Genome Evolution after Whole Genome Triplication: the Subgenome Dominance in *Brassica rapa*

9

Feng Cheng, Jian Wu, Bo Liu and Xiaowu Wang

Abstract

Subgenome dominance is widely existed in plant species that experienced allopolyploidization. Subgenome dominance represents a series of biased pehnomenons as that one subgenome retains more genes, more dominantly expressed genes, less functional mutations etc., over the other subgenomes. *Brassica rapa*, which experienced a whole genome triplication (WGT) event ~11 million years ago, exhibits significant subgenome dominance, with the LF (the least fractionated) subgenome retains ~1.5 times more genes in average than the other two subgenomes MF1 and MF2 (more fractionated 1 and 2). Furthermore, paralogous genes in LF are always expressed to higher levels and accumulated less functional mutations than that located at MF1 and MF2. Further research found that small RNA mediated methylation of transposons that distributed at genes' flanking regions plays an important role in the formation of subgenome dominance. Finally, based on these findings, a two-step process was proposed to illustrate the WGT event in *B. rapa*.

9.1 Introduction

Whole genome duplication (WGD) or polyploidization is a common feature that is widely spread among the plant species, and many even experience several rounds of WGD. WGD occurred frequently in the evolutionary history of plants; it improved the tolerance of plants to mutations and provided abundant genetic material for the evolution of new features. This allowed plants to survive challenging external factors, such as environmental fluctuations and/or habitat changes. *Brassica rapa* is a crop species that evolved through rounds of polyploidization, such as the γ triplication, and α and β duplications. These three polyploidization events are shared by *B. rapa* and the model plant *Arabidopsis thaliana*, and most of the species in Brassicaceae. In addition to the three rounds of

F. Cheng · J. Wu · B. Liu · X. Wang (✉)
Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences,
Beijing 100081, China
e-mail: wangxiaowu@caas.cn

© Springer-Verlag Berlin Heidelberg 2015

X. Wang and C. Kole (eds.), *The Brassica rapa Genome*, Compendium of Plant Genomes,
DOI 10.1007/978-3-662-47901-8_9

polyploidization, *B. rapa* also experienced an extra whole genome triplication (WGT) event that is shared by all *Brassica* crops.

The WGT of *B. rapa* serves as a good model to study the subgenome's evolution in polyploids. Polyploidization plays important roles in the evolution of plant species. However, the mechanism of subsequent evolution following polyploidization has not been sufficiently investigated. The lack of whole genome sequences for appropriate polyploid species has restrained related progress. The whole genome sequencing of *B. rapa* offered a good chance to investigate genome evolution following polyploidization. The WGT event in *B. rapa* occurred approximately 11 million years ago, which is long enough for the three subgenomes generated from the WGT to become well fractionated and differentiated. However, the event is young enough that the three subgenomes still can be separated unambiguously. The genomic differentiation and reshuffling has not fractionated the three subgenomes too heavily to be recognized, and most genes are clearly identifiable in the outgroup, *A. thaliana*. Owing to the genome sequences of *B. rapa*, the subgenome dominance phenomenon was observed and the mechanism regulating the phenomenon investigated. Based on this information, a two-step theory of polyploidization was hypothesized to explain the process of WGT that occurred in *B. rapa*.

This chapter will introduce studies of the structural and functional evolution of the *B. rapa* genome following WGT, mainly focusing on the subgenome dominance phenomenon. Subgenome dominance is believed to occur in allotetraploids (Garsmeur et al. 2014). The phenomenon is an integration of a set of bias features among subgenomes produced by a polyploidization event: (1) biased fractionation: one subgenome retains more genes, while others retain less; (2) dominant expression: one subgenome has more genes expressed at higher levels than their paralogs in other subgenomes; (3) biased mutation: genes from one subgenome tend to be more resistant to functional mutations and accumulate fewer such mutations than genes from other subgenomes. There is evidence that transposable elements (TEs) play an important role in the formation of subgenome dominance.

9.2 Reconstruction of the Three Subgenomes in *B. rapa*

To investigate the subgenome evolution in *B. rapa*, the accurate partitioning of the three subgenomes is the first and also the most important step. By multiple genome comparisons, researchers found that genomes of species in Brassicaceae can be divided into 24 genomic blocks (GBs; labeled from A to X) (Parkin et al. 2005; Schranz et al. 2006; Cheng et al. 2013). The genome that has eight chromosomes and one set of 24 GBs was suggested to be the ancestral common karyotype (ACK) of family Brassicaceae. The examples for extant species that keep the ACK genome structure are *Arabidopsis lyrata* and *Capsella rubella* (Hu et al. 2011; Slotte et al. 2013). It is the rearrangement sometimes accompanied with WGDs that gave birth to all the Brassicaceae species. The genomic fragments corresponding to these 24 GBs were defined in the genome of *A. thaliana*, using two genes to denote the boundaries of each GB. It is a useful resource for comparative genomic analyses in *Brassicaceae*.

The genome of *A. thaliana* was used as the reference to determine the distributions of GBs in the genome of *B. rapa*. Since *B. rapa* experienced an extra round of WGT compared with the diploid species in Brassicaceae as ACK, it should have three copies of 24 GBs, i.e., 72 GBs in total. Since the triplication event that occurred in the early stage of *Brassica* origin is much more recent than that of *Carica papaya* or *Vitis vinifera* ($\sim$80 Mya) or the most recent tetraploidy (α duplication) for *A. thaliana* lineage, the syntenic genes in *A. thaliana* and *B. rapa* can be determined at a high accuracy (Cheng et al. 2012a). In total, 7813, 5439, and 1675 genes were determined to have 1, 2, and 3 syntenic copies, respectively, of *B. rapa* genes in *A. thaliana*. The GB information in *A. thaliana* can be easily transferred to *B. rapa* based on these syntenic

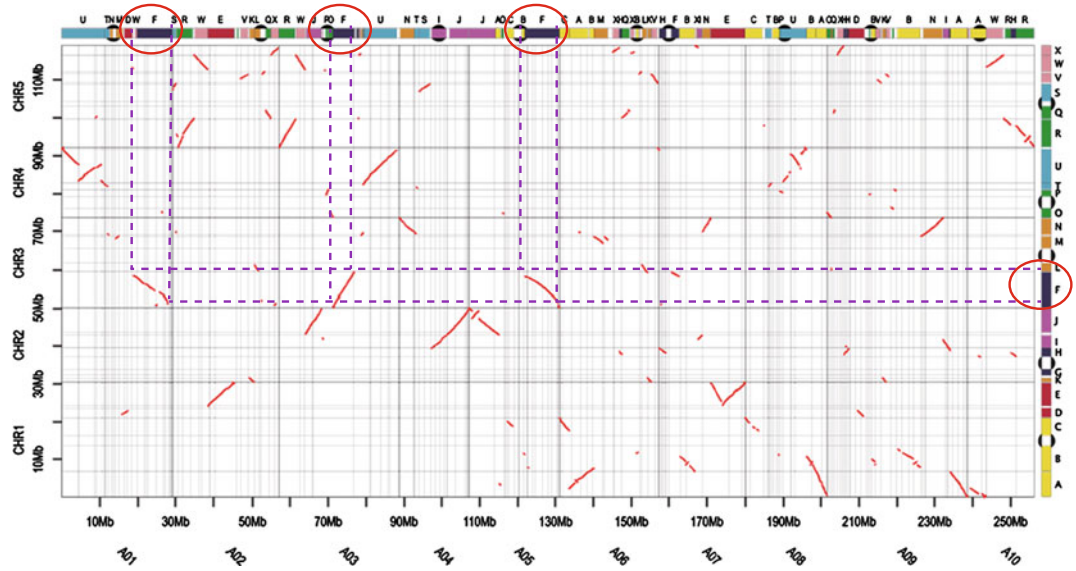

Fig. 9.1 Syntenic gene pairs in the genomes of *Arabidopsis thaliana* and *Brassica rapa* were used as anchors to transfer the genomic blocks information from the genome of *A. thaliana* to *B. rapa*. Block F was used as an example (*purple dashed lines*) to show how the positions of the three copies of F block were determined in the genome of *B. rapa*. Revised from figure in Wang et al. (2011)

gene pairs (Fig. 9.1), thus determining the distribution, as well as the fractionation information for the 72 GBs (actually 71 GBs detected, one copy of G block completely lost) across 10 chromosomes of *B. rapa*.

The three subgenomes can be accurately separated by comparing syntenic fragments of *B. rapa* with the diploid ancestral genomes. The continuous syntenic fragments of *B. rapa* to *A. thaliana* can be selected and ordered along the seven chromosomes of the diploid ancestor translocation Proto-Calepineae Karyotype (tPCK) (Cheng et al. 2013) or any other diploid genome. The three copies of syntenic fragments in *B. rapa* have different breakpoints compared with their ancestral tPCK genome; thus it is quite straight forward to place them along the chromosomes of tPCK, just like placing jigsaw puzzle pieces (Supplementary Fig. 1 in Cheng et al. 2014). After all these syntenic fragments were correctly placed, for each of the seven ancestral chromosomes in tPCK, we should obtain three copies corresponding to the three subgenomes in *B. rapa*.

9.3 Biased Gene Fractionation Among the Three Subgenomes of *B. rapa*

After WGD, subgenomes that coexist in one nucleus are always differentiated. This differentiation is easily detected by comparing the gene density among subgenomes generated from WGD (Thomas et al. 2006; Schnable et al. 2011). This bias in gene densities among subgenomes is called as biased gene fractionation. It has been observed in *A. thaliana* and maize (Thomas et al. 2006; Schnable et al. 2009; Woodhouse et al. 2010), and may be a common feature in species with ancient polyploid genomes (Sankoff et al. 2010).

Biased gene fractionation was observed among the three subgenomes of *B. rapa*. After the reconstruction of the three copies of chromosomes for all seven chromosomes of the ancestral genome tPCK, the gene loads for each chromosome copy could be easily investigated. For each of the seven chromosomes of tPCK,

Table 9.1 Numbers of dominantly expressed genes in the subgenomes LF, MF1, and MF2 of *Brassica rapa*, revised from table of Cheng et al. (2012b)

Organism/Accession	Twofold dominance		
	LF	MF1 (LF/MF1)	MF2 (LF/MF2)
Leaf	393	262 (1.50)	233 (1.69)
Stem	362	258 (1.40)	228 (1.59)
Root	356	273 (1.30)	221 (1.61)
Chiifu	363	253 (1.43)	216 (1.68)
L58	355	229 (1.55)	194 (1.83)

these is always one copy of a reconstructed chromosome from the *B. rapa* genome that has significantly more genes than the other two copies (~1.5-fold higher in gene density). While the other two copies of chromosomes with fewer genes contained slightly different numbers of genes. The significant difference in gene content cannot occur accidently at the chromosome level; thus it represents the real gene variation among the three subgenomes. Therefore, one set of seven chromosomes with the highest gene density for each were grouped and called the LF subgenome. Another set of seven chromosomes that had moderate numbers of genes were grouped to be the more fractionated subgenome 1 (MF1), while the remaining copies of the seven chromosomes that had the least genes represented the more fractionated subgenome 2 (MF2).

9.4 Dominant Gene Expression Among Subgenomes of *B. rapa*

Besides the variation of gene fractionation, paralogous genes in different subgenomes also show variant levels of expression. The subgenome that retains more genes is the one that has more genes expressed at higher levels than their paralogs in the other subgenomes. This phenomenon is called genome dominance. It has been observed in many genomes with recent polyploidization, such as maize (Schnable et al. 2011), the allotetraploids of *A. thaliana* and *A. arenosa* (Wang et al. 2006), and the natural allotetraploid *Tragopogon miscellus* (Buggs et al. 2010), as well as an allotetraploid cotton species (Senchina et al. 2003).

Genome dominance was observed in *B. rapa* across many accessions and by means of different statistical methods. Several whole genome transcriptome sequencing datasets were used to confirm the genome dominance phenomenon among subgenomes of *B. rapa*, such as mRNA-Seq data for different organs and different accessions or varieties of *B. rapa* (Wang et al. 2011, 2012; Cheng et al. 2012b). Two methods was used to evaluate the dominance expression among paralogs extracted from a confident fully retained homologs set (Cheng et al. 2012b). The first one is the twofold rule, in which only when a gene is expressed at least twofold higher than the other two homologs, is it considered to be dominantly expressed. Using this method, genome dominance was observed from all the available expression datasets. The subgenome LF, which has the most retained genes, always has more dominantly expressed genes (1.30–1.83 times) than that of subgenomes MF1 and MF2 (Table 9.1). Subgenome MF1 has slightly more dominantly expressed genes than subgenome MF2. The second method is the horse race experiment; a gene winning by any fraction of expression was considered the dominantly expressed one. Similar patterns of genome dominance were observed under the horse race experiment compared with that of the twofold rule.

The more strictly the method was applied, the more significant the observed genome dominance was among subgenomes of *B. rapa*. When determining the dominantly expressed genes using a more than twofold higher rule, such as a

three or fivefold rule, more significant genome dominance was observed. For instance, the subgenome LF will have relatively more dominant genes than that of MF1 and MF2. Furthermore, the median difference in syntenic gene pairs in which LF expressed at a higher level was marginally higher than the median difference for the pairs in which either MF1 or MF2 expressed at a higher level. These results supported the hypothesis that subgenome dominance effects should be stable and are controlled by factors in the genome of *B. rapa*.

9.5 Biased Distribution of Functional Mutations Among Subgenomes of *B. rapa*

Another difference among subgenomes of polyploids is the variation of functional mutations. The dominant subgenome, which retains more genes and has more genes expressed at higher levels, is also the subgenome whose genes have better defenses against functional mutations. This phenomenon was observed in maize (Schnable et al. 2011) and is also true in the genome of *B. rapa* (Cheng et al. 2012b).

Genes in subgenome LF accumulated fewer functional mutations than those of subgenomes MF1 and MF2. Biased distributions of functional mutations among subgenomes of *B. rapa* could also be considered as ongoing biased gene fractionation because many functional mutations resulted in pseudogenes. Using the resequencing data from different accessions of *B. rapa*, it was found that genes in the subgenome LF accumulate significantly fewer nonsynonymous single nucleotide polymorphisms (SNPs) and frameshift InDels than the two MF subgenomes. Take *B. rapa* strains L144, a rapid cycling laboratory accession, and VT117, a vegetable turnip accession, as examples (Cheng et al. 2012b), the former contains 561,367 SNPs and 45,995 InDels, and the latter contains 562,935 SNPs and 60,003 InDels in the 23,716 confident genes determined by their resequencing data. Among them, genes in subgenome LF always had fewer nonsynonymous SNPs (4574 and 4103 for L144 and VT117, respectively) and frameshift InDel

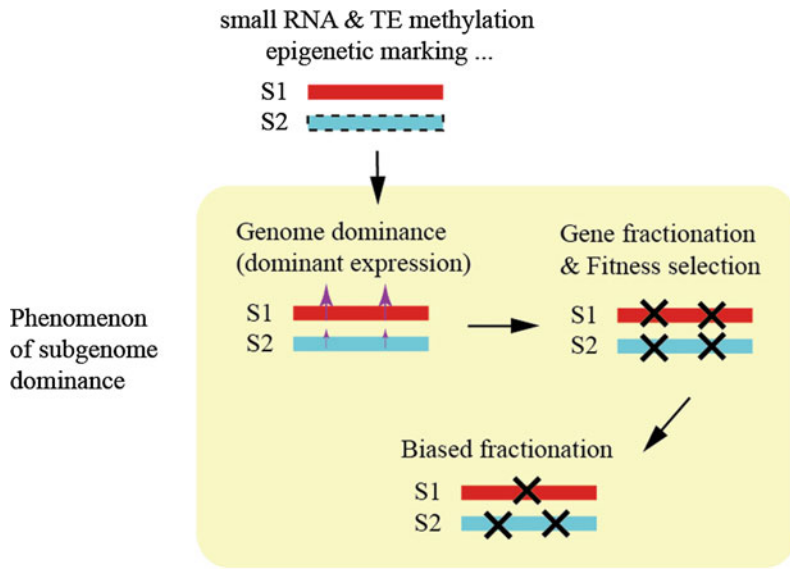

Fig. 9.2 Proposed formation of subgenome dominance in the genome of *Brassica rapa*. Phenomenon of dominant expression, ongoing biased functional mutation, and biased gene fractionation observed in *B. rapa* were united by the aim of improving plant fitness. Meanwhile, the gene dominant expression is likely to be regulated by the small RNA mediated methylation of transposable elements

(52 and 60 for L144 and VT117, respectively) mutations than the genes in the MFs (4834 and 4866 SNPs in the MF1 of L144 and turnip, respectively, and 4620 and 4530 SNPs in the MF2 of L144 and VT117, respectively; 72 and 77 InDels in the MF1 of L144 and VT117, respectively, and 73 and 83 InDels in the MF2 of L144 and VT117, respectively) of both L144 and VT117.

The three features of subgenome dominance —biased gene fractionation, dominant expression of genes and biased functional mutations— in *B. rapa* can be united in improving the fitness of plants. In this explanation system (Fig. 9.2), genes that are expressed to higher levels than their paralogs are more important for the biological activity of the plant. Thus, functional mutations of these genes would be more significant in reducing the plant's fitness than mutations of their syntenic paralogs. Therefore, natural selection conserves these dominantly expressed genes against functional mutations, whereas their paralogs accumulate more mutations and eventually fractionated, resulting in a higher gene density in the dominant subgenome and lower gene density in the dominated subgenomes. This explanation was first suggested in the maize genome and subsequently in the genome of *B. rapa* (Schnable et al. 2011, 2012; Cheng et al. 2012b). However, this explanation of the subgenome dominance phenomenon still leaves unanswered questions: what element controls the biased distribution of dominantly expressed genes among subgenomes?

9.6 Small RNA-Mediated Methylation of TEs Regulates Genome Dominance

Epigenetic modifications of TEs play important roles in regulating gene expression. In *A. thaliana*, Hollister and Gaut (2009), Hollister et al. (2011) found that the methylated TEs could suppress the expression of nearby genes. They suggested that there was a dynamic balance between gene expression and the activity of neighboring TEs in plants (Hollister and Gaut 2009). The activation of TEs reduces the stability of the plant genome, which is harmful, while the methylation of TEs will inactive them, which is beneficial to the plant. However, the methylated TEs will also suppress nearby gene expression, leading to the reduced fitness of the host. Thus, there should be a trade-off between the methylation of TEs and gene expression. By analyzing the small RNA data of *B. rapa*, it was revealed that small RNAs also play an important role in the formation of subgenome dominance in *B. rapa*.

Small RNAs regulate the dominance expression among the subgenomes of *B. rapa* through the methylation of TEs in the flanking regions of genes (Woodhouse et al. 2014). Based on the analysis of small RNA-Seq data, it was discovered that 24 bp small RNAs were mapped primarily to TE sequences located at the 5′ and 3′ regions of genes and more small RNAs mapped to genes located in the MF1 and MF2 subgenomes than in the LF. The biased targeting of small RNAs to TE sequences was much more significant when comparing dominantly expressed genes with their paralogs. These data suggest that small RNA-targeted TEs play an important role in the formation of subgenome dominance (Fig. 9.2). It is likely that the 24 bp small RNA-mediated TE methylation suppressed the expression of nearby genes, and its biased distribution in the subgenomes of *B. rapa* then led to subsequent subgenome dominance.

9.7 Theory of Two-Step Polyploidization and Its Relationship to Subgenome Dominance

A two-step theory was hypothesized to explain the polyploidization process of WGT that occurred in the ancestor of *B. rapa* (Wang et al. 2011; Cheng et al. 2012b; Tang et al. 2012). In the first step, the two tPCK genomes (precursors of MF1 and MF2) merged. A first round of genomic reshuffling and gene fractionation gave

birth to a new diploid (consisting of subgenomes MF1 and MF2). Since there is no significant genome dominance detected between MF1 and MF2 in the current genome of *B. rapa*, autotetraploidization cannot be excluded as a possible process for the first tetraploidization. However, considering that more small deletions occurred recently in the MF1subgenome than MF2 (Tang et al. 2012), allotetraploidization is favored as the first duplication event. In the second step, another tPCK genome (LF) was added to the fractionated diploid genome (MF1 and MF2). Then, a second round of genomic shuffling and gene fractionation resulted in the mesohexaploid ancestor of *B. rapa*. The "two" merged genomes (LF and MFs) are different, indicating that the second step was a process of alloploidization, resulting in the biased gene fractionation and dominant gene expression phenomenon.

References

Buggs RJ, Chamala S, Wu W, Gao L, May GD et al (2010) Characterization of duplicate gene evolution in the recent natural allopolyploid *Tragopogon miscellus* by next-generation sequencing and Sequenom iPLEX MassARRAY genotyping. Mol Ecol 19(Suppl1):132–146

Cheng F, Wu J, Fang L, Wang X (2012a) Syntenic gene analysis between Brassica rapa and other Brassicaceae species. Front Plant Sci 3:198

Cheng F, Wu J, Fang L, Sun S, Liu B et al (2012b) Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. PLoS One 7:e36442

Cheng F, Mandakova T, Wu J, Xie Q, Lysak MA et al (2013) Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. Plant Cell 25:1541–1554

Cheng F, Wu J, Wang X (2014) Genome triplication drove the diversification of Brassica plants. Hortic Res 1:14024

Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A et al (2014) Two evolutionarily distinct classes of paleopolyploidy. Mol Biol Evol 31:448–454

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 19:1419–1428

Hollister JD, Smith LM, Guo YL, Ott F, Weigel D et al (2011) Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. Proc Natl Acad Sci USA 108:2322–2327

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF et al (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet 43:476–481

Parkin IA, Gulden SM, Sharpe AG, Lukens L, Trick M et al (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. Genetics 171:765–781

Sankoff D, Zheng C, Zhu Q (2010) The collapse of gene complement following whole genome duplication. BMC Genom 11:313

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci USA 108:4069–4074

Schnable JC, Wang X, Pires JC, Freeling M (2012) Escape from preferential retention following repeated whole genome duplications in plants. Front Plant Sci 3:94

Schranz ME, Lysak MA, Mitchell-Olds T (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci 11:535–542

Senchina DS, Alvarez I, Cronn RC, Liu B, Rong J et al (2003) Rate variation among nuclear genes and the age of polyploidy in Gossypium. Mol Biol Evol 20:633–643

Slotte T, Hazzouri KM, Agren JA, Koenig D, Maumus F et al (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. Nat Genet 45:831–835

Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS et al (2012) Altered patterns of fractionation and exon deletions in *Brassica rapa* support a two-step model of paleohexaploidy. Genetics 190:1563–1574

Thomas BC, Pedersen B, 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

Wang J, Tian L, Lee HS, Wei NE, Jiang H et al (2006) Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. Genetics 172:507–517

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

Wang F, Li L, Li H, Liu L et al (2012) Transcriptome analysis of rosette and folding leaves in Chinese cabbage using high-throughput RNA sequencing. Genomics 99:299–307

Woodhouse MR, Schnable JC, Pedersen BS, LyonsE, Lisch D et al (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol 8: e1000409

Woodhouse MR, Cheng F, Pires JC, Lisch D, Freeling M et al (2014) Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. Proc Natl Acad Sci USA 111:5283–5288. doi: 10.1073/pnas.1402475111