Impacts of Whole-Genome Triplication on *MIRNA* Evolution in *Brassica rapa*

Chao Sun[1,2], Jian Wu[1], Jianli Liang[1], James C. Schnable[3], Wencai Yang[2], Feng Cheng[1,*], and Xiaowu Wang[1,*]

[1]Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Zhongguancun Nandajie, Beijing, People's Republic of China

[2]Beijing Key Laboratory of Growth and Developmental Regulation for Protected Vegetable Crops, Department of Vegetable Science, China Agricultural University, Yuanmingyuan Xilu, Beijing, People's Republic of China

[3]Department of Agronomy and Horticulture, University of Nebraska, Lincoln

*Corresponding author: E-mail: wangxiaowu@caas.cn; chengfeng@caas.cn.

Accepted: October 19, 2015

Abstract

MicroRNAs (miRNAs) are a class of short non-coding, endogenous RNAs that play essential roles in eukaryotes. Although the influence of whole-genome triplication (WGT) on protein-coding genes has been well documented in *Brassica rapa*, little is known about its impacts on *MIRNA*s. In this study, through generating a comprehensive annotation of 680 *MIRNA*s for *B. rapa*, we analyzed the evolutionary characteristics of these *MIRNA*s from different aspects in *B. rapa*. First, while *MIRNA*s and genes show similar patterns of biased distribution among subgenomes of *B. rapa*, we found that *MIRNA*s are much more overretained than genes following fractionation after WGT. Second, multiple-copy *MIRNA*s show significant sequence conservation than that of single-copy *MIRNA*s, which is opposite to that of genes. This indicates that increased purifying selection is acting upon these highly retained multiple-copy *MIRNA*s and their functional importance over singleton *MIRNA*s. Furthermore, we found the extensive divergence between pairs of miRNAs and their target genes following the WGT in *B. rapa*. In summary, our study provides a valuable resource for exploring *MIRNA* in *B. rapa* and highlights the impacts of WGT on the evolution of *MIRNA*.

Key words: *MIRNA*, whole-genome triplication, retention, coevolution, *Brassica rapa*.

Introduction

MicroRNAs (miRNAs) are a class of endogenous small RNA sequences, mainly 20–22 nt in length, and regulate gene expression at the transcriptional or posttranscriptional level, by cleaving the transcripts of target genes or inhibiting their translation (Chen 2009; Voinnet 2009; Vazquez et al. 2010; Axtell 2013). MiRNAs are derived from single-stranded RNA precursors which are transcribed by RNA polymerase II, and further processed by the DICER-like complex to generate self-complementary fold-back structures (stem-loop, hairpin or pri-miRNAs) (Ghildiyal and Zamore 2009; Czech and Hannon 2011; Ameres and Zamore 2013). Mature miRNAs are incorporated into RNA-induced silencing complexes, which targets mRNAs and other larger RNAs based on sequence complementarity (Vazquez et al. 2010; Cuperus et al. 2011). Plant miRNAs play crucial regulatory roles in diverse biological processes related to plant development (Mallory and Vaucheret 2006) such as: seed germination (Liu et al. 2007), organ formation (Moxon et al. 2008), developmental timing and patterning (Wu et al. 2013), phase transition (Aukerman and Sakai 2003), and biotic/abiotic stress resistance (Ruiz-Ferrer and Voinnet 2009; Jeong et al. 2011), etc. In many cases, *MIRNA*s (precursor genes of mature miRNAs) can be grouped into families of different loci that encode similar or identical mature miRNAs (Meyers et al. 2008). Previous research reported that only a small part of annotated *MIRNA* families are well conserved in multiple species, whereas the majority are lineage or species-specific (Axtell 2008; Cuperus et al. 2011; Jones-Rhoades 2012). Some plant *MIRNA* families are conserved for hundreds of millions of years in multiple lineages, they have higher expression than less conserved or species-specific miRNAs, possess more paralogous loci, and have a higher diversity of target genes. Unlike highly conserved *MIRNA*s, young *MIRNA* families are often weakly expressed, processed imprecisely, they lack targets, and display patterns of neutral variation. Based on these

© The Author(s) 2015. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

features, previous reports suggest that these young *MIRNA*s are rapidly changing and evolutionarily transient (Fahlgren et al. 2010; Ma et al. 2010), with only a subset of them stabilizing by integrating into regulatory networks (Fahlgren et al. 2007).

Whole-genome duplication (WGD) is widespread in plant genomes and may result in extensive genetic redundancy (Wang, Wang, et al. 2012). At least five ancient WGD events characterize the evolutionary history of the model plant, *Arabidopsis thaliana* (*A. thaliana*) (Cheng, van den Bergh, et al. 2013). Being from the Brassicaceae, the *Brassica* genus experienced the same events, with an additional and relatively recent whole-genome triplication (WGT, hexaploidy) event (Br-α) (Wang et al. 2011). Since this triplication, both the number and expression of genes differentiated among the three subgenomes. The subgenomes that had fewer genes (more highly fractionated) are enriched for genes that were expressed at relatively lower mRNA levels compared with the paralogs (Cheng, Wu, Fang, Sun, et al. 2012). Such phenomenon was defined as "subgenome dominanace," which was also observed in maize (Schnable et al. 2011).

Brassica rapa is a widely cultivated and economically important vegetable and oil crop around the world. The functional analysis of miRNAs in *B. rapa* has been the focus of several recent studies, which is potentially important for the genetic improvement of *B. rapa* crops (Yu et al. 2012). A less conserved miRNA, bra-miR1885, can regulate specific TIR-NBS-LRR transcripts in response to pathogen stress, particularly to turnip mosaic virus (TuMV) infection (He et al. 2008). Bra-miR398a and bra-miR398b are conserved miRNAs involved in heat response (Yu et al. 2012). Bra-miR156 affects the expression of *BraSPL9-2,* which controls the heading time in Chinese cabbage (Wang et al. 2014), whereas bra-miR319a targets TCP genes to modulate head shape in Chinese cabbage (Mao et al. 2014). Considering the important role of miRNAs in the regulation of development of *B. rapa*, several miRNAs and their putative target genes have been identified in *B. rapa*, but an analysis of the evolutionary relationship between miRNAs and their target genes is still lacking. At the time of this study, 157 mature miRNA sequences (127 unique sequences), representing 96 (Yu et al. 2012; Jiang et al. 2014) *MIRNA*s have been released in the latest version of miRBase (version 21, June 2014; http://www.mirbase.org/, last accessed November 04, 2015) (Griffiths-Jones et al. 2008; Kozomara and Griffiths-Jones 2010). Apart from the miRNAs recorded in miRBase, two studies (Kim et al. 2012; Wang, Li, et al. 2012) annotated of 587 nonredundant miRNAs in *B. rapa* by small RNA sequencing and analyzing relative sequenced reads. However, only a limited number of studies have inferred that many *MIRNA*s vary in paralogous gene copy number (deletion of one or more redundant paralogs) in *B. rapa* (Kim et al. 2012; Yu et al. 2012). The additional WGT experienced by *B. rapa* may have important consequences for the evolution of *MIRNA*s, and thus a systematic and comprehensive study is needed to elucidate the subsequent changes.

In this study, we annotated *B. rapa MIRNA*s globally by analyzing small RNA sequencing data, comparing the data with previous studies, and performed microsynteny predictions to the well annotated *MIRNA*s in Arabidopsis. We then investigated the patterns of preservation/conservation of *MIRNA*s from different subgenomes of the palehexaploid genome of *B. rapa*. Our results provide a valuable resource to explore the function of *MIRNA*s and highlight the evolutionary consequences of *MIRNA*s following WGT.

Materials and Methods

Small RNA Sequencing and Annotation

Small RNA sequencing data (Woodhouse et al. 2014) was generated from three tissues (root, stem, and leaf) in *B. rapa* cultivar Chiifu-401/42 and used to annotate *MIRNA*s. We used "cutadapt" to trim both 5′- and 3′-adaptors from sequencing data (Martin 2011) as described (Woodhouse et al. 2014). Clean reads were aligned to the *B. rapa* genome (V1.5; http://brassicadb.org, last accessed November 04, 2015) (Cheng et al. 2011) using Bowtie (Langmead et al. 2009). Only reads that matched perfectly to less than 30 genomic regions in the *B. rapa* genome were used for *MIRNA* annotation. The mapped regions were subjected to secondary hairpin structure prediction using "Mireap" (http://sourceforge.net/projects/mireap/, last accessed November 04, 2015). The predicted structures with less than -18 kcalmol^{-1} free energy, less than 300 nt spaces between the miRNA and its complementary sequence, greater than 16 matched nucleotides, and less than 4 nt bulge of miRNA or its complementary sequence were subjected to further analysis. Only the secondary hairpin structures without any overlap of the annotated coding region and RNAs (tRNA, rRNA, snRNA, and snoRNA) were considered reliable *MIRNA*s.

Integration of *MIRNA* Annotation Data Sets

The mature and precursor miRNA sequences of *B. rapa* from four previous studies (Kim et al. 2012; Wang, Li, et al. 2012; Yu et al. 2012; Jiang et al. 2014) were aligned to the *B. rapa* genome using BLASTN with e-value of 0.001. BLASTN hits with less than 80% coverage of the precursor were removed. If the precursors could be mapped to multiple positions in the genome, the best hit of the precursor was considered as the candidate locus of the *MIRNA*. The position of mature miRNAs was then determined by the position of their precursors. Altogether, we combined four data sets (Kim et al. 2012; Wang, Li, et al. 2012; Yu et al. 2012; Jiang et al. 2014) and these annotation results with our aforementioned data set into a comprehensive *MIRNA* atlas of *B. rapa*.

Microsynteny Comparisons among Species to Identify *B. rapa MIRNAs*

For the microsynteny-based identification of *B. rapa MIRNA*s, the mature miRNA sequences of *A. thaliana*, *Arabidopsis lyrata* were downloaded from the latest version (21.0, June 2014) of miRBase (Griffiths-Jones et al. 2008). These sequences were mapped to the *B. rapa* genome by "PatMan" (Prüfer et al. 2008) with a maximum of two mismatches allowed. The syntenic genes between *B. rapa* and each outgroup, *A. thaliana* and *A. lyrata,* were predicted by SynOrths (Cheng, Wu, Fang, Wang, et al. 2012). Twenty protein-coding genes in the respective upstream and downstream regions of the miRNA queries and mapped miRNA loci were extracted and compared. The syntenic miRNA pairs, defined with at least four syntenic gene pairs among the 40 flanking gene windows, were analyzed further. Then, we extended the flanking genomic sequences (300 nt) around each mapped locus in *B. rapa* genome and submitted them to MIRcheck to determine secondary hairpin structures (Jones-Rhoades and Bartel 2004). The mapped loci which passed the criterion of MIRcheck were considered accurate. The method was also used to determine orthologous *MIRNA*s of annotated *B. rapa MIRNA*s in *A. thaliana*, *A. lyrata*, and *Schrenkiella parvula*, respectively. In this analysis, the *A. thaliana* genome was downloaded from The *Arabidopsis* Information Resource (TAIR10; http://www.Arabidopsis.org/index.jsp, last accessed November 04, 2015) (*Arabidopsis* Genome Initiative 2000). The genomic data set for *A. lyrata* was obtained from the Joint Genome Initiative database (http://genome.jgi-psf.org/Araly1/Araly1.home.html, last accessed November 04, 2015; Hu et al. 2011) and *S. parvula* data set was retrieved from Dassanayake et al. (2011).

MIRNA Nucleotide Divergence Analysis

Syntenic pairs of *MIRNA* precursors between *A. thaliana* and *B. rapa* were divided into five regions: 5′-region, miRNA-5p, region between miRNA-5p and miRNA-3p, miRNA-3p, and 3′-region. The 5′-region was defined as 60 nt upstream of miRNA-5p, whereas the 3′-region was defined as 60 nt downstream of miRNA-3p. Using the *A. thaliana MIRNA*s as reference, we used the strategy reported by Ma et al. (2010) to analyze the nucleotide divergence of single-copy *MIRNA*s, multiple-copy *MIRNA*s, and *MIRNA*s among subgenomes. We divided each of the five regions into seven equal-length bins and calculated the average variation ratio by counting mismatches in the bins between pairwise *MIRNA*s based on MUSCLE alignments (Edgar 2004).

The Prediction of miRNA Targets Genes

MiRNAs bind to target genes with perfect or near-perfect complementary, this nature motivates the prediction of potential targets through computational approaches. We utilized multiple methods to reduce false positive predictions of target genes (Dai et al. 2011; Wang and Adams 2015). Three tools for plant-specific miRNA target prediction, TargetFinder (Fahlgren et al. 2007, 2010), psRNAtarget (Dai and Zhao 2011), and psRobot (Wu et al. 2012) were used to identify the target genes of miRNAs in *A. thaliana* and *B. rapa* under default parameters. The final miRNA target genes were predicted by at least two of the three tools. Specifically, the predicted target genes of miRNAs from *MIR5658*, *MIR5021*, and *MIR414* were discarded because of the massively amplified trinucleotide repeats in the mature sequences (UGA, GAA, and UCA, respectively).

Results

A Comprehensive Annotation of *MIRNA* Genes in the *B. rapa* Genome

To obtain a complete data set of *B. rapa MIRNA*s, we utilized a comprehensive approach to annotate *MIRNA*s, including an analysis of small RNA sequencing reads, integrating the previous *MIRNA*s data sets of *B. rapa*, and a microsynteny relationship to well-annotated *MIRNA*s in *Arabidopsis*.

Previous sRNA-seq data (Woodhouse et al. 2014) from three tissues (root, stem, and leaf) of the reference genome *B. rapa* Chiifu was used to annotate *B. rapa MIRNA*s. After adaptor trimming, the numbers of unique and total reads were calculated and mapped to the *B. rapa* genome (supplementary table S1, Supplementary Material online). In summary, the MIRNA size distributions share similar patterns between unique reads (supplementary fig. S1*A*, Supplementary Material online) and total reads (supplementary fig. S1*B*, Supplementary Material online), with 24 nt small RNAs being the most abundant, followed by 23 and 21 nt small RNAs, consistent with observations in *Brassica napus* (Huang et al. 2013). Approximately 3–4.6 million reads from each tissue were mapped perfectly to the *B. rapa* genome and were used for further analysis. After performing secondary structure predictions with the help of Mireap, we filtered the secondary hairpin structures of coding region and RNAs (tRNA, rRNA, snRNA, and snoRNA) in *B. rapa* (Materials and Methods). Finally, a total of 326 *MIRNA*s in the *B. rapa* genome were found with perfect hairpin structures, as expected by miRNA. Among these *MIRNA*s, 161 (49.4%), 50 (15.3%), and 115 (35.2%) are expressed as mature miRNAs in all three tissues, two tissues and only one tissue, respectively. The percentage of expressed tissue-specific *MIRNA*s in three tissues is 9.2–13.8%, with the leaf tissue expressing the most miRNAs.

We merged our annotated *MIRNA*s with previously reported data sets to include miRNAs that are expressed in other tissues or in different conditions. We compared our results with four previously reported *MIRNA* data sets in *B. rapa* (Kim et al. 2012; Wang, Li, et al. 2012; Yu et al. 2012; Jiang et al. 2014) and integrated them with our data set into a comprehensive *MIRNA*s annotation (Materials and

Methods, supplementary table S2, Supplementary Material online). In total, 670 *MIRNA*s were combined into the *MIRNA*s annotation (fig. 1). While almost one-third of the *MIRNA*s, 202 out of 670, were present in at least two data sets, only 20 (2.99%) were in all five data sets, indicative of spatial-, time-, or condition-specific expression.

As a complement to de novo *MIRNA* annotation based on sequence, interspecific microsynteny comparisons of *MIRNA* loci have also been widely used in *MIRNA* analysis (Maher et al. 2006; Fahlgren et al. 2010; Ma et al. 2010; Shen et al. 2014). The numerous studies of *MIRNA* in *A. thaliana* and *A. lyrata* provide abundant information to annotate *B. rapa MIRNA*s orthologues. Candidate *MIRNA* loci in *B. rapa* were found using the microsynteny relationships of *B. rapa*, *A. thaliana*, and *A. lyrata.* The candidates *MIRNA*s were then submitted to MIRcheck to identify the secondary structures (Jones-Rhoades and Bartel 2004). The *MIRNA*s that remained after MIRcheck were considered as reliable. A total of 181 orthologous syntenic *MIRNA*s between *B. rapa* and *Arabidopsis* were detected (supplementary fig. S2, Supplementary Material online). Intriguingly, only ten *MIRNA*s were not present in the integrated annotation, but raise the total number of annotated *MIRNAs* in *B. rapa* to 680. Among these ten *MIRNA*s, mature miRNAs of three (*MIR4245a*, *MIR840a*, and *MIR848a*) were expressed under a specific condition or treatment in *Arabidopsis* (Rajagopalan et al. 2006; Fahlgren et al. 2007; Moldovan et al. 2010; Zhang et al. 2012; Vidal et al. 2013).

Altogether, we curated a comprehensive annotation of 680 *MIRNA*s (supplementary table S2, Supplementary Material online) in *B. rapa*, with precise and accurate genome locations that, to our knowledge, constitutes the most extensive effort to characterize the complete *B. rapa MIRNA* genes to date.

Retention Bias of *B. rapa MIRNAs* Following WGT

The complete *MIRNA* annotation was subjected to an interspecific microsynteny comparison with three diploid species *A. lyrata*, *A. thaliana*, and *S. parvula*. These three species were selected as outgroups due to their prominence in the evolutionary history of Brassicaceae, as *A. lyrata* and *S. parvula* were reported to represent the ancestral karyotypes of Brassicaceae family (Lysak et al. 2006; Schranz et al. 2006; Mandáková and Lysak 2008) and *Brassica* tribe (Cheng, Mandáková, et al. 2013). *Arabidopsis thaliana* is also part of the Brassicaceae and a model dicot system with thorough *MIRNA*s studies. All of the aforementioned annotated *B. rapa MIRNA*s were used for microsynteny comparisons among these four species. From these 680 *MIRNAs*, we found that 204 retained a syntenic relationship to 97 nonredundant *MIRNA* loci from the three diploid species. The 204 *MIRNAs* belong to 178 *B. rapa MIRNA* loci, with 42 *MIRNA*s compressed into 16 *MIRNA* tandem arrays (Maher et al. 2006) (supplementary table S3, Supplementary Material online). Among the 97 nonredundant

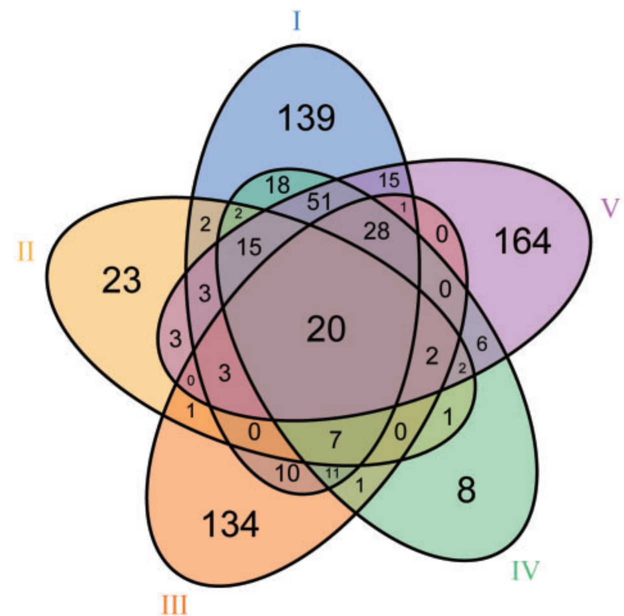

Fig. 1.—Distribution of integrated *B. rapa MIRNA*s among five data sets. Distribution of *B. rapa MIRNA*s in the five data sets (I, our data set; II, data set from Jiang et al. [2014]; III, data set from Wang, Li, et al. [2012]; IV, data set from Yu et al. [2012]; V, data set from Kim et al. [2012]). The numbers of *MIRNA*s present in each data set are given in the individual sections.

MIRNA loci in the diploid species, 40 had one, 33 had two, and 24 had three syntenic orthologs (1*40 + 2*33 + 3*24 = 178 orthologs) from the three subgenomes of *B. rapa*. The ratio of single-copy to multiple-copy (duplicated/tripled) *MIRNA* (0.70; 40/57) was significantly lower (P value = 0.002, χ^2 test) than that of genes in the whole genome (1.34, 9,175/6,836. Wang et al. 2011; Li et al. 2014), indicating that *MIRNA*s were more highly retained following the fractionation subsequent to WGT. We then compared the retention of *MIRNA*s to their neighbor gene loci (tandem genes were considered as a locus). As shown in figure 2, it is clear that *MIRNA*s have a higher ratio of retention than that of their neighbor gene loci. We also found that the *MIRNA*s' retention was associated with the retention of local genomic fragments. Taking the first gene adjacent to an individual *MIRNA* as an example, of the 114 genes flanking the 57 aforementioned multiple copy *MIRNA*s, 47 (41.2%) retained multiple copies and 67 (58.8%) had only one copy. For the 80 genes flanking the 40 single-copy *MIRNA*s, 26 (32.5%) were multiple-copy and 54 (67.5%) were single-copy. The retention ratio of genes flanking the multiple-copy *MIRNA*s was higher than that of genes flanking the single-copy *MIRNA*s.

*MIRNA*s showed a biased distribution in the three subgenomes of *B. rapa*. Among the 178 syntenic *MIRNA* loci in *B. rapa*, 75 *MIRNA* loci were located in subgenome LF (the least fractionated subgenome), 54 at MF1 (the more fractionated subgenome one), and 49 at MF2 (the more fractionated

subgenome two) (fig. 3). Subgenome LF retained more *MIRNA*s than the other two subgenomes, which is similar (*P* value = 0.807) to the distribution of genes in the three subgenomes (LF: 17504; MF1: 12543; MF2: 10354) (Wang et al. 2011; Li et al. 2014). It indicates that subgenome fractionation did not discriminate between *MIRNA*s and genes (Cheng, Wu, Fang, Sun, et al. 2012), which provides additional support for subgenome dominance following WGT in *B. rapa*.

Multiple-Copy *MIRNAs* Are More Conserved than Single-Copy *MIRNAs*

The nucleotide variation of *MIRNA* in *B. rapa* was estimated by sequence comparisons with orthologous *MIRNA*s in *A. thaliana*. Nucleotide divergence was measured independently for five regions of *MIRNA*s (fig. 4*A*): 5′-region, miRNA-5p, region between miRNA-5p and miRNA-3p, miRNA-3p, and the 3′-region. Each region was divided into seven bins with equal sizes to perform the variation analysis. The lowest divergence was observed within the miRNA-5p/miRNA-3p region for single-copy and multiple-copy *MIRNA*s, reflecting purifying selection to maintain the stem-loop secondary structure and complementary sequences with target genes. In contrast, the other *MIRNA* three regions were relatively unconstrained with higher levels of divergence over miRNA-5p/miRNA-3p, indicating that these parts were under fewer evolutionary constraints than that of miRNA-5p/miRNA-3p (fig. 4*B*).

Although the patterns of divergence in the regions of single-copy and multiple-copy *MIRNA*s were qualitatively similar, there was clearly more divergence in the single-copy *MIRNA*s, especially in sequences of mature miRNA-5p (fig. 4*B*). This observation indicates that the single-copy *MIRNA*s may have less constraint to keep the function of mature miRNA sequences, which is in contrast to the genic observation (De Smet et al. 2013). In another analyses, no significant difference in sequence variability was found for *MIRNA*s located in different subgenomes (fig. 4*C*), indicating constraint following WGT. In conclusion, our results reveal that multiple-copy *MIRNA*s are evolutionarily more constrained than single-copy *MIRNA*s.

Functional Divergence of miRNA Target Genes

MiRNAs have been found to target the genes of transcription factors (Willmann and Poethig 2007; Chen 2009; Voinnet 2009; Zhang et al. 2009; Li et al. 2010; Luo et al. 2013), and thus can have a broad influence on the biological functions in *B. rapa*. Therefore, it is important to identify the miRNA target genes to characterize function. Plant miRNAs have perfect or near-perfect complementary sequences to their target genes (Mallory and Vaucheret 2006; Axtell 2013) which allow many computational approaches to predict target genes. The combination of different computational tools to predict miRNA targets is beneficial in that false results are minimized (Dai et al. 2011; Wang and Adams 2015). Here, we generated reliable pairs of miRNAs and target genes with replicated reports from at least two plant-specific target prediction tools. Based on this criterion, a total of 2,559 pairs of miRNA family target genes were identified and submitted to further analysis, including 2,396 unique putative genes targeted by 689 miRNAs that belong to 255 miRNA families. Among the 2,396 predicted target genes, 1,995 (83.3%) have functional InterPro annotations (Zdobnov and Apweiler 2001). We found that 72 miRNA families corresponded to 227 miRNAs targeted to 277 transcription factors genes (supplementary table S4, Supplementary Material online). This suggests that these miRNA families play important roles in posttranscriptional regulation and transcription networks. The other targeted genes were involved in diverse physiological and metabolic processes, including protein kinases, ATPase, Cyclin-like F-box, Zinc finger, DNA/RNA helicase, DEAD-like helicase, and Cytochrome P450, etc (supplementary table S4, Supplementary Material online).

To gain a better understanding of the functional roles of the predicted miRNA target genes in *B. rapa*, we annotated miRNAs with Gene Ontology (GO) categories for the current *B. rapa* genome version (1.5). Of the predicted targets, 1,431 (59.7 %) had GO assignments. We found that miRNA families preferentially target genes involved in a wide spectrum of regulatory functions and biological processes, including gene expression/transcription, metabolism, transport, signal transduction, and translation, etc (supplementary table S5, Supplementary Material online).

Coevolution of *MIRNAs* and Target Genes

The *MIRNA*s and their target genes are fractionated in *B. rapa* after the WGT event. The gain and loss of pairwise relationships between miRNAs and target genes are two important processes in the coevolution of miRNAs and their targets (Guo et al. 2008; Wang and Adams 2015). The systematic annotation of *MIRNA*s and the determination of their target genes in *B. rapa* provide an opportunity to study the gain/loss dynamics of their pairwise relationships following a WGT event. The 1,236 pairs of miRNA family targets in *A. thaliana* were used as a reference to assess their conservation in *B. rapa*. Orthologs of *A. thaliana* miRNAs and their target genes were determined in *B. rapa* with SynOrths (Cheng, Wu, Fang, Wang, et al. 2012). In total, out of the 1,236 *A. thaliana* miRNA-target pairs, only 111 (~10%) were retained in *B. rapa*, with 71 pairs whose target genes have paralogs and 40 with single-copy target genes (fig. 5). These results show the fast evolution and fractionation of this interactive relationship between miRNAs and target genes in *B. rapa*. As mentioned above, among the 71 multiple-copy target genes, 41 pairs keep miRNA binding sites on all retained paralogs, suggesting that these genes most developed their miRNA binding sites

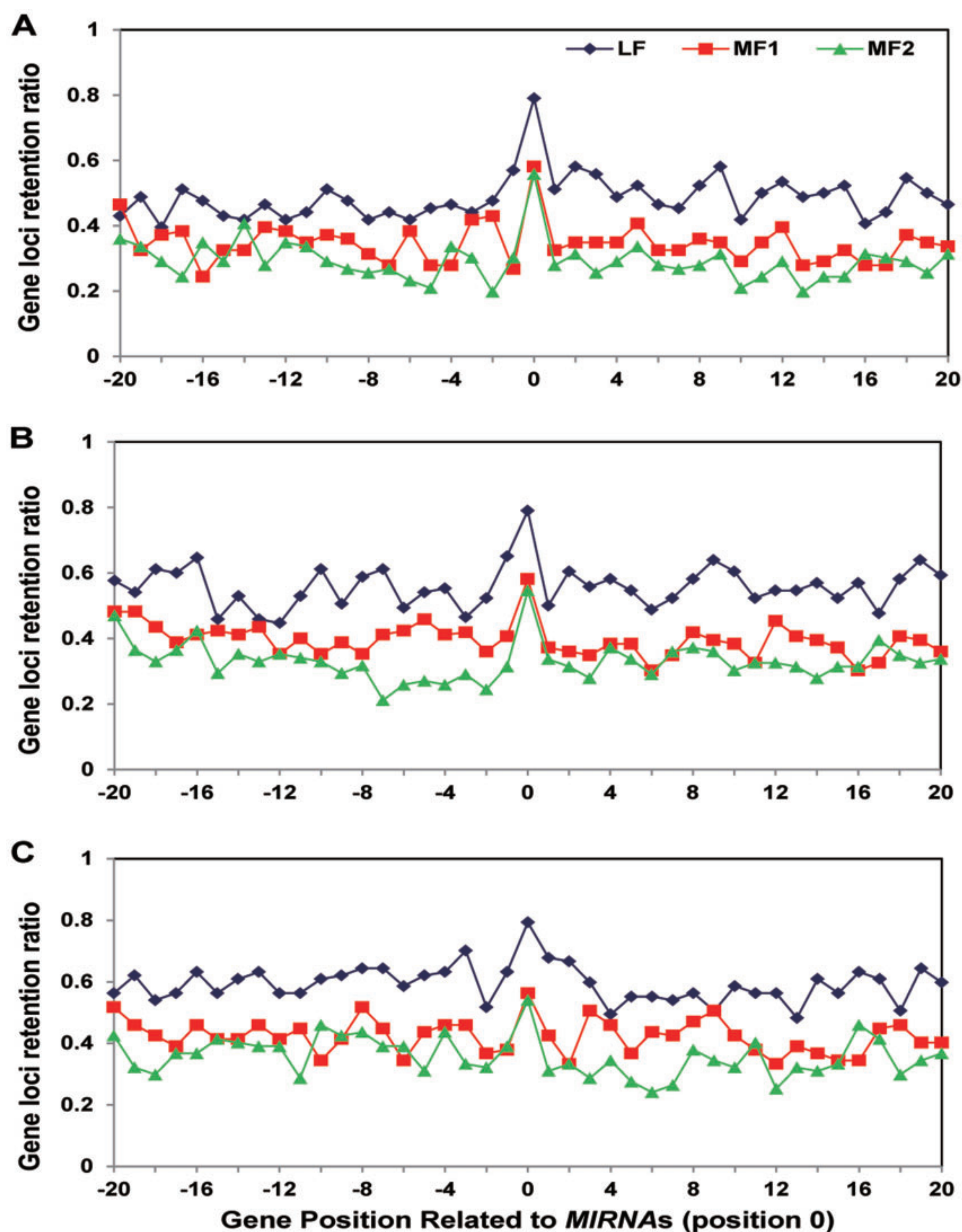

Fig. 2.—Retention rates of *MIRNA*s and their neighboring gene loci (20 flanking genes on either side of the *MIRNA*s) in *B. rapa*. Retention of orthologs among the *MIRNA*s (position 0) and their immediate neighbors in the three subgenomes of *B. rapa* using *MIRNA*s of *A. lyrata* (*A*), *A. thaliana* (*B*), and *S. parvula* (*C*) as reference. Blue line indicates subgenome LF, red line indicates subgenome MF1, and green line indicates subgenome MF2.

prior to WGT and maintained them subsequently. For the remaining 30 multiple-copy genes, only one or two (not all) paralog(s) were targeted by miRNAs, indicative of the gain or loss of miRNA target sites after WGT. These observations reveal the coevolutionary relationship between miRNAs and their target genes following WGT in *B. rapa*, and the rapid divergence of miRNA binding sites during this process.

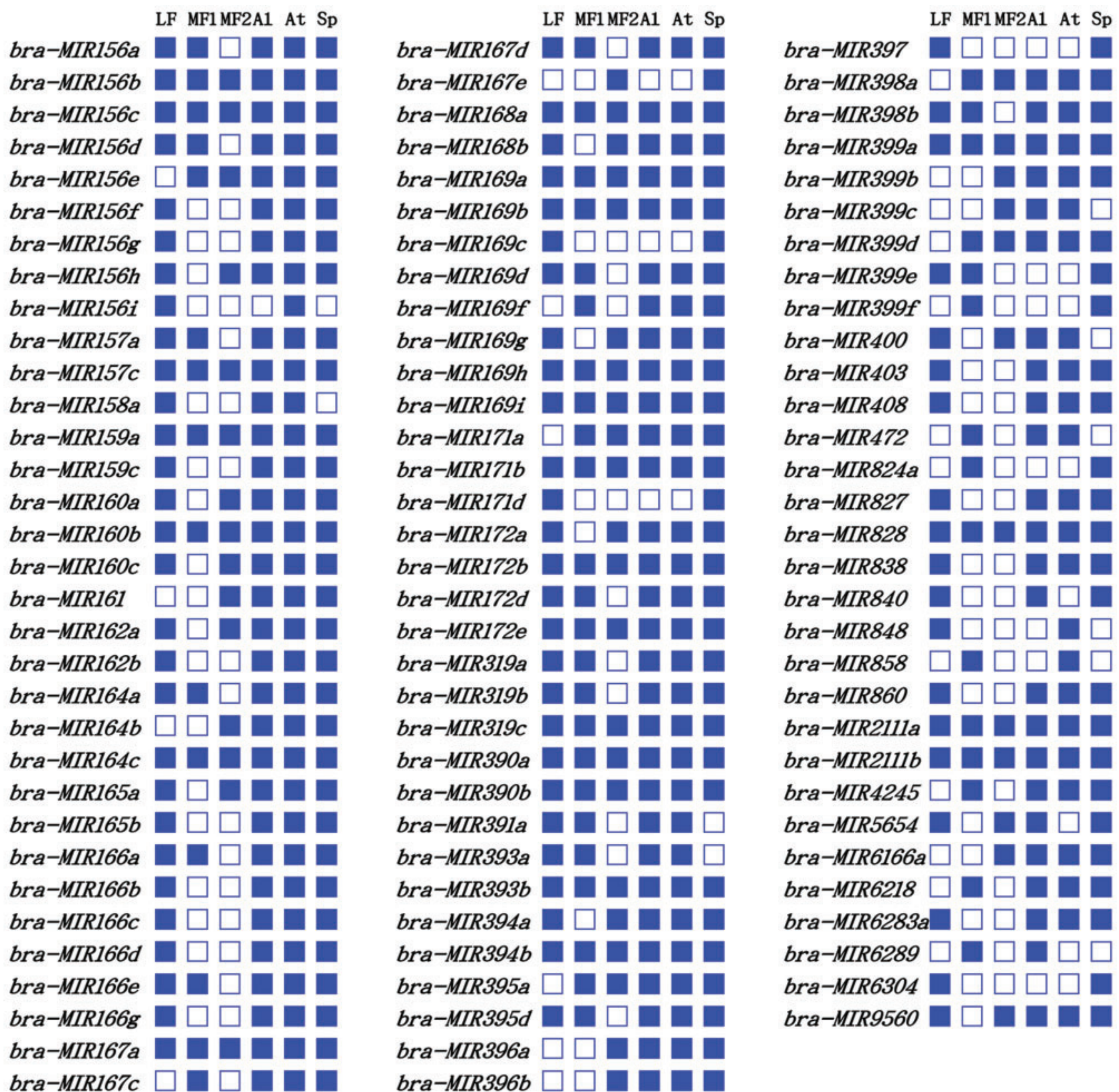

Fig. 3.—Subgenome distribution of syntenic *MIRNA*s among *B. rapa*, *A. lyrata* (Al), *A. thaliana* (At), and/or *S. parvula* (Sp). A solid square indicates that the *MIRNA* exists in the relative subgenomes or species. The hollow square indicates that the *MIRNA* does not exist in the relative subgenomes or species.

Discussion

In this study, we carried out a comprehensive annotation of *B. rapa MIRNA*s through high-throughput small RNA sequencing, integrating previously reported *B. rapa MIRNA*s, and interspecies microsynteny comparisons to well-annotated *MIRNA*s in *Arabidopsis*. A complete data set of *B. rapa MIRNA*s was generated, which serves as a valuable resource to promote the studies of *MIRNA*s in *B. rapa*. Based on this, we further performed a systematic analysis of *MIRNA* evolution in the mesohexaploidized genome of *B. rapa*.

MiRNAs are essential regulators of gene expression in both plants and animals (Axtell et al. 2011). According to the latest miRBase database (Release 21), 325 and 205 *MIRNA*s were annotated in genomes of *A. thaliana* and *A. lyrata*, respectively. *Brassica rapa* is a paleohexaploid with a much larger and more complex genome than that of *Arabidopsis*. In this study, we annotated 969 miRNAs (765 unique sequences) from 680 *MIRNA* genes in *B. rapa*, which were about twice the number in *A. thaliana*. This result is consistent with the nature of the larger polyploid genome of *B. rapa*. Among the 680 *MIRNA*s,

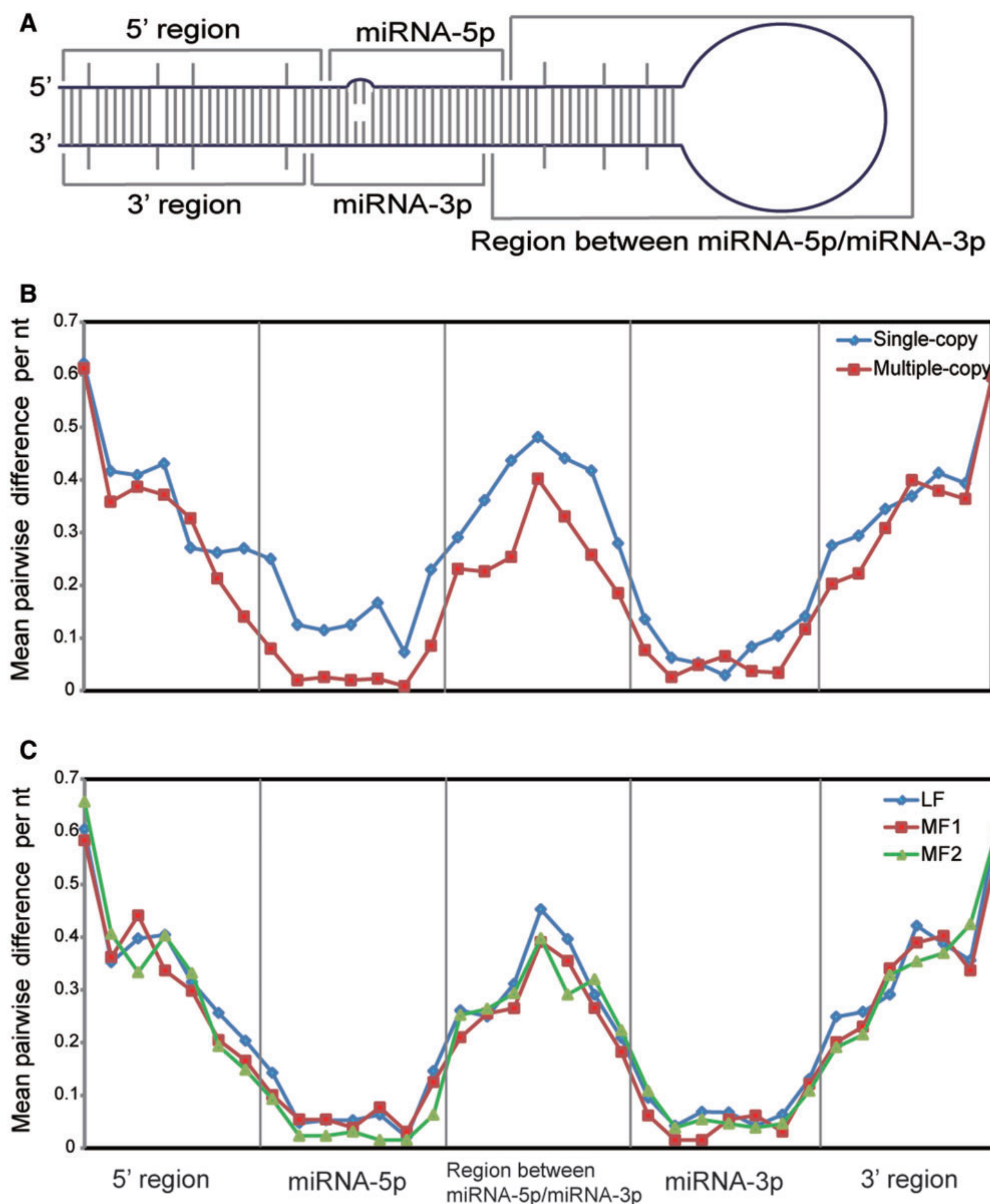

Fig. 4.—Nucleotide divergence within fold-backs of *MIRNA*s in *B. rapa*. (*A*) *MIRNA* fold-backs were divided into five regions, and miRNA can be miRNA-5p or miRNA-3p. (*B*) Average sequence divergence between single-copy and multiple-copy *MIRNA* fold-backs. Single-copy *MIRNA* fold-backs are more diverged than that of multiple-copy *MIRNA* in both miRNA-5p region and region between miRNA-5p and miRNA-3p. (*C*) Average sequence divergence among *MIRNA* fold-backs located at different subgenomes of *B. rapa*. Similar variation patterns of *MIRNA* fold-backs were found across all the five regions. The fold-backs of orthologous *MIRNA*s in *A. thaliana* were used as references.

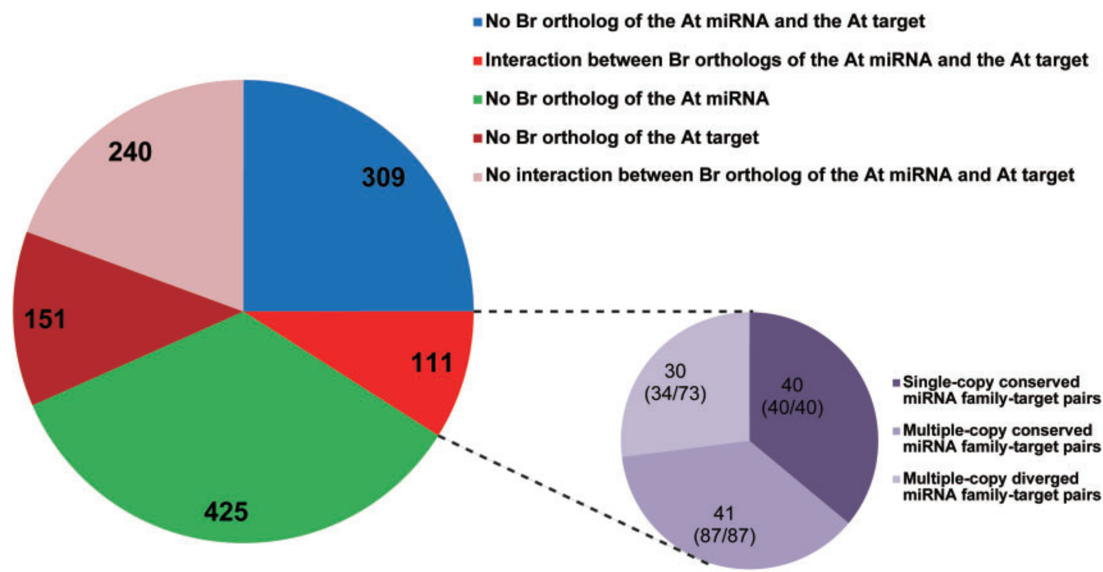

Fig. 5.—Conservation of miRNA and target gene pairs in *B. rapa*, the orthologous pairs of miRNA and target genes in *A. thaliana* (At) were used as reference data set. The left chart indicates the retention of miRNA-target genes following WGT in *B. rapa* (Br). The right pie chart indicates the detailed changes (conservation or divergence) of retained pairs of miRNA family target genes. The numbers in the brackets show the ratio of conserved copies to the total copies of target genes.

only 204 (~30%) *MIRNA*s had orthologous *MIRNA*s in *A. thaliana*, *A. lyrata*, or *S. parvula*. The large number of *B. rapa*-specific *MIRNA*s may indicate the expansion of recently evolved *MIRNA*s that undergoing frequent births and deaths (Fahlgren et al. 2007, 2010; Ma et al. 2010) after the *Arabidopsis–Brassica* divergence and *Schrenkiella–Brassica* divergence.

As a frequently used complement to small RNA sequencing annotation, microsynteny-based methods have been widely used in the identification of plant *MIRNA*s (Maher et al. 2006; Ma et al. 2010; Shen et al. 2014). It detected *MIRNA*s with low expression or condition-specific expression in specific pathways. In our study, three *MIRNA*s (*bra-MIR4245a*, *bra-MIR840a*, and *bra-MIR848a*) that were not detected by the small RNA sequencing were found using microsynteny comparisons. In *Arabidopsis*, miR4245 plays a role in the regulation of the jasmonate pathway by targeting AT1G57620, AT5G11020 (emp24/gp25L/p24 family protein, kinase), and AT5G46470 (disease resistance protein, TIR-NBS-LRR class) (Zhang et al. 2012). MiR840 targets the WHIRLY transcription factor *AT2G02740* (Rajagopalan et al. 2006; Fahlgren et al. 2007) and may take part in the process of nitrate regulation, coordination of carbon, and nitrogen metabolism (Vidal et al. 2013). MiR848 plays a role in gene regulation and developmental responses to hypoxia, which is likely to be dependent on mitochondrial function (Moldovan et al. 2010). Based on the syntenic relationships between *B. rapa* and *Arabidopsis*, complete or similar functions can be inferred for *MIRNA*s in *B. rapa*. With the development of sequencing technology, more and more plant genomes will be sequenced, and the abundant information of annotated *MIRNA*s in model species can be applied to predict *MIRNA*s of newly sequenced species using interspecies microsynteny comparisons.

Following WGT, the deletion of one or more *MIRNA* paralogs among subgenomes resulted in differential conservation *MIRNA* copies (Kim et al. 2012). Our observation of biased fractionation in *MIRNA*s among the three subgenomes in *B. rapa* genome provided further evidence for the subgenome dominance phenomenon in polyploids (Schnable et al. 2011; Wang et al. 2011; Cheng, Wu, Fang, Sun, et al. 2012). This findings is consistent with previous studies of grass *MIRNA*s (Abrouk et al. 2012) and recent studies on duplicated gene retention (some miRNA targets) following whole-genome duplications in many plant genomes (Thomas et al. 2006; Freeling 2009; Throude et al. 2009; Schnable et al. 2011, 2012; Cheng, Wu, Fang, Sun, et al. 2012). Compared with total genes in the whole genome or flanking genes of *MIRNA*s, the *MIRNA*s were highly retained after WGT, suggestive of the higher conservation and important regulatory functions of *MIRNA*s. In *B. rapa*, genes in the least fractionated subgenome (LF) were dominantly expressed over the genes in more fractionated subgenomes (MFs: MF1 and MF2) (Cheng, Wu, Fang, Sun, et al. 2012). However, the expression among *MIRNA* paralogs in three subgenomes could not be distinguished due to their identical or highly similar mature sequences. Besides, the nucleotide variation of paralogous *MIRNA*s was not significantly different among three subgenomes, although it was different from genes (Cheng, Wu, Fang, Sun, et al. 2012). One possible explanation is that the *MIRNA*s are more sensitive to mutations than genes

and need higher conservation to maintain their regulatory roles.

Because multiple-copy *MIRNA*s could produce identical miRNAs to regulate the same target gene, whereas a single-copy miRNA could regulate expression of multiple target genes, it is complex to study the coevolution of *MIRNA*s and their target genes, especially in polyploid genomes. To simplify the analysis of this coevolutionary relationship between miRNAs and their target genes in *B. rapa*, we only extracted genes regulated by miRNAs from a single *MIRNA* locus in the 97 nonredundant loci. We found that target genes of *MIRNA* multiples tend to be retained as multiples rather than singletons, whereas target genes of *MIRNA* singletons tend to be retained as singletons. Such a tendency was coincident with previous research in soybean (Zhao et al. 2015), which suggested coevolution between *MIRNA*s and their target genes. Additionally, the massive gain and/or loss of miRNA binding sites among paralogs demonstrated the divergence of the miRNA-target genes in *B. rapa*, which may be an important factor that affects gene divergence following WGT (Guo et al. 2008; Wang and Adams 2015). These analyses highlight the role of miRNA regulation in the evolutionary fate of duplicated genes following WGT.

Conclusion

Our work provides a valuable resource for *B. rapa MIRNA*s and systematic information of *MIRNA* evolution following WGT in *B. rapa*. The combined *MIRNA* annotation led to the development of a comprehensive data set of 969 miRNAs from 680 *MIRNA*s in *B. rapa*. Furthermore, *MIRNA*s are more likely to be retained than genes following WGT, while biased distributions of *MIRNA*s among subgenomes is similar to that of genes. Interestingly, compared with singleton *MIRNA*s, the multiple-copy *MIRNA*s are under strong purifying selection during evolution, indicating their functional importance to be highly retained and more conserved than the singleton *MIRNA*s. Together, our comprehensive annotation and evolutionary analysis of *MIRNA*s in *B. rapa* contributes a clear picture of species-specific *MIRNA* evolution and a greater understanding of the impact that whole-genome duplication has on *MIRNA* evolution.

Supplementary Material

Supplementary figures S1 and S2 and tables S1–S5 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the 973 program 2012CB113900 to X.W. and F.C.; the 973 program 2013CB127000 and the 863 program 2012AA100101 to J.W.; the National Natural Science Foundation of China NSFC grant 31301771 to F.C., 31272179 to J.W. and 31301784 to J.L. Research was carried out in the Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture, P. R. China and Key Laboratory of Biology and Genetic Improvement of Root and Tuber Crops, Ministry of Agriculture, P. R. China. The work was also funded by the Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences. The authors thank Bo Liu, Xiaobo Wang, and Yuan Ma for their helpful suggestions.

Literature Cited

Abrouk M, et al. 2012. Grass microRNA gene paleohistory unveils new insights into gene dosage balance in subgenome partitioning after whole-genome duplication. Plant Cell 24:1776–1792.

Ameres SL, Zamore PD. 2013. Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol. 14:475–488.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796.

Aukerman MJ, Sakai H. 2003. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes. Plant Cell 15:2730–2741.

Axtell MJ. 2008. Evolution of microRNAs and their targets: are all microRNAs biologically relevant? Biochim Biophys Acta 1779:725–734.

Axtell MJ. 2013. Classification and comparison of small RNAs from plants. Annu Rev Plant Biol. 64:137–159.

Axtell MJ, Westholm JO, Lai EC. 2011. Vive la difference: biogenesis and evolution of microRNAs in plants and animals. Genome Biol. 12:221.

Chen X. 2009. Small RNAs and their roles in plant development. Annu Rev Cell Dev Biol. 25:21–44.

Cheng F, et al. 2011. BRAD, the genetics and genomics database for *Brassica* plants. BMC Plant Biol. 11:136.

Cheng F, Mandáková T, et al. 2013. Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. Plant Cell 25:1541–1554.

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

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

Cheng S, van den Bergh E, et al. 2013. The Tarenaya hassleriana genome provides insight into reproductive trait and genome evolution of crucifers. Plant Cell 25:2813–2830.

Cuperus JT, Fahlgren N, Carrington JC. 2011. Evolution and functional diversification of MIRNA genes. Plant Cell 23:431–442.

Czech B, Hannon GJ. 2011. Small RNA sorting: matchmaking for Argonautes. Nat Rev Genet. 12:19–31.

Dai X, Zhao PX. 2011. psRNATarget: a plant small RNA target analysis server. Nucleic Acids Res. 39:W155–W159.

Dai X, Zhuang Z, Zhao PX. 2011. Computational analysis of miRNA targets in plants: current status and challenges. Brief Bioinform. 12:115–121.

Dassanayake M, et al. 2011. The genome of the extremophile crucifer *Thellungiella parvula*. Nat Genet. 43:913–918.

De Smet R, et al. 2013. Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci U S A. 110:2898–2903.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Fahlgren N, et al. 2007. High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS One 2:e219.

Fahlgren N, et al. 2010. MicroRNA gene evolution in *Arabidopsis lyrata* and *Arabidopsis thaliana*. Plant Cell 22:1074–1089.

Freeling M. 2009. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. Annu Rev Plant Biol. 60:433–453.

Ghildiyal M, Zamore PD. 2009. Small silencing RNAs: an expanding universe. Nat Rev Genet. 10:94–108.

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. 2008. miRBase: tools for microRNA genomics. Nucleic Acids Res. 36:D154–D158.

Guo X, et al. 2008. Selection and mutation on microRNA target sequences during rice evolution. BMC Genomics 9:454.

He XF, Fang YY, Feng L, Guo HS. 2008. Characterization of conserved and novel microRNAs and their targets, including a TuMV-induced TIR–NBS–LRR class R gene-derived novel miRNA in *Brassica*. FEBS Lett. 582:2445–2452.

Hu TT, et al. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet. 43:476–481.

Huang D, Koh C, Feurtado JA, Tsang EW, Cutler AJ. 2013. MicroRNAs and their putative targets in *Brassica napus* seed maturation. BMC Genomics 14:140.

Jeong DH, et al. 2011. Massive analysis of rice small RNAs: mechanistic implications of regulated microRNAs and variants for differential target RNA cleavage. Plant Cell 23:4185–4207.

Jiang J, Lv M, Liang Y, Ma Z, Cao J. 2014. Identification of novel and conserved miRNAs involved in pollen development in *Brassica campestris* ssp. chinensis by high-throughput sequencing and degradome analysis. BMC Genomics 15:146.

Jones-Rhoades MW. 2012. Conservation and divergence in plant microRNAs. Plant Mol Biol. 80:3–16.

Jones-Rhoades MW, Bartel DP. 2004. Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell. 14:787–799.

Kim B, et al. 2012. Identification and profiling of novel microRNAs in the *Brassica rapa* genome based on small RNA deep sequencing. BMC Plant Biol. 12:218.

Kozomara A, Griffiths-Jones S. 2010. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 39(Database issue):D152–157.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.

Li J, et al. 2014. Expression profiling reveals functionally redundant multiple-copy genes related to zinc, iron and cadmium responses in *Brassica rapa*. New Phytol. 203:182–194.

Li YF, et al. 2010. Transcriptome-wide identification of microRNA targets in rice. Plant J. 62:742–759.

Liu PP, et al. 2007. Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages. Plant J. 52:133–146.

Luo Y, Guo Z, Li L. 2013. Evolutionary conservation of microRNA regulatory programs in plant flower development. Dev Biol. 380:133–144.

Lysak MA, et al. 2006. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. Proc Natl Acad Sci U S A. 103:5224–5229.

Ma Z, Coruh C, Axtell MJ. 2010. *Arabidopsis lyrata* small RNAs: transient MIRNA and small interfering RNA loci within the *Arabidopsis* genus. Plant Cell 22:1090–1103.

Maher C, Stein L, Ware D. 2006. Evolution of *Arabidopsis* microRNA families through duplication events. Genome Res. 16:510–519.

Mallory AC, Vaucheret H. 2006. Functions of microRNAs and related small RNAs in plants. Nat Genet. 38:S31–S36.

Mandáková T, Lysak MA. 2008. Chromosomal phylogeny and karyotype evolution in x= 7 crucifer species (Brassicaceae). Plant Cell 20:2559–2570.

Mao Y, et al. 2014. microRNA319a-Targeted *Brassica rapa* ssp. pekinensis TCP genes modulate head shape in chinese cabbage by differential cell division arrest in leaf regions. Plant Physiol. 164:710–720.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17:10–12.

Meyers BC, et al. 2008. Criteria for annotation of plant MicroRNAs. Plant Cell 20:3186–3190.

Moldovan D, et al. 2010. Hypoxia-responsive microRNAs and trans-acting small interfering RNAs in *Arabidopsis*. J Exp Bot. 61:165–177.

Moxon S, et al. 2008. Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. Genome Res. 18:1602–1609.

Prüfer K, et al. 2008. PatMaN: rapid alignment of short sequences to large databases. Bioinformatics 24:1530–1531.

Rajagopalan R, Vaucheret H, Trejo J, Bartel DP. 2006. A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. Genes Dev. 20:3407–3425.

Ruiz-Ferrer V, Voinnet O. 2009. Roles of plant small RNAs in biotic stress responses. Annu Rev Plant Biol. 60:485–510.

Schnable JC, Freeling M, Lyons E. 2012. Genome-wide analysis of syntenic gene deletion in the grasses. Genome Biol Evol. 4:265–277.

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

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.

Shen D, et al. 2014. Identification and characterization of microRNAs in oilseed rape (*Brassica napus*) responsive to infection with the pathogenic fungus Verticillium longisporum using *Brassica* AA (*Brassica rapa*) and CC (*Brassica oleracea*) as reference genomes. New Phytol. 204:577–594.

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

Throude M, et al. 2009. Structure and expression analysis of rice paleo duplications. Nucleic Acids Res. 37(4):1248–1259.

Vazquez F, Legrand S, Windels D. 2010. The biosynthetic pathways and biological scopes of plant small RNAs. Trends Plant Sci. 15:337–345.

Vidal EA, et al. 2013. Integrated RNA-seq and sRNA-seq analysis identifies novel nitrate-responsive genes in *Arabidopsis thaliana* roots. BMC Genomics 14:701.

Voinnet O. 2009. Origin, biogenesis, and activity of plant microRNAs. Cell 136:669–687.

Wang F, Li L, et al. 2012. High-throughput sequencing discovery of conserved and novel microRNAs in Chinese cabbage (*Brassica rapa* L. ssp. pekinensis). Mol Genet Genomics. 287:555–563.

Wang S, Adams KL. 2015. Duplicate gene divergence by changes in microRNA binding sites in *Arabidopsis and Brassica*. Genome Biol Evol. 7:646–655.

Wang X, et al. 2011. The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet. 43:1035–1039.

Wang Y, Wang X, Paterson AH. 2012. Genome and gene duplications and gene expression divergence: a view from plants. Ann N Y Acad Sci. 1256:1–14.

Wang Y, Wu F, Bai J, He Y. 2014. BrpSPL9 (*Brassica rapa* ssp. pekinensis SPL9) controls the earliness of heading time in Chinese cabbage. Plant Biotechnol J. 12:312–321.

Willmann MR, Poethig RS. 2007. Conservation and evolution of miRNA regulatory programs in plant development. Curr Opin Plant Biol. 10:503–511.

Woodhouse MR, et al. 2014. Origin, inheritance, and gene regulatory consequences of genome dominance in polyploids. Proc Natl Acad Sci U S A. 111:5283–5288.

Wu HJ, Ma YK, Chen T, Wang M, Wang XJ. 2012. PsRobot: a web-based plant small RNA meta-analysis toolbox. Nucleic Acids Res. 40(Web Server issue):W22–28.

Wu L, et al. 2013. Regulation of FLOWERING LOCUS T by a microRNA in *Brachypodium distachyon*. Plant Cell 25:4363–4377.

Yu X, et al. 2012. Identification of conserved and novel microRNAs that are responsive to heat stress in *Brassica rapa*. J Exp Bot. 63:1025–1038.

Zdobnov EM, Apweiler R. 2001. InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17:847–848.

Zhang B, Jin Z, Xie D. 2012. Global analysis of non-coding small RNAs in *Arabidopsis* in response to jasmonate treatment by deep sequencing technology. J Integr Plant Biol. 54:73–86.

Zhang L, et al. 2009. A genome-wide characterization of microRNA genes in maize. PLoS Genet. 5:e1000716.

Zhao M, Meyers BC, Cai C, Xu W, Ma J. 2015. Evolutionary patterns and coevolutionary consequences of MIRNA Genes and microRNA targets triggered by multiple mechanisms of genomic duplications in soybean. Plant Cell 27:546–562.

Associate editor: Kenneth Wolfe