthetic allotetraploid tobacco. New Phytol. **186,** 135-147.

Piednoel, M., Sousa, A., and Renner, S.S. (2015). Transposable elements in a clade of three tetraploids and a diploid relative, focusing on Gypsy amplification. Mob. DNA **6,** 5.

Rapp, R.A., Udall, J.A., and Wendel, J.F. (2009). Genomic expression dominance in allopolyploids. BMC Biol. **7,** 18.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26,** 139-140.

Ryan, D.P., and Ehninger, D. (2014). Bison: bisulfite alignment on nodes of a cluster. BMC bioinformatics **15,** 337.

Shen, Y., Sun, S., Hua, S., Shen, E., Ye, C.Y., Cai, D., Timko, M.P., Zhu, Q.H., and Fan, L. (2017). Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid *Brassica napus* uncovers key roles for small RNAs. Plant J. **91,** 874-893.

Slater, G.S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics **6,** 31.

Smit, A., Hubley, R., and Green, P. (2015). RepeatMasker Open-4.0.

Smit, A.F., and Hubley, R. (2008). RepeatModeler Open-1.0.

Soltis, D.E., Visger, C.J., Marchant, D.B., and Soltis, P.S. (2016). Polyploidy: Pitfalls and paths to a paradigm. Am. J. Bot. **103,** 1146-1166.

Soltis, P.S., and Soltis, D.E. (2009). The role of hybridization in plant speciation. Annu. Rev. Plant Biol. **60,** 561-588.

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

Song, Q., and Chen, Z.J. (2015). Epigenetic and developmental regulation in plant polyploids. Curr. Opin. Plant Biol. **24,** 101-109.

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics **24,** 637-644.

Szadkowski, E., Eber, F., Huteau, V., Lode, M., Coriton, O., Jenczewski, E., *et al.* (2011). Polyploid formation pathways have an impact on genetic rearrangements in resynthesized *Brassica napus*. New Phytol. **191,** 884-894.

Szadkowski, E., Eber, F., Huteau, V., Lode, M., Huneau, C., Belcram, H., *et al.* (2010). The first meiosis of resynthesized *Brassica napus*, a genome blender. New Phytol. **186,** 102-112.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25,** 1105-1111.

U, N. (1935). Genome analysis in *Brassica* with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. J. Jpn. Bot. **7,** 389-452.

Ungerer, M.C., Strakosh, S.C., and Stimpson, K.M. (2009). Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred

from phylogenetic data. BMC Biol. **7,** 40.

Van de Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. Nat. Rev. Genet. **18,** 411.

Waminal, N.E., and Kim, H.H. (2012). Dual-color FISH karyotype and rDNA distribution analyses on four Cucurbitaceae species. Hortic. Environ. Biote. **53,** 49-56.

Wang, X., Wang, H., Wang, J., Sun, R., Wu, J., Liu, S., *et al.* (2011). The genome of the mesopolyploid crop species *Brassica rapa*. Nat. Genet. **43,** 1035-1039.

Wendel, J.F. (2000). Genome evolution in polyploids. Plant Mol. Biol. **42,** 225-249.

Wendel, J.F. (2015). The wondrous cycles of polyploidy in plants. Am. J. Bot. **102,** 1753-1756.

Wendel, J.F., Jackson, S.A., Meyers, B.C., and Wing, R.A. (2016). Evolution of plant genome architecture. Genome Biol. **17,** 37.

Wyman, S.K., Jansen, R.K., and Boore, J.L. (2004). Automatic annotation of organellar genomes with DOGMA. Bioinformatics **20,** 3252-3255.

Xiong, Z., Gaeta, R.T., and Pires, J.C. (2011). Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proc. Natl. Acad. Sci. U.S.A. **108,** 7908-7913.

Yang, X., and Li, L. (2011). miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics **27,** 2614-2615.

Zang, C.Z., Schones, D.E., Zeng, C., Cui, K.R., Zhao, K.J., and Peng, W.Q. (2009). A clustering approach for identification of enriched

domains from histone modification ChIP-Seq data. Bioinformatics **25,** 1952-1958.

Zhan, Z.X., Nwafor, C.C., Hou, Z.K., Gong, J.F., Zhu, B., Jiang, Y.F., *et al.* (2017). Cytological and morphological analysis of hybrids between *Brassicoraphanus*, and *Brassica napus* for introgression of clubroot resistant trait into *Brassica napus* L. Plos One **12**.

Zhang, H., Bian, Y., Gou, X., Zhu, B., Xu, C., Qi, B., *et al.* (2013). Persistent whole-chromosome aneuploidy is generally associated with nascent allohexaploid wheat. Proc. Natl. Acad. Sci. U.S.A. **110,** 3447-3452.

Zhang, M., Liu, X.K., Fan, W., Yan, D.F., Zhong, N.S., Gao, J.Y., *et al.* (2018). Transcriptome analysis reveals hybridization-induced genome shock in an interspecific F1 hybrid from *Camellia*. Genome **61,** 477-485.

Zhong, S., Joung, J.G., Zheng, Y., Chen, Y.R., Liu, B., Shao, Y., *et al.* (2011). High-throughput illumina strand-specific RNA sequencing library preparation. Cold Spring Harb. Protoc. **2011,** 940-949.

CHAPTER 2

Divergence of *cis*- and *trans*-regulatory elements drives reconstruction of transcriptome network in x*Brassicoraphanus*

ABSTRACT

Genome hybridization and subsequent duplication is a major force in the evolution of many plant species, producing novel characteristics to increase fitness to the new environment. x*Brassicoraphanus* is an allotetraploid between Chinese cabbage (*Brassica rapa*) and radish (*Raphanus sativus*), which is a rare case of genetic merging two different genus species. x*Brassicoraphanus* displays intermediate or novel phenotypes of the parents due presumably to incomplete dominance and enormous changes in transcriptome profiles. A large portion of duplicated genes was shown to undergo significant changes in expression levels, indicating that massive reconstruction of transcription control network has occurred after a merger of two divergent genomes. The majority of the duplicated genes are adjusted to similar levels in x*Brassicoraphanus*, for which the sequence similarity at *cis*-elements and sharing common transcription factor are important for transcriptional regulations. This study suggest that compatibility of regulatory system might modulate homoeologous expression and reconstruct transcriptome network.

INTRODUCTION

Polyploidy and whole genome duplication are common to many organisms (Wendel, 2000; Moon et al., 2004; Comai, 2005; Semon and Wolfe, 2007; Soltis and Soltis, 2009; Mable et al., 2011; Hittinger, 2013). This phenomenon is especially common in plants, with most of the flowering plants thought to have undergone at least one polyploidization event (Cui et al., 2006; Challacombe et al., 2011; Soltis et al., 2014). The allopolyploidization between divergent genomes is thought to trigger phenotypic changes through massive gene expression alterations. A Massive genome-wide transcriptional change referred to as “transcriptome shock” is a distinctive phenomenon observed in hybrid and allopolyploid species (Hegarty et al., 2006; Buggs et al., 2011). A major feature of transcriptome change in allopolyploids is non-additive gene expression, in which the expression level of the hybrid is not equal to the average of their parental expression levels (Yoo et al., 2014), which often contributes ecological competitiveness, adaptation, and evolutionary plasticity relative to their parent species (Chen, 2007; Abbott et al., 2013).

In newly formed allopolyploids, two divergent parental genomes reside in a single nucleus, and the expression levels of duplicated genes are adjusted to this new environment. Previous studies have examined changes in the expression patterns of hybrids compared to the parents, and found that the total expression of homeologous pairs changed to be similar to the expression of the single gene in one of the parents, a phenomenon called "expression-level dominance" (ELD) (Rapp et al., 2009; Chelaifa et al., 2010; Bardil et al., 2011; Grover et al., 2012; Yoo et al., 2013; Wu et al., 2016; Zhang et al., 2016). However, the detailed mechanisms underlying expression changes have not been elucidated. Transcriptional changes between the parents and the progeny result in part from differences in *cis* and *trans* regulators. Such effects are reported in hybrid species of yeast, Drosophila, Arabidopsis, maize, and rice (Springer and Stupar, 2007; Tirosh et al., 2009; McManus et al., 2010; Shi et al., 2012; Wu et al., 2016).

Intergeneric hybridization is a rare case of genetic merging between two extremely divergent parental species. Compared to interspecific hybridization, the more extreme divergence is expected to compromise genome stability. Since the first such study in 1826 by Sageret (Oost, 1984), occasional crosses between *Brassica* and *Raphanus* have been reported, but it is hard to produce stable progenies (Karpechenko, 1928; Mcnaughton,

1973; Dolstra, 1982; Lee et al., 2011). The recently developed x*Brassicoraphanus* koranhort cv. BB1 (x*B*; AARR; 2n = 4x = 38, also known as Baemoochae) is a successful intergeneric hybrid between *Brassica rapa* (*Br*; AA; 2n = 2x = 20) and *Raphanus sativus* (*Rs*; RR; 2n = 2x = 18) (Lee et al., 2011). The stabilized x*B* individual was produced by embryo rescue after crossing and subsequent microspore culture with induced mutation of F1. Finally, self-fertilization over 30 generations has produced fertile and genetically stable lines (Lee et al., 2011).

In this study, I analyzed the gene expression changes in x*B* relative to parental species. Genome-wide transcriptional changes were observed in x*B*, showing transcription network has been reconstructed after allopolyploidization. I found that *cis*-elements and transcription factors between *Br* and *Rs* contribute to expression change of duplicated genes in x*B*, and large portion of differentially expressed homoeologous pairs were shown to be adjusted to similar level in x*B* owing to less divergent *cis*-element and sharing common transcriptome factors. These data suggests that divergence of *cis*- and *trans*-regulatory in parental species is one of the major effects on massive transcription changes observed in hybrid genome.

.

MATERIALS AND METHODS

Plant materials

x*Brassicoraphanus koranhort* cv. BB1, *Brassica rapa* L. cv. Chiifu-401-42, and *Raphanus sativus* L. cv. WK10039 were grown on 1 × MS medium (Duchefa) in a growth chamber under 16 hr of fluorescent light at 20 ± 10 μmol m^{-2} s^{-1}, 22°C for 14 days. The whole seedlings including shoots and roots and leaf, hypocotyl and root tissues from the seedlings were harvested for RNA-seqs. The three different developmental stages of petal tissues (early immature, breaker, and opened flower) were sampled from a greenhouse for RNA-seq. For vernalization treatment, at nine days after germination, the seedling were moved to a 4 ℃ glass window refrigerator for 34 days.

Orthologous and homoeologous gene pairs

To identify orthologous gene pairs between parental genomes (A_{Br} vs. R_{Rs}), the reciprocal best BLAST hit was performed with >80% of identity and >80% of coverage. Syntenic regions were defined as contiguous

regions containing at least five homologous gene pairs in Br and Rs genomes, and the pairs in syntenic regions were determined as orthologous gene pairs. Homoeologous gene pairs between the parental genomes (A_{xB} vs. R_{xB}) were defined following the same standard.

Transcriptome size estimation

Leaves from 8-week plants of *Br*, *Rs*, and x*B* grown in a growth room were grounded, and divided into six tubes containing 0.1 g grounded tissues. Three tubes were used for DNA extraction with DNeasy Mini kit (Qiagen) and the other tubes were used for RNA extraction with RNeasy Mini kit (Qiagen). Average of cell numbers in the 0.1 g grounded tissues were calculated by dividing the mass of extracted genomic DNA by the mass of 1 x genome size (*Br*, 485 Mb; *Rs*, 510 Mb; x*B*, 995 Mb). Transcriptome sizes were estimated by dividing the total RNA in the 0.1 g grounded tissues by the cell number of the 0.1 g grounded tissues.

RNA-seq analysis

Total RNA was extracted with RNeasy plant kit (Qiagen) as the manufacture's protocol. The DNase treated RNA samples, including two replicates for each of seedling, leaf, hypocotyl and flower, and one replicate

for root and vernalized seedling of x*B* and its parents, were used for constructing RNA-seq libraries (Zhong et al., 2011). RNA sequencing was performed with Illumina HiSeq 2000 platform. The obtained raw reads were filtered using FASTX-Toolkit and low quality reads (Q < 20) were removed. The filtered reads were mapped on *Br* genome (https://plants.ensembl.org/Brassica_rapa), *Rs* genome (http://radish-genome.org/), and x*B* genome using Tophat (Trapnell et al., 2009) with default parameters. The mapped read counts were calculated using HTSeq (Anders et al., 2015). Statistical tests of differentially expressed genes (DEGs) were performed using EdgeR (Robinson et al., 2010) with false discovery rate (FDR) < 0.05 and fragments per kilobase of transcript per million mapped reads (FPKM) $\log_2$ fold change (FC) > 1.

Categorization of additive and non-additive expression patterns

Statistical tests of differential expression among x*B* (average of the FPKM in both A_{xB} and R_{xB} homoeologs), *Br*, and *Rs* was performed using an ANOVA test (FDR < 0.1) followed by the Duncan's multiple range test. Expression patterns of *Br*, *Rs* and xB was classified into 12 groups with six categories "Additive" (I and XII), "Maternal ELD" (II and XI), "Paternal ELD" (IV, IX), "Transgressive down" (III, VII, and X), "Transgressive up"

(V, VI, and VIII), and “No change” according to previously study (Rapp et al., 2009).

Assignment of cis- and trans- regulatory divergence

For assignment of the *cis*- and *trans*-regulatory divergence, gene pairs were classified as following seven categories according to previous report (McManus et al., 2010):

Cis only: Significant differential expression in A_{Br} vs. R_{Rs} and A_{xB} vs. R_{xB}. No significant differential expression in A_{Br} vs. A_{xB} and R_{Rs} vs. R_{xB}.

Trans only : Significant differential expression in A_{Br} vs. R_{Rs}, but not A_{xB} vs. R_{xB}. Significant differential expression in A_{Br} vs. A_{xB} or R_{Rs} vs. R_{xB}.

Cis + *trans*: Significant differential expression in A_{Br} vs. R_{Rs} and A_{xB} vs. R_{xB}. Significant differential expression in A_{Br} vs. A_{xB} or R_{Rs} vs. R_{xB}. The $\log_2$-expression ratio of homeologous genes in A_{Br} vs. R_{Rs} and A_{xB} vs. R_{xB} has the same sign, and expression difference between homeologous genes is more diverged in A_{Br} vs. R_{Rs} than A_{xB} vs. R_{xB} (synergistic *cis*- and *trans*-acting differences).

Cis × *trans*: Significant differential expression in A_{Br} vs. R_{Rs} and A_{xB} vs. R_{xB}. Significant differential expression in A_{Br} vs. A_{xB} or R_{Rs} vs. R_{xB}.

The $\log_2$-expression ratio of homeologous genes in A_{Br} vs. R_{Rs} and A_{xB} vs. R_{xB} has the opposite sign, or expression difference between homeologous genes is more diverged in A_{xB} vs. R_{xB} than A_{Br} vs. R_{Rs} (antagonistic *cis*- and *trans*-acting differences).

Compensatory: Significant differential expression in A_{xB} vs. R_{xB}, but not A_{Br} vs. R_{Rs}. Significant differential expression in A_{Br} vs. A_{xB} or R_{Rs} vs. R_{xB}.

Conserved: No significant differential expression in A_{Br} vs. R_{Rs}, A_{xB} vs. R_{xB}, A_{Br} vs. A_{xB} or R_{Rs} vs. R_{xB}.

Ambiguous: All other patterns of case, which have no clear biological interpretation.

Gene ontology analysis

The Gene ontology (GO) terms of x*B* genome were annotated by Blast2Go using the non-redundant sequence database from NCBI with < 1e-15 of e-value parameter. The statistical comparison of GO term accumulation in *trans* or *cis*-only category was conducted with TopGo in R package (Alexa and Rahnenfuhrer, 2010).

RESULTS

Divergence of *B. rapa* and *R. sativus* genomes

It is assumed that speciation between *Br* and *Rs* has occurred earlier than that between *Br* and *B. oleracea* (*Bo*), although exact speciation times are controversial (Mitsui et al., 2015; Jeong et al., 2016; Kim et al., 2018). Pairwise comparison of coding sequences (CDS) of all orthologs revealed 95.74% of identity between *Br* and *Bo* within the same genus but 91.91% (*Br* and *Rs*) and 92.03% (*Bo* and *Rs*) across the genera (Figure 2-1). The same analysis in tribe Camelineae also showed similar sequence divergence for interspecific (93.52% for *A. thaliana* vs. *A. lyrata*) and intergeneric relationships (89.72% for *A. lyrata* vs. *Capsella rubella*, and 90.39% for *A. thaliana* vs. *C. rubella*) (Figure 2-1). Such high sequence divergence allowed to clearly distinguish between *Br*- and *Rs*-specific transcripts in the x*B* transcriptome data for expression analysis. In transcriptome analysis, on average, 93.58% and 84.57% of *Br* and *Rs* reads were uniquely mapped onto *Br* and *Rs* genome (referred to as A_{Br} and R_{Rs} hereafter), respectively (Table2-1). x*B* reads also mapped onto x*B* genome and 90.62% of x*B* reads were uniquely mapped onto the single subgenome of x*B* (Table2-1), not to

both subgenomes, showing distinguishability of transcript originality. This data suggests that CDS similarity between *Br* and *Rs* is sufficient to distinguish between *Br*- and *Rs*-originated transcripts.

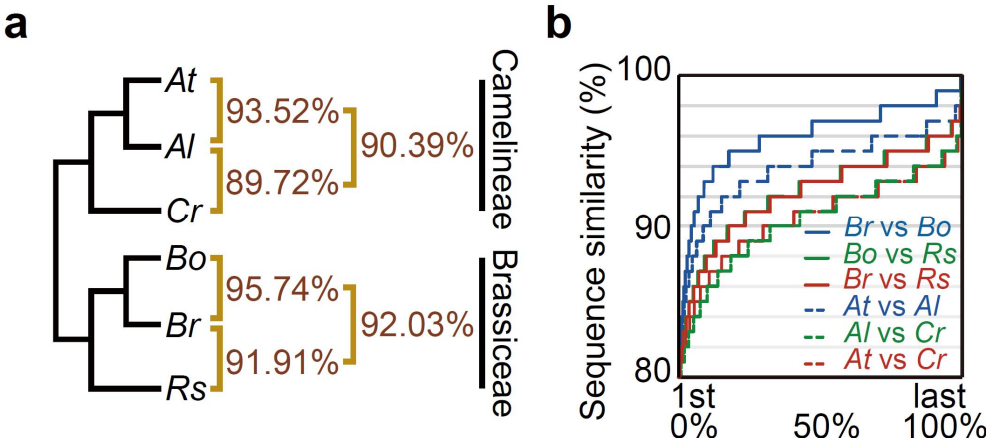


Figure 2-1. Phylogenetic tree and coding sequence similarities between Brassicaceae species.

(a) Phylogenetic relationship and sequence divergence in Camelineae and Brassiceae tribes. Percentages between species represent their CDS similarity of orthologous gene pairs. (b) Distribution of sequence similarities of interspecific/intergeneric orthologs. Horizontal axis indicate orthologous gene pairs sorted in ascending order of sequence similarity. *At*, *A. thaliana*; *Al*, *A. lyrata*; *Cr*, *C. rubella*.

Table 2-1. Number of mapped reads.

Library	Raw read pairs	Total clean read pairs	Clean read pairs mapped on reference genomes	Uniquely mapped read pairs (Mapped read pairs/clean read pairs)
Br seedling 1	16,138,135	15,059,654 (93.3%)	14,054,071 (93.3%)	13,240,295 (94.2%)
Br seedling 2	17,537,316	16,361,858 (93.2%)	15,289,977 (93.4%)	14,454,420 (94.5%)
Rs seedling 1	17,069,315	15,971,644 (93.5%)	14,287,216 (89.4%)	12,050,531 (84.3%)
Rs seedling 2	16,374,890	15,359,919 (93.8%)	13,744,462 (89.4%)	11,593,976 (84.4%)
xB seedling 1	15,724,234	14,739,092 (93.7%)	14,086,494 (95.5%)	12,619,395 (89.6%)
xB seedling 2	17,611,000	16,445,026 (93.3%)	15,680,255 (95.3%)	14,103,034 (89.9%)
Br root	5,191,770	4,803,041 (92.5%)	4,385,507 (91.3%)	4,142,104 (94.4%)
Rs root	5,207,134	4,773,740 (91.6%)	4,151,900 (86.9%)	3,497,863 (84.2%)
xB root	5,381,620	4,948,712 (91.9%)	4,649,741 (93.9%)	4,273,419 (91.9%)
Br leaf 1	7,628,613	7,008,079 (91.8%)	6,489,362 (92.5%)	5,998,612 (92.4%)
Br leaf 2	7,432,059	6,797,630 (91.4%)	6,301,456 (92.7%)	5,916,863 (93.9%)
Rs leaf 1	7,200,455	6,600,605 (91.6%)	5,731,343 (86.8%)	4,859,405 (84.8%)
Rs leaf 2	8,023,031	7,321,129 (91.2%)	6,436,288 (87.9%)	5,448,354 (84.7%)
xB leaf 1	8,212,175	7,506,981 (91.4%)	7,108,879 (94.6%)	6,418,793 (90.3%)
xB leaf 2	7,570,050	6,946,904 (91.7%)	6,578,069 (94.6%)	5,920,761 (90.0%)
Br hypocotyl 1	8,222,674	7,492,816 (91.1%)	6,804,994 (90.8%)	6,276,729 (92.2%)
Br hypocotyl 2	6,988,471	6,354,836 (90.9%)	5,820,593 (91.5%)	5,455,483 (93.7%)
Rs hypocotyl 1	7,587,264	6,859,736 (90.4%)	6,079,804 (88.6%)	5,217,556 (85.8%)
Rs hypocotyl 2	9,368,947	8,528,311 (91.0%)	7,505,769 (88.0%)	6,410,867 (85.4%)

xB hypocotyl 1	7,914,349	7,186,570 (90.8%)	6,804,881 (94.6%)	6,240,590 (91.7%)
xB hypocotyl 2	3,853,431	3,509,583 (91.0%)	3,328,276 (94.8%)	3,047,640 (91.6%)
Br flower 1	7,544,467	6,723,436 (89.1%)	6,181,933 (91.9%)	5,676,763 (91.8%)
Br flower 2	7,716,816	7,041,132 (91.2%)	6,549,725 (93.0%)	6,057,816 (92.5%)
Rs flower 1	13,725,611	12,446,822 (90.6%)	10,872,535 (87.3%)	9,098,013 (83.7%)
Rs flower 2	12,621,385	11,425,466 (90.5%)	9,958,890 (87.1%)	8,353,195 (83.9%)
xB flower 1	12,619,565	11,467,102 (90.8%)	10,786,906 (94.0%)	9,738,581 (90.3%)
xB flower 2	6,187,110	5,694,003 (92.0%)	5,338,603 (93.7%)	4,836,798 (90.6%)

Maintenance of parental transcriptome size in x*Brassicoraphanus*

For RNA sequencing data, read counts in a gene were generally normalized by reads per kilobase per million reads (RPKM) or FPKM to compare relative expression between two samples. Gene expression is the relative expression per transcriptome size of sample, and gene expression difference between two samples can be comparable when two samples have similar transcriptome size. As a result, information of transcriptome size was required to compare gene expression changes between two divergent parental species and x*B*. Transcriptome sizes of *Br*, *Rs*, and x*B* were estimated through the calculation of RNA abundance per cell (Coate and Doyle, 2010; Coate and Doyle, 2015). DNA and RNA were extracted from 0.1g of grounded tissues from *Br*, *Rs*, and x*B*. The number of cells was predicted by dividing mass of extracted DNA into mass of 1x genome size, and then, total RNA mass of one cell was calculated. I found that cells of *Br* and *Rs* had similar transcriptome sizes, and the sum of the parental transcriptome sizes was similar to the x*B* transcriptome sizes (*Br* = 0.52±0.04; Rs = 0.47±0.02; xB = 1.00±0.10) (Table 2-2). Transcriptome size of two subgenomes in x*B* were estimated with proportion of transcripts mapped onto two subgenomes of x*B* (referred to as A_{xB} and R_{xB} hereafter). In x*B* seedling transcriptome, about half of the reads (51.43%) were

assigned to A_{xB} and the other half to R_{xB} (48.56%), indicating that both subgenomes equally contribute to x*B* transcriptome, and similar portions of A_{xB} and R_{xB} transcripts were also present in four different tissues examined (Figure 2-2). In consequence, these data showed that transcriptome size of *Br* and *Rs* were similar and their transcriptome size would be maintained in each subgenome of x*B* after allopolyploidization.

Table 2-2. Transcriptome size estimation.

	Equation	*B .rapa*	*xBrassicoraphanus*	*R .sativus*
Total DNA in 0.1 g grounded tissues (ng/g)	a	1,053.0 ±173.2	606.6 ±76.1	809.9 ±31.1
Genome size (Mb)	b	485	995	510
Number of cells in 0.1 g grounded tissues	c=a/(mass of 1x genome)	2,152,683.0 ±354,215.7	604,540.2 ±75,837.3	1,559,340.3 ±59,977.1
Total RNA in 0.1 g grounded tissues (μg/g)	d	62.97±5.54	37.09±3.77	45.13±2.38
Total RNA in one cell (pg)	e=d/c	32.01±2.57	61.36±6.24	28.94±1.53
Relative transcriptome size	e/61.36	0.52±0.04	1.00±0.10	0.47±0.02

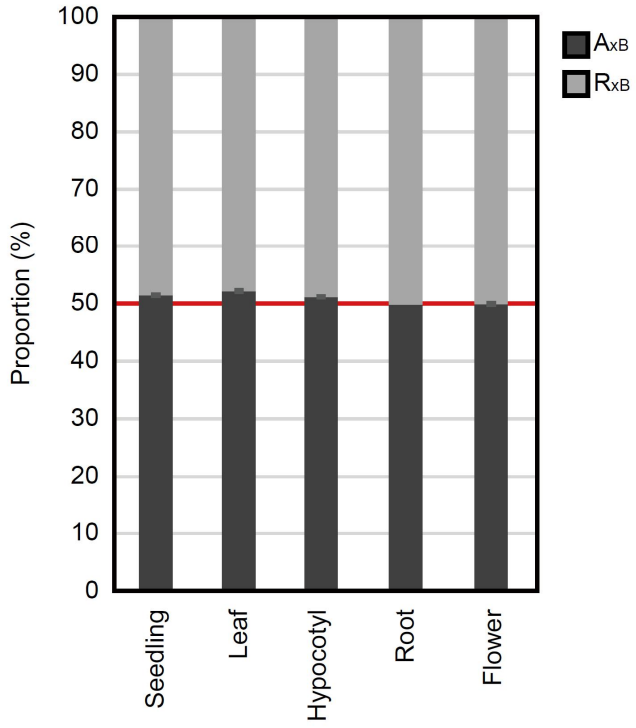


Figure 2-2. Proportions of subgenomic transcriptome size in x*Brassicoraphanus*.

Proportions of reads mapped to the A_{xB} (black) and R_{xB} (grey) subgenomes in various tissues (seedling, leaf, hypocotyl, root and flower). Error bars represent standard deviations of two replicates.

Expression similarity of homoeologous gene pairs in x*Brassicoraphanus*

Orthologous genes between progenitors become homoeologous to each other in the hybrid (Figure. 2-3). When two parental genome are merged, not all homoeologs are additively expressed in the hybrids. Rather, many of homoeologs show non-additive expression, for which interactions between the subgenomes are likely involved through yet unidentified mechanisms (Yoo et al., 2014; Ding and Chen, 2018; Hu and Wendel, 2019).

A total of 28,751 genes were commonly annotated in *Br* and x*B*, among which 2,703 (9.40%) were differentially expressed in seedlings (1,251 up-DEGs and 1,452 down-DEGs in x*B*) (Figure 2-4a). Moreover, 21,905 genes simultaneously expressed in *Br* and x*B* seedlings showed a positive correlation (Pearson's correlation: r = 0.9367, P < 2.2e-16) (Figure 2-4b). By contrast, 4,767 (20.96%) out of total 22,741 annotated genes in *Rs* and x*B* were identified as DEGs in seedlings (2,395 up-DEGs and 2,372 down-DEGs in x*B*) (Figure 2-4a), and 17,158 commonly expressed genes showed a less positive correlation (r = 0.8403, P < 2.2e-16) (Figure 2-4b). Gene expressions levels in other tissues also changed more in R subgenomes than A subgenomes (Figure 2-4c). These findings indicate that *Br*-derived genes more tend to maintain their original expression levels than *Rs*-derived genes, albeit the majority of genes retain parental gene

expression levels in x*B*. In other words, the A_{xB} subgenome displays "expression level dominance" (Rapp et al., 2009; Grover et al., 2012) over the R_{xB} subgenome in x*B*, in which expression of A_{xB} subgenome is similar to its parental A_{Br} genome whereas the other is not.

A total of 15,376 syntenic homoeologous pairs were identified in *Br* and *Rs*, among which 5,701 (37.07%) were differentially expressed in seedlings (2,440 up- and 3,261 down-DEGs in *Br*) (Figure 2-5a). Notably, 12,320 homoeologous genes simultaneously expressed genes in x*B* showed a fairly low positive correlation ($r = 0.7843$, $P < 2.2e\text{-}16$) (Figure 2-5b). In x*B* seedling, however, homoeologous expression were more positively correlated each other ($r = 0.8658$, $P < 2.2e\text{-}16$), with fewer number of homoeologs (3,655; 23.77%) differentially expressed (1,553 up- and 2,102 down-DEGs in A_{xB}) once the two genomes were merged into x*B* (Figures Figure 2-5a and 2-5b). This suggests that distinct expression patterns of many of the orthologous genes in each progenitor are adjusted to similar levels in the context of homoeologous relationships in the hybrid genome. Such expression adjustment was also observed in tissue-specific expression profiles of x*B* (Figure 2-5c).

Previous studies generally analyzed the changes of gene expression levels in allopolyploids relative to the parents, focusing on additive and non-

additive expressions of duplicated genes (Yoo et al., 2014; Ye et al., 2016; Zhang et al., 2016; Wu et al., 2018; Zhang et al., 2018b). Using the classification of expression pattern in allopolyploids, relative expression levels of x*B* were categorized as additivity, expression-level dominance, and transgressivity (Rapp et al., 2009; Grover et al., 2012; Yoo et al., 2013). Out of 15,376 gene pairs, 2,352 (15.2%) and 1,790 (11.6%) gene pairs were binned in additivity (I and XII) and transgressivity (III, VI, V, VII, VIII and X), respectively (Table 2-3). Among transgressively expressed genes, transgressive up-regulation (1,271 genes, 8.2%) was highly observed compared to transgressive down-regulation (519 genes, 3.3%) (Figure 2-6). Almost one-third of gene pairs (4,811 genes, 31.2%) showed expression-level dominance (II, VI, IX, and XI), with bias toward the A genome (3,058 genes, 19.9%) more frequent than toward the R genome (1,753 genes, 11.4%) (Table 2-3), indicating expression level dominance of A subgenomes.

This three-point classification is suitable to identify changes in the sum of duplicated gene expression levels in an allopolyploid nucleus, but it is hard to distinguish expression changes between homoeologs. For a comprehensive comparison of the relative expression level between x*B* and its progenitors, I monitor changes in expression levels of homoeologous

pairs in the A_{xB} and R_{xB} subgenomes from their parental genomes A_{Br} and R_{Rs} (Figure 2-7a). Out of 12,150 orthologous/homoelogous pairs, 7,631 (62.80%) pairs were expressed at similar levels in every genome context (grey in Figure 2-7a). Expressions of 1,435 (11.81%) pairs whose expression levels were significantly different ($|\log_2 FC| \geq 1$) between the parental genomes A_{Br} and R_{Rs} were maintained in the hybrid subgenomes A_{xB} and R_{xB} (blue in Figure 2-7a). Interestingly, at least either one of the pairs of 3,084 (25.38%) orthologs exhibited altered expression in x*B*. It is notable that expressions of 63.91% of them (1,971/3,084) were adjusted to similar levels each other in both subgenomes A_{xB} and R_{xB} (red in Figure 2-7a), where more R_{Rs}-originated genes (n = 1,483) shifted the expression levels to those of A_{Br}-originated homoeologous counterparts than the opposite (n = 316). This "convergent expression" may result from *trans*-acting regulation of gene expression of the homoeologs sharing similar *cis*-regulatory elements. I analyzed sequence similarity in CDS and *cis*-element between the two groups of homoeologous genes that showed convergent (red in Figure 2-7a) and biased (blue in Figure 2-7a) expressions, respectively (Figure 2-7b). Both groups showed a high level of sequence identity between homoeologous pairs in CDS, but the group of homoeologs displaying convergent expression shared higher sequence similarity than the

other (Figure 2-7b). This implies that the *cis*-elements in *Br* and *Rs* have maintained the sequence similarity enough to allow the transcriptional control under the same *trans*-acting factors in the hybrid genome of x*B*. This also suggests that degrees of divergence in *cis*- and *trans*-elements of *Br* and *Rs* would influence on reconstruction of the transcriptional network in x*B*.

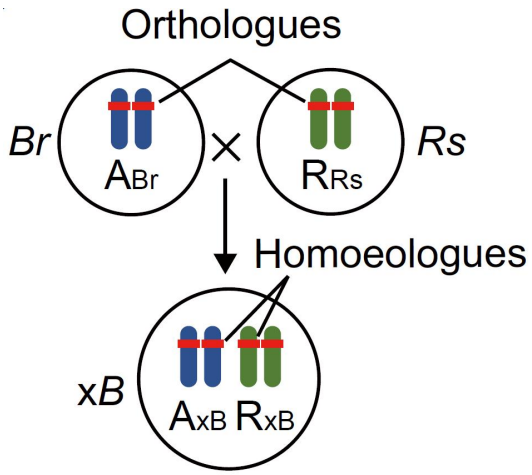


Figure 2-3. Relationship between orthologous and homoeologous genes in x*B* and its progenitors.

Blue and green bars depict *Br* and *Rs* genomes, and gene. A_{Br} and R_{Rs} represent the *Br* and *Rs* genomes. A_{xB} and R_{xB} are two subgenomes of x*B* originated from their parental genomes.

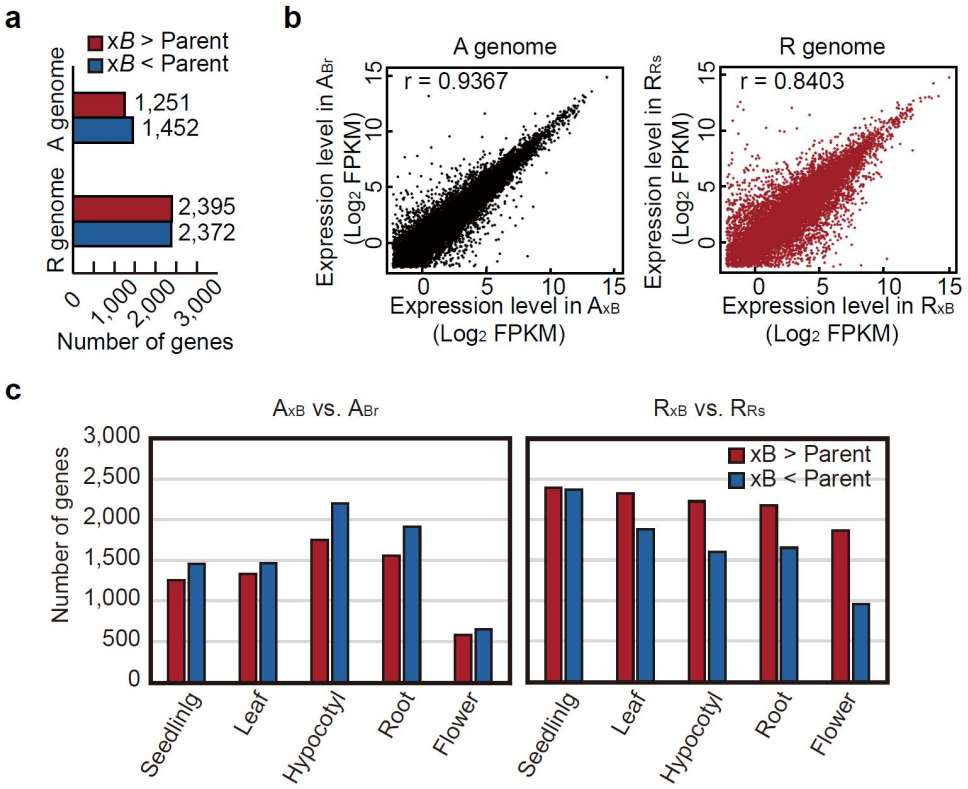


Figure 2-4. Gene expression difference between subgenomes of x*Brassicoraphanus* and its corresponding parental genomes

(a) Number of DEGs in x*B* relative to the progenitors (A_{Br} vs. A_{xB} and R_{Rs} vs. R_{xB}). (b) Scatter plots comparing gene expression levels between A_{Br} and A_{xB} (black), and R_{Rs} and R_{xB} (red). (c) The numbers of DEGs between A_{xB} and A_{Br} (left) and between R_{xB} and R_{Rs} (right) in seedling, leaf, hypocotyl, root and flower.

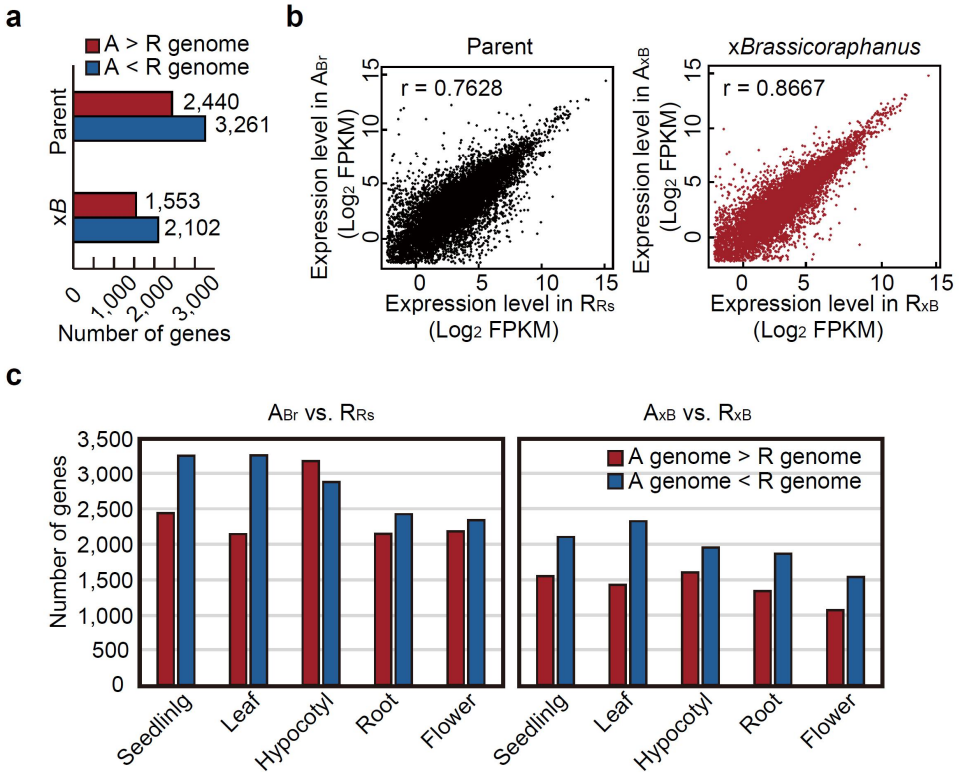


Figure 2-5. Gene expression difference between *B. rapa* and *R. sativus* genomes and between x*Brassicoraphanus* subgenomes.

(a) Number of DEGs in x*B* relative to the progenitors (A_{Br} vs. R_{Rs} and A_{xB} vs. R_{xB}). (b) Scatter plots comparing gene expression levels between A_{Br} and R_{Rs} (black), and A_{xB} and R_{xB} (red). (c) The numbers of DEGs between A_{Br} and R_{Rs} (left) and between A_{xB} and R_{xB} (right) in seedling, leaf, hypocotyl, root and flower.

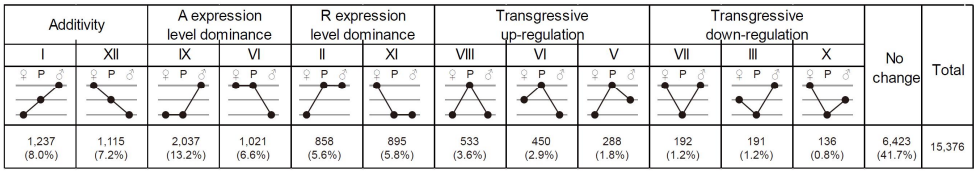


Figure 2-6. The 12 possible classifications of differential expression.

The 12 possible classifications of differential expression between x*B* and the progenitors, based on the report by *Rapp et al.* (2009). ♀, *B. rapa*; ♂, *R. sativus*; P, x*Brassicoarphanus*.

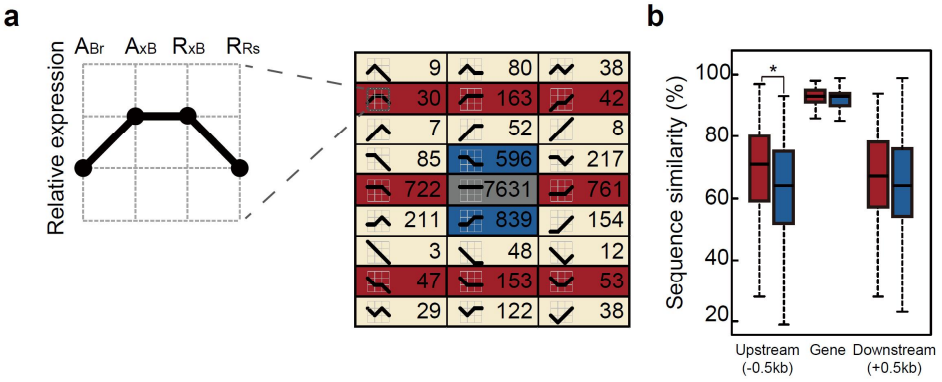


Figure 2-7. Homoeologous gene expression patterns in the x*Brassicoraphanus* relative to progenitors.

(a) Expression patterns of homoeologs in the x*B* relative to parental orthologs. The grey block indicates the gene pairs expressed at similar levels in every genome context. Blue and red blocks represent genes showing biased and convergent expressions, respectively. (h) Sequence similarities of genic and adjacent upstream/downstream regions (0.5 kb) of orthologous genes showing convergent (red) and biased (blue) expressions in x*B* subgenomes (Wilcoxon's rank-sum test, $^{*}P < 2.2\text{e-}10$).

Orthologous gene expression differences by divergence of *trans*-elements of parental species provokes transcriptional changes in x*Brassicoraphanus*

The expression of homoeologous gene pairs could be largely influenced by divergence of *cis*- and *trans*-regulatory elements when they are combined in a single nucleus. (Rockman and Kruglyak, 2006; Williams et al., 2007; Tirosh et al., 2009; McManus et al., 2010; He et al., 2012; Shi et al., 2012). The *cis*-regulatory divergence has allele-specific effects on gene expression, whereas, *trans*-regulatory divergence could equally influence on the expression level of both alleles in the hybrid. Thus, if difference between gene expression levels of two parental orthologs is induced by only *cis*-regulatory divergence, expression bias will be maintained in F1 hybrids. Conversely, if *trans*-regulatory divergence only induce difference between gene expression levels of two parental orthologs, both homoeologs will be expressed equally in F1 hybrids.

The effects of *cis*- and *trans*-regulatory divergence on gene expression difference between homoeologs in x*B* were estimated by calculating the ratio of orthologous gene pairs of the parental species and homeologous gene pairs of the subgenomes [$\log_2(A_{Br}/R_{Rs})$ - $\log_2(A_{xB}/R_{xB})$], and categorized based on the previous classification (McManus et al., 2010),

in which the patterns of *cis* and *trans* effects were divided into seven groups (*cis* only, *trans* only, *cis* + *trans*, *cis* x *trans*, compensatory, conserved, and ambiguous). Out of the 15,376 gene pairs, 7,631 (49.6%) gene pairs had no evidence of *cis*- or *trans*-regulatory divergence, without expression divergence between the parental species and expression changes after allopolyploidization (Figure 2-8a). Expression of 1,933 (12.6%) pairs, whose gene expression differences were removed in x*B*, were classified into *trans*-only, and 1,435 (9.3%) gene pairs showing maintained expression divergence in x*B* were categorized into *cis*-only (Figure 2-8a). The gene pairs with both *cis*- and *trans*-effect were subdivided into three categories of "*cis* + *trans*" (*cis*- and *trans*-effect on the same direction of expression divergence, enhancing effect), "*cis* x *trans*" (*cis*- and *trans*-effect on opposite directions of expression divergence, compensating effect), and "compensatory" (*cis* and *trans*-effect on opposite directions resulting in complete compensation with no significant expression difference between progenitors, compensating effect). Transcriptional regulation was slightly more affected by *trans* only effect than *cis* only effect in x*B*. As shown in Figure 2-7b, gene pairs regulated by *cis*-only effect have lower sequence similarities than *trans*-only effected gene pairs, showing that sequence diversifications of upstream regions have effects on gene expression

divergence (Figure 2-8b). The absolute magnitude of expression divergence between parental homoeologs was also measured for comparing the impacts on expression divergence by *cis* or *trans* effects. I found that *trans*-only effect showed significantly higher divergence of homoeologous expression than *cis*-only effect (P=3.932e-12) (Figure 2-8b). These data suggest that divergences of *trans*-elements between orthologous gene pairs in parental species had more effects on gene expression divergence than those of *cis*-elements, and their pre-existed transcriptional divergences generate massive gene expression changes, resulting in convergent expression in x*B*.

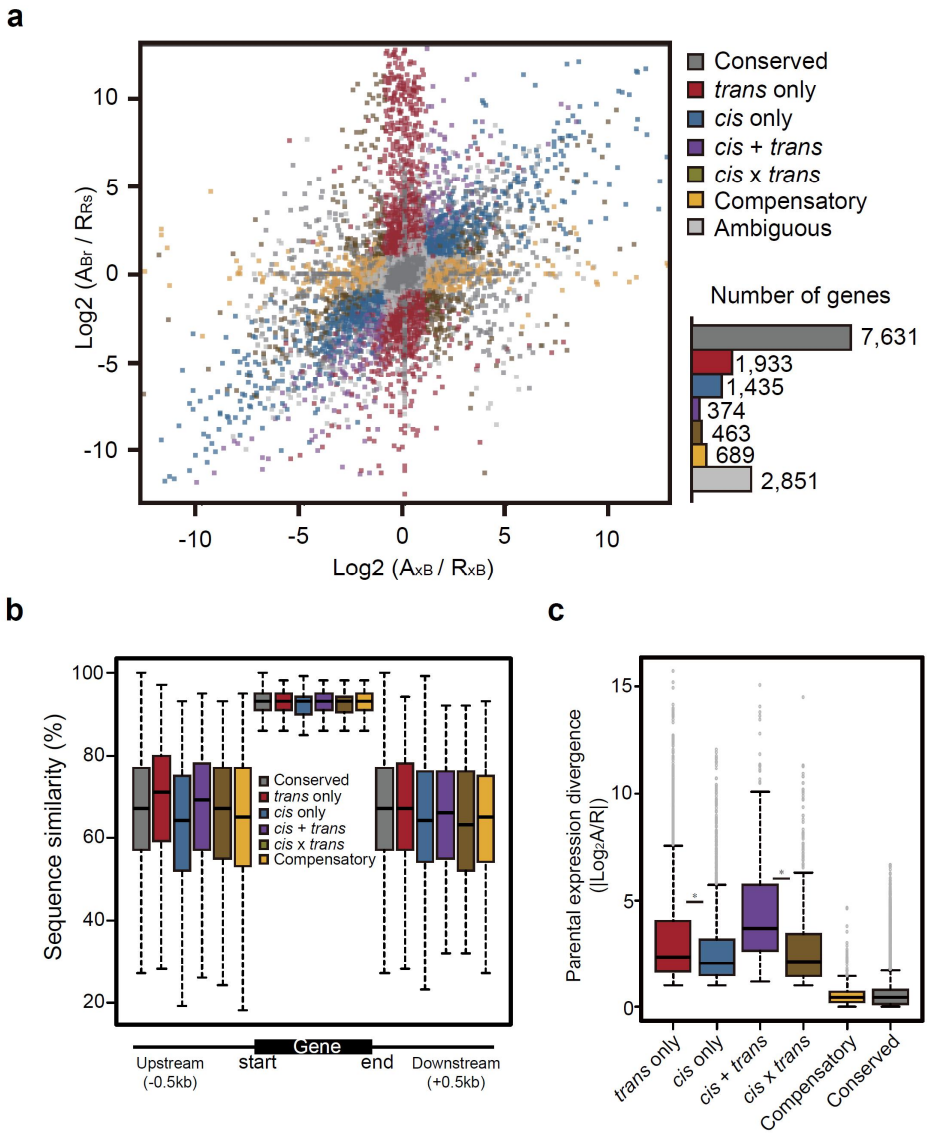


Figure 2-8. Classification of *cis* and *trans*-regulation effects on gene expression changes in x*Brassicoraphanus*.

(a) Scatter plot of homoeologous gene expression fold changes in the progenitor genomes and in x*B* subgenomes. The relative expression levels of A_{xB} and R_{xB} (x axis) was plotted against the relative expression levels

between A_{Br} and R_{Rs} (y axis). Each point is color-coded according to the categories of expression patterns. The bar graph (bottom right) depicts the number of genes in each category. (b) Sequence similarities of genic and adjacent upstream/downstream regions (0.5 kb) of gene pairs showing conserved, *cis*-only, *trans*-only, *cis* + *trans*, *cis* x *trans* and compensatory pattern. (c) Absolute magnitude (fold-change) of expression divergence between parents resulting from *cis*-only, *trans*-only, *cis* + *trans*, *cis* x *trans*, compensatory and conserved pattern. The colors used for boxplot were represented according to (a). Expression divergence between progenitor species by *trans* only was larger than that of *cis* only (2.36 median fold change for *trans*-only and 2.03 median fold change for *cis*-only, Wilcoxon's rank-sum test, P=3.93e-12), and expression divergence by *cis* + *trans* were larger than those of any other effects (3.69 median fold change for *cis* + *trans*, 2.11 median fold change for *cis* x *trans*, 0.44 median fold change for compensatory, and 0.41 median fold change for conserved pattern).

Simultaneous regulation of cellular response in x*Brassicoraphanus* subgenomes

The *cis*- and *trans*-regulatory divergences between parental species are important for reconstruction of transcription network in x*B*. Gene ontology (GO) analysis revealed that homoeologous gene pairs showing convergent expression were significantly associated with cellular responses (Tables 2-3 and 2-4). This implies that cellular signaling by external stimulus could affect on expressions of both homoeologous gene pairs which have similar *cis*-element identity.

To investigate whether the transcription level between parental genomes could be commonly regulated in hybrid genome by external stimulus, changes of gene expression by long term cold stress were measured in x*B* and progenitors. A total of 1,579 and 2,378 genes were differentially expressed in A_{Br} and R_{Rs} by cold stress, respectively (Figure 2-9a). Only 273 genes (17.28% in DEGs of A_{Br} and 11.48% in DEGs of R_{Rs}) were shared indicating clear differences of cold response in *Br* and *Rs* (Figure 2-9a). In cold stress treated x*B* seedling, expression levels of 1,431 and 1,226 genes were significantly changed and notably, 639 genes were commonly regulated (44.65% of DEGs in A_{xB} and 52.12% of DEGs in R_{xB}) (Figure 2-9a). In addition, I analysed the response of gene expression

changes in homoeologous gene pairs showing biased and convergent expressions. There is no correlation between expression change of A_{Br} and R_{Rs} genes showing biased (*cis*-only) and convergent (*trans*-only) expressions (Figure 2-9b), but homoeologous gene pairs in x*B* showing convergent expression were simultaneously regulated with a positive correlation, contrary to those showing biased expression (Figure 2-9c). This observation suggests that environmental stimulus could commonly regulate the transcriptions of homoeologous gene pairs and transcriptional network might be co-regulated by shared transcription factors in hybrid x*B* genome.

Table 2-3. Gene ontology analysis of *cis*-only regulated genes.

GO ID	Term	Annotated[a]	Significant[b]	Expected[c]	p-value	level
GO:0055086	nucleobase-containing small molecule metabolic process	474	65	42.36	0.0002	4
GO:0006753	nucleoside phosphate metabolic process	459	63	41.02	0.0003	5
GO:0009117	nucleotide metabolic process	459	63	41.02	0.0003	6
GO:0019637	organophosphate metabolic process	708	89	63.27	0.0004	4
GO:0046916	cellular transition metal ion homeostasis	33	10	2.95	0.0004	9
GO:0055114	oxidation-reduction process	458	62	40.93	0.0004	4
GO:0006260	DNA replication	205	33	18.32	0.0006	6
GO:0055076	transition metal ion homeostasis	53	13	4.74	0.0006	9

[a]Number of genes mapped to GO id in all homeologous pairs

[b]Number of genes mapped to GO id in *cis*- or *trans*-only category genes

[c]Expected number of genes mapped to GO id

Table 2-4. Gene ontology analysis of *trans*-only regulated genes (Top 20).

GO ID	Term	Annotateda	Significantb	Expectedc	p-value	level
GO:0009611	response to wounding	234	83	31.16	1.5E-18	4
GO:1901700	response to oxygen-containing compound	1624	318	216.23	3.1E-16	4
GO:0010243	response to organonitrogen compound	292	86	38.88	1.2E-13	4
GO:0010200	response to chitin	283	84	37.68	1.5E-13	5
GO:0050896	response to stimulus	3614	584	481.2	6.2E-13	2
GO:0010033	response to organic substance	1860	338	247.66	2.3E-12	4
GO:0009719	response to endogenous stimulus	1318	255	175.49	5.2E-12	3
GO:0042221	response to chemical	2319	402	308.77	7.3E-12	3
GO:0006950	response to stress	2319	398	308.77	5.1E-11	3
GO:0001101	response to acid chemical	1117	219	148.73	8.2E-11	4
GO:0009628	response to abiotic stimulus	1696	306	225.82	1.3E-10	3
GO:0002679	respiratory burst involved in defense response	102	39	13.58	1.8E-10	4
GO:0045730	respiratory burst	102	39	13.58	1.8E-10	4
GO:0009415	response to water	302	80	40.21	3.3E-10	4
GO:0009753	response to jasmonic acid	349	88	46.47	6.4E-10	5
GO:0009414	response to water deprivation	297	78	39.55	8.5E-10	4
GO:0071215	cellular response to abscisic acid stimulus	254	68	33.82	4.6E-09	6
GO:0009738	abscisic acid-activated signaling pathway	250	67	33.29	5.7E-09	6
GO:0006952	defense response	938	183	124.89	7.1E-09	4
GO:0009605	response to external stimulus	1293	237	172.16	9.9E-09	3

[a]Number of genes mapped to GO id in all homeologous pairs

[b]Number of genes mapped to GO id in cis- or trans-only caterogy genes

[c]Expected number of genes mapped to GO id

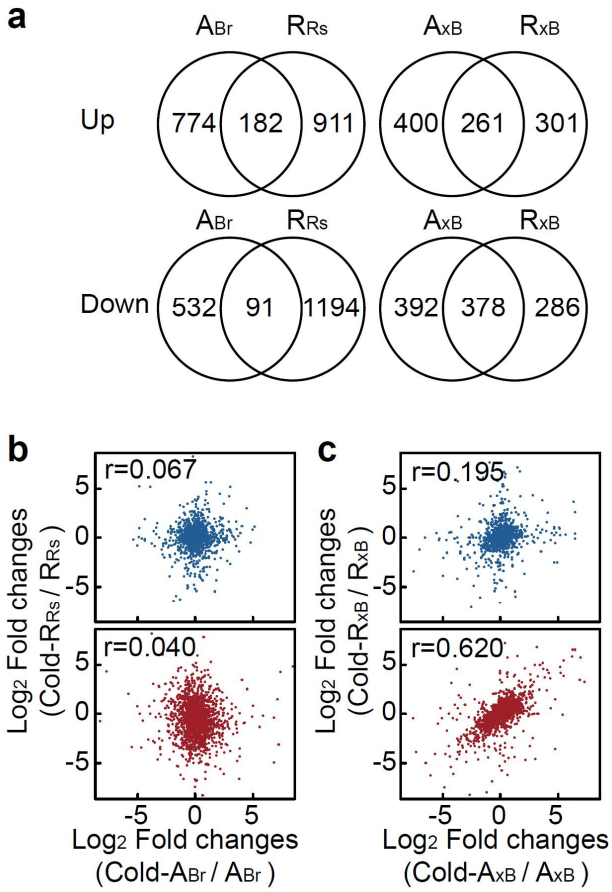


Figure 2-9. Comparison of cold response in *B. rapa*, *R. sativus* and x*Brassicoraphanus*.

(a) Venn diagram of up- and down-regulated genes by long term cold treatment between *Br* and *Rs* genomes (left) and between x*B* subgenomes (right). (b) Scatter plot of expression fold changes with cold treatment of biased (blue) and convergently expressed (red) homoeologous gene pairs between *Br* and *Rs* genome. (c) Scatter plot of expression fold changes with cold treatment of biased (blue) and convergently expressed (red) homoeologous gene pairs between x*B* subgenomes.

DISCUSSION

Genome hybridization coinciding with polyploidization is often observed in natural species, suggesting that it is one of the driving forces of speciation and thereby evolution. In particular, the genus *Brassica* is well known for allopolyploidization within the same genus. Three allotetraploid species, *B. napus* (AACC), *B. juncea* (AABB) and *B. carinata* (BBCC), were generated by hybridization of three diploid species *Br* (AA), *B. nigra* (BB), and *Bo* (CC), and *Rs* can be hybridized with *Br* or *Bo*, generating two type of x*B* (AARR, x*Brassicoraphanus*; RRCC, *Raphanobrassica*, 2n=4x=36). As an extreme case of intergeneric allopolyploids, transcriptome analysis of x*B* would provide a comprehensive view of genome-wide transcriptional changes in allopolyploid.

Expression level dominance between species of U's triangle including *R. sativus*

In the studies of interspecific hybrids and allopolyploids, homoeologous gene expression could be represented by "additive expression" as the expression conservation of parental genes or "non-

additive expression" as altered expression in allopolyploids compared with parental expressions (Rapp et al., 2009; Grover et al., 2012; Yoo et al., 2014). Patterns of non-additive expression were investigated in various interspecific hybrid and allopolyploid plants using RNA-seq technology (Yoo et al., 2013; Combes et al., 2015; Wu et al., 2016; Ye et al., 2016; Zhang et al., 2016; Wu et al., 2018; Zhang et al., 2018b), and ELD has been the most remarkable phenomenon in almost allopolyploid plant. In consistence with previous studies, I found that 6,601 (42.9%) of genes were non-additively expressed, and 4,811 (31.2%) genes of them showed ELD in x*B* (Table 2-3). More ELD-A (3,058 genes, 19.8%) than ELD-R (1,753 genes, 11.4%) gene pairs were observed (Table 2-3), and parental expression profile was more maintained in A_{xB} than R_{xB} (Figures 2-4 and 2-7), indicating that homoeologous gene from A genome has more dominance to maintain intrinsic expression than R genome (AA > RR).

B. napus and *Raphanbrassica* are good materials for investigating direction of expression level dominance in species of *Br* (AA), *Bo* (CC) and *Rs* (RR). In recent transcriptome studies, more ELD-A than ELD-C gene pairs were found in the resynthesized allotetraploid *B. napus*, indicating ELD bias toward A genome (AA > CC) (Wu et al., 2018). In addition, *Raphanbrassica* had relatively more gene pairs of ELD-R than those of

ELD-C, displaying ELD bias towards R genome (RR > CC) (Ye et al., 2016). In this study, ELD-A was dominantly observed compared to ELD-R (AA > RR). Taken together, the three genomes might have a hierarchy of ELD (A > R > C) and the preference of transcriptional regulation on subgenomes in Brassicaceae family may be existed.

A hierarchical level of expression was also found in nucleolar dominance in species of the U's triangle (U, 1935). Nucleolar dominance is the phenomenon in interspecific hybrids, which only rRNA genes inherited from one parent are expressed due to the silencing of other parental rRNA genes. Previous studies in species of the U's triangle revealed that rRNA genes derived from B genome were more expressed in *B. juncea* (BB > AA) and *B. carinata* (BB > CC), and rRNA genes derived from A genome were more expressed in *B. napus* (AA > CC), showing a hierarchical level of nucleolar dominance in *Brassica* species (BB > AA > CC) (Chen and Pikaard, 1997; Chen, 2007). This coincidence of the hierarchy between nucleolar dominance and expression level dominance suggest that the preference of transcriptional regulation on subgenomes in Brassicaceae family may be existed.

Divergence of *cis*- and *trans*-elements between parental species determined the appearance of transcriptome shock in allopolyploids.

Gene expression changes in hybrids could be determined by the magnitude of *cis*- and *trans*-regulatory divergence between their two parental species (Tirosh et al., 2009; McManus et al., 2010; Shi et al., 2012; Combes et al., 2015). Previous studies showed that the amount of *cis*-effect and *trans*-effect were different in their species, representing relatively higher contributions of *cis*-effects (Shi et al., 2012; Yang et al., 2016; Zhang et al., 2018a) or higher contributions of *trans*-effect (McManus et al., 2010; Xu et al., 2014; Combes et al., 2015). Several studies with intra or interspecific hybrids suggested that *cis*-regulatory divergence increases along with divergence time between species (Coolon et al., 2014), and would more influence on expression differences between species than within species (Emerson et al., 2010). In addition, *trans*-regulatory divergence may predominantly contribute to expression divergence of genes responding environment (Tirosh et al., 2009).

In transcriptome data of x*B*, the proportion of homoeologs with *trans*-only effect (1,933 genes, 12.8%) is higher than that of *cis*-only effect (1,435 genes 9.3%), and *trans*-only effect more contribute to expression changes in x*B* than *cis*-only effect (Figure 2-8), despite of their intergeneric

relationship. This result indicates that divergence of *trans* acting factor might be more preferred than *cis*-element divergence for expression divergence between *Br* and *Rs*, and less diverged *cis*-elements and sharing common transcription factor would contribute to adjustment of expression of homoeologous genes at similar levels in x*B*.

GO term enrichment analysis in x*B* showed that terms associated with "basic cellular functions" were enriched in *cis*-only regulated genes, whereas the GO terms with "cellular responses" were significantly represented in *trans*-only regulated genes (Tables 2-3 and 2-4). This result suggests that expression changes may be required to adapt environment in ancient *Brassica* species, and expression divergence of genes responding environment might be more accelerated by *trans*-regulatory divergence than *cis*-element divergence in the speciation of *Br* and *Rs*. In addition, genes regulated by compensating *cis* and *trans* effect are more observed than that of enhancing *cis* and *trans* effect in xB (Figure 2-8). Enhancing *cis* and *trans* effects would be accelerated by disruptive selection, increasing expression divergence between species, whereas compensating *cis* and *trans* effects could be promoted by stabilizing selection, reducing expression divergence (Tirosh et al., 2009). These suggest that stabilizing selection

rather than disruptive selection may be dominantly induced in the speciation of *Br* and *Rs*.

In this study, I found that expression levels of duplicated genes in intergeneric allotetraploid x*B* were regulated by the balance of *cis*- and *trans*-element divergence between parental species. In the speciation of two species, changes of *cis*- and *trans*-elements were accumulated as time passes, and expression difference resulting from alteration of regulatory elements would contribute to survival in nature. This divergence of regulatory elements between parental species eventually leads to reconstruction of transcriptome network immediately after the two genome merged, while displaying massive transcriptional changes as “transcriptome shock” (Hegarty et al., 2006; Buggs et al., 2011). This instantaneous genome-wide transcriptional alteration would contribute to possibility for ecological competitiveness and adaptation in hybrid plants for plant evolution, and this process can be further expanded to the intergeneric level.

REFERENCES

Abbott, R., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J.E., Bierne, N., *et al.* (2013). Hybridization and speciation. J. Evol. Biol. **26,** 229-246.

Alexa, A., and Rahnenfuhrer, J. (2010). topGO: enrichment analysis for gene ontology. R package version **2,** 2010.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics **31,** 166-169.

Bardil, A., de Almeida, J.D., Combes, M.C., Lashermes, P., and Bertrand, B. (2011). Genomic expression dominance in the natural allopolyploid *Coffea arabica* is massively affected by growth temperature. New Phytol. **192,** 760-774.

Buggs, R.J.A., Zhang, L.J., Miles, N., Tate, J.A., Gao, L., Wei, W., *et al.* (2011). Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. Curr. Biol. **21,** 551-556.

Challacombe, J.F., Eichorst, S.A., Hauser, L., Land, M., Xie, G., and Kuske, C.R. (2011). Biological consequences of ancient gene acquisition and duplication in the large genome of *Candidatus Solibacter usitatus* Ellin6076. Plos One **6,** e24882.

Chelaifa, H., Monnier, A., and Ainouche, M. (2010). Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina x townsendii* and *Spartina anglica* (Poaceae). New Phytol. **186,** 161-174.

Chen, Z.J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu. Rev. Plant Biol. **58,** 377-406.

Coate, J.E., and Doyle, J.J. (2010). Quantifying whole transcriptome size, a prerequisite for understanding transcriptome evolution across species: an example from a plant allopolyploid. Genome Biol. Evol. **2,** 534-546.

Coate, J.E., and Doyle, J.J. (2015). Variation in transcriptome size: are we getting the message? Chromosoma **124,** 27-43.

Comai, L. (2005). The advantages and disadvantages of being polyploid. Nat. Rev. Genet. **6,** 836-846.

Combes, M.C., Hueber, Y., Dereeper, A., Rialle, S., Herrera, J.C., and Lashermes, P. (2015). Regulatory divergence between parental alleles determines gene expression patterns in hybrids. Genome Biol. Evol. **7,** 1110-1121.

Coolon, J.D., McManus, C.J., Stevenson, K.R., Graveley, B.R., and Wittkopp, P.J. (2014). Tempo and mode of regulatory evolution in *Drosophila*. Genome Res. **24,** 797-808.

Cui, L., Wall, P.K., Leebens-Mack, J.H., Lindsay, B.G., Soltis, D.E., Doyle, J.J., *et al.* (2006). Widespread genome duplications throughout the history of flowering plants. Genome Res. **16,** 738-749.

Ding, M., and Chen, Z.J. (2018). Epigenetic perspectives on the evolution and domestication of polyploid plant and crops. Curr. Opin. Plant Biol. **42,** 37-48.

Dolstra, O. (1982). Synthesis and fertility of *Brassicoraphanus* and ways of transferring *Raphanus* characters to *Brassica* (Pudoc).

Emerson, J.J., Hsieh, L.C., Sung, H.M., Wang, T.Y., Huang, C.J., Lu, H.H.S., *et al.* (2010). Natural selection on *cis* and *trans* regulation in yeasts. Genome Res. **20,** 826-836.

Grover, C.E., Gallagher, J.P., Szadkowski, E.P., Yoo, M.J., Flagel, L.E., and Wendel, J.F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. New Phytol. **196,** 966-971.

He, F., Zhang, X., Hu, J.Y., Turck, F., Dong, X., Goebel, U., *et al.* (2012). Genome-wide analysis of *cis*-regulatory divergence between species in the *Arabidopsis* Genus. Mol. Biol. Evol. **29,** 3385-3395.

Hegarty, M.J., Barker, G.L., Wilson, I.D., Abbott, R.J., Edwards, K.J., and Hiscock, S.J. (2006). Transcriptome shock after interspecific hybridization in *Senecio* is ameliorated by genome duplication. Curr. Biol. **16,** 1652-1659.

Hittinger, C.T. (2013). Saccharomyces diversity and evolution: a budding model genus. Trends Genet. **29,** 309-317.

Hu, G., and Wendel, J.F. (2019). *Cis-trans* controls and regulatory novelty accompanying allopolyploidization. New Phytol. **221,** 1691-1700.

Jeong, Y.M., Kim, N., Ahn, B.O., Oh, M., Chung, W.H., Chung, H., *et al.* (2016). Elucidating the triplicated ancestral genome structure of radish based on chromosome-level comparison with the *Brassica* genomes. Theor. Appl. Genet. **129,** 1357-1372.

Karpechenko, G.D. (1928). Polyploid hybrids of *Raphanus sativus* L. x*Brassica oleracea* L. Mol. Gen. Genet. **48,** 1-85.

Kim, C.K., Seol, Y.J., Perumal, S., Lee, J., Waminal, N.E., Jayakodi, M., *et al.* (2018). Re-exploration of U's triangle *Brassica* species based on chloroplast genomes and 45S nrDNA Sequences. Sci. Rep. **8**.

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. Theor. Appl. Genet. **122,** 885-891.

Mable, B.K., Alexandrou, M.A., and Taylor, M.I. (2011). Genome duplication in amphibians and fish: an extended synthesis. J. Zool. **284,** 151-182.

McManus, C.J., Coolon, J.D., Duff, M.O., Eipper-Mains, J., Graveley, B.R., and Wittkopp, P.J. (2010). Regulatory divergence in *Drosophila* revealed by mRNA-seq. Genome Res. **20,** 816-825.

Mcnaughton, I.H. (1973). Synthesis and Sterility of *Raphanobrassica*. Euphytica **22,** 70-88.

Mitsui, Y., Shimomura, M., Komatsu, K., Namiki, N., Shibata-Hatta, M., Imai, M., *et al.* (2015). The radish genome and comprehensive gene expression profile of tuberous root formation and development. Sci. Rep. **5,** 10835.

Moon, C.D., Craven, K.D., Leuchtmann, A., Clement, S.L., and Schardl, C.L. (2004). Prevalence of interspecific hybrids amongst asexual fungal endophytes of grasses. Mol. Ecol. **13,** 1455-1467.

Oost, E. (1984). x*Brassicoraphanus* Sageret or x*Raphanobrassica* Karpechenko? Cruciferae Newsletter. **9,** 11-12.

Rapp, R.A., Udall, J.A., and Wendel, J.F. (2009). Genomic expression dominance in allopolyploids. BMC Biol. **7,** 18.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26,** 139-140.

Rockman, M.V., and Kruglyak, L. (2006). Genetics of global gene expression. Nat. Rev. Genet. **7,** 862-872.

Semon, M., and Wolfe, K.H. (2007). Consequences of genome duplication. Curr. Opin. Genet. Dev. **17,** 505-512.

Shi, X.L., Ng, D.W.K., Zhang, C.Q., Comai, L., Ye, W.X., and Chen, Z.J. (2012). *Cis*- and *trans*-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. Nat. Commun. **3,** 950

Soltis, D.E., Visger, C.J., and Soltis, P.S. (2014). The Polyploidy Revolution Then And Now: Stebbins Revisited. Am. J. Bot. **101,** 1057-1078.

Soltis, P.S., and Soltis, D.E. (2009). The role of hybridization in plant speciation. Annu. Rev. Plant Biol. **60,** 561-588.

Springer, N.M., and Stupar, R.M. (2007). Allelic variation and heterosis in maize: How do two halves make more than a whole? Genome Res. **17,** 264-275.

Tirosh, I., Reikhav, S., Levy, A.A., and Barkai, N. (2009). A yeast hybrid provides insight into the evolution of gene expression regulation. Science **324,** 659-662.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25,** 1105-1111.

Wendel, J.F. (2000). Genome evolution in polyploids. Plant Mol. Biol. **42,** 225-249.

Williams, R.B., Chan, E.K., Cowley, M.J., and Little, P.F. (2007). The influence of genetic variation on gene expression. Genome Res. **17,** 1707-1716.

Wu, J., Lin, L., Xu, M., Chen, P., Liu, D., Sun, Q., Ran, L., *et al.* (2018).

Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. BMC Genomics **19,** 586.

Wu, Y., Sun, Y., Wang, X.T., Lin, X.Y., Sun, S., Shen, K., *et al.* (2016). Transcriptome shock in an interspecific F1 triploid hybrid of *Oryza* revealed by RNA sequencing. J. Integr. Plant Biol. **58,** 150-164.

Xu, C.M., Bai, Y., Lin, X.Y., Zhao, N., Hu, L.J., Gong, Z.Y., *et al.* (2014). Genome-Wide Disruption of Gene Expression in Allopolyploids but Not Hybrids of Rice Subspecies. Mol. Biol. Evol. **31,** 1066-1076.

Yang, M., Wang, X., Huang, H., Ren, D., Su, Y., Zhu, P., *et al.* (2016). Natural variation of H3K27me3 modification in two *Arabidopsis* accessions and their hybrid. J. Integr. Plant Biol. **58,** 466-474.

Ye, B., Wang, R., and Wang, J. (2016). Correlation analysis of the mRNA and miRNA expression profiles in the nascent synthetic allotetraploid *Raphanobrassica*. Sci. Rep. **6,** 37416.

Yoo, M.J., Szadkowski, E., and Wendel, J.F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity **110,** 171-180.

Yoo, M.J., Liu, X., Pires, J.C., Soltis, P.S., and Soltis, D.E. (2014). Nonadditive gene expression in polyploids. Annu. Rev. Genet. **48,** 485-517.

Zhang, D.W., Pan, Q., Tan, C., Liu, L.L., Ge, X.H., Li, Z.Y., *et al.* (2018a). Homoeolog expression is modulated differently by different subgenomes in *Brassica napus* hybrids and allotetraploids. Plant Mol. Biol. Rep. **36,** 387-398.

Zhang, H., Gou, X., Zhang, A., Wang, X., Zhao, N., Dong, Y., Li, L., and Liu, B. (2016). Transcriptome shock invokes disruption of parental expression-conserved genes in tetraploid wheat. Sci. Rep. **6,** 26363.

Zhang, M., Liu, X.K., Fan, W., Yan, D.F., Zhong, N.S., Gao, J.Y., *et al*. (2018b). Transcriptome analysis reveals hybridization-induced genome shock in an interspecific F1 hybrid from *Camellia*. Genome **61,** 477-485.

Zhong, S., Joung, J.G., Zheng, Y., Chen, Y.R., Liu, B., Shao, Y., *et al*. (2011). High-throughput illumina strand-specific RNA sequencing library preparation. Cold Spring Harb. Protoc. **2011,** 940-949.

ABSTRACT IN KOREAN

이질배수화 현상은 서로 다른 종간의 교잡에 의하여 염색체 쌍이 배가되어 한 핵 속에 존재하게 되는 현상으로, 즉각적인 종 분화를 발생시키며 새로운 형질을 창출하는 등의 생물 진화에 매우 중요한 요소이다. 이러한 현상은 특히 식물에서 많이 보고되어 왔으며, 개화식물의 다양한 종의 분화에 기여했을 것으로 여겨진다. 그러나 초기 이질배수체의 형성은 일반적으로 임성의 감소나 불임에 의하여 저해되며, 이는 합성초기 관찰되는 배수체의 유전체의 불안정성에 의하여 발생한다고 알려져 있다. 반면, 배무채는 유전적 거리가 먼 배추와 무의 속간교배를 통해 만든 배수체로, 대부분의 새로 합성된 초기 배수체 식물과는 달리 임성이 존재하며 유전적으로 안정적인 특성을 보인다. 이 연구에서는 원연 간 교배로 합성된 유전체의 안정화 기작에 대한 연구를 위하여 배추와 무, 그리고 배무채의 유전체, 후성유전체 및 전사체에 대한 비교 분석을 실시하였다. 유전체 연구를 위하여 배무채의 염기서열 분석을 통해 유전체 서열을 해독하였으며, 해독된 유전체를 기반으로 양친의 유전체 구조가 안정적으로 보존되어 있다는 것을 확인하였다.

그리고 양친 유전체의 유전적 차이가 감수분열시 발생할 수 있는 양친 유래 염색체간의 재조합의 가능성을 낮추어 염색체 이수성 현상을 저해 할 수 있으며, 이를 안정적인 유전체 구조의 원인으로 제안하였다. 또한 후성유전체 연구를 통해 이동성 유전자 내에 DNA 메틸화가 증가되어 있음을 확인하였으며, 양친 유전체간 small RNA 상호작용에 의한 DNA 메틸화 증가 기작을 제안하였다. 전사체 분석을 통해, 배무채로 합성됨에 따라 발생하는 전반적인 유전자 발현 변화 양상을 파악하였으며, 양친에서 기인한 유전자의 *cis* 인자의 유사성과 핵 내 전사조절인자의 공통작용에 의하여 배무채에서 중복으로 존재하는 유전자가 비슷한 발현을 가지도록 변화하였음을 확인하였다. 이 연구를 통해서 양친 유전체의 유전적 차이가 배수화 현상 이후 발생하는 유전체 쇼크를 억제하는데 기여할 수 있음을 보였으며, 양친의 발현 조절인자들의 호환성이 배수체의 전사조절 네트워크의 재정립에 중요한 요소임을 확인하였다.