

Neofunctionalization of Duplicated *Tic40* Genes Caused a Gain-of-Function Variation Related to Male Fertility in *Brassica oleracea* Lineages¹[W][OPEN]

Xiaoling Dun², Wenhao Shen, Kaining Hu, Zhengfu Zhou³, 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 (*BraA10.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; Fligel 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).

[W] The online version of this article contains Web-only data.

[OPEN] 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* 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., 2007; Xia et al., 2012). The male sterility in this system is controlled by two genes, a recessive male sterile gene *Bnms3* and an epistatic gene *BnRf* (Zhou et al., 2012). *Bnms3* was recently reported to be a homolog to *A. thaliana Tic40* (Dun et al., 2011). *Tic40* was identified as a member of the TIC (for translocon at the inner envelope membrane of chloroplasts) complex that functions as a cochaperone to coordinate *Tic110* and the stromal chaperone heat shock protein93 (*Hsp93*) (Chou et al., 2003, 2006). The *A. thaliana tic40* mutant displayed a chlorotic phenotype throughout development (Chou et al., 2003) but a male-fertile phenotype with mature pollen grains (Dun et al., 2011). Interestingly, one allele of *Bnms3*, *BnaC.Tic40*, can rescue the fertility of the *B. napus* male-sterile line 7365A (Dun et al., 2011).

In this study, we identified and characterized *Tic40* homologs in *B. napus* and three basic diploid *Brassica* spp. We suggested that neofunctionalization 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 *bnac9.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 paleoploid *Brassica* spp.

RESULTS

Identification of *Tic40* Homologs in *B. napus* and Three Basic Diploid *Brassica* spp.

A. thaliana contains a single *Tic40* locus (*At5g16620*) located on the short arm of chromosome 5. The complementary DNA (cDNA) sequence of *AtTic40* was used to query the *Brassica* spp. EST database in GenBank (National Center for Biotechnology Information). Overall, 53 *Brassica* spp. sequences were identified that were aligned together with cDNAs of *Tic40* homologs from several Brassicaceae spp. with available whole-genome sequence data: *B. rapa* (Wang et al., 2011), *B. oleracea* (Liu et al., 2014), *Arabidopsis lyrata* (Hu et al., 2011), *Capsella rubella* (Slotte et al., 2013), and *Thellungiella halophila* (Yang et al., 2013). Degenerate primers were designed using the conserved sequences identified from this alignment (for details of degenerate primers, see Supplemental Table S1; for expected annealing positions of degenerate primers, see Supplemental Fig. S1). From these primers, seven combinations were assayed in PCR to amplify *Tic40* homologs in *A. thaliana* and three diploid *Brassica* spp., *B. rapa*, *B. oleracea*, and *B. nigra*. For each primer combination, one PCR fragment was obtained in *A. thaliana*, while two to four fragments were observed in the diploid *Brassica* spp. genotypes (Fig. 1), demonstrating multiple *Tic40* copies in *Brassica* spp. genomes. For each amplified *Brassica* spp. genotype, 60 sequence reads obtained from PCR products of three primer combinations (6L/14R1, 8L/14R2, and 10L/12R) were assembled. Within genotypes, each unique sequence from one primer combination could be well overlapped with a unique sequence from another primer combination (excluding rare PCR recombination events), suggesting that all *Tic40* copies in each genotype were amplified. This result supports the suitability of these primer combinations for amplifying *Tic40* genes in *Brassica* spp. Based on sequence divergence, these assembled sequences were condensed into two groups in all amplified *Brassica* spp. genotypes, indicating the existence of two *Tic40* loci in the diploid *Brassica* spp. (Table I). This successful PCR cloning approach was also used to isolate *Tic40* homologs from homozygous *B. napus* genotypes, the wild-type 7365B and the male-sterile 7365A from the *B. napus* near-isogenic line 7365AB. A total of four *Tic40* loci were identified in *B. napus*, and the fertility-restoring gene *BnaC.Tic40* and its allele were identified at the *Bnms3* locus in linkage group N19 of 7365B and 7365A, respectively, which has been reported by Dun et al. (2011).

Genome Syteny of the *Tic40* Genomic Regions in Several Brassicaceae Genomes

To study the evolutionary features of genomic regions around *Tic40* in *A. thaliana* and *Brassica* spp., genes flanking the *Tic40* locus in *A. thaliana* were used to identify its syntenic regions in the genomes of several Brassicaceae spp. As a result, one syntenic region was identified in the genomes of *A. lyrata*, *C. rubella*,

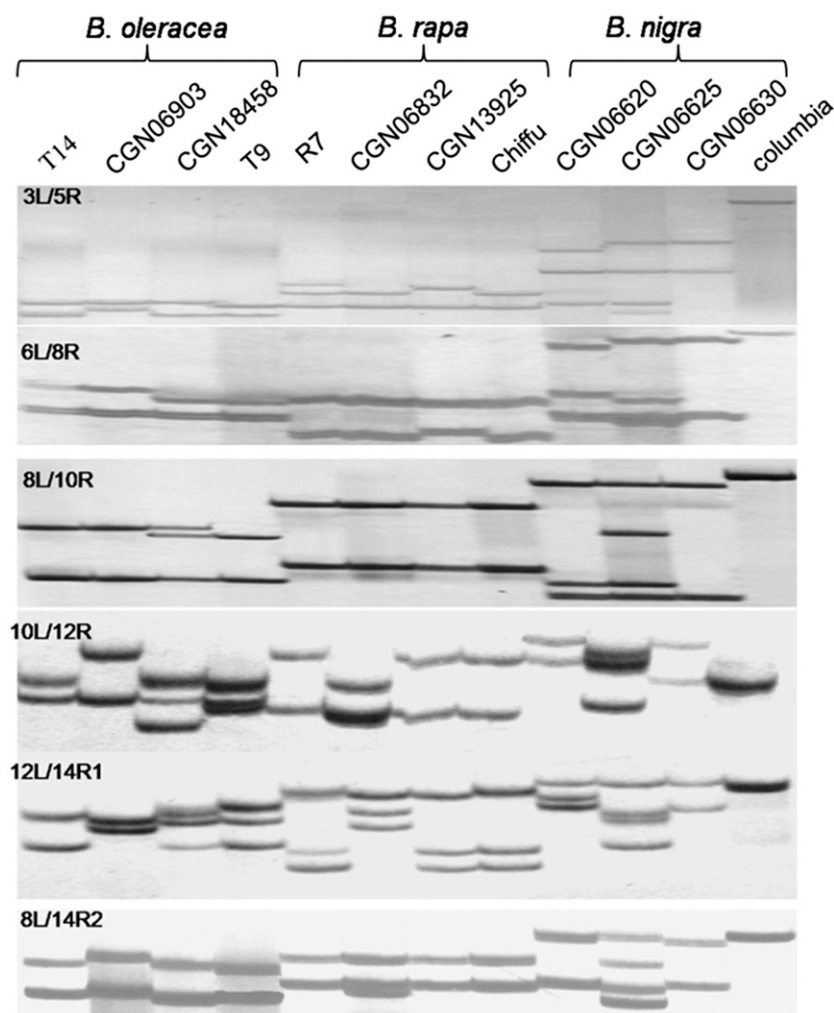


Figure 1. Amplification profiles of *Tic40* genes from three diploid *Brassica* spp. Seven primer combinations, 3L/5R, 6L/8R, 8L/10R, 10L/12R, 12L/14R1, 8L/14R2, and 6L/14R1, were used to amplify DNA from genotypes of *B. oleracea* (T14, CGN06903, and CGN18458, T9), *B. rapa* (R7, CGN06832, CGN13925, and Chiffu), *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).

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*)

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

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

^aFragments from within a diploid *Brassica* spp. with sequence identity over 90% were considered alleles at a *Tic40* locus.

^bDeduced number

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

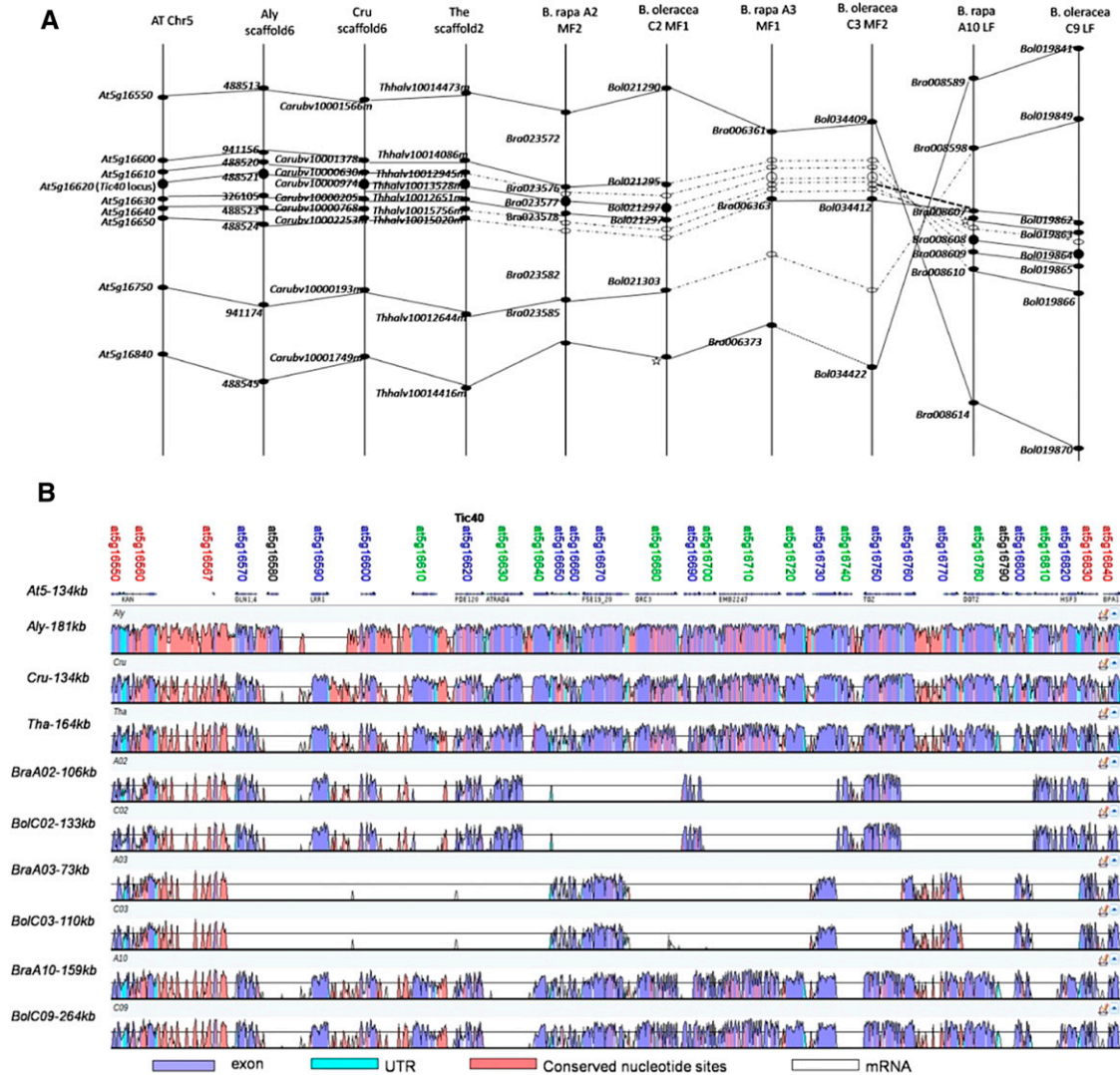


Figure 2. Synteny of *Tic40* genomic regions in Brassicaceae genomes. A, Alignment of homologous *Tic40* regions from six Brassicaceae genomes. The loci depicted by black circles are *Tic40* homologs, and ellipses represent homologs of *Tic40* flanking genes. White circles and ellipses represent the 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*, *BniB.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*, *BraA10.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 *BraA10.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 homozygous genotypes *B. rapa* Chiffu and *B. oleracea* T9 (Fig. 4A), and 18 of those (from IPJH4 to IPJH22) detected segregating loci for which the two homozygous *B. napus* genotypes 7365A and 7365B had

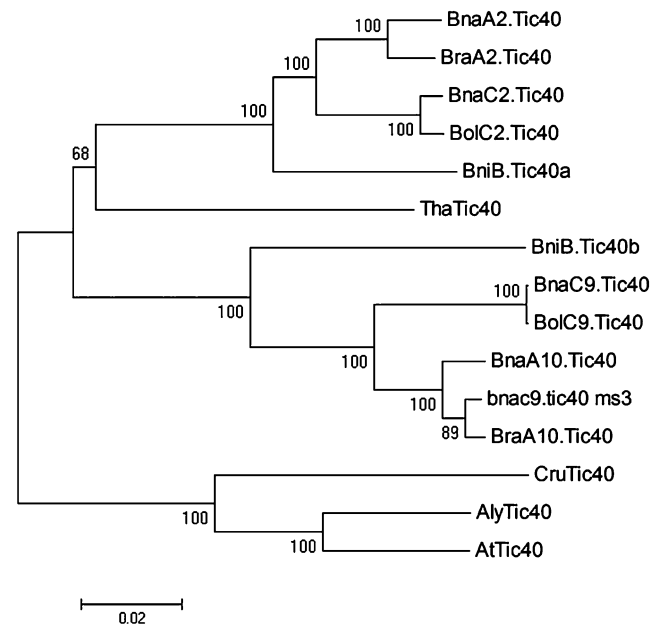


Figure 3. A neighbor-joining (NJ) distance tree for *Tic40* homologs in the Brassicaceae. The full-length ORF sequences of *Tic40* homologs were used for phylogenetic analysis using MEGA4.0 (Tamura et al., 2007). Bna, *B. napus*; Bol, *B. oleracea*; Bra, *B. rapa*; Bni, *B. nigra*; At, *A. thaliana*; Aly, *A. lyrata*; Cru, *C. rubella*; Tha, *T. halophila*. The numbers on nodes are bootstrap values.

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 IPJH22 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*, *BnaC2.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*, *BraA2.Tic40* and *BraA10.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 Tetratricopeptide

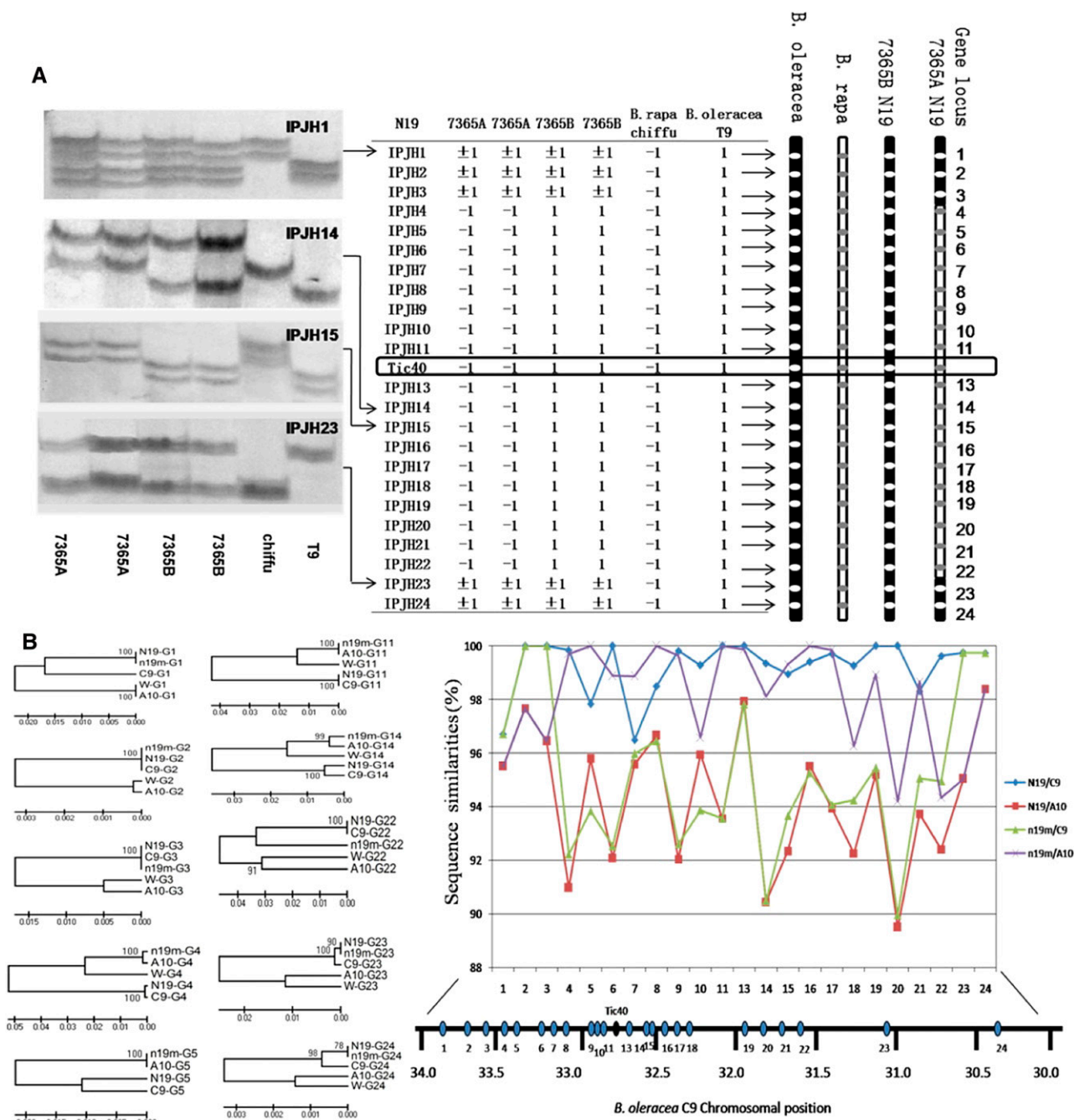


Figure 4. Analysis of the subgenome origin of genomic regions around *BnaC9.Tic40* and *bnac9.tic40* in *B. napus*. A, Segregation of IP molecular markers on linkage group N19 of *B. napus* 7365A and 7365B. Twenty-three IP molecular markers (IPJH1–IPJH24) corresponded to 23 loci flanking *Tic40* in *B. 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* C9. 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 St1p/Hop/Hip (Hop, known as Hsp70 and Hsp90 organizing protein, and also known as St1p; 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 St1p/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, AΔN 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 AΔN 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 St1p/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 St1p/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 St1p/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 St1p/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, the 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 group1, 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 from 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).

Table II. Results of genetic complementation assays for 12 *Tic40* genes introduced into the 7365A mutant and the *atitc40* mutant

+ indicates that the phenotype of the mutant was restored; – indicates that the phenotype of the mutant was not restored. Numbers in parentheses represent number of restored plants/number of plants assayed. na, Not applicable, as the genotype 7365A already harbors the *Tic40* homologs introduced.

Transformed Genotype	Introduced <i>Tic40</i> Gene											
	<i>AtTic40</i>	<i>BolC9.Tic40</i>	<i>BolC2.Tic40</i>	<i>BraA10.Tic40</i>	<i>BraA2.Tic40</i>	<i>BniB.Tic40a</i>	<i>BniB.Tic40b</i>	<i>BnaC9.tic40</i>	<i>BnaC9.tic40</i>	<i>BnaA10.tic40</i>	<i>BnaC2.tic40</i>	<i>BnaA2.tic40</i>
7365A	– (0/13)	+ (8/11)	– (0/17)	– (0/10)	– (0/14)	– (0/11)	– (0/13)	+ (11/13)	na	na	na	na
<i>atitc40</i>	+ (7/7)	+ (5/5)	+ (8/9)	+ (5/5)	+ (5/5)	+ (6/6)	+ (7/7)	+ (14/14)	+ (11/11)	+ (6/6)	+ (8/8)	+ (10/10)

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.

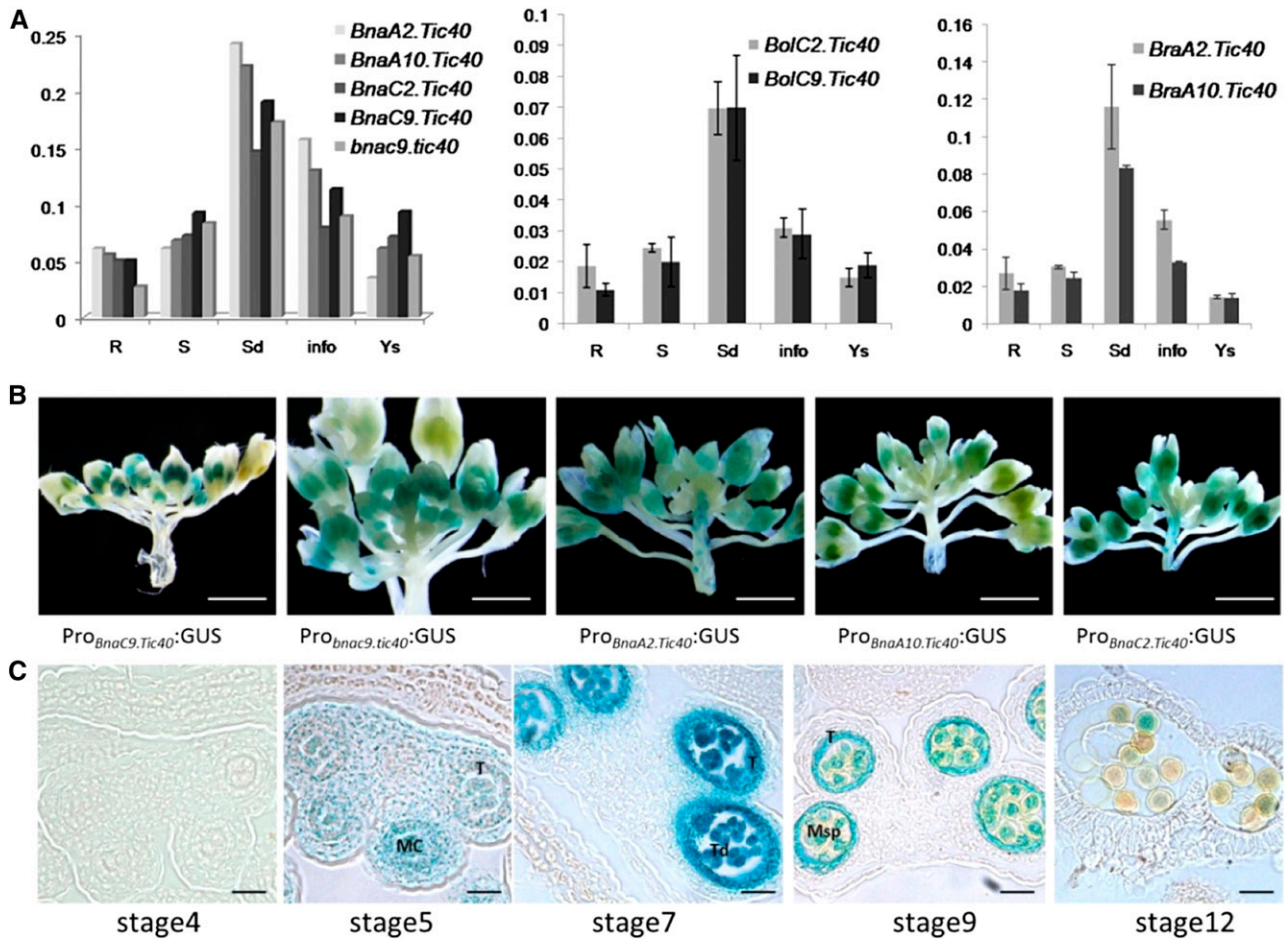


Figure 5. Expression patterns of *Brassica* spp. *Tic40* genes. A, Spatial and temporal expression analyses of *Brassica* spp. *Tic40* genes in selected tissues using real-time PCR. R, Root; S, stem; Sd, seedling; info, inflorescence; Ys, young silique. Expression patterns of *Brassica* spp. *Tic40* genes are shown for *B. napus* (left), *B. oleracea* (middle), and *B. rapa* (right). B, Histochemical analysis of the GUS activities (blue staining) directed by individual *B. napus* *Tic40* promoters in transgenic *A. thaliana* plants. GUS expression was driven by the five promoters of *BnaC9.Tic40*, *bnac9.tic40*, *BnaA2.Tic40*, *BnaA10.Tic40*, and *BnaC2.Tic40*. At least three independent transgenic plants for each construct displayed similar GUS activities in the anthers of young buds, one of which is shown in each image. C, GUS expression patterns in the anthers of the *B. napus* *Tic40* promoter-GUS transgenic line. Transgenic lines of the five *B. napus* *Tic40* promoters displayed similar GUS staining patterns; thus, only anther sections of one *Pro_{bnac9.tic40}*-GUS transgenic line is shown. The developmental stages of anthers were described according to the previous report by Ma (2005). GUS staining signals were obviously detectable in the tapetum, tetrads, and microspores during anther development. MC, Microsporocyte; Msp, microspore; T, tapetum; Td, tetrad. Bars = 5 mm in B and 100 μ m in C.

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 At*Tic40* 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 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.

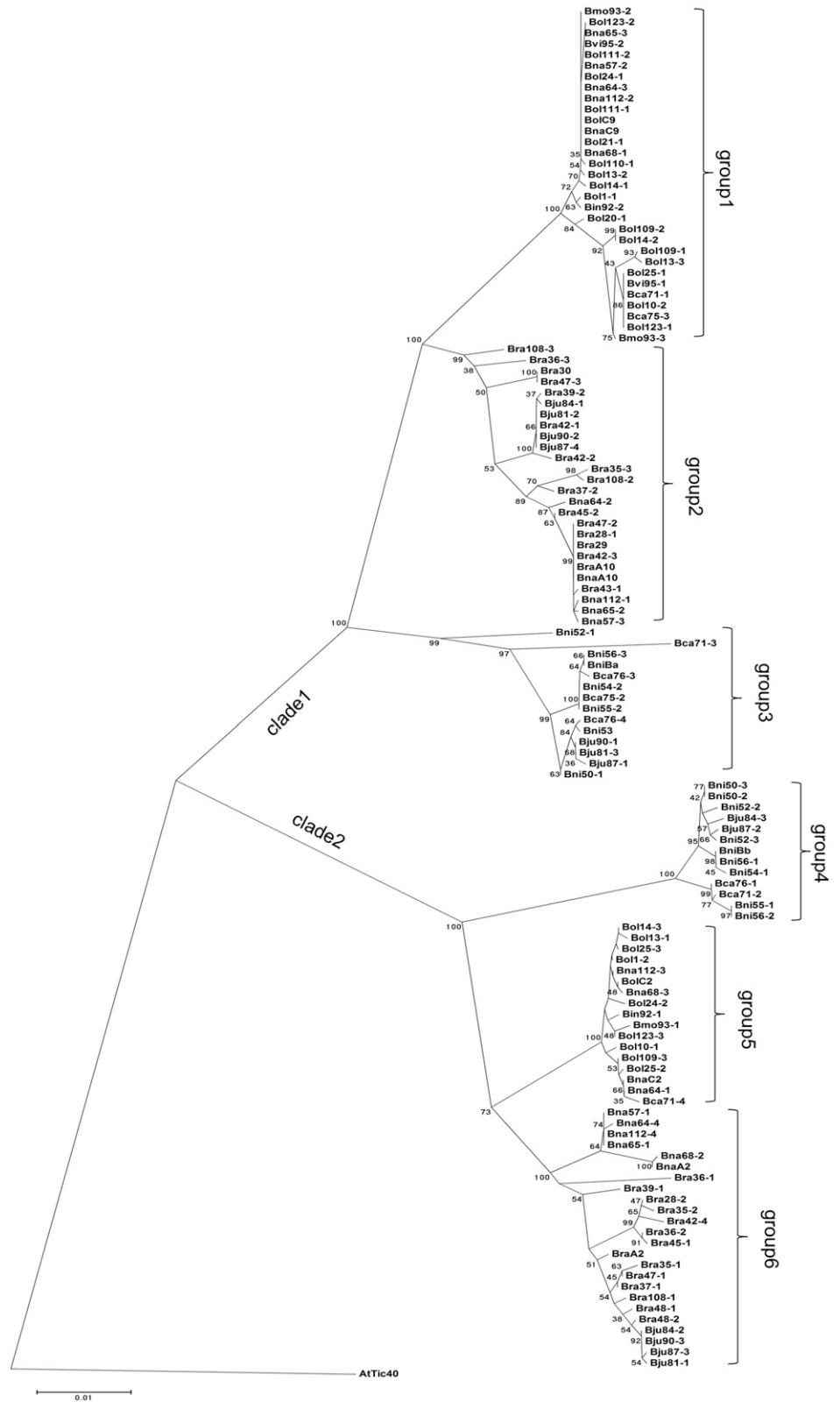


Table III. Nucleotide and amino acid variations at 12 replacement sites of 115 sequences orthologous (group 1 in Fig. 7) and homologous (groups 2–6 in Fig. 7) to *BolC9.Tic40*

Group	Sequence Type	Replacement Amino Acid Sites in <i>BolC9.Tic40</i>											
		261	294	304	307	311	321	343	361	378	386	408	422
Group 1 in Figure 7	Nucleotide sequence	GAA	GTT/GGT	TGG/GGG	GGG/GCC	GTT	TTT/GTT	CGT	GAG	CCG	GGA	TCA	GAC
	Amino acid	Gln	Phe/Ala	Trp/Gly	Gly	Val	Phe/Val	Leu	Gln	Leu	Gly	Ser	Asp
Groups 2 to 6 in Figure 7	Nucleotide sequence	GCA	GTT	GGC	TTA/CTG/TTG	GCT/GCC	GTT/GTC/GTG	CTT	GGG/GGA	CTG/CTT/CTA/CTC/TTG	CAA/CAG	GCA/GCC / GCT/GCC	GAA
	Amino acid	Ala	Phe	Gly	Leu	Ala	Val	Arg	Gly	Pro	Gln	Ala	Gln

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 α -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

but losing its original function (Liu and Adams, 2010). Within that scenario, our data suggest that *BolC9.Tic40* and its ortholog *BnaC9.Tic40* gained the novel function related to male fertility but also retained an ancestral function similar to *AtTic40*, whereas its paralogs and homologs in *Brassica* spp. only retained the ancestral function.

Sequence analysis of homologous *Tic40* genes revealed that the distribution of the gain-of-function *Tic40* copy was strictly restricted to the *B. oleracea* lineages and its derivatives, suggesting that the neofunctionalization occurred after the divergence of *Brassica* spp. A, B, and C genomes. Molecular dating supports that the *B. rapa* lineage (A genome) and the *B. oleracea* lineage (C genome) diverged from each other less than 3 million years ago (Inaba and Nishio, 2002; Navabi et al., 2013; Arias et al., 2014). The separation of the *B. nigra* (B genome) clade from that of the *B. oleracea*-*B. rapa* clade occurred much earlier, around 20 million years ago (Arias et al., 2014). Due to their close relationships, comparisons between the three *Brassica* spp. genomes revealed extensive conservation of gene content and sequence identity (Panjabi et al., 2008; Cheng et al., 2013). However, analysis of the tissue-specific alteration of gene expression mediated by transposable element insertions suggests a potential for rapid functional divergence of orthologous genes between the A and C genomes (Zhao et al., 2013). Although polyploidy events provide a large number of new genes that could potentially undergo functional divergence (Blanc and Wolfe, 2004; Chen et al., 2011), examples of substantial neofunctionalization or subfunctionalization of paralogous and homologous genes from the A, B, and C genomes are rare (Liu and Adams, 2010). Our results here suggest that *BolC9.Tic40* diverged functionally from its homologous genes, *BraA10.Tic40* and *BniB.Tic40b*, and the paralogous gene *BolC2.Tic40* and that this divergence occurred recently, after the separation of the *Brassica* spp. A and C genomes.

Evolutionary Dynamics of the Gain-of-Function *Tic40* Gene in *Brassica* spp. Lineages

Although the evolutionary novelty of *Tic40* by neofunctionalization 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 Chiffu, and *Brassica nigra* genotypes CGN06620, CGN06625, and CGN06630 (Table 1) 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; Slotte et al., 2013; Yang et al., 2013; Liu et al., 2014). The genomic multiple sequences alignments were performed using the

progressive alignment tool BLASTZ (http://www.bx.psu.edu/miller_lab/Harris, 2007). The results were visualized using the VISTA tool, a synteny browser for comparative sequences from multiple species (<http://genome.lbl.gov/vista/index.shtml>). Gene annotation of the *A. thaliana Tic40* region was produced using the Web-based gene prediction program FGENE-SH Arabidopsis (<http://www.softberry.com/berry.phtml>).

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 homozygous *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 59 *Brassica* spp. genotypes were isolated using the primer combination 6L/14R1.

Sequence Alignment and Phylogenetic Analysis of *Tic40* Homologs

For full-length cDNA amplification, total RNA prepared from young buds of 7365A, 7365B, Chiffu, T9, and CGN06620 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'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 (Chou et al., 2003).

Validation of Homologous Chromosomal Rearrangements

We obtained information of 23 flanking gene models around *BolC9.Tic40* and *BraA10.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* Chiffu. The primer combinations with identical bands in 7365A and 7365B but different bands between *B. rapa* Chiffu and *B. oleracea* T9, or with stable co-dominant polymorphic bands between 7365A and 7365B, were considered to amplify actual flanking sequences around *BnaC9.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* CGN06620, *B. oleracea* T9, *B. rapa* Chiffu, 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 *bnac9.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 *bnac9.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 7365A 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* Chiffu, 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 $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Approximately 1,500- to 2,000-bp upstream regions of the five *B. napus Tic40* genes were amplified from 7365AB and cloned into the binary vector pBI201. 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 data libraries under accession numbers *Fragaria vesca Tic40*, XP_004307173; *Citrus clementina Tic40*, ESR43817; *Cucumis sativus Tic40*, XP_004148914; *Populus trichocarpa Tic40*, XP_002305876; *Ricinus communis Tic40*, XP_002531917; *Solanum lycopersicum Tic40*, XP_004250413; *Vitis vinifera Tic40*, XP_002282574; *Theobroma cacao Tic40*, EOY03909; *Glycine max Tic40*, XP_003538352; *Cicer arietinum Tic40*, XP_004500418; *Pisum sativum Tic40*, CAB50925; *Zea mays Tic40*, DAA37691; *Setaria italica Tic40*, XP_004975683; *Sorghum bicolor Tic40*, XP_002447878.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. The *BnaC.Tic40* gene structure, expected annealing positions of degenerate primers used to amplify *Brassica* spp. *Tic40* fragments, and an alignment of deduced amino acid sequences of *Tic40* homologs from several monocot and eudicot plant species.

Supplemental Table S1. Primers used for PCR amplification in this study.

Supplemental Table S2. Sequence identity analysis of the full-length ORFs of *Tic40* homologs by ClustalW.

Supplemental Table S3. The orthologs of the 23 flanking genes analyzed in *A. thaliana*, *B. oleracea*, and *B. rapa* and their physical genomic locations in *B. oleracea* and *B. rapa*.

Supplemental Table S4. The information for 59 diverse *Brassica* spp. accessions from different global regions.

ACKNOWLEDGMENTS

We thank Hanhui Kuang (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. oleracea*, and the Centre for Genetic Resources (The Netherlands; <http://www.wageningenur.nl/>) for providing the Brassicaceae germplasm.

Received July 9, 2014; accepted August 22, 2014; published September 2, 2014.

LITERATURE CITED

Arias T, Beilstein MA, Tang M, McKain 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

- Bédard J, Kubis S, Bimanadham S, Jarvis P (2007) Functional similarity between the chloroplast translocon component, Tic40, and the human co-chaperone, Hsp70-interacting protein (Hip). *J Biol Chem* **282**: 21404–21414
- Beilstein MA, Al-Shehbaz 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**: 851–865
- 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, Mandáková 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, Drou 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 transit-peptide release and ATP hydrolysis by a cochaperone 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 co-chaperone homolog in the chloroplast protein translocon. *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, Soltis DE, Doyle JJ, Soltis 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) BnaC. Tic40, a plastid inner membrane translocon originating from *Brassica oleracea*, is essential for tapetal function and microspore development in *Brassica napus*. *Plant J* **68**: 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 resynthesized *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 (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. *Nat Genet* **43**: 476–481
- Huang Z, Chen Y, Yi B, Xiao L, Ma C, Tu J, Fu T (2007) Fine mapping of the recessive genic male sterility gene (*Bnms3*) 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*. *Theor Appl Genet* **105**: 1159–1165
- Lagercrantz U (1998) Comparative mapping between *Arabidopsis thaliana* and *Brassica nigra* indicates that *Brassica* genomes have evolved through extensive genome replication accompanied by chromosome fusions and frequent rearrangements. *Genetics* **150**: 1217–1228
- Lagercrantz U, Lydiat DJ (1996) Comparative genome mapping in *Brassica*. *Genetics* **144**: 1903–1910
- Lamb JR, Tugendreich S, Hieter P (1995) Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem Sci* **20**: 257–259
- Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IAP, Zhao M, Ma J, Yu J, Huang S, et al (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. *Nat Commun* **5**: 3930
- 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
- Livak 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, Udall J, Pires JC, Schranz ME, Osborn T (2004) Genome redundancy and plasticity within ancient and recent *Brassica* crop species. *Biol J Linn Soc Lond* **82**: 675–688
- Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassicaceae. *Genome Res* **15**: 516–525
- 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 triplication. *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
- Nicolas SD, Le Mignon G, Eber F, Coriton O, Monod H, Clouet V, Huteau V, Lostanlen A, Delourme 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**: 487–503
- Osborn TC, Buttrille DV, Sharpe AG, Pickering KJ, Parkin IAP, Parker JS, Lydiat DJ (2003) Detection and effects of a homeologous reciprocal transposition in *Brassica napus*. *Genetics* **165**: 1569–1577
- Panjabi P, Jagannath A, Bisht NC, Padmaja KL, Sharma S, Gupta V, Pradhan AK, Pental D (2008) Comparative mapping of *Brassica juncea* and *Arabidopsis thaliana* using intron polymorphism (IP) markers: homoeologous relationships, diversification and evolution of the A, B and C *Brassica* genomes. *BMC Genomics* **9**: 113
- Parkin IAP, Gulden SM, Sharpe AG, Lukens L, Trick M, Osborn TC, Lydiat DJ (2005) Segmental structure of the *Brassica napus* genome based on comparative analysis with *Arabidopsis thaliana*. *Genetics* **171**: 765–781
- Parkin IAP, Lydiat 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, Lydiat 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, Quijada PA, Lukens LN, Osborn TC (2004) Flowering time divergence and genomic rearrangements in resynthesized *Brassica* polyploids (Brassicaceae). *Biol J Linn Soc Lond* **82**: 675–688
- Rodríguez-Trelles F, Tarrío R, Ayala FJ (2003) Convergent neofunctionalization 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-Olds T (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. *Trends Plant Sci* **11**: 535–542
- Slotte T, Hazzouri KM, Ågren JA, Koenig D, Maumus F, Guo YL, Steige K, Platts AE, Escobar JS, Newman LK, et al (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet* **45**: 831–835
- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis 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 rapa*. *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