Neofunctionalization of Duplicated Tic40 Genes Caused a Gain-of-Function Variation Related to Male Fertility in Brassica oleracea Lineages

Xiaoling Dun, Wenhao Shen, Kaining Hu, Zhengfu Zhou, Shengqian Xia, Jing Wen, Bin Yi, Jinxiong Shen, Chaozhi Ma, Jinxing Tu, Tingdong Fu, and Ulf Lagercrantz

National Key Laboratory of Crop Genetic Improvement, National Center of Rapeseed Improvement, Huazhong Agricultural University, Wuhan 430070, China (X.D., W.S., K.H., Z.Z., S.X., J.W., B.Y., J.S., C.M., J.T., T.F.); and Department of Ecology and Genetics, Evolutionary Biology Center, Uppsala University, Uppsala SE-75236, Sweden (U.L.)

Gene duplication followed by functional divergence in the event of polyploidization is a major contributor to evolutionary novelties. The Brassica genus evolved from a common ancestor after whole-genome triplication. Here, we studied the evolutionary and functional features of Brassica spp. homologs to Tic40 (for translocon at the inner membrane of chloroplasts with 40 kDa). Four Tic40 loci were identified in allotetraploid Brassica napus and two loci in each of three basic diploid Brassica spp. Although these Tic40 homologs share high sequence identities and similar expression patterns, they exhibit altered functional features. Complementation assays conducted on Arabidopsis thaliana tic40 and the B. napus male-sterile line 7365A suggested that all Brassica spp. Tic40 homologs retain an ancestral function similar to that of AtTic40, whereas BolC9.Tic40 in Brassica oleracea and its ortholog in B. napus, BnaC9.Tic40, in addition, evolved a novel function that can rescue the fertility of 7365A. A homologous chromosomal rearrangement placed bnaC9.tic40 originating from the A genome (BnaA10.Tic40) as an allele of BnaC9.Tic40 in the C genome, resulting in phenotypic variation for male sterility in the B. napus near-isogenic two-type line 7365AB. Assessment of the complementation activity of chimeric B. napus Tic40 domain-swapping constructs in 7365A suggested that amino acid replacements in the carboxyl terminus of BnaC9.Tic40 cause this functional divergence. The distribution of these amino acid replacements in 59 diverse Brassica spp. accessions demonstrated that the neofunctionalization of Tic40 is restricted to B. oleracea and its derivatives and thus occurred after the divergence of the Brassica spp. A, B, and C genomes.

Polyploidy or whole-genome duplication is thought to be a prominent evolutionary force in eukaryotes (Wolfe, 2001; Udall and Wendel, 2006), especially for flowering plants (Blanc and Wolfe, 2004; Van de Peer et al., 2009). Almost 95% of angiosperms show evidence of having undergone at least one round of whole-genome duplication in their evolutionary history, suggesting that most extant diploid flowering plants have evolved from ancient polyploids (Cui et al., 2006; Soltis et al., 2009). Gene duplications in the event of polyploidization provide sources for evolutionary novelties that could benefit plants (Lukens et al., 2004; Chen, 2007). Divergence after gene duplication could result in three primary evolutionary fates of duplicated genes: pseudogenization, neofunctionalization, and subfunctionalization (Force et al., 1999; Conant and Wolfe, 2008; Liu and Adams, 2010). Pseudogenization implies that duplicated genes with redundant functions lose their function by accumulating negative mutations; neofunctionalization denotes that the redundant gene evolves a new adaptive function; while subfunctionalization causes the duplicated genes to adopt a different part of the function of an ancestral gene (Rodríguez-Trelles et al., 2003; Flagel and Wendel, 2009; Liu and Adams, 2010).

The Brassica genus consists of three basic diploid species: Brassica rapa (AA; n = 10), Brassica nigra (BB; n = 8), and Brassica oleracea (CC; n = 9), and their derivative allotetraploid species: Brassica juncea (AABB; n = 18), Brassica napus (AACC; n = 19), and Brassica carinata (BBCC; n = 17; Beilstein et al., 2006). Comparative genetic mapping demonstrated that these Brassica spp., which diverged from Arabidopsis thaliana approximately 20 to 40 million years ago (Lagercrantz and Lydiate, 1996; Blanc et al., 2003; Town et al., 2006), descended from a common ancestor after whole-genome triplication (Parkin et al., 2002). Collinear comparison showed that for each of 24 ancestral genomic blocks defined in the ancestral karyotype in A. thaliana, three syntenic copies were identified in each of the diploid Brassica spp. genomes, with only one exception (Schranz et al., 2006; Wang et al., 2011; Cheng et al., 2013).

1 This work was supported by the National Natural Science Foundation of China (grant no. 31130040) and the High-Tech Research and Development Program 863 (grant no. 2011AA10A104).
2 Present address: Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China.
3 Present address: Wheat Research Institute, Henan Academy of Agricultural Science, Zhengzhou, Henan 450002, China.
* Address correspondence to tujx@mail.hzau.edu.cn.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Jinxing Tu (tujx@mail.hzau.edu.cn).

[OPEN] [W] The online version of this article contains Web-only data. [OPEN] [W] Articles can be viewed online without a subscription. www.plantphysiol.org/cgi/doi/10.1104/pp.114.246470
Fractionation (gene loss from homologous genomic regions) and chromosomal rearrangements were prevalent in the diploidization process of the hexaploid Brassica spp. common ancestor (Lagercrantz, 1998; Town et al., 2006; Ziolkowski et al., 2006; Mun et al., 2009). Based on gene density differences caused by varying gene loss rates in the three collinear genomic block copies, these genomic blocks were classified into three subgenomes in the diploid Brassica spp. genomes: the least fractionated (LF), the medium fractionated (MF1), and the most fractionated (MF2) subgenomes (Wang et al., 2011; Tang and Lyons, 2012; Cheng et al., 2013). The different rates of gene loss of the three subgenomes support a two-step origin of the Brassicaceae ancestral genome involving a tetraploidization process followed by substantial fractionation of the subgenomes MF1 and MF2 and more recently hybridization with the third subgenome LF to form a hexaploid (Wang et al., 2011; Tang et al., 2012). Although collinearity and changes in genomic structure, including duplications, deletions, and rearrangements of the Brassica spp. and A. thaliana Tic40 homologs in several Brassicaceae spp. are well studied, there is limited knowledge of the molecular and functional divergence of duplicated or homologous genes in the Brassica spp.

To utilize heterosis in B. napus breeding, hybrid production is based mainly on male sterility. Currently, the recessive epistatic genic male-sterile three-type line system 7365ABC is widely used for oilseed heterosis due to its advantages, producing 100% sterile offspring for realizing the triple-cross hybrid (Huang et al., 2012). Although collinearity and changes in genomic structure, including duplications, deletions, and rearrangements of the Brassica spp. and A. thaliana Tic40 are well studied, there is limited knowledge of the molecular and functional divergence of duplicated or homologous genes in the Brassica spp.

In this study, we identified and characterized Tic40 homologs in B. napus and three basic diploid Brassica spp. We suggested that neo-functionalization of Bnac9.Tic40 after the divergence of the Brassica spp. A, B, and C genomes was caused by amino acid replacements in the C terminus of BnaC9.Tic40. In addition, we validated that the allelic genes Bnac9.Tic40 (equivalent to BnaC.Tic40 as described by Dun et al. [2011]) and bna9.9.tic40 originate from the C and A genomes, respectively, and became allelic due to a homologous chromosomal rearrangement in B. napus. These results provide further knowledge for the effective utilization of the restoring gene of 7365A and a better insight into the functional divergence of homologous duplicated genes in paleopoloid Brassica spp.
and *T. halophila*, and three syntenic regions were identified in the genomes of *B. rapa* and *B. oleracea* (Fig. 2A). Notably, the genomes of *B. rapa* and *B. oleracea* each harbor two Tic40 loci, and no Tic40 homologs are present at the proposed syntenic regions of *B. rapa* A3 and *B. oleracea* C3 (Fig. 2A), which is consistent with the results presented above. Interestingly, gene deletions were prevalent in the three subgenomes of *Brassica* spp., as evident from the absence of several prospective homologous genes (Fig. 2A).

Tic40 genomic regions orthologous to the *A. thaliana* Tic40 genomic region (from At5g16550 to At5g16840)

![Amplification profiles of Tic40 genes from three diploid Brassica spp.](image)

**Figure 1.** Amplification profiles of Tic40 genes from three diploid *Brassica* spp. Seven primer combinations, 3L/5R, 6L/8R, 8L/10R, 10L/12R, 12L/14R, 8L/14R1, and 6L/14R1, were used to amplify DNA from genotypes of *B. oleracea* (T14, CGN06903, and CGN18458, T9), *B. rapa* (R7, CGN06832, CGN13925, and Chifu), *B. nigra* (CGN06620, CGN06625, and CGN06630), and *A. thaliana* (Columbia). For 6L/14R1, the amplification profile is not shown because of the low resolution for large amplified fragments (about 1,500 bp).

**Table 1. Identification of Tic40 loci in three diploid Brassica spp.**

| Genotype | Taxonomic Group | No. of Polymorphism Fragments | Sequence Identity between Polymorphism Fragments | No. of Tic40 Loci<sup>a</sup> |
|----------|----------------|-------------------------------|--------------------------------------------------|-------------------------------|
| T14      | *B. oleracea* group Borecole | 2                             | 74%                                              | 2                             |
| CGN06903 | Wild-type *B. oleracea*      | 3                             | 98%, 75%, 75%                                   | 2                             |
| CGN18458 | *B. oleracea* group Kohlrabi | 3                             | 95%, 73%, 75%                                   | 2                             |
| T9       | Wild-type *B. oleracea*      | 3                             | 96%, 74%, 75%                                   | 2                             |
| R7       | *B. rapa* group Broccoleto   | 2                             | 74%                                              | 2                             |
| CGN06832 | Wild-type *B. rapa* group    | 2                             | 93%, 76%, 77%                                   | 2                             |
| CGN13925 | *B. rapa* group Pak Choy     | 3                             | 96%, 74%, 75%                                   | 2                             |
| Chifu    | *B. rapa* group Chinese Cabbage | 2                 | 74%                                              | 2                             |
| CGN06620 | *B. nigra* var. Abyssinica   | 3                             | 96%, 70%, 70%                                   | 2                             |
| CGN06625 | *B. nigra* var. Abyssinica   | 4                             | 99%, 99%, 70%, 69%, 70%, 70%                    | 2                             |
| CGN06630 | *B. nigra* var. Abyssinica   | 2                             | 70%                                              | 2                             |

<sup>a</sup>Fragments from within a diploid *Brassica* spp. with sequence identity over 90% were considered alleles at a Tic40 locus. <sup>b</sup>Deduced number of loci based on number of polymorphic fragments and sequence identity.
in each genome were aligned using BLASTZ. The resulting VISTA plot (Fig. 2B) reveals a high degree of conservation among genomic regions, although a noncontinuous pattern of conserved subsets of genes are present in *Brassica* spp. Tic40 genomic regions due to a number of deletion events. The regions from *A. lyrata*, *C. rubella*, and *T. halophila* showed high similarity to that of *A. thaliana*. The apparent absence of homologs of At5g16580 and At5g16790 from all *Brassica* spp. as well as *T. halophila* might be the result of insertions in the *A. thaliana* lineage after the separation of the *A. thaliana* lineage and the one leading to the genera *Brassica* and *Thellungiella*. The degree of conservation between *A. thaliana* and the different *Brassica* spp. was considerably lower. Out of the other 29 gene models in the region, only five genes (17.24%) are represented in all six *Brassica* spp. paralogous regions; 14 genes (48.28%, including Tic40) have conserved gene models in four *Brassica* spp. paralogous regions, and 10 genes (34.48%) have one homolog in the diploid

**Figure 2.** Synteny of Tic40 genomic regions in Brassicaceae genomes. A, Alignment of homologous Tic40 regions from six Brassicaceae genomes. The loci depicted by black circles are Tic40 homologs, and ellipses represent homologs of Tic40 flanking genes. White circles and ellipses represent the absence of prospective homologous genes. The star indicates that the transcript of a corresponding homolog was absent in the data but the homologous genomic sequence was present. Solid lines connect homologs, and broken lines were used when at least one homolog was absent. B, Genomic multiple alignment of 10 Brassicaceae Tic40 genomic regions aligned using BLASTZ and visualized with a VISTA plot. The names of the 31 gene models in this *A. thaliana* region (from At5g16550 to At5g16840) are depicted with different colors to indicate the number of homologs detected in the two diploid *Brassica* spp. genomes: red, three homologs in each *Brassica* spp. genome; blue, two homologs; green, one homolog; and black, loss of homologs in these *Brassica* spp. genomes. Conserved regions in the similarity plots are colored to illustrate annotation of the region. At, *A. thaliana*; Aly, *A. lyrata*; Cru, *C. rubella*; The, *T. halophila*; Bra, *B. rapa*; Bol, *B. oleracea*; UTR, untranslated region.
*Brassica* spp. genomes (Fig. 2B). These results suggest numerous deletion events in the diploidization process of the *Brassica* spp. ancestor. Furthermore, in agreement with a previous report (Cheng et al., 2013), deletion events in the Tic40 genomic regions from *B. rapa* A10 and *B. oleracea* C9 subgenome LF were much fewer than those from *B. rapa* A02 and *B. oleracea* C02 subgenome MF2 and the *B. rapa* A03 and *B. oleracea* C03 subgenome MF1 (Fig. 2B). In summary, these analyses demonstrate that duplication and deletion of Tic40 homologs in the *Brassica* spp. genomes are a result of genome triplication following diploidization in the *Brassica* spp. common ancestor.

According to their location in the *Brassica* spp. chromosomes, the identified Tic40 genes were named *BraA2.Tic40* and *BraA10.Tic40* in *B. rapa*, *BolC2.Tic40* and *BolC9.Tic40* in *B. oleracea*, *BnaB.Tic40a* and *BniB.Tic40b* in *B. nigra*, and *BnaA2.Tic40*, *BnaA10.Tic40*, *BnaC2.Tic40*, and *BnaC9.Tic40* (equivalent to the fertility-recovering gene *BnaC.Tic40*) in *B. napus*. The allele of *BnaC9.Tic40* in the male-sterile line 7365A was named *bnac9.tic40*.

### Evolutionary Analysis of *Brassica* spp. Tic40 Homologs

In *B. napus*, the A and C genomes were considered to be homologous subgenomes due to their evolutionary origins. Because of this complex genome background, Tic40 homologs in *B. napus* were classified into paralogs (duplicated genes in the same subgenome) and homologs (orthologous genes in different subgenomes).

The full-length cDNAs and genomic sequences of the identified *Brassica* spp. Tic40 genes were isolated and compared with other Brassicaceae Tic40 genes. The phylogenetic tree constructed using the full-length Tic40 open reading frames (ORFs) revealed that these Brassicaceae Tic40 genes fell into three groups: *AtTic40*, *AlyTic40*, and *CruTic40* in group 1; *BnaC9.Tic40*, *BnaA10.Tic40*, *BolC9.Tic40*, *BnaC9.tic40*, and *BniB.Tic40a* in group 2; and *ThaTic40*, *BnaA2.Tic40*, *BnaC2.Tic40*, *BolC2.Tic40*, *BraA2.Tic40*, and *BniB.Tic40b* in group 3, which is consistent with the hypothesis stating a duplication within the *Brassica* spp. ancestor after the divergence from the lineage leading to *A. thaliana* (Fig. 3). With the exception of *bnac9.tic40*, each Tic40 gene existing in *B. napus* clustered with its corresponding diploid progenitor ortholog (Fig. 3). Sequence identities shared between *Brassica* spp. Tic40 orthologous gene pairs (98%–100%; e.g. *BnaC9.Tic40* and *BolC9.Tic40*) were higher than those displayed between homologous gene pairs (92%–96%; e.g. *BnaC9.Tic40* and *BnaA10.Tic40*; Supplemental Table S2). Interestingly, *bnac9.tic40* showed only 94% sequence identity with *BolC9.Tic40*, less than that with *BnaA10.Tic40* (98%) and *Bra10.Tic40* (99%), while its allele *BnaC9.Tic40* displayed 100% sequence identity with *BolC9.Tic40* (Supplemental Table S2). Phylogenetic tree analysis also suggested that *bnac9.tic40* was evolutionarily closer to *BraA10.Tic40* than *BolC9.Tic40* (Fig. 3). These data support the conclusion that *BnaC9.Tic40* originated from *BolC9.Tic40*, but its allele *bnac9.tic40* was derived from *BraA10.Tic40*, most likely due to homologous recombination between the A and C genomes (see below).

### Homologous Chromosomal Rearrangement around BnaC9.Tic40 in *B. napus*

Due to high synteny between the A and C genomes in *B. napus*, chromosomal rearrangements caused by homologous recombination events between these two related genomes are not uncommon (Udall et al., 2005). We hypothesized that the genes *BnaC9.Tic40* and *bnac9.tic40*, originating from the C and A genomes, respectively, became allelic on N19 as the result of an earlier chromosomal rearrangement between homologous Tic40 genomic regions of linkage groups N10 and N19 in *B. napus*. To confirm this deduction, we developed 23 intron polymorphism (IP) molecular markers using information of flanking genes spanning an approximately 3.5-Mb region around *BolC9.Tic40*. These 23 IP markers (from IPJH1 to IPJH24, without IPJH12 because the gene locus 12 is condescending to Tic40 locus) detected polymorphic fragments between the highly homoygous genotypes *B. rapa* Chiffu and *B. oleracea* T9 (Fig. 4A), and 18 of them (from IPJH14 to IPJH22) detected segregating loci for which the two homozygous *B. napus* genotypes 7365A and 7365B had

![Image](image-url)
different allelic fragments (Fig. 4A). In these cases, 7365A had the polymorphic fragment coincident to that detected by the same marker in B. rapa Chiffu and 7365B had the other polymorphic fragment corresponding to that in B. oleracea T9 (Fig. 4A). For other IP markers, IPJH1, IPJH2, IPJH3, IPJH23, and IPJH24, 7365A and 7365B had the fragments from both Chiffu and T9 (Fig. 4A). The most likely explanation for this pattern is that the genomic region from IPJH4 to IPJH24 on N19 in 7365A is derived from B. rapa (see outline of linkage groups in Fig. 4A).

Furthermore, partial sequences (300–800 bp) of 23 flanking genes around BnaC9.Tic40 and bnac9.tic40 were sequenced, and the resulting sequences were aligned with their orthologs in B. rapa A10 and B. oleracea C09. Based on these alignments, we constructed phylogenetic trees of homologs of these flanking genes and compared the sequence divergence between these homologs. As shown in Figure 3B, both the phylogenetic trees and sequence divergence analysis suggested that flanking genes 1, 2, 3, 23, and 24 of 7365A and all analyzed flanking genes of 7365B were evolutionarily closer to their homologs from B. oleracea C09, whereas flanking gene models 4 to 22 of 7365A were more similar to their homologs from B. rapa A10 rather than B. oleracea C09. These data showed that an approximately 2-Mb fragment (from gene models 4 to 22) surrounding bnac9.tic40 is derived from B. rapa A10 (Fig. 4B). These results collectively support our hypothesis that the different origins of the allelic genes BnaC9.Tic40 and bnac9.tic40 were due to a homologous chromosomal rearrangement around the Tic40 locus in B. napus N19. Thus, BnaC9.Tic40 and bnac9.tic40 are evolutionary homologs, although they are allelic in B. napus.

BolC9.Tic40 and BnaC9.Tic40 Gained a Novel Function Related to Male Fertility

BnaC9.Tic40 has been reported as a fertility-restorer gene of the B. napus male-sterile line 7365A, the sterility of which is controlled by two genes, Bnms3 (bnac9.tic40) and BnRf, whereas loss of Tic40 in A. thaliana causes a chlorotic phenotype throughout development but does not affect male fertility (Chou et al., 2003; Dun et al., 2011). The A. thaliana tic40 mutant and the B. napus male-sterile 7365A mutant provide good systems for studying the functional features of Tic40 family members. To pursue a better understanding of the functional characteristics of Brassica spp. Tic40 genes, genomic fragments of BnaC9.Tic40, bnac9.tic40, BnaA10.Tic40, BnaA2.Tic40, BnaA2.Tic40, six diploid Brassica spp. Tic40 genes, and Attic40, including approximately 2,000 bp of the putative upstream promoter region and 700 bp of the downstream region, were cloned into the pCAMBIA1305.1 binary vector. The 12 resulting constructs were transformed into both the A. thaliana attic40 mutant and the B. napus 7365A male-sterile line. Genetic complementation assays suggested that all Tic40 genes could restore the attic40 mutant to the wild type (Table II). We carefully investigated the phenotypes of more than 10 T0 transgenic B. napus 7365A plants for each Tic40 construct at the flowering stage. For BolC9.Tic40 and BnaC9.Tic40, eight out of 11 and 11 out of 13 transgenic lines displayed recovered fertility when grown to maturity, respectively (Table II). However, for all other Tic40 constructs, no 7365A transgenic plants exhibited a fertility-recovered phenotype (Table II). These results indicated that bnac9.tic40 is not a loss-of-function mutant but that the dominant gene BnaC9.Tic40 and its ortholog BolC9.Tic40 have gained a novel function relating to male fertility.

Expression Profiles of Tic40 Homologs in B. napus, B. rapa, and B. oleracea

To determine whether the expression patterns of Brassica spp. Tic40 homologs have diverged, we compared the expression patterns of bnac9.tic40, BnaC9.Tic40, BnaC2.Tic40, BnaA2.Tic40, and BnaA10.Tic40 in B. napus, BnaA2.Tic40 and BnaA10.Tic40 in B. rapa, and BolC2.Tic40 and BolC9.Tic40 in B. oleracea by real-time PCR analysis. The data revealed that these Tic40 homologs are expressed in various organs, such as roots, seedlings, stems, young siliques, and inflorescences (Fig. 5A). Furthermore, expression patterns were similar for all Tic40 homologs throughout development (Fig. 5A).

We also analyzed five B. napus Tic40 homolog promoter activities in anther tissues by constructing GUS reporter gene fusion vectors. Promoter fragments of about 2,000 bp from five B. napus Tic40 genes were inserted into the pBI201-GUS vector. These fusion constructs were introduced into wild-type A. thaliana by Agrobacterium tumefaciens-mediated transformation, and GUS histochemical assays were performed on inflorescences of A. thaliana transgenic plants. Analysis of transformed plants showed that transcriptional activities of these Tic40 genes were high in the anthers of young buds, and those spatial and temporal expression patterns of the five B. napus Tic40 homologs were similar during anther development (Fig. 5, B and C). These results suggested that the novel function related to male fertility probably arose through altered protein function rather than through changes in gene expression patterns.

Functional Divergence of Brassica spp. Tic40 Homologs Resulted from Amino Acid Sequence Differences in the C Terminus

Analysis of the identified Brassicaceae Tic40 deduced amino acid sequences revealed the presence of several functional domains, a predicted chloroplast/plastid targeting transit peptide, and a transmembrane domain at their N-terminal region, mediating a protein-protein interaction domain Tetratricopeptid...
Figure 4. Analysis of the subgenome origin of genomic regions around BnaC9.Tic40 and bnac9.tic40 in B. napus. A, Segregation of IP molecular markers on linkage group N19 of B. napus 7365A and 7365B. Twenty-three IP molecular markers (IPJH1–IPJH24) corresponded to 23 loci flanking Tic40 in B. rapa A10 and B. oleracea C9 (Supplemental Table S3). For IPJH1, IPJH14, IPJH15, and IPJH23, amplification patterns are shown on the left. For each molecular marker, numbers were assigned to polymorphic fragments: −1 if they were identical to those from B. rapa Chiffu and +1 if they were identical to those from B. oleracea T9. The presence of these fragments in B. napus 7365A and 7365B is shown in the center table, and the result is visualized on the right, with white ovals representing C genome alleles and gray ovals representing A genome alleles. For gene loci IPJH1, IPJH2, IPJH3, IPJH23, and IPJH24, 7365A and 7365B displayed the expected pattern with fragments from both Chiffu and T9 (e.g. IPJH1 and IPJH23). However, for loci IPJH4 to IPJH22, B. rapa-specific fragments were present in 7365A; conversely, for IPJH13 and IPJH14, B. oleracea-specific fragments were present in 7365B. B, Sequence divergence of homologs of 23 flanking genes. The approximate genomic locations of the flanking genes at B. oleracea C9 are indicated by ovals (bottom right). In both 7365A and 7365B, partial sequences (300–800 bp) of 23 flanking genes were aligned with their orthologs in B. rapa A10 and B. oleracea C09. Phylogenetic trees of homologs of several flanking genes constructed by MEGA4.0 are shown on the left. Sequence divergences between these homologs are shown on
repeats (TPR) and a Sti1p/Hop/Hip (Hop, known as Hsp70 and Hsp90 organizing protein, and also known as Sti1p; Hip, Hsp70 interacting protein) domain known to bind to the ATPase domain of Hsp93 at the C-terminal region (Fig. 6A; Chou et al., 2003). The predicted Brassicaceae protein sequences share high sequence similarity (over 81%). Further analysis of the predicted protein sequences revealed 19 unique amino acid replacements in the gain-of-function Tic40 proteins as compared with other Brassicaceae Tic40 homologs. Four of these were found before the transmembrane domain, three were found between the transmembrane domain and the TPR domain, 10 were found inside the TPR domain, and the rest were found inside the Sti1p/Hop/Hip domain (Fig. 6A). This high degree of variation makes it difficult to confirm which mutation is responsible for the fertility-recovered phenotype. To get a better insight into the functional importance of different parts of BnaC9.Tic40, four binary constructs were made using the promoters and the full-length cDNAs of BnaC9.Tic40 and bnaC9.tic40. BΔN was constructed using the promoter and the cDNA of BnaC9.Tic40, but with the N terminus before the TPR domain substituted by the corresponding fragment from bnaC9.tic40. Conversely, ΔΔN included the promoter of bnaC9.tic40, the N terminus of BnaC9.Tic40, and the C terminus (TPR and Hop domains) of bnaC9.tic40. Likewise, in the BΔC and AΔC constructs, the C termini of BnaC9.Tic40 and bnaC9.tic40 were replaced by each other.

These constructs were used to transform 7365A plants. More than 21 transformants were identified for each of the four constructs. The constructs BΔN and AΔC both complemented the 7365A plants, restoring a wild-type-like appearance with visible pollen grains in the anthers (Fig. 6D). By contrast, none of the ΔΔN and BΔC transformants resulted in a complemented phenotype (Fig. 6D). These results suggested that amino acid substitutions present in the C terminus of BolC9.Tic40 and BnaC9.Tic40, comprising the TPR and Sti1p/Hop/Hip domains, caused the observed functional divergence of Brassica spp. Tic40 homologs.

In addition to A. thaliana and several Brassicaceae spp., we identified additional Tic40 homologs by BLASTp searches in various plant species, including several monocots and eudicots. Despite their sequence divergence, comparisons show a similar structure of all studied Tic40 homologs with conserved functional domains, a transit peptide and a transmembrane domain at the N terminus, a TPR domain, and a Sti1p/Hop/Hip domain at the C terminus (Supplemental Fig. S1). Furthermore, direct sequence comparison of these Tic40 homologs revealed six amino acid substitutions unique to BnaC9.Tic40 and BolC9.Tic40 at their C-terminal regions (Supplemental Fig. S1). Five of these were found in the TPR domain at positions 307, 321, 343, 378, and 386. The sixth substitution occurred at position 408 in the Sti1p/Hop/Hip domain. These sites are highly conserved in other Tic40 homologs with amino acids identical to bnac9.tic40 (Supplemental Fig. S1). These data support that mutations in the TPR and/or Sti1p/Hop/Hip domains of BnaC9.Tic40 and BolC9.Tic40 are the causal mutations resulting in the restoration of fertility to 7365A.

Neofunctionalization of Tic40 in B. oleracea Lineages after Divergence of the Brassica spp. A, B, and C Genomes

To investigate the diversity of Tic40 in Brassica spp., the primer combination 6L/14R1 as described above was used to isolate partial Tic40 sequences, encoding TPR and Hop domains, from a set of 59 diverse Brassica spp. accessions (Supplemental Table S4). To avoid the effects of modern breeding, these Brassica spp. accessions were chosen from collections in the wild, landraces and breeders varieties from regions around the world in order to represent a large part of the existing genetic diversity (Supplemental Table S4). Due to the genome triplication of the Brassica spp. ancestor and the heterozygosity in these accessions, more than one copy with high sequence similarity to AtTic40 was present in any specific accession. Thus, different copies obtained from an accession were labeled with the suffix 1, 2, or 3 and so on. As a result, we obtained 115 Tic40 sequences from these accessions. These Tic40 homologs were combined with previously identified Brassica spp. Tic40 homologs for further analysis. An NJ distance tree was constructed using Tic40 from A. thaliana as an outgroup (Fig. 7). As expected, the Brassica spp. Tic40 genes were divided into two main clusters, clade 1 and clade 2, with an NJ bootstrap value of 100% (Fig. 7). This dichotomy strongly supports that two Tic40 loci existed in the common ancestor to extant Brassica spp. The Brassica spp. Tic40 genes were further divided into six major groups, in which these Tic40 genes were grouped together with a Brassica spp. Tic40 homolog of three basic Brassica diploid species, respectively (Fig. 7). For example, for group 1, 30 sequences with an NJ bootstrap value of 100% were grouped together with the

Figure 4. (Continued.)

the right. A10, Homologs from B. rapa; C9, homologs from B. oleracea; G1 to G24, gene loci 1 to 24 flanking Tic40; N19, homologs from 7365B; n19m, homologs from 7365A; W, homologs existing in both 7365B and 7365A, representing homologs of other syntenic regions. For genes G1, G2, G3, G23, and G24, sequences from N19 and n19m cluster together with the sequence from C9. For the other genes, only N19 sequences cluster with C9 sequences, while n19m sequences cluster with those from A10, supporting the conclusion that the region around Tic40 on N19 in 7365A is derived from N10 (corresponding to A10).
BolC9.Tic40 gene, and all of these sequences are from accessions belonging to *B. oleracea* and its derivatives, like *Brassica montana*, *Brassica villosa*, *B. carinata*, and *B. napus* (Fig. 7; Supplemental Table S4).

Furthermore, we analyzed the prevalence of the 12 amino acid substitution sites, identified in the C-terminal domain of BnaC9.Tic40 and BolC9.Tic40, in the set of 115 Tic40 sequences. In our collection, all Tic40 homologs from group 2 to group 6 carried identical amino acids with bnaC9.tic40 at these 12 unique substitution sites (Table III). In contrast, most Tic40 homologs from group 1 carried identical amino acids with BnaC9.Tic40 and BolC9.Tic40 at these sites, although some Tic40 homologs from group 1 carried amino acids identical to bnaC9.tic40 at three sites, positions 294, 304, and 321 (Table III). These results indicate that the amino acid replacements in the C terminus of Tic40 that are related to functional divergence are confined to *B. oleracea* lineages and its derivatives. Overall, based on the evolutionary and functional analysis of *Brassica* Tic40 homologs, we proposed that BolC9.Tic40 and its ortholog BnaC9.Tic40 underwent neofunctionalization after the separation of the *B. oleracea* lineage from the other diploid *Brassica* spp. lineages, although the Tic40 duplication occurred much earlier.

**DISCUSSION**

**Two Ancient Tic40 Loci in Brassica spp.**

In this study, we identified Tic40 homologs in several representative Brassicaceae spp. The *A. thaliana* genome that represents the ancestral ploidy state for the Brassicaceae family has one Tic40 locus, as do the closely related genomes of *A. lyrata* and *C. rubella*. The genomes of *B. rapa*, *B. oleracea*, and *B. nigra* are generally expected to contain three Tic40 loci due to the whole-genome triplication of the *Brassica* spp. ancestor after divergence from *A. thaliana* (Lysak et al., 2005; Parkin et al., 2005; Cheng et al., 2013). *B. napus* is expected to contain all of the loci present in *B. rapa* and *B. oleracea*, which merged to form *B. napus* (Cheung et al., 2009). However, only four Tic40 loci were identified in *B. napus*, and their genome origins from the A and C genomes were determined by phylogenetic tree and sequence identity analysis. The recent genome sequencing of *B. rapa* and *B. oleracea* also uncovered two Tic40 homologs in each genome (Wang et al., 2011), demonstrating that our results represent correct information of Tic40 homologs in *B. napus* and three *Brassica* diploid species.

The loss of one Tic40 locus took place on the MF1 subgenome as defined in *B. rapa* and *B. oleracea*. In addition, consistent with the estimated gene loss rate in three subgenomes of diploid *Brassica* spp. (Cheng et al., 2013), greater gene loss was observed in the Tic40 genomic regions in MF1 and MF2 subgenomes than in the LF subgenome. These data indicate that deletion of one Tic40 locus occurred in the common ancestor of *Brassica* spp. due to the diploidization process.
Functional Conservation and Divergence of Tic40 Genes

Initial analyses of the attic40 mutant demonstrated that Tic40 is required for the vegetative development of plants and functions as part of the protein translocon complex to assist protein translocation across the chloroplast inner membrane (Chou et al., 2003, 2006). The high conservation of Tic40 homologs in flowering plants, including monocots and eudicots, implies a general importance of Tic40 for plant development. The C-terminal domain of Tic40 shares sequence similarity with the C-terminal domains of the mammalian cochaperones Hip and Hop, acting as Hsp70-interacting protein and Hsp70/Hsp90-organizing protein, respectively (Bédard et al., 2007), and the Saccharomyces cerevisiae cochaperone Sti1p (Chou et al., 2003). Notably, the complementation efficiency of the Tic40:Hip fusion construct with the Sti1p/Hop/Hip domain of Tic40 exchanged with the human counterpart showed functional equivalence between the Sti1p/Hop/Hip domain of AtTic40 and that of human Hip (Bédard et al., 2007). Furthermore, among Tic40 sequences, this region is the most conserved in various plant species, suggesting that the function from Saccharomyces spp. (Sti1p) to human (Hop and Hip) to plant (Tic40) is evolutionarily conserved (Fig. 4, A and B; Bédard et al., 2007). Tic40 homologs in plants exhibit high conservation of additional functional domains, including the highly conserved predicted chloroplast/plastid-targeting transit peptide, the transmembrane domain at their N-terminal region.
Figure 6. Functional analysis of *B. napus* Tic40 complementation constructs. A, ClustalW alignment and structural features of the deduced Tic40 proteins. The conserved domains were predicted by comparison with the structure of *A. thaliana* Tic40. TP, Transit peptide; TM, transmembrane domain; TPR, tetratricopeptide domain; Sti1p/Hop/Hip, Sti1p/Hop/Hip domain. B, Schematic drawings showing the positions of protein domains of BnaC9.Tic40 and bnac9.tic40 and the organization of the Tic40 N- and C-terminal domain-swap constructs. C, Typical anthers from four representative 7365A lines carrying the indicated *B. napus* Tic40 chimeric complementation constructs, alongside wild-type 7365B and untransformed 7365A controls.
Figure 7. An NJ distance tree for Tic40 homologs from Brassica spp. Tic40 from A. thaliana (AtTic40) was used as an outgroup. The numbers on nodes are bootstrap values. Bvi, B. villosa; Bmo, B. montana; Bin, Brassica incana; Bca, B. carinata; Bju, B. juncea; Bna, B. napus; Bol, B. oleracea; Bra, B. rapa; Bni, B. nigra. B. incana, B. villosa, and B. montana belong to the B. oleracea species complex.
and the TPR. The highly conserved protein structure suggests that Tic40 homologs should maintain a similar biochemical and cellular function in protein translocation across the chloroplast/plastid inner membrane, and the four functional domains of Tic40 presented here should be important for plant development.

In this study, we have characterized the structure and function of a set of Tic40 genes in *Brassica* spp. These *Brassica* spp. Tic40 genes showed low levels of sequence divergence but did show functional diversity. BolC9.Tic40 and its ortholog BnaC9.Tic40, identified as gain-of-function Tic40 copies in the *B. oleracea* lineages, confer a fertility-restoring effect on the *B. napus* male-sterile mutant 7365A, while other representative *Brassica* Tic40 copies display an identical function with AtTic40. The analysis of deduced amino acid sequences and the assessment of complementation activity of different *B. napus* Tic40 constructs in the 7365A mutant suggested that amino acid replacements in the TPR and/or the Sti1p/Hip/Hop domains are responsible for this functional divergence. Furthermore, six amino acids specific to BnaC9.tic40 and BolC9.tic40 are highly conserved among other plant Tic40 homologs carrying amino acids identical to bnac9.tic40, suggesting that they are important for protein function. And five of them were found inside the TPR domain. As mentioned above, the Sti1p domain has a conserved function in eukaryotes, although the two domains from the C terminus of Tic40 and HsHip were only 35% homologous (Bédard et al., 2007). The TPR domain comprises three or more motifs that form a pair of anti-parallel \(\alpha\)-helices that interact with different target proteins. The target proteins that bind to the TPR domains rely on the primary sequences of the TPR domain (Lamb et al., 1995), implying that variation of the TPR primary sequences should contribute to changes in the target-binding proteins and result in the functional divergence of Tic40 genes.

Neofunctionalization of Tic40 in *B. oleracea* Lineages Caused Functional Divergence of Homologous Genes from *Brassica* spp. A, B, and C Genomes

Two copies of Tic40 are retained in diploid *Brassica* spp. genomes after the whole-genome triplication. In addition to gene loss, possible evolutionary fates of paralogous genes created by polyploidy followed by diploidization include maintaining the ancestral function, silencing, and functional divergence (Whittle and Krochko, 2009; Liu and Adams, 2010). Functional divergence after gene duplication can hypothetically result in two alternative evolutionary fates: neofunctionalization and subfunctionalization (Liu and Adams, 2010). A previous study with the paralogs SHORT SUSPENSOR (SSP) and Brassinosteroid Kinase1 (BSK1) formed by a polyploidy event in the Brassicaceae family also showed evidence of neofunctionalization after duplication, as described by BSK1 retaining the ancestral expression pattern and function and SSP gaining a new function.

### Table III. Nucleotide and amino acid variations at 12 replacement sites of 113 sequences orthologous (group 1 in Fig. 7) and homologous (groups 2–6 in Fig. 7) to BolC9.Tic40

| Group | Replacement Amino Acid Sites in BolC9.Tic40 | Group 1 in Figure 7 | Amino Acid | Group 2 to 6 in Figure 7 | Amino Acid |
|-------|---------------------------------------------|---------------------|------------|--------------------------|------------|
|       | Nucleotide | GAA | Gln | GCA | Ala | Val | Gly | Leu | Pro | Gly | Leu | Gly | Ala | Ala | Phe | Ala |
|       | Amino Acid | GTT | Phe | GGG | Trp | Val | Leu | Gln | Val | Pro | Gln | Ala | Ala | Phe | Gln | Ala | Phe |

Plant Physiol. Vol. 166, 2014
Evolutionary Dynamics of the Gain-of-Function Tic40 Gene in Brassica spp. Lineages

Although the evolutionary novelty of Tic40 by neo-functionalization is well supported, the underlying causes for the emergence and maintenance of the novel Tic40 gene that provides the new function related to male fertility are unknown. Most analyzed diploid B. oleracea lineage accessions carried the gain-of-function Tic40 copies, suggesting its importance for the adaptive evolution of B. oleracea lineages. The novel function of Tic40 can restore the male fertility of the B. napus male-sterile line 7365A, the sterility of which results from the dominant gene BnRf. Thus, it seems likely that the emergence and maintenance of the novel function of Tic40 was coupled to the evolution of Rf in B. oleracea lineages. This scenario suggests intermolecular coevolution and that the BolC9.Tic40 protein might directly interact with the Rf protein.

Homologous Chromosomal Rearrangements Contribute to Allelic and Phenotypic Diversity

In the amphidiploid B. napus with merged homologous A and C genomes, cytological observations and genetic mapping with molecular markers demonstrated that the A and C genomes have retained sufficient homology to allow chromosome pairing, which may result in occasional homologous exchanges (Parkin et al., 1995; Nicolas et al., 2007). Chromosomal rearrangements caused by such homologous exchanges are prevalent in B. napus, as evidenced by genetic mapping analysis of four B. napus segregating double haploid populations (Udall et al., 2005). Further studies demonstrated that homologous nonreciprocal transpositions may cause qualitative changes in the expression of specific homologous genes and phenotypic variation in the resynthesized amphidiploid B. napus (Gaeta et al., 2007). Studies of the effects of homologous chromosomal rearrangements in polyploidy have revealed several causes of phenotypic variation: changes in the expression of parental genes can occur by altering methylation or regulation, for example, flowering time variation with the altered expression of parental FLOWERING LOCUS C genes (Pires et al., 2004), and some genomic exchanges exhibit a heterosis-like effect due to the formation of intergenome heterozygosity, which was interpreted by altering variation in seed yield coupled to homologous chromosomal rearrangements (Osborn et al., 2003).

In this study, we show that changes in protein function of homologous genes that evolved after duplication, followed by homologous exchange, gave rise to allelic and phenotypic diversity related to male sterility. Thus, homologous chromosomal rearrangements may be an important mechanism creating novel allele combinations and phenotypic diversity in polyploid species with highly syntenic genomes.

MATERIALS AND METHODS

Plant Materials

The plants used in this study were grown in soil under natural conditions. Genomic DNAs of Brassica oleracea genotypes T14, CGN06903, CGN18458, and T9, Brassica rapa genotypes R7, CGN06832, CGN13925, and Chiifu, and Brassica nigra genotypes CGN06620, CGN06625, and CGN06630 (Table I) were used for the isolation of diploid Brassica spp. Tic40 homologs by PCR. The self-cross homozygous wild-type progeny and the male-sterile individual from the 7365AB near-isogenic line were used to isolate Brassica napus Tic40 homologs. The Arabidopsis thaliana tic40 mutant (transfer DNA insertion mutant SALK_028413) was purchased from the Arabidopsis Biological Resource Center mutant collection (http://abrc.osu.edu/). Fifty-nine genotypes of nine Brassica spp. were collected from different regions of the globe that are considered to represent great genetic diversity (Supplemental Table S4).

Genome Synteny Analysis

Whole-genomic information for A. thaliana, Arabidopsis lyrata, Capsella rubella, Thellungiella halophila, B. rapa, and B. oleracea was downloaded from public databases (www.phytozome.com and http://www.arabidopsis.org/; Hu et al., 2011; Wang et al., 2011; Slotta et al., 2013; Yang et al., 2013; Liu et al., 2014). The genomic multiple sequences alignments were performed using the
Cloning of Tic40 Genes from B. napus and Three Diploid Brassica spp.

The cDNA sequence of AtTic40 was used to query the Brassica spp. EST database in GenBank (National Center for Biotechnology Information) using the MEGABLAST program. Identified Brassica spp. Tic40 sequences were aligned together with cDNAs of several above-mentioned Brassicaceae spp. Tic40 homologs. Conserved regions in the exons identified in this alignment were selected to design degenerate primers to amplify Tic40 homologs from various Brassica spp. (for details of degenerate primers, see Supplemental Table S1). To validate the efficacy of these degenerate primers, multiple primer combinations were used to amplify diverse Tic40 homologs in three diploid Brassica spp. and A. thaliana. PCR products were separated on a 6% (w/v) denaturing polyacrylamide gel. Twenty clones from each PCR product were sequenced. A set of three primer combinations resulting in overlapping sequences was selected to amplify Brassica spp. Tic40 genes. Tic40 copies of homoygous B. napus genotypes 7365B and 7365A were amplified using the same set of primer combinations. The full-length genomic sequences of these Tic40 homologs were validated by PCR. Subsequently, partial sequences containing the TPR and Hop domains of Tic40 from 5′9 to 1418R1 were used to amplify the corresponding cDNA using degenerate primers.

Sequence Alignment and Phylogenetic Analysis of Tic40 Homologs

For full-length cDNA amplification, total RNA prepared from young buds of 7365A, 7365B, Chifflu, T9, and CNG06620 were used in reverse transcription-PCR. Then, the full-length ORFs of Tic40 homologs were aligned using ClustalX (Thompson et al., 1997), and nucleotide identities between them were calculated simultaneously. Bootstrapped NJ phylogenetic trees were constructed using Kimura (1980)'s two-parameter model, and bootstrap values (1,000 replications) were calculated using MEGA4.0 (Tamura et al., 2007). Homologous protein sequences of Tic40 in flowering plants were identified by BLASTp from the nonredundant protein sequences databases at the National Center for Biotechnology Information. Multiple protein sequence alignments were performed by MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/); and visualized by GeneDoc (http://www.nrbsc.org/gfx/genedoc/). Domain structures of Tic40 proteins were predicted according to the topology of AtTic40 as described by Chou et al. (2003).

Validation of Homologous Chromosomal Rearrangements

We obtained information of 23 flanking gene models around BolaC9.Tic40 and BolaA10.Tic40 from the genomes of B. oleracea C9 and B. rapa A10, respectively, using orthologs in A. thaliana as controls. Conserved primer combinations spanning one or more introns in each flanking gene were designed to carry out the PCR assay in 7365A, 7365B, B. oleracea T9, and B. rapa Chifflu. The primer combinations with identical bands in 7365A and 7365B but different bands between B. rapa Chifflu and B. oleracea T9, or with stable co-dominant polymorphic bands between 7365A and 7365B, were considered to amplify actual flanking sequences around BolaC9.Tic40 and bnaC9.Tic40. PCR products were separated on a 6% (w/v) denaturing polyacrylamide gel. Ten clones from each product were sequenced. Sequencing reads of 300 to 800 bp were assembled using DNAStar (http://www.dnastar.com/). Homologous sequences were aligned and sequence divergences were calculated by MEGA4.0 (Tamura et al., 2007). If successive flanking sequences in 7365B and 7365A originated from B. oleracea C9 and B. rapa A10, respectively, the homologous chromosomal rearrangement around BnaC9.Tic40 could be validated.

Genetic Complementation

The genomic fragments of 12 Tic40 homologs from B. napus 7365A and 7365B, B. nigra CNG06620, B. oleracea T9, B. rapa Chifflu, and A. thaliana Columbia were amplified using high-fidelity PCR (for details of primers, see Supplemental Table S1). The fragments were cloned into the pCAMBIA1305.1 binary vector. cDNA fragments and 2,000-bp promoter regions of BnaC9.Tic40 and bna9.1.Tic40 were used to generate chimeric complementation constructs. Constructs were assembled by gene splicing by overlap extension PCR (for details of primers, see Supplemental Table S1). Chimeric cDNA fragments were inserted downstream of the BnaC9.Tic40 and bna9.1.Tic40 promoters in pCAMBIA1305.1. These constructs were introduced into the host cells Agrobacterium tumefaciens GV3101. A. tumefaciens-mediated transformation of the attic40 mutant was performed, using the floral dip method (Clough and Bent, 1998), and transformation of the B. napus male-sterile 7365A5 was performed as described previously by Dun et al. (2011).

Expression Patterns of Brassica spp. Tic40 Homologs

Total RNA isolated from various tissues including roots, stems, seedlings, inflorescences, and young siliques of heterozygous 7365B, B. rapa Chifflu, and B. oleracea T9 were used for real-time PCR analysis. First-stand cDNAs were diluted 20-fold and amplified according to the instructions for RealMasterMix (SYBR Green [FP202]; TIANGEN; for details of primers, see Supplemental Table S1). The assay was performed in triplicate with the CFX96 real-time system (Bio-Rad), and BnACT7 (EV220887.1) was used as a control to normalize the expression data. The results were analyzed using CFX Manager Software according to the 2−ΔΔCT method (Livak and Schmittgen, 2001). Approximately 1,500- to 2,000-bp upstream regions of the five B. napus Tic40 genes were amplified from 7365A5 and cloned into the binary vector pBl201. The promoter-GUS fusion constructs were introduced into A. thaliana wild-type plants by A. tumefaciens-mediated transformation. GUS activity was visualized by staining inflorescences from transgenic lines.

Sequence data from this article can be found in the GenBank/EMBL database with the following Accession Numbers: Fragaria vesca (XP_0043071); Vitis vinifera (XP_002285274); Solanum lycopersicum (XP_004250413); Populus trichocarpa (XP_002305876); Ricinus communis (XP_002531917); Zea mays (XP_0043071); Glycine max (XP_003538352); Vitis vinifera (XP_002285274); Theobroma cacao (XP_0043071); Convolvulaceae (XP_0043071); Cupressaceae (XP_0043071); Poaceae (XP_0043071); Cucurbitaceae (XP_0043071); Brassicaceae (XP_0043071); Dicotyledoneae (XP_0043071); Caryophyllales (XP_0043071); Portulacaceae (XP_0043071); Dilleniidaeae (XP_0043071); Fabeae (XP_0043071). Neofunctionalization of Duplicated Tic40 Genes

We thank Hanhui Huang (Huazhong Agricultural University) for research guidance, Dr. Shengyi Liu (Olicrops Research Institute, Chinese Academy of Agricultural Sciences) for providing the complete genomic information for B. oleraceae, and the Centre for Genetic Resources (The Netherlands; http://www.wageningenur.nl/) for providing the Brassicaceae germplasms. Received July 9, 2014; accepted August 22, 2014; published September 2, 2014.

LITERATURE CITED

Arias T, Beilstein MA, Tang M, McKin MR, Pires JC (2014) Diversification times among Brassica (Brassicaceae) crops suggest hybrid formation after 20 million years of divergence. Am J Bot 101: 86–91
Lamb JR, Kubis S, Bimanadham S, Jarvis P (2007) Functional similarity between the chloroplast translocase component, Tic40, and the human.co-chaperone, Hsp70-interacting protein (Hip). J Biol Chem 282: 21404–21414

Beilstein AM, Al-Sherbaz IA, Kellogg EA (2006) Brassicaceae phylogeny and trichome evolution. Am J Bot 93: 607–619

Blanc G, Hokamp K, Wolfe KH (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Res 13: 137–144

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. Plant Cell 16: 1679–1691

Chen X, Truksa M, Snyder CL, El-Mezawy A, Shah S, Weselake RJ (2011) Three homologous genes encoding sn-glycerol-3-phosphate acyltransferase 4 exhibit different expression patterns and functional divergence in Brassica napus. Plant Physiol 155: 851–866

Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol 58: 377–406

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

Cheung F, Trick M, Droz N, Lim YP, Park JY, Kwon SJ, Kim JA, Scott R, Pires JC, Paterson AH, et al (2009) Comparative analysis between homoeologous genome segments of Brassica napus and its progenitor species reveals extensive sequence-level divergence. Plant Cell 21: 1912–1928

Chou ML, Chu CC, Chen LJ, Akita M, Li HM (2006) Stimulation of transitpeptide release and ATP hydrolysis by a co-chaperone during protein import into chloroplasts. J Cell Biol 175: 893–900

Chou ML, Fitzpatrick LM, Tu SL, Budziszewski G, Potter-Lewis S, Akita M, Levin JZ, Keegstra K, Li HM (2003) Tic40, a membrane-anchored chaperone homolog in the chloroplast protein translocase. EMBO J 22: 2970–2980

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J 16: 735–743

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9: 938–950

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Solitis DE, Doyle JJ, Solitis PS, Carlson JE, Arumuganathan K, Barakat A, et al (2006) Widespread genome duplications throughout the history of flowering plants. Genome Res 16: 738–749

Dun X, Zhou Z, Xia S, Wen J, Yi B, Shen J, Ma C, Tu J, Fu T (2011) BnaC4, a plastid inner membrane translocon originating from Brassica oleracea, is essential for tapetal function and microspore development in Brassica rapa. Plant J 67: 532–545

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytol 183: 557–564

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151: 1531–1545

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in mesohexaploid Brassica napus and their effect on gene expression and phenotype. Plant Cell 19: 3403–3417

Harris RS (2007) Improved pairwise alignment of genomic DNA. PhD thesis. Pennsylvania State University, University Park, PA

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, et al (2009) Precise mapping of homeologous reciprocal translocations provides evidence for extensive gene rearrangement within the Brassica genome. Nat Genet 41: 835–840

Huang Z, Chen Y, Bi X, Xiao L, Ma C, Tu J, Fu T (2007) Fine mapping of the recessive genetic male sterility gene (Bmns3) in Brassica napus L. Theor Appl Genet 115: 113–118

Inaba R, Nishio T (2002) Phylogenetic analysis of Brassicaceae based on the nucleotide sequences of the S-locus related gene, SLR1, indicates that Brassica rapa and B. nigra in relation to homologous regions of Arabidopsis thaliana, B. rapa and B. oleracea. BMC Evolutionary Biology 14: 250

Jung HJ, Wu SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, Kim JS, Baek S, Choi BS, et al (2009) Genome-wide comparative analysis of the Brassica rapa gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome duplication. Genome Biol 10: R111

Liu SL, Adams KL (2010) Dramatic change in function and expression pattern of a gene duplicated by polyploidy created a paternal effect gene in the Brassicaceae. Mol Biol Evol 27: 2817–2828

Livas KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25: 402–408

Lukens L, Quijada PA, Uddall J, Pires JC, Schranz ME, Osborn T (2004) Genome redundancy and plasticity within ancient and recent Brassica crop species. Biol J Linnean Soc Lond 82: 675–688

Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassicaceae. Genome Res 15: S16–S25

Ma H (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. Annu Rev Plant Biol 56: 393–434

Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA, Lim MH, Kim JS, Baek S, Choi BS, et al (2009) Genome-wide comparative analysis of the Brassica rapa gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome duplication. Genome Biol 10: R111

Navabi ZK, Huebert T, Sharpe AGO, O’Neill CM, Bancroft I, Parkin IA (2013) Conserved microstructure of the Brassica B genome of Brassica nigra in relation to homologous regions of Arabidopsis thaliana, B. rapa and B. oleracea. BMC Genomics 14: 250

Nicas JD, Le Mignon N, Eber F, Coriton O, Monod H, Cloutet V, Huteau B, Lostenan A, Deloumee R, Chalhoub B, et al (2007) Homeologous recombination plays a major role in chromosome rearrangements that occur during meiosis of Brassica napus haploids. Genetics 175: 847–903

Osborn TC, Buttrulle DV, Sharpe AG, Pickering KJ, Parkin IAP, Parker JS, Lydiate DJ (2003) Mapping and effects of a homeologous reciprocal transposition in Brassica napus. Genetics 165: 1569–1577

Panjabi P, Jagn Nath A, Bisht NC, Padmaja KL, Sharma S, Gupta V, Pradhan AK, Parkin IA (2007) Genomic and evolutionary analysis of homeologous gene pairs in Brassica nigra. Genome 50: 835–840

Parkin IAP, Lydiate DJ, Trick M (2002) Assessing the level of collinearity between Arabidopsis thaliana and Brassica napus for A. thaliana chromosome 5. Genome 45: 356–366

Parkin IAP, Sharpe AG, Keith DJ, Lydiate DJ (1995) Identification of the A and C genomes of amphidiploid Brassica napus (oilseed rape). Genome 38: 1122–1131

Pires JC, Zhao JW, Schranz ME, Leon EJ, Osborn PA, Lukens LN, Osborn TC (2004) Flowering time divergence and genomic rearrangements in mesohexaploid Brassica napus (Brassicaceae). Nat Genet 36: 825–828

Rodriguez-Trelles F, Tarrio R, Ayala FJ (2003) Convergent neo-functionalization by positive Darwinian selection after ancient recurrent duplications of the xanthine dehydrogenase gene. Proc Natl Acad Sci USA 100: 13413–13417

Schranz ME, Lysak MA, Mitchell-Olta D (2006) The ABC’s of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci 11: 535–542

Slotté T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo YL, Steiger K, Platts AE, Escobar JS, Newman LK, et al (2013) The Capsella rupelila genome and the genomic consequences of rapid mating system evolution. Nat Genet 45: 831–835

Solitis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Solitis PS (2009) Polyploidy and angiosperm diversification. Am J Bot 96: 336–348

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (Mega) software version 4.0. Mol Biol Evol 24: 1596–1599

Tang H, Lyons E (2012) Unleashing the genome of Brassica rapa. Front Plant Sci 3: 172

Tang H, Woodhouse MR, Cheng F, Schnable JC, Pedersen BS, Conant G, Wang X, Freeling M, Pires JC (2012) Altered patterns of fractionation and exon deletions in Brassica rapa support a two-step model of paleohexaploidy. Genetics 190: 1563–1574

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25: 4876–4882
Town CD, Cheung F, Maiti R, Crabtree J, Haas BJ, Wortman JR, Hine EE, Althoff R, Arbogast TS, Tallon LJ, et al (2006) Comparative genomics of *Brassica oleracea* and *Arabidopsis thaliana* reveal gene loss, fragmentation, and dispersal after polyploidy. Plant Cell 18: 1348–1359

Udall JA, Quijada PA, Osborn TC (2005) Detection of chromosomal rearrangements derived from homologous recombination in four mapping populations of *Brassica napus* L. Genetics 169: 967–979

Udall JA, Wendel JF (2006) Polyploidy and crop improvement. Crop Sci (Suppl) 46: S3–S14

Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. Nat Rev Genet 10: 725–732

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun JH, Bancroft I, Cheng F, et al (2011) The genome of the mesopolyploid crop species *Brassica napus*. Nat Genet 43: 1035–1039

Whittle CA, Krochko JE (2009) Transcript profiling provides evidence of functional divergence and expression networks among ribosomal protein gene paralogs in *Brassica napus*. Plant Cell 21: 2203–2219

Wolfe KH (2001) Yesterday’s polyploids and the mystery of diploidization. Nat Rev Genet 2: 333–341

Xia S, Cheng L, Zu F, Dun X, Zhou Z, Yi B, Wen J, Ma C, Shen J, Tu J, et al (2012) Mapping of *BnMs4* and *BnRf* to a common microsyntenic region of *Arabidopsis thaliana* chromosome 3 using intron polymorphism markers. Theor Appl Genet 124: 1193–1200

Yang R, Jarvis DE, Chen H, Beilstein MA, Grimwood J, Jenkins J, Shu S, Prochnik S, Xin M, Ma C, et al (2013) The reference genome of the halophytic plant *Eutrema salsugineum*. Front Plant Sci 4: 46

Zhao M, Du J, Lin F, Tong C, Yu J, Huang S, Wang X, Liu S, Ma J (2013) Shifts in the evolutionary rate and intensity of purifying selection between two *Brassica* genomes revealed by analyses of orthologous transposons and relics of a whole genome triplication. Plant J 76: 211–222

Zhou Z, Dun X, Xia S, Shi D, Qin M, Yi B, Wen J, Shen J, Ma C, Tu J, et al (2012) *BnMs3* is required for tapetal differentiation and degradation, microspore separation, and pollen-wall biosynthesis in *Brassica napus*. J Exp Bot 63: 2041–2058

Ziolkowski PA, Kaczmarek M, Babula D, Sadowski J (2006) Genome evolution in *Arabidopsis/Brassica*: conservation and divergence of ancient rearranged segments and their breakpoints. Plant J 47: 63–74