REVIEW ARTICLE

Genome triplication drove the diversification of *Brassica* plants

Feng Cheng, Jian Wu and Xiaowu Wang

The genus *Brassica* belongs to the plant family Brassicaceae, which includes many important crop species that are used as oilseed, condiments, or vegetables throughout the world. *Brassica* plants comprise many diverse species, and each species contains rich morphotypes showing extreme traits. *Brassica* species experienced an extra whole genome triplication (WGT) event compared with the model plant *Arabidopsis thaliana*. Whole genome sequencing of the *Brassica* species *Brassica rapa*, *Brassica oleracea* and others demonstrated that WGT plays an important role in the speciation and morphotype diversification of *Brassica* plants. Comparative genomic analysis based on the genome sequences of *B. rapa* and *A. thaliana* clearly identified the WGT event and further demonstrated that the translocated Proto-Calepine Karyotype (tPCK, *n* = 7) was the diploid ancestor of the three subgenomes in *B. rapa*. Following WGT, subsequent extensive genome fractionation, block reshuffling and chromosome reduction accompanied by paleocentromere descent from the three tPCK subgenomes during the rediploidization process produced stable diploid species. Genomic rearrangement of the diploid species and their hybridization then contributed to *Brassica* speciation. The subgenome dominance effect and biased gene retention, such as the over-retention of auxin-related genes after WGT, promoted functional gene evolution and thus propelled the expansion of rich morphotypes in the *Brassica* species. In conclusion, the WGT event initiated subsequent genomic and gene-level evolution, which further drove *Brassica* speciation and created rich morphotypes in each species.

*Horticulture Research* (2014) 1, 14024; doi:10.1038/hortres.2014.24; published online: 21 May 2014

**GENUS BRASSICA IN BRASSICACEAE**

Plants of the genus *Brassica* are grouped into the tribe Brassiceae, which belongs to the plant family Brassicaceae. Brassicaceae comprises a large family of plants that exhibit common and distinct features in their flowers. The flowers have cruciform petals and comprise a large family of plants that exhibit common and distinct features in their flowers. The flowers have cruciform petals and are homologous with the nine chromosomes of *B. oleracea*.

**THE U’S TRIANGLE MODEL DESCRIBES THE RELATIONSHIP AMONG BRASSICA CROPS**

Six species of the genus *Brassica* are used widely throughout the world as oilseed, condiments, fodder or vegetable crops. Three of these species are diploid (*Brassica rapa*, *n* = 10; *B. nigra*, *n* = 8; and *B. oleracea*, *n* = 9), whereas the other three are allotetraploids (*B. juncea*, *n* = 18; *B. napus*, *n* = 19; *B. carinata*, *n* = 17) derived from each pair of the three diploid species. The genetic relationships of these species were identified and confirmed by extensive experimental crosses between tetraploid and/or diploid plants as well as karyotyping or microscopic inspection at the synopsis stage of meiosis in these crosses. For example, when crosses were performed between *B. napus* and *B. rapa*, F1 plants with a chromosome number of *n* = 29 were generated. When there were two sets of chromosomes from the *B. rapa* genome, only nine normal pairs of synaptic chromosomes at the synopsis stage in meiotic cells of the F1 plants were observed by microscopic inspection. The other 10 chromosomes of *B. napus* that can never form normal pairs of synaptic chromosomes in the F1 were affected by the chromosomes of the A genome, *B. rapa*. This supports the theory that these 10 chromosomes of *B. napus* are homologous to the 10 chromosomes in *B. rapa*. Similar experiments using crosses of *B. napus* and *B. oleracea* showed that the other nine chromosomes of *B. napus* are homologous with the nine chromosomes of *B. oleracea*. Taken together, these results lead us to conclude that *B. napus* is a tetraploid of *B. rapa* and *B. oleracea*. Based on experimental evidence, the relationships of the six species were simply described by the U’s triangle model, in which the three diploid species *B. rapa*, *B. nigra* and *B. oleracea* are considered to be the basic genomes A, B and C, respectively, and are placed at the three vertices of the triangle. The three allotetraploids, *B. juncea*, *B. napus* and *B. carinata*, which are hybrids of AB, AC and BC, respectively, are placed in the middle of the three edges of the triangle. U’s triangle has been successfully applied to aid in understanding the relationships among *Brassica* crops and has fostered the genetic study of these species.

**THE RICH DIVERSITY OF BRASSICA PLANTS**

*Brassica* plants have rich diversity with respect to both speciation and the abundant morphotypes in each *Brassica* species. *Brassica* crops described by U’s triangle are close relatives, and many traits are shared but developed independently and in parallel, such as heading leaves and enlarged roots (Figure 1). One of the important vegetables of *B. rapa* is Chinese cabbage, which has the distinct feature of a leafy head; this feature is also observed in *B. oleracea* and *B. juncea*. Turnip, another morphotype of *B. rapa*, develops enlarged roots as storage organs, and this feature is also found in *B. juncea* and *B. napus*. Furthermore, each *Brassica* species has evolved multiple morphotypes, including leafy heads and enlarged roots, other enlarged organs of stems and inflorescences, oilseeds, sarsons and even ornamental features. Different morphotypes...
have different usages. In *B. rapa* (Figure 1a), heading Chinese cabbage and pak choi are consumed as leafy vegetables. Chinese cabbage is distinct for its large leafy head, whereas pak choi has relatively smaller leaves and does not develop heading leaves. Turnip has an enlarged root that is eaten and occasionally used as fodder. Caixin and purple caitai bolt rapidly and generate long, tender stems used as food. Morphotypes of oilseed *B. rapa* produce large, full seeds for oil extraction and sarsons produce seed pods that are eaten in India. Some morphotypes of *B. rapa* develop beautiful leaf patterns and colors, thus are used as ornamental plants. *B. oleracea* also has an abundance of morphotypes, as shown in Figure 1b. Heading *B. oleracea* is consumed as a leaf vegetable, whereas oilseed *B. oleracea* produces edible oil. Cauliflower and broccoli, special morphotypes of *B. oleracea*, have developed enlarged inflorescences that are eaten as vegetables. Other *Brassica* crops, such as *B. juncea*, have even greater morphotype richness than *B. rapa* and *B. oleracea*. In addition to these cultivated crops, there are many wild relatives of the species in U’s triangle that have greatly diversified phenotypes, further extending the diversity of *Brassica* plants.

The WGT event was important to the speciation and the expansion of rich morphotypes in the genus *Brassica*. The subsequent genomic rearrangement and gene evolution initiated by WGT promoted the appearance of a variety of *Brassica* plants.

**CHROMOSOME EVOLUTION AFTER WGT PROMOTED BRASSICA SPECIATION**

Genomic blocks (GBs) are collinear chromosome fragments conserved among different genomes. The genomes of Brassicaceae species are composed of 24 GBs labeled A to X. These blocks were defined by a comparative genomic analysis among the genomes of many Brassicaceae species, such as *B. napus*, *Arabidopsis thaliana*, *A. lyrata* and *Capsella rubella*. These GBs formed the basic units in ancestral chromosome reshuffling that generated the present-day species. Parkin et al. constructed a high-density linkage map of *B. napus*, which is the allotetraploid of *B. rapa* and *B. oleracea*. Based on this map, they defined 21 GBs shared between *B. napus* and *A. thaliana*. Subsequently, Schranz and co-workers combined the *B. napus* linkage map with comparative mapping results from

---

**Figure 1.** Rich morphotypes of *Brassica* plants. (a) Morphotypes of *B. rapa*; top two lines from left to right: pak choi, heading *B. rapa*, turnip, oilseed, purple pak choi, caixin, mizuna, purple caitai and takucai; the third line shows additional morphotypes or varieties of the previous morphotypes. (b) Morphotypes of *B. oleracea*; top two lines from left to right: heading cabbage, Brussels sprouts, broccoli, cauliflower, purple cabbage, purple cauliflower, collard; the third line shows additional morphotypes or varieties. Some of the pictures were collected from the Internet.
A. lyrata and C. rubella, both of which have eight chromosomes and are defined as the ancestral common karyotype (ACK or AK) of Brassicaceae, resulting in the definition of 24 GBs (A–X) as the basic units of all Brassicaceae genomes. Various combinations of these 24 blocks, occasionally accompanied by whole genome duplication, compose all the genomes of Brassicaceae species. Genomes that contain only one set of the 24 GBs are considered diploid species. There are many such genomes in Brassicaceae; the main karyotypes include ACK (n = 8), Proto-Calepine karyotype (PCk, n = 7), translocated PCK (tPCk, n = 7) and A. thaliana (n = 5).12,13 Previous studies based on phylogenetics and comparative chromosome painting showed that ACK was the ancestral karyotype of Brassicaceae.12,13 ACK has eight chromosomes with GBs ordered from A to X across chromosomes one to eight. Examples of extant ACK species include A. lyrata and C. rubella.14,16,17 Both PCk and tPCk have seven chromosomes and differ in one inter-chromosomal translocation.13 Conringia orientalis has a PCK chromosome order,13 whereas Schrenkiella parvula has a tPCK genome.18 Genomes having more than one set of the 24 GBs in Brassicaceae are considered paleopolyploid species, including all Brassica crop species, which experienced a WGT event.

The genomic structure of the triplicated-genome species in the Brassica genus as well as their ancestral genome evolution were first studied in detail after the whole genome sequencing of the genus B. rapa3 followed by the C genome B. oleracea.19 The sequencing of other Brassica species is now underway. Genome datasets of the Brassica species are maintained and continuously updated within the Brassica database (http://brassicadb.org).20 For B. rapa, the genome size was estimated to be 485 Mb based on K-mer analysis, and gene prediction suggested 41,020 protein-coding gene models. For B. oleracea, the genome size was estimated to be 630 Mb with approximately 47,588 gene models.

Comparative genomic analysis between B. rapa and A. thaliana clearly indicated the WGT event experienced by B. rapa.3 Syntetic gene analysis between the triplicated genome of B. rapa and the diploid genome of A. thaliana using tool SynOrths showed that most genes inherited from their nearest common diploid ancestor were shared by both species (80.2% and 73.8% for B. rapa or A. thaliana, respectively).21,22 After WGT, the genomic fragments were reshuffled and fractionated. However, the local gene order was conserved, and syntetic genomic fragments can be clearly observed in both B. rapa and A. thaliana. Furthermore, for each GB in A. thaliana, three corresponding syntetic GBs in B. rapa were detected, which were generated by the WGT event.2,21

Genomic synteny analysis between B. oleracea and A. thaliana as well as between B. rapa and B. oleracea showed that B. oleracea has good genomic collinearity with genomes of A. thaliana and B. rapa (Figure 2). It was found that as observed in B. rapa, B. oleracea has three copies of each GB found in A. thaliana, thus confirming at the whole genome sequence level that B. oleracea also experienced the extra WGT event.19 A previous comparative genomic study of B. juncea and other Brassicas based on genetic maps showed that the genome of B. nigra also shared the WGT event.23 Based on the comparison of GB distribution in the Brassica A, B and C genomes, we provided a framework for the comparative genomic study for Brassica species.

The diploid ancestor of B. rapa before WGT had seven chromosomes, which resembles the block arrangement of tPCk.24 In B. rapa, there should be three sets of the 24 GBs, ideally 72 GBs in total, that resulted from WGT. Using A. thaliana as a reference, the genomic fragments of the three copies of each GB were clearly identified (with the exception of one copy of block G) in the genome of B. rapa.24 For a certain GB, the three corresponding copies in B. rapa are not always all associated with a same GB. Some block

![Figure 2](image-url)  
Figure 2. Framework for the comparative genomic analysis of Brassica plants. (a) Chromosomal synteny between B. rapa and B. oleracea determined based on whole chromosome sequences. Block associations for each chromosome are listed above or below the chromosome bars. The numbers 1, 2 and 3 placed after each block label (A–X) denote subgenomes LF, MF1 and MF2, respectively. (b) Evolutionary relationships between the chromosomes of B. rapa and B. nigra. Block information for B. nigra was extracted from the genetic map of B. napus.11,12 Block information is shown below each chromosome bar, and the syntenic chromosomes of B. rapa to B. nigra are listed above the chromosome bars.
associations are found in the ACK genome, such as block A, which associates with block B (block association A/B). However, some block associations do not exist in ACK, suggesting unique block associations for the diploid ancestor of *B. rapa* but not ACK; this lack of associations may result from genomic reshuffling after WGT. Based on this block distribution information, the block association relationships across the 10 chromosomes of *B. rapa* were observed. The breakage and formation of block associations occur independently; thus, the probability that an ancestral block association was broken more than twice after WGT is low, and the probability that a newly derived block association formed more than twice after WGT is also low. Based on this rule, by counting the copy numbers of all block associations in genome of *B. rapa* and comparing these numbers with the extant diploid karyotypes in *Brassicaceae*, such as ACK, PCK, tPCK and *A. thaliana*, Cheng and co-workers found that the diploid ancestor of *B. rapa* had a tPCK-like karyotype. Furthermore, block association analysis of *B. oleracea* or *B. napus* and *R. sativus* based on genetic maps showed that these species also evolved from an ancestor having a tPCK genome.

The distribution patterns of transposable elements (TEs) support the positions of the 21 tPCK paleocentromeres in the genome of *B. rapa*. It is well known that TEs are enriched in the flanking regions of centromeres; this configuration has been observed in the genomes of many species such as *A. thaliana*, maize and soybean. TE sequences continue to show a relatively high density surrounding the positions of the 21 paleocentromeres in *B. rapa* millions of years after rediploidization following WGT. After reconstructing the three subgenomes of *B. rapa* along the seven tPCK ancestral chromosomes, using a method similar to playing a jigsaw puzzle (Supplementary Fig. S1), we plotted the TEs as a function of their density along the 21 reconstructed tPCK chromosomes (Figure 3).

This plot clearly shows that the TE distribution variation reflects the locations of the 10 inherited centromeres in *B. rapa* as well as the 11 inactivated paleocentromeres. The supported locations of the 21 paleocentromeres accurately match the centromere regions of tPCK, which are positioned between the block associations B/C, G/H, I/J, S/T, P/W, M/E and D/V for tPCK chromosomes one to seven, respectively. Other TE-rich regions, such as the distal end of AK2/5/6/8 in subgenome LF (the least fractionated subgenome) (Figure 3), could represent traces of paleocentromeres from more ancient genome duplications. In addition, as shown in Figure 3, gene fractionations or large genomic fragmental deletions are generally more concentrated near the paleocentromere regions compared to the genomic background.

Chromosomal reduction together with paleocentromere descent from the primal hexaploid ancestor (tPCK × 3, n = 21) is important for the speciation of *Brassica* plants. After WGT, extensive chromosome reshuffling during rediploidization contributed to the origin of closely related species in *Brassica*. As mentioned above, in the cross between *B. napus* and *B. rapa*, more than two copies of homologous chromosomes in the synopsis stage of meiosis will result in abnormal synaptonemal complexes, thereby decreasing the fertility of gametes. Logically, natural selection drives the rediploidization process with chromosomal rearrangement that removes the extra homologous chromosomes. Further rounds of genomic reshuffling of the rediploid ancestor at different evolutionary timepoints then created the different species in *Brassica*. In the *B. rapa* genome, the number of chromosomes and paleocentromeres was reduced from 21 to 10. The chromosomes were reduced by multi-chromosome translocation, fusion, and inter-/intrachromosomal recombination. Taking chromosomes A03 and A08 as examples, A03 evolution involved six chromosomes of tPCK (Figure 4a), AK2/5, AK4, AK6, AK7, AK1, and AK2/5/6/8. The positions of the paleocentromeres are shown by large black dots. The frequency of TEs is shown by colors: green for >60% of the TE sequences, yellow for 30% of the TE sequences, and black for 0 sequences.

**Figure 3.** Distribution of TEs supporting the positions of the 21 paleocentromeres in the three tPCK subgenomes of the *B. rapa* genome. The color of each bin represents the ratio of TE sequences in the flanking region of a given gene used to reconstruct the tPCK subgenomes. The x-axis shows the position of the reconstructed chromosomes; the y-axis shows the three copies of the seven chromosomes in tPCK, which are the three subgenomes of *B. rapa*. 
AK7, AK2/5/6/8, AK3, AK6/8 and AK4 (a proposed chromosomal rearrangement process is shown in Figure 4b), whereas A08 was generated from several rounds of interchromosomal translocation of two tPCK chromosomes, AK1 and AK7 (Figure 4c). The circle model for block associations of M/N/T/U/D and V/K/L/Q/X (Figure 4b) or T/U (Figure 4c) has been used in previous reports. However, the chromosomal rearrangement explained by the circle model (Supplementary Fig. S2a) can also be achieved through an alternative process of chromosome translocation and fusion (Supplementary Fig. S2b).

**GENE EVOLUTION AFTER WGT PROPELLED THE EXPANSION OF RICH MORPHOTYPES FOR BRASSICA SPECIES**

Subgenome dominance has been detected among the three tPCK subgenomes in *B. rapa*. The subgenome dominance effect resulted in the differentiation of paralogous genes and featured the following characteristics: (i) one subgenome retained more genes than the other two through gene fractionation after WGT; (ii) genes located in the subgenome with high gene density are always expressed at higher levels than their paralogs in the other two subgenomes; and (iii) genes in the dominant subgenome accumulated fewer non-synonymous mutations than did the other subgenomes. Gene density differentiation is clearly observed when counting the number of genes within the reconstructed tPCK subgenomes: the subgenome LF has approximately 1.6 times more genes than the other two subgenomes MF1 and MF2 (the more fractionated subgenomes one and two). Using mRNA-Seq data generated for different organs of *B. rapa*, a comparison of paralogous gene pairs showed that a greater number of genes located in subgenome LF are expressed at a higher level (i.e., either those showing at least two-fold greater expression or ‘horserace’ winners) than their paralogs in the MF subgenomes. This subgenome dominance effect has also been observed in the genome of maize.

The three aspects of the dominance effect among the subgenomes of *B. rapa* are united by the rule of improving the fitness of the plant. Under this rule, genes that are expressed at higher levels than their paralogs should be more important for the biological function of the plant. Thus, functional mutations of these dominantly expressed genes would be more significant in reducing the plant’s fitness than mutations of their syntenic paralogs.
Therefore, natural selection drives the conservation of the dominantly expressed genes against functional mutations, whereas their paralogs accumulate more mutations and eventually become fractionated, resulting in a higher gene density in the dominant subgenome and lower gene density in the dominated subgenomes. This explanation was first suggested following an analysis of the maize genome and subsequently for the genome of *B. rapa*.

Short homologous sequence-mediated deletion regulates gene fractionation in *B. rapa*. By investigating the fractionated genes in the *B. rapa* genome, it was found that genes were lost individually rather than via the simultaneously deletion of many genes located in a large fragment. Short repeated sequence-mediated individual gene fractionation has been observed in maize.

First, a pair of small direct repeats appear near the gene coding region before fractionation. The small repeated sequences then form a loop for intrachromosome recombination, and the gene sequence located in the middle of the two homologous repeat sequences is deleted. This mechanism was also found to function in the process of gene fractionation in *B. rapa*.

The 24-bp small RNA-targeted TE methylation that suppressed the expression of nearby genes as well as its biased distribution among the subgenomes of *B. rapa* led to subsequent subgenome dominance. Small RNA-Seq data analysis showed that dominantly expressed genes in *B. rapa* always have fewer 24-bp RNA-targeted TEs in their 1-kb flanking regions compared with their paralogs. Previous reports on *A. thaliana* showed that small RNA-targeted TEs were subjected to methylation and the methylated TEs then suppressed nearby gene expression. All of these observations suggest that the biased distribution of small RNA-targeted TEs played an important role in the formation of the subgenome dominance effect.

WGT provided a bulk of genes that served as both the raw materials and a buffer pool for multicopy genes to evolve disparate or new functions (subfunctionalization or neofunctionalization), whereas the subgenome dominance effect facilitated this process by differentiating the multicopy genes. These newly evolved functions further promoted the evolution of rich morphotypes in *Brassica*. In *A. thaliana*, after several rounds of whole genome duplication, many duplicated genes were subfunctionalized and/or neofunctionalized. For example, in *A. thaliana*, some genes from extra duplications have subfunctionalized compared with those in *Carica papaya*, such as the enzymes *CYP79A* and *CYP79B* that catalyze the first step of glucosinolate synthesis. Some genes have neofunctionalized to develop extra biosynthetic pathways for indole and methionine-derived aliphatic glucosinolates in *A. thaliana*, which are not detected in *C. papaya*. Analysis of glucosinolate genes in the genome of *B. rapa* showed that most of these genes, multiple copies were retained after WGT. These over-retained genes would be undersubfunctionalization or neofunctionalization to develop new biological functions related to glucosinolate metabolism in *B. rapa*, as in *A. thaliana*. It is expected that there are many more such examples of other over-retained genes in *B. rapa*. The subgenome dominance effect may aid in this evolutionary process by conserving one copy of the dominant gene and letting the other copies differentiate or develop new roles. Finally, these differentiated genes will contribute to the different traits of *B. rapa*.

Biased gene retention after WGT promoted the morphotype diversification of *Brassica* plants. Phytohormones, especially auxin, play important roles in plant morphogenesis. By comparing the gene contents in *A. thaliana* and other sequenced genomes, such as *Carica papaya* and *Vitis vinifera*, it was found that auxin-related genes were expanded in the *B. rapa* genome. Furthermore, by comparing the number of gene categories that retained only one or multiple copies, genes involved in the response to phytohormone signaling were found to be significantly over-retained via gene fractionation following WGT in the genomes of *B. rapa* and *B. oleracea*.

**TWO-STEP THEORY TO ILLUSTRATE THE WGT PROCESS IN BRASSICA**

From a genome evolution perspective, a two-step theory of polyploidization was suggested to illustrate the process of WGT in *Brassica* plants. Based on the results of comparative subgenome analysis in *Brassica*, mainly in *B. rapa* as summarized above, we proposed that the WGT event occurred as two genome duplication steps (Figure 5). In the first step, the two tPCK genomes MF1 and MF2 were merged together. Subsequently, a round of genomic reshuffling and gene fractionation resulted in a new diploid. No significant genome dominance is observed between the MF1 and MF2 subgenomes of *B. rapa* now and thus autotetraploidization cannot be excluded as a possible process for the first duplication. However, based on the observation that there were a greater

![Figure 5](image-url)  
*Figure 5.* Two-step polyploidization theory for the WGT event experienced by *Brassica* plants.
CONFLICT OF INTEREST
relationship of different morphotypes in mechanism of the parallel evolution of similar traits that developed Brassica nomic important traits of in U’s triangle then generated the stable allotetraploid species based on Brassica ciation and generated an abundance of rich morphotypes of the morphotypes in each ciation, biased gene retention through gene fractionation after the speciation of diploid both the speciation and expansion of rich morphotypes for each species should be extensively studied to address the following aspects genes,46 phytohormone-related genes,47 certain transcript factor families48 and other genes,49–54 were accurately determined indi-

B. rapa genes belonging to different gene families or having important biological functions have been systematically analyzed. Based on the genome sequencing and comparative genomic study of B. rapa, the evolution of many gene families or categories such as circadian clock genes,43 resistance genes,52 stress response genes,43–45 glucosinolate genes,48 anthocyanin biosynthesis genes,46 phytohormone-related genes,47 certain transcript factor families48 and other genes,49–54, were accurately determined individually and studied in detail. Moreover, some functionally important genes related to self-incompatibility,55,56 male sterility,57 flowering regulation,56–60 leaf heading61 or color,62 have been identified or cloned, and functional studies have been conducted for some of these genes. These follow-up studies in B. rapa helped to further elucidate the evolution of specific genes after the WGT event.

CONCLUSIONS AND DISCUSSION
WGT promoted the diversification of Brassica plants with respect to both the speciation and expansion of rich morphotypes for each species. First, WGT promoted genomic reshuffling, i.e., rediploidization, to stabilize the genome and the meiosis process. Genomic reshuffling accompanied by chromosome reduction contributed the speciation of diploid Brassica plants, such as B. rapa, B. nigra and B. oleracea. Genomic differentiation of the three basic genomes in U’s triangle then generated the stable allotetraploid species B. carinata, B. napus and B. juncea. Second, subgenome differentiation, biased gene retention through gene fractionation after WGT, and further multicopy gene subfunctionalization or neofunc-
tionalization promoted the parallel evolution of many different morphotypes in each Brassica species. Therefore, WGT with subsequent genomic and gene-level evolution drove Brassica spec-
ation and generated an abundance of rich morphotypes of the Brassica species.

In the future, additional research should be conducted to invest-
igate the morphotype evolution of Brassica plants. Previous studies based on de novo genome sequencing determined the genome- and gene-level evolution of only one accession of B. rapa. Subsequently, additional B. rapa accessions or other Brassica species should be extensively studied to address the following aspects regarding the Brassica population: (i) the origins and phylogenetic relationship of different morphotypes in Brassica species; (ii) the mechanism of the parallel evolution of similar traits that developed independently in different Brassica species (e.g., the leafy head in B. rapa and B. oleracea); and (iii) the genes involved in the develop-
ment of different morphotypes or genes that regulate the agro-
nomic important traits of Brassica crops. This knowledge will increase our understanding of Brassica morphotype diversification and ultimately leverage the benefits of genomic studies for the genetic improvement of Brassica crops.

CONFICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGMENTS
This work was funded by the 973 program (2012CB113900 and 2013CB127000), the 863 program (2012AA100101) and a National Natural Science Foundation of China NSFC grant (31301771). Research was conducted at the Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture, P. R. China, and the Sino-Dutch Joint Lab of Horticultural Genomics Technology.

REFERENCES
1 Warwick SI, Francis A, Al-Shehbaz IA. Brassicaceae: species checklist and database on CD-Rom. Plant Syst Evol 2006; 259: 249–258.
2 Warwick SI, Mummenhoff K, Sauer CA et al. Closing the gaps: phylogenetic relationships in the Brassicaceae based on DNA sequence data of nuclear ribosomal ITS region. Plant Syst Evol 2010; 285: 209–232.
3 Wang X, Wang H, Wang J et al. The genome of the mesopolyploid crop species Brassica rapa. Nat Genet 2011; 43: 1035–1039.
4 Beilstein MA, Nagalingum NS, Clements MD et al. Dated molecular phylogenies indicate a Miocene origin for Arabidopsis thaliana. Proc Natl Acad Sci USA 2010; 107: 18724–18728.
5 Lukens LN, Quijada PA, Udall J et al. Phylogenomic redundancy and plasticity within ancient and recent Brassica crop species. Biol J Linnean Soc 2004; 82: 665–674.
6 Lysak MA, Koch MA, Pecinka A et al. Chromosomal triplication found across the tribe Brassicaceae. Genome Res 2005; 15: 516–525.
7 Arias T, Beilstein MA, Tang M et al. Diversification times among Brassica (Brassicaceae) crops suggest hybrid formation after 20 million years of divergence. Am J Bot 2014; 101: 86–91.
8 Nagaharu U. Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. Jpn J Bot 1935; 7: 389–452.
9 Morinaga T. Interspecific hybridization in Brassica I. The cytology of F1 hybrids of B. napelle and various other species with 10 chromosomes. Cytologia 1929; 1: 16–27.
10 Zhao J, Wang X, Deng B et al. Genetic relationships within Brassica rapa as inferred from AFLP fingerprints. Theor Appl Genet 2005; 110: 1301–1314.
11 Parkin IA, Gulden SM, Sharpe AG et al. Segmental structure of the Brassica napus genome based on comparative analysis with Arabidopsis thaliana. Genetics 2005; 171: 765–781.
12 Schranz ME, Lysak MA, Mitchell-Olids T. The ABC’s of comparative genetics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci 2006; 11: 535–542.
13 Mandakova T, Lysak MA. Chromosomal phylogeny and karyotype evolution in x = 7 crucifer species (Brassicaceae). Plant Cell 2008; 20: 2559–2570.
14 Lysak MA, Berr A, Pecinka A et al. Mechanisms of chromosome number reduction in Arabidopsis thaliana and related Brassicaceae. Proc Natl Acad Sci USA 2006; 103: 5224–5229.
15 Koch MA, Kiefer M. Genome evolution among cruciferous plants: a lecture from the comparison of the genetic maps of three diploid species—Capsella rubella, Arabidopsis lyrata subsp. petraea, and A. thaliana. Am J Bot 2005; 92: 761–767.
16 Hu TT, Pattyn P, Bakker EG et al. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat Genet 2011; 43: 476–481.
17 Slotte T, Hazzouri KM, Agren JA et al. The Capsella rubella genome and the genomic consequences of rapid mating system evolution. Nat Genet 2013; 45: 831–835.
18 Dassanayake M, Oh DH, Haas JS et al. The genome of the extremophile crucifer Thellungiella parvula. Nat Genet 2011; 43: 913–918.
19 Liu S, Liu Y, Yang X et al. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nature Communication 2014; in press.
20 Cheng F, Liu S, Wu J et al. BRAD, the genetics and genomics database for Brassica plants. BMC Plant Biol 2011; 11: 136.
21 Cheng F, Wu J, Fang L et al. Biased gene fractionation and dominant gene expression among the subgenomes of Brassica rapa. PLoS ONE 2012; 7: e36442.
22 Cheng F, Wu J, Fang L et al. Syntenic gene analysis between Brassica rapa and other Brassicaceae species. Front Plant Sci 2012; 3: 198.
23 Panjabi P, Jagannath A, Bishc NC et al. Comparative mapping of Brassica juncea and Arabidopsis thaliana using Intron Polymorphism (IP) markers: homeologous relationships, diversification and evolution of the A, B and C Brassica genomes. BMC Genomics 2008; 9: 113.
24 Cheng F, Mandakova T, Wu J et al. Deciphering the diploid ancestral genome of the Mesohexaploid Brassica rapa. Plant Cell 2013; 25: 1541–1554.
25 Initiative AG. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 2000; 408: 796–815.
26 Schnable PS, Ware D, Fulton RS et al. The B73 maize genome: complexity, diversity, and dynamics. Science 2009; 326: 1112–1115.
27 Schmutz J, Cannon SB, Schlueter J et al. Genome sequence of the palaeopolyploid soybean. Nature 2010; 463: 178–183.
28 Mun JH, Kwon SJ, Seol YJ et al. Sequence and structure of Brassica rapa chromosome A3. Genome Biol 2010; 11: R94.
29 Trick M, Kwon SJ, Choi SR et al. Complexity of genome evolution by segmental rearrangement in Brassica rapa revealed by sequence-level analysis. BMC Genomics 2009; 10: 539.
Genome triplication drove *Brassica* diversification

F Cheng et al

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

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

32 Schnable JC, Wang X, Pires JC et al. Escape from preferential retention following repeated whole genome duplications in plants. Front Plant Sci 2012; 3: 94.

33 Woodhouse MR, Cheng F, Pires JC, Lisch D, Freeling M, Wang X. Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. Proc Natl Acad Sci USA 2014 Apr 8; 111(14): 5283–5288.

34 Hollister JD, Smith LM, Guo YL et al. Glucosinolate biosynthetic genes in *Brassica* rapa. Gene 2011; 477: 135–142.

35 Santner A, Estelle M. Recent advances and emerging trends in plant hormone signalling. Nature 2009; 459: 1071–1078.

36 Woodhouse MR, Schnable JC, Pedersen BS et al. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 2009; 19: 1419–1428.

37 Bekaert M, Edger PP, Hudson CM, Pires JC et al. Metabolic and evolutionary costs of herbivory defense: systems biology of glucosinolate synthesis. New Phytol 2012; 196: 596–605.

38 Wang H, Wu J, Sun S et al. Glucosinolate biosynthetic genes in *Brassica* rapa. Gene 2011; 478: 135–142.

39 Hu L, Li S, Gao W. Expression, divergence and evolution of the calceolin gene family in *Brassica* rapa. Arch Biol Sci 2013; 65: 863–876.

40 Liu Z, Zhang M, Kong L et al. Genome-wide identification, phylogeny, duplication, and expression analyses of two-component system genes in Chinese cabbage (*Brassica rapa* ssp. pekinensis). DNA Res; e-pub ahead of print 27 February 2014; doi:10.1093/dnares/dsu004.

41 Peng FY, Weselake RJ. Genome-wide identification and analysis of the B3 superfamily of transcription factors in *Brassica*ceae and major crop plants. Theor Appl Genet 2013; 126: 1305–1319.

42 Wu P, Shao ZQ, Wu XZ. Loss/retention and evolution of NBS-encoding genes upon whole genome triplication of *Brassica* rapa ssp. pekinensis. Mol Genet Genomics 2013; 286: 939–951.

43 Xia D, Zhao JJ, Hou XL et al. The *Brassica* rapa FLC homologue FLC2 is a key regulator of flowering time, identified through transcriptional co-expression networks. J Exp Bot 2013; 64: 4503–4516.

44 Lee SC, Lim MH, Yu JG et al. Genome-wide characterization of the CBF/DREB1 gene family in *Brassica* rapa. Plant Physiol Biochem 2012; 54: 179–191.

45 Ma J, Wang F, Li M et al. Genome-wide analysis of the NAC transcription factor family in Chinese cabbage to elucidate responses to temperature stress. Sci Hort 2014; 165: 82–90.

46 Zhu B, Wang Y, Zhan Y et al. Chalcone synthase family genes have redundant roles in anthocyanin biosynthesis and in response to blue/UV-A light in turnip (*Brassica rapa*; *Brassicaeae*). Am J Bot 2013; 100: 2458–2467.

47 Liu Z, Kong L, Zhang M et al. Genome-wide identification, phylogeny, evolution and expression patterns of AP2/ERF genes and cytokinin response factors in *Brassica rapa* ssp. pekinensis. PLoS ONE 2014; 8: e83444.

48 Peng FY, Weselake RJ. Genome-wide identification and analysis of the B3 superfamily of transcription factors in *Brassicaeae* and major crop plants. Theor Appl Genet 2013; 126: 1305–1319.

49 Song XM, Liu TK, Duan WK et al. Genome-wide analysis of the GRAS gene family in Chinese cabbage (*Brassica rapa* ssp. pekinensis). Genomics 2013; 103: 135–146.

50 Sampaio P, Lee SC, Lee J et al. Characterization of a new high copy Stowaway family MITE, BRAM1-1 in *Brassica* genome. BMC Plant Biol 2013; 13: 56.

51 Liu Z, Zhang M, Kong L et al. Genome-wide identification, phylogeny, duplication, and expression analyses of two-component system genes in Chinese cabbage (*Brassica rapa* ssp. pekinensis). DNA Res; e-pub ahead of print 27 February 2014; doi:10.1093/dnares/dsu004.

52 Song X, Liu G, Duan W et al. Genome-wide identification, classification and expression analysis of the heat shock transcription factor family in Chinese cabbage. Mol Genet Genomics e-pub ahead of print 8 March 2014; doi:10.1007/s00438-014-0833-5.

53 Hu L, Li S, Gao W. Expression, divergence and evolution of the calceolin gene family in *Brassica* rapa. Arch Biol Sci 2013; 65: 863–876.

54 Hu L, Yin W, Chen Y et al. Functional divergence and evolutionary dynamics of the putative AAAP gene family in *Brassica* rapa. Plant Mol Biol Rep 2014; 32: 517–530.

55 Indriolo E, Tharamapalan P, Wright SI et al. The ARC1 E3 ligase gene is frequently deleted in self-compatible Brassicaceae species and has a conserved role in *Arabidopsis lyrata* self-pollen rejection. Plant Cell 2012; 24: 4607–4620.

56 Lao X, Suwabe K, Niikura S et al. Physiological and genetic analysis of CO2-induced breakdown of self-incompatibility in *Brassica* rapa. J Exp Bot 2014; 65: 939–951.

57 Xu X, Sun X, Zhang J et al. Identification of candidate genes associated with male sterility in CMS7311 of heading Chinese cabbage (*Brassica campestris* L. ssp. pekinensis). Acta Physiol Plant 2013; 35: 3265–3270.

58 Imminik RG, Pose D, Ferrario S et al. Characterization of SOC1’s central role in flowering by the identification of its upstream and downstream regulators. Plant Physiol 2012; 160: 433–449.

59 Xia D, Zhao JJ, Hou XL et al. The *Brassica* rapa FLC homologue FLC2 is a key regulator of flowering time, identified through transcriptional co-expression networks. J Exp Bot 2013; 64: 4503–4516.

60 Wu J, Wei K, Cheng F et al. A naturally occurring Indel variation in BraFLC2 (BrFLC2) associated with flowering time variation in *Brassica* rapa. BMC Plant Biol 2012; 12: 151.

61 Liu Z, Jia L, Wang H et al. HYL1 regulates the balance between adaxial and abaxial identity for leaf flattening via miRNA-mediated pathways. J Exp Bot 2011; 62: 4367–4381.

62 Lee S, Lee SC, Byun DH et al. Association of molecular markers derived from the BrCristo1 gene with prolycopeno-enriched orange-colored leaves in *Brassica* rapa. Theor Appl Genet 2014; 127: 179–191.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/

Supplemental Information for this article can be found on the *Horticulture Research* website (http://www.nature.com/hortres).