Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids

Misook Ha[a,b,c], Jie Lu[a,b], Lu Tian[a], Vanitharani Ramachandran[d], Kristin D. Kasschau[e], Elisabeth J. Chapman[e], James C. Carrington[e], Xuemei Chen[d], Xiu-Jie Wang[f], and Z. Jeffrey Chen[a,b,c,g,1]

[a]Section of Molecular Cell and Developmental Biology, [b]Institute for Cellular and Molecular Biology, [c]Center for Computational Biology and Bioinformatics, and [g]Section of Integrative Biology, The University of Texas, Austin, TX 78712; [d]Department of Botany and Plant Sciences, University of California, Riverside, CA 92521; [e]Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR 97331; and [f]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, 5 Datun Road, Chaoyang District, Beijing 100101, China

Communicated by Wen-Hsiung Li, University of Chicago, Chicago, IL, and approved July 1, 2009 (received for review March 19, 2009)

Small RNAs, including microRNAs (miRNAs), small interfering RNAs (siRNAs), and *trans*-acting siRNAs (tasiRNAs), control gene expression and epigenetic regulation. Although the roles of miRNAs and siRNAs have been extensively studied, their expression diversity and evolution in closely related species and interspecific hybrids are poorly understood. Here, we show comprehensive analyses of miRNA expression and siRNA distributions in two closely related species *Arabidopsis thaliana* and *Arabidopsis arenosa*, a natural allotetraploid *Arabidopsis suecica*, and two resynthesized allotetraploid lines (F_1 and F_7) derived from *A. thaliana* and *A. arenosa*. We found that repeat- and transposon-associated siRNAs were highly divergent between *A. thaliana* and *A. arenosa*. *A. thaliana* siRNA populations underwent rapid changes in F_1 but were stably maintained in F_7 and *A. suecica*. The correlation between siRNAs and nonadditive gene expression in allopolyploids is insignificant. In contrast, miRNA and tasiRNA sequences were conserved between species, but their expression patterns were highly variable between the allotetraploids and their progenitors. Many miRNAs tested were nonadditively expressed (deviating from the mid-parent value, MPV) in the allotetraploids and triggered unequal degradation of *A. thaliana* or *A. arenosa* targets. The data suggest that small RNAs produced during interspecific hybridization or polyploidization serve as a buffer against the genomic shock in interspecific hybrids and allopolyploids: Stable inheritance of repeat-associated siRNAs maintains chromatin and genome stability, whereas expression variation of miRNAs leads to changes in gene expression, growth vigor, and adaptation.

expression regulation | microRNAs | polyploidy | hybrid vigor

Polyploidy is formed by whole genome duplication within species (autopolyploidy) or between species (allopolyploidy) (1). Both allopolyploids and autopolyploids are prevalent in nature, suggesting an evolutionary advantage of having multiple sets of genetic material for adaptation and development (2). Moreover, heterozygosity and intergenomic interactions in allopolyploids give rise to phenotypic variation and growth vigor (1, 3), which may explain why many crops such as wheat, cotton, and canola are allopolyploid.

Arabidopsis suecica is a natural allotetraploid derived, ≈12,000 to 300,000 years ago, from *Arabidopsis thaliana* and *Arabidopsis arenosa* (4), which diverged ≈6-million years ago (Mya) (5), similar to the divergence time between human and chimpanzee (6). *Arabidopsis* allotetraploids are readily resynthesized by pollinating tetraploid *A. thaliana* with *A. arenosa* (7, 8). They are genetically stable, resemble natural allotetraploid *A. suecica*, and display growth vigor (Fig. S1) (1, 3).

Newly formed interspecific hybrids and resynthesized allopolyploids may undergo "genomic shock," as predicted by B. McClintock (9), leading to genome-wide relaxation of gene expression including transposon activation. Indeed, DNA demethylation and retroelement reactivation were observed in marsupial interspecific hybrids (10), and abnormal imprinting was found in rodent interspecific hybrids (11). In *Arabidopsis* 15–45% genes are expressed differently between the related species *A. thaliana* and *A. arenosa* (7). Among 3,900 (up to ≈11,000) genes that are differentially expressed between the two species, ≈68% are nonadditively expressed in the allotetraploids. Nonadditive expression is defined as the expression level of a gene in the allotetraploid that is not equal to the sum of two parental loci or mid-parent value (MPV) (null hypothesis: 1 + 1 = 2), leading to activation (>2) or repression (<2) (1). Many microRNA (miRNA) targets were nonadditively expressed in the allotetraploids (7), suggesting a role for miRNAs in buffering genetic clashes between species (12). Small RNAs, including miRNAs (13), small interfering RNAs (siRNAs) (14), and *trans*-acting siRNAs (tasiRNAs) (15, 16), mediate posttranscriptional regulation, RNA-directed DNA methylation, and chromatin remodeling. Here, we sequenced small RNAs from 10 libraries prepared from leaves and flowers of three allotetraploids and two parents and characterized small RNA distributions in these lines. Sequence reads, miRNA microarrays, and small RNA blots were used to analyze miRNA abundance in the allotetraploids and their progenitors. We further studied the expression of miRNA biogenesis genes and tested the roles for target transcript abundance and miRNA-triggered target cleavage in nonadditive expression of several targets in allotetraploids.

Results

Dynamic Changes for Small RNAs in Allotetraploids and Their Progenitors. To test how small RNAs buffer the genetic clashes in allopolyploids, we generated ≈1.5-million small RNA sequences by pyrosequencing in 10 libraries, each from leaves and flower buds of *A. suecica*, resynthesized F_1 and F_7 allotetraploids, and their parents *A. thaliana* and *A. arenosa* (Fig. S1). After removal of adaptors and contaminants and the sequences identical to known cellular RNAs and plastid and mitochondrial DNA (*SI Materials and Methods*), we identified a total of 818,000 small RNA sequences of 20–25-nt in length. Among them, 296,231 (≈36%) perfectly matched *A. thaliana* Columbia or L*er* genomic sequences and belonged to 112,255 distinct small RNAs that were unique and nonredundant (Table S1). The small RNAs originating in *A. arenosa* and *A. suecica* but not perfectly matching the *A. thaliana* genome were excluded from

Author contributions: Z.J.C. designed research; M.H., J.L., L.T., V.R., and E.J.C. performed research; M.H., K.D.K., J.C.C., X.C., and X.-J.W. contributed new reagents/analytic tools; M.H. and J.L. analyzed data; and M.H. and Z.J.C. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE15443)

[1]To whom correspondence should be addressed. E-mail: zjchen@mail.utexas.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0907003106/DCSupplemental.

GENETICS

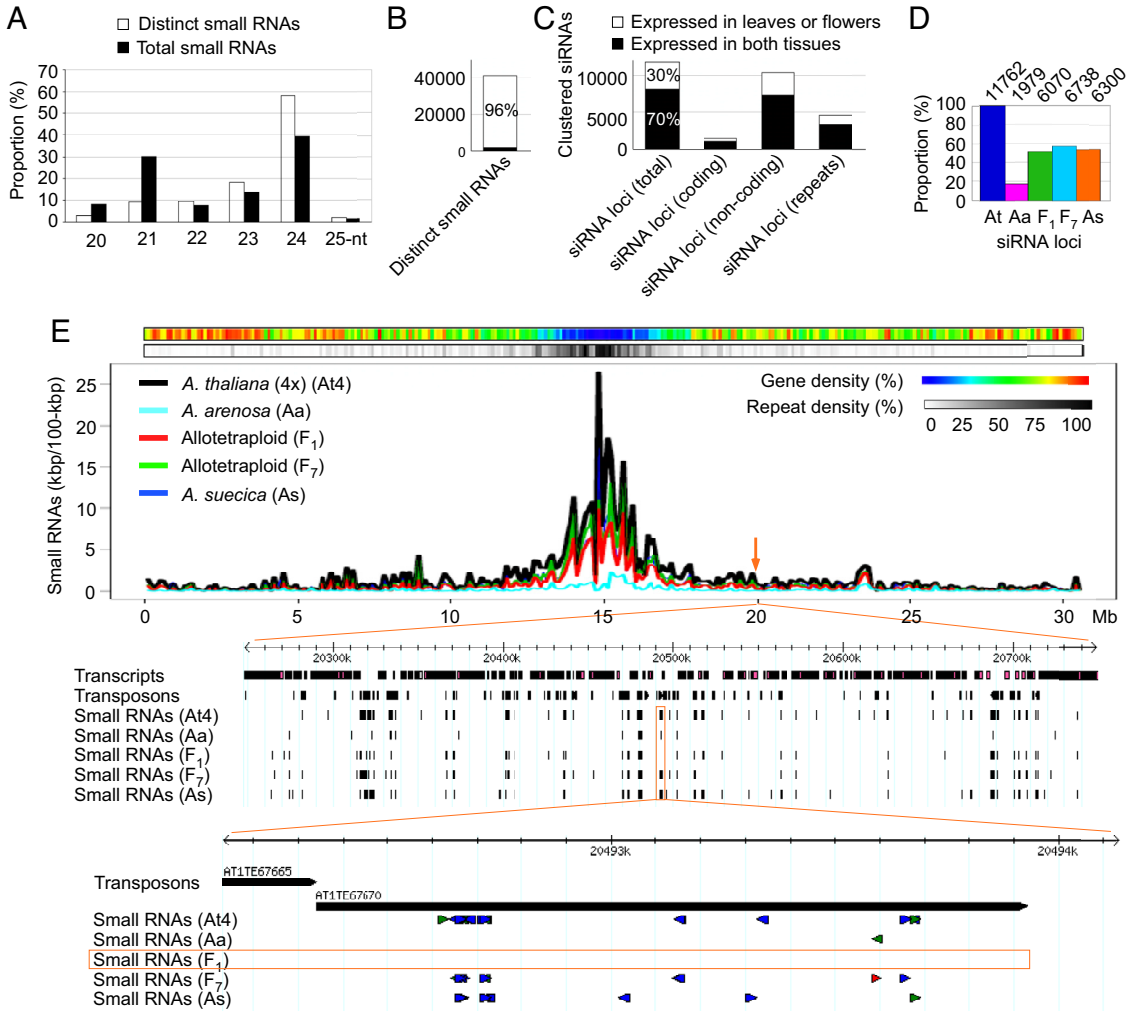

Fig. 1. Small RNA distributions in allopolyploids and their progenitors. (*A*) Size distribution (%) of total small RNAs (black) and distinct small RNAs (unfilled) in *A. thaliana*. (*B*) Total distinct small RNAs in both leaves and flowers (black) and in either one tissue (unfilled) in *A. thaliana*. (*C*) Distribution of clustered siRNA loci in total, coding and noncoding sequences, and repeats and transposons. (*D*) Percentage of *A. thaliana* clustered siRNAs loci in *A. thaliana* (blue), *A. arenosa* (red), F_1 (green), F_7 (cyan), and *A. suecica* (orange). (*E*) Small RNA density distributions in 100-kb sliding windows with a 50-kb slide on chromosome 1. Black, cyan, red, green, and blue indicate siRNA densities in *A. thaliana*, *A. arenosa*, F_1, F_7, and *A. suecica*, respectively. Gradient color and gray bars indicate percentages of the gene and repeat densities, respectively. (*Middle*) Zoom-in view of small RNA distributions in the region (arrow). (*Bottom*) *A. thaliana* siRNAs from transposon AT1TE67670 were absent in F_1 (boxed) and restored in F_7 and *A. suecica*.

most analyses because the complete *A. arenosa* genome sequence is unavailable. Thus, the analysis is limited to *A. thaliana* small RNAs.

In *A. thaliana*, 24-nt small RNAs were most abundant, representing 58% of the unique small RNAs, followed by 23-nt (18%), 22-nt (10%), and 21-nt (9%) small RNAs (Fig. 1*A*). miRNAs and tasiRNAs were conserved and easily identified mainly in the group of 21-nt RNAs (≈30% of total reads). siRNAs were 23- to 24-nt in size and derived from repeats, transposons, and intergenic and some genic regions (17, 18). The overall frequencies of miRNAs (27%), tasiRNAs (4%), and siRNAs (68%) in this study were similar to the data published in refs. 17–19, indicating a good representation of small RNAs in these libraries.

Most distinct small RNA sequences (≈96%) were present in either leaves or flowers in *A. thaliana*, and only 4% were found in both tissues (Fig. 1*B*). To examine how siRNAs are distributed, we assigned a clustered siRNA locus to a genomic region that contains overlapping or closely linked siRNAs (≤250 bp) with two or more reads (*SI Materials and Methods*). Among 11,762 clustered siRNA loci identified, 8,170 (≈70%) were found in both leaves and flowers, and only 3,592 (≈30%) were present in either leaves or flowers (Fig. 1*C* and Fig. S2*A*). In a comparative analysis using data published in ref. 17, ≈96.5% of 137,671 siRNA reads were present in either leaves or flowers, and only ≈3.4% were present in both leaves and flowers. These siRNAs were grouped into 17,010 clustered siRNA loci, including ≈69.4% small RNAs in both leaves and flowers and ≈30.6% in either flowers or leaves. However, the diversity of common siRNA loci is similar in comparison with leaves and between leaves and flowers. The data suggest that diverse siRNAs are generated from a genomic region in both leaves and flowers in *A. thaliana*.

Consequently, clustered siRNA loci were used to compare siRNA diversity and conservation among species and in allopolyploids. The siRNA densities are normalized using total distinct siRNA loci within a genomic region in a 100-kb sliding window. Except for a small amount of siRNAs found in coding regions, most siRNAs originated from noncoding sequences and repeats (Fig. 1*C*). Interestingly, 52–57% clustered siRNA loci in *A. thaliana* were found in *A. suecica* and resynthesized F_1 and F_7 allotetraploids (Fig. 1*D*), consistent with an allotetraploid genome (100%) that consists of 50% *A. thaliana* and 50% *A. arenosa* sequences. The siRNA locus densities in F_1, F_7, and natural

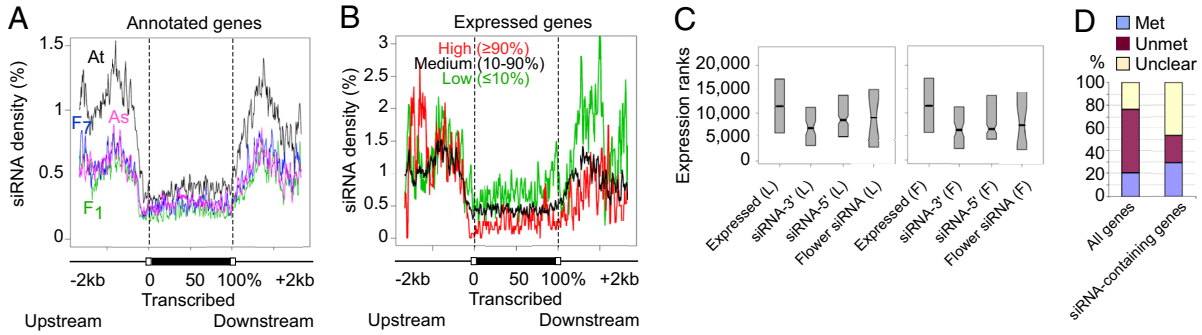

Fig. 2. siRNAs and gene expression in allotetraploids and *A. thaliana*. (*A*) Distribution of siRNAs in transcribed, 5′ and 3′ regions in *A. thaliana* (black), F_1 (green), F_7 (blue), and *A. suecica* (red). (*B*) Distribution of *A. thaliana* siRNAs in 5′ and 3′ regions of the genes that are expressed at high (≥90%, red), medium (10–90%, black), and low (≤10%, green) levels. (*C*) Distribution of *A. thaliana* siRNAs relative to genes expression levels in leaves (*L*) and flowers (*F*). (*Left*) Expressed (L), all expressed genes in leaves; siRNA-3′ and siRNA-5′ (L), expressed genes with siRNA in the 3′ end and 5′ end, respectively; and Flower siRNA (L), the siRNA-associated genes in flowers only were poorly expressed in leaves. (*Right*) Similar to the above notes except for flowers (F); Flower siRNA (F), the siRNA-associated genes in flowers only were poorly expressed in flowers. (*D*) Percentage of DNA methylation present in all genes and siRNA-containing genes. Met, methylated; Unmet, unmethylated; Unclear, methylation status unclear.

allopolyploids were approximately half of that in *A. thaliana* (Fig. 1*E* and Fig. S3). The data suggest that *A. thaliana* siRNAs are stably maintained in the allotetraploids. Among a few abundant siRNA loci examined, the *A. arenosa* and *A. thaliana* siRNAs were conserved and equally present in the stable allotetraploids (Fig. S2*B*).

Interestingly, siRNA densities in each window were significantly lower in F_1 than F_7 allotetraploids (paired Wilcoxon ranks sum test, $P \approx 0$) (Fig. S4 *A–D*), whereas the difference of siRNA densities between F_7 and *A. suecica* were insignificant (Paired Wilcoxon ranks sum test, $P = 0.7$). Likewise, siRNA distributions near the genes were significantly lower in F_1 than in F_7 and *A. suecica*. ($\chi^2 = 15{,}683$, $P \approx 0$). The number of siRNA-containing repeats was also lower in F_1 than in F_7 ($P \approx 0$). Most *A. thaliana* siRNAs absent in F_1 were present in F_7 and *A. suecica* (Fig. 1*E* and Fig. S4*E*). Among 6,000 siRNA-generating transposons in *A. thaliana*, 5,123 (≈85%) also produced siRNAs in one or more allotetraploids. Compared with *A. thaliana*, 329 transposons lost siRNAs, and 12 gained siRNAs in F_1, although this may also be related to sequencing depth. Among 329 siRNA-free transposons in F_1, 311 and 259 produced siRNAs in F_7 and *A. suecica*, respectively, resulting in the siRNA distributions of F_7 and *A. suecica* more similar to *A. thaliana* than to F_1.

In *A. thaliana*, siRNA accumulated at high levels in the upstream and downstream regions of pseudogenes associated with repeats (PsG + R) and transposable elements (TEs) (Fig. S4 and Table S2). Over 55% (377/679) of siRNA-generating genes contained TEs within genic regions or introns. Moreover, pseudogenes without repeats did not produce abundant siRNAs. The data suggest that repeat regions are predisposed for siRNA biogenesis.

At the 5′ and 3′ ends of the transcribed regions, the siRNA distribution trends were similar in *A. thaliana* and resynthesized and natural allotetraploids (Fig. 2*A*). The siRNAs predominated in the 5′ and 3′ ends with peak densities near 1,000 bp upstream and downstream of the transcribed regions. Again, siRNA densities in F_1 allotetraploids were slightly lower than those in F_7 and *A. suecica*.

We classified gene expression levels as high (≥90%), medium (10–90%), and low (≤10%) using microarray data in *Arabidopsis* leaves (20). The siRNA levels were high in the 3′ ends of the poorly expressed genes but low in the highly expressed genes (Fig. 2*B*), suggesting a potential role for siRNAs in gene repression via siRNA regulation in the 3′ ends.

To test siRNA effects on gene expression, we identified 679 protein-coding genes that generated siRNAs in *A. thaliana*, *A. suecica*, and resyntheized allopolyploids (Table S2). The gene expression microarray data (20) were normalized to yield similar distributions in both leaves and flowers (Fig. 2*C*). The genes containing siRNAs in flowers but not in leaves were expressed at lower levels in flowers than in leaves, suggesting that siRNAs are responsible for repression of these genes. The genes containing siRNAs in the 5′ and 3′ ends (within 1,000 bp) were expressed at significantly lower levels in leaves ($P = 2 \times 10^{-7}$ and 3×10^{-7} for 5′ and 3′ ends) and flowers ($P = 9 \times 10^{-12}$ and 1×10^{-7} for 5′ and 3′ ends) than all genes expressed in the corresponding tissues. Gene repression correlated more with the siRNAs generated in the 3′ than in the 5′ region. Moreover, the siRNA-generating genes (≈30%) are preferentially methylated (Pearson's $\chi^2 = 92$, $P = 2.2e^{-16}$) (Fig. 2*D*), suggesting a mechanism for siRNA-mediated DNA methylation and transposon silencing in interspecific hybrids and allotetraploids as in diploids (14, 21).

siRNAs and Nonadditive Gene Regulation in Allopolyploids. We further examined a relationship between siRNAs in *A. thaliana* and nonadditively expressed genes in the allotetraploids. In an F_6 allotetraploid, the same line used for small RNA sequencing, 277 and 899 genes were up- and down-regulated in leaves relative to their parents, respectively (*SI Materials and Methods*). The proportion of the genes that gained or lost siRNAs in the allotetraploid was insignificantly different from that in up- or down-regulated genes. This suggests a minor role for siRNAs in nonadditive gene expression in the allotetraploids, although some siRNAs may affect the expression of these target genes.

Sequence Conservation and Expression Divergence Among miRNAs in Allopolyploids. Unlike siRNAs, many miRNA and tasiRNA loci are conserved in sequences. As a result, miRNAs and tasiRNAs were commonly identified in the allotetraploids and their progenitors (Fig. S2*C*). Among 153 miRNA loci reported (22), 152 miRNA loci were found in our sequencing data; 140 were identified in *A. thaliana*, 107 in *A. arenosa*, 114 in allotetraploid F_1, 127 in F_7, and 121 in *A. suecica*. The miRNAs absent in allotetraploids and *A. arenosa* may be poorly expressed or related to sequencing depth. Similarly, all previously characterized tasiRNAs present in the allotetraploids were also found in their progenitors. Except for miR163, each canonical miRNA had the same sequence in *A. thaliana* as in *A. arenosa* and allotetraploids.

Despite sequence conservation, miRNA expression levels were highly variable in *A. thaliana*, *A. arenosa*, *A. suecica*, and resynthesized allotetraploids (Fig. 3 *A–D*). Among 85 distinct miRNAs and 23 tasiRNAs on the microarrays, 69 miRNAs and 17 tasiRNAs were expressed above the detection level (Table S3), which was confirmed by sequencing (Table S2). In leaves, 35 (≈51%) miRNAs and 8 (≈47%) tasiRNAs were expressed differently between

GENETICS

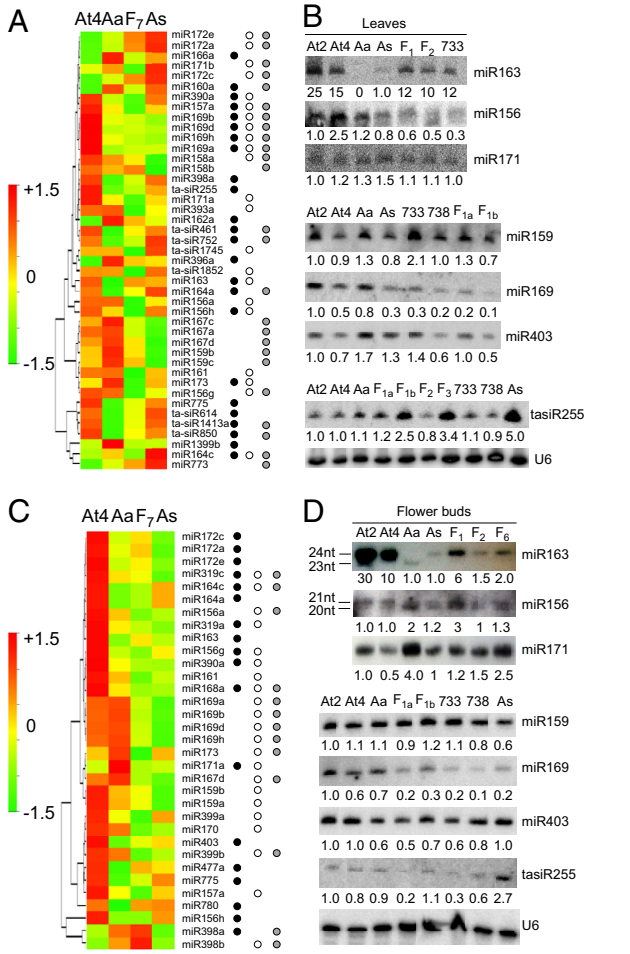

Fig. 3. Expression diversity of miRNAs and their targets between closely relates species and in allopolyploids. (*A*) Hierarchical cluster analysis of miRNA expression variation in leaves in At, Aa, F_7, and As (see *SI Materials and Methods*). Black, open, and gray circles indicate the targets that were expressed differently between At and Aa, mid-parent value (MPV) and F_7, and MPV and As, respectively. (*B*) Small RNA blot analysis of miRNA expression variation in the leaves of As, F_1, F_2 F_6, and two allotetraploids in F_7 generation (733 and 738), and their progenitors *A. arenosa* and *A. thaliana* (At2, diploid; At4, tetraploid). Two F_1 individuals (F_{1a} and F_{1b}) were used in some blots. The fold-changes were averaged ($n = 2$). (*C*) Hierarchical cluster analysis of miRNA expression variation in flower buds. The colors, symbols, and lines were the same as in *C*. (*D*) Small RNA blot analysis of miRNA expression variation in flower buds ($n = 2$).

A. thaliana and *A. arenosa* or between an allotetraploid (F_7 or *A. suecica*) and MPV (Fig. 3*A*), whereas in flowers 33 (≈40%) miRNAs were differentially expressed between these comparisons (Fig. 3*C*). The nonadditive accumulation of miRNAs is reminiscent of nonadditive expression of many protein-coding genes in the resynthesized allotetraploids (7). For example, 12 miRNAs were expressed at higher levels in *A. thaliana* than in *A. arenosa*, and many were down-regulated in F_7 allotetraploids. The data suggest a trend of repressing miRNAs of the *A. thaliana* origin in resynthesized allotetraploids, a direction consistent with the repression of *A. thaliana* rRNA and protein-coding genes in the allopolyploids (1). Interestingly, many miRNAs accumulated differently between F_7 and *A. suecica*, indicating a role for miRNAs in allopolyploid evolution.

Although the number of differentially expressed miRNAs was fewer in flowers than in leaves, miRNA abundance in flowers varied dramatically between the two related species. *A. thaliana* is inbreeding and has small and white flowers, whereas *A. arenosa* is outcrossing and has large and pink flowers (7). However, miRNA expression levels in flowers were relatively similar between F_7 and *A. suecica*, suggesting a role for miRNAs in maintaining flower morphology and development in stable allotetraploids.

Among the miRNAs and tasiRNAs examined, the expression levels estimated from microarrays (Table S3) and sequencing (Table S2) were closely related ($r = 0.52$, $P = 9.8 \times 10^{-7}$, Fig. S5*A*) and matched those examined by small RNA blots (Fig. 3 *B* and *D*). However, the normalized reads (per 100,000 miRNAs) displayed a larger variation than the expression levels detected by microarrays and small RNA blots probably because of sensitivity and/or technical variability in sequencing (17).

The 24-nt miR163, a recently evolved miRNA (23), was abundant in *A. thaliana* but undetectable in *A. arenosa* leaves using both *A. thaliana* and *A. arenosa* probes (Fig. 3*B*). In flowers, a 23-nt RNA was detected at a level of 30-fold lower in *A. arenosa* than in *A. thaliana* (Fig. 3*D*), suggesting sequence divergence and expression diversity of this young miRNA. In leaves miR159 and 403 were expressed at higher levels in *A. arenosa* than in *A. thaliana* and allotetraploids. tasiR255 accumulated at higher levels in F_1, F_3 and *A. suecica* than in other lines, and tasiR255 and miR159 expression levels were altered in two F_1 siblings, F_3, and *A. suecica*, suggesting rapid and dynamic changes of miRNA expression levels in allopolyploids. Although expression variation in flowers was generally low, miR156 and miR171 accumulated at higher levels in the flowers than in the leaves in the allotetraploids. Except for miR156 that showed a positive correlation between ploidy and expression levels, miRNA abundance was inversely correlated with ploidy levels in *A. thaliana* diploids (At2) and autotetraploids (At4), suggesting negative dosage effects on miRNA expression in autopolyploids (1).

Nonadditive Expression of Target Genes and miRNA Target Preference in Allopolyploids. Among 10 miRNA target genes that were nonadditively expressed in the F_6 allotetraploids under both common and per-gene variances (7), log-fold changes in miRNA and target mRNA levels were negatively correlated ($r = -0.76$, $P = 0.01$, $R^2 = 0.58$) (Fig. 4*A*). Nonadditive accumulation of these miRNAs explained ≈58% of expression changes in the allopolyploids (7, 12). This suggests that differential regulation of miRNAs plays a role in interspecific variation of protein-coding gene expression (12). When the per-gene variance is used (7), the number of nonadditively expressed miRNA targets is increased to 71 that include multiple targets of a miRNA family. The negative correlation between miRNA and target mRNA levels was low but still significant ($P = 0.04$; $R^2 = 0.07$), suggesting that other mechanisms such as transcriptional regulation of miRNA targets may also account for target expression variation.

Nonadditive accumulation of miRNAs may be caused by nonadditive expression of miRNA biogenesis genes such as *DCL1* and *AGO1* (13). Indeed, *DCL1* was nonadditively expressed in the allotetraploids relative to the progenitors in the leaves and flowers, whereas *AGO1* and *AGO2* expression was nearly additive (Fig. S5*B*), suggesting different roles for these genes in miRNA biogenesis in allopolyploids. *DCL4* and *AGO4* are primarily involved in siRNA biogenesis or effector function (14). The expression of *DCL4* was generally additive, whereas *AGO4* expression is nonadditive in leaves but additive in flowers (Fig. S5*B*).

miRNA target binding sites were highly conserved in *A. thaliana* and *A. arenosa*, as predicted in other plants (24). We analyzed cleavage frequencies of the targets that were nonadditively expressed (7) and negatively correlated with miRNA abundance in allotetraploids. Among six miRNA targets tested, the cleavage sites were identical in two related species. However, in the allotetraploid three miRNAs (miR163, 164 and 168) triggered degradation of *A. thaliana* targets more than of *A. arenosa* homoeologs, whereas miR159a triggered degradation of *A. arenosa* targets more than of

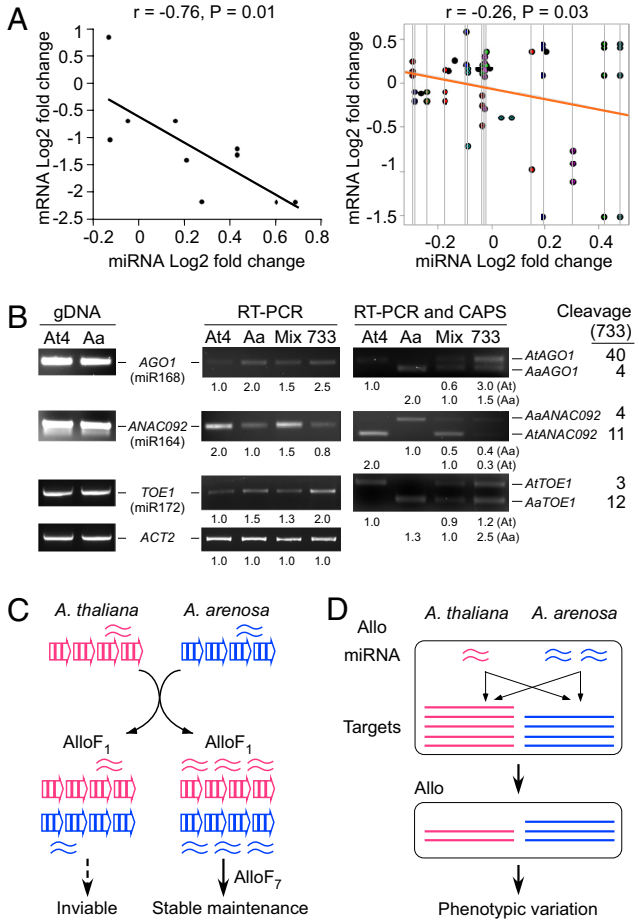

Fig. 4. Regulation of miRNA targets in allopolyploids and models for miRNAs and siRNAs in genomic stability and phenotypic variation in allopolyploids. (*A*) Inverse expression correlations between a subset of miRNAs and 10 targets (*Left*) and 71 targets (*Right*) that are nonadditively expressed in allotetraploid Allo733. Multiple targets of a single miRNA are indicated by a vertical line. (*B*) Nonadditive expression of homoeologous target loci and biased target cleavage in the allotetraploids. (*Left*) Target loci originating in *A. thaliana* (At4) and *A. arenosa* (Aa) were amplified in PCR using genomic DNA (gDNA) or cDNA in RT-PCR. Mix, an equal mixture of At4 and Aa; 733, allotetraploid Allo733 (n = 2). (*Right Middle*) RT-PCR and CAPS analysis showing relative expression levels of At and Aa transcripts in Allo733 relative to the parents (n = 2). (*Right*) Sequencing frequencies of the cleaved At and Aa target mRNA in Allo733. (*C*) rasiRNAs were well maintained in stable allotetraploids, whereas loss of siRNAs may be related to genomic instability and lethality in F1 allotetraploids. Blue and red arrows indicate *A. arenosa* and *A. thaliana* repeats. Short curved lines indicate siRNAs. (*D*) miRNA-triggered degradation of *A. thaliana* (red lines) or *A. arenosa* (blue) targets by miRNAs, leading to gene expression and phenotypic variation.

A. thaliana homoeologs (Fig. 4*B* and Fig. S6). miR168a guided comparable cleavage of targets in *A. thaliana* and *A. arenosa*, but in the allotetraploid the ratio of cleaved *A. thaliana* (At) over *A. arenosa* (Aa) targets was 40:4. Likewise, the cleaved target ratios (At:Aa) of miR163 and miR164 were 8:0 and 11:4, respectively. On the contrary, the ratio (At:Aa) of miR159-guided cleaved targets was 6:12, suggesting a preference for targeting of *A. arenosa* transcripts. The cleavage ratios (At:Aa) of miR172a targets were 12:3 (At5g60120) and 3:12 (At2g28550). Target preference is unlikely associated with PCR amplification bias because relatively equal numbers of both targets were detected whan an equal mixture of *A. thaliana* and *A. arenosa* RNA was used in the assays (*SI Materials and Methods*).

We examined target transcript levels using RT-PCR and cleaved amplified sequence polymorphism (CAPS) assays that discriminate between *A. thaliana* and *A. arenosa* loci (7). *AGO1*, a miR168 target (25), was expressed at higher levels in *A. arenosa* than in *A. thaliana*. In allotetraploid 733, *A. thaliana AGO1* mRNA accumulated ≈2-fold higher than *A. arenosa AGO1* mRNA, consistent with the high level of *A. thaliana AGO1* mRNA cleaved products (Fig. 4*B*). In the allotetraploid *A. arenosa* miR172 target (*TOE1*) was expressed ≈2-fold higher than *A. thaliana TOE1*, and *A. arenosa TOE1* mRNA was preferentially cleaved. This suggests that nonadditive accumulation of homoeologous target transcripts partially explains biased cleavage, although the fold changes in the expression assays were lower than in the cleavage ratios. However, the correlation between miR164 target (*ANACO92*) transcript levels and cleavage ratios was not positive. *A. thaliana ANACO92* and *A. arenosa* transcript levels were relatively equal in the allotetraploid, but *A. thaliana ANACO92* transcripts were preferentially cleaved, suggesting a role for miRNAs in biased degradation of homoeologous targets in the allotetraploids.

Discussion

Roles for miRNAs and siRNAs in Nonadditive Gene Regulation in Allopolyploids. We found that miRNAs and tasiRNAs are associated with nonadditive expression of target genes in the allotetraploids. Nonadditive expression of the targets is partly mediated by transcriptional regulation. The genome merger in the allotetraploids induces genetic and epigenetic changes (1), leading to nonadditive expression of miRNA targets and the miRNA primary transcripts. At the posttranscriptional level, nonadditive expression of miRNA biogenesis genes results in nonadditive accumulation of miRNAs. For example, *DCL1*, *AGO1*, and *AGO2* expression levels were higher in *A. suecica* than in resynthesized allotetraploids in leaves but not in flowers, consistent with the miRNA accumulation levels in the corresponding tissues. Moreover, differential expression of *A. thaliana* and *A. arenosa* miRNAs and their targets in the allotetraploids leads to the biased target degradation probably because target mRNA suppression depends on a threshold of miRNA concentration (26). Alternatively, the secondary structures of homoeologous target transcripts may affect the efficiency of miRNA-triggered degradation (27).

The correlation between siRNA-generating genes and the genes that are nonadditively expressed in the allotetraploids is insignificant, which is consistent with a few genes that are affected by DNA hypomethylation in *A. suecica* (28). This is probably because siRNAs are tightly regulated for heterochromatin maintenance and genome stability, or siRNA-containing transposons and repeats are underrepresented in microarrays (7).

Tissue-Specific siRNAs and Gene Expression Variation. Many distinct siRNA loci (~30%) may be tissue-specific in leaves or flowers. One possibility is that transcriptional control for these siRNAs is differentially regulated during leaf and flower development probably through the activities of RNA polymerases IV and/or V (30) or DCL and Argonaute proteins. Uniparental expression of PolIV-dependent siRNAs was observed in developing endosperm (31). A consequence of siRNAs in different genomic regions may affect the degree of genomic stability and transposon activities, as observed in pollen (32).

Although the majority of siRNAs do not correlate with gene expression variation in leaves or flowers, siRNAs are associated with gene repression mainly through the 3′-end of the genes, consistent with the notion that siRNAs generated in the 3′ downstream regions effectively repress gene expression, as observed in *FLC* locus (33). The correlation between siRNA and gene expression was not found in a previous study (18) possibly because the genes were not split into high, medium, and low expression levels. Moreover, one-third of siRNA-generating genes studied are preferentially methylated, suggesting that these genes are subjected to siRNA-directed DNA methylation (14, 21).

siRNAs Maintains Genome Stability, Whereas miRNAs Mediate Gene Expression Diversity. The data suggest that siRNAs maintains genomic stability in response to the genomic shock (9) in allopolyploids (Fig. 4*C*) (12). Although the proportion of repeat-associated siRNAs (rasiRNAs) is low in F_1, the number of miRNA reads is high in F_1, indicating rapid and dynamic changes of siRNAs and miRNAs in early stages of allopolyploid formation. rasiRNAs (21) diverged rapidly among closely related species. Some rasiRNAs are associated with gene repression in diploids but weakly with gene expression changes between the related species or in allotetraploids. rasiRNAs are enriched in centromeres, consistent with siRNA accumulation and DNA hypermethylation of *A. thaliana* homoeologous centromeres in *A. suecica* (28). Many siRNAs absent in F_1 are restored in late and natural allotraploids, indicating that several generations are needed to establish stable expression patterns of siRNAs as of protein-coding genes (29), leading to heterochromatin and genome stability. Reduction of siRNAs in F_1 may activate some transposable elements in response to "genomic shock" (9) in marsupial interspecific hybrids (10) and induce genome instability and infertility in *Arabidopsis* allotetraploids (8, 29). Over time, the genome stability is restored through regeneration of rasiRNAs in genetically stable allotetraploids.

The repression of *A. thaliana* homoeologous loci (7) and accumulation of *A. thaliana* centromeric siRNAs (28) may be similar to the repression of transposons through maternal transmission of endogenous siRNAs in *Drosophila* (34). Indeed, interspecific hybrids and allotetraploids can only be produced using *A. thaliana* as the maternal parent (8, 29).

In contrast, miRNAs and tasiRNAs mediate expression diversity between closely related species and in the allotetraploids (Fig. 4*D*). Progenitors' miRNA loci may diverge and gain new expression patterns, as a consequence of genetic and epigenetic changes (1). A combination of diverged progenitors' loci in interspecific hybrids and new allopolyploids lead to *cis*- and *trans*-acting effects on expressing miRNAs and their biogenesis genes. As a result, the miRNA accumulation levels are nonadditive (1), leading to nonadditive expression of their targets in interspecific hybrids and allotetraploids (7). This nonadditivity is partly associated with transcriptional regulation of miRNA loci and their targets (1, 12). In addition, nonadditive expression of miRNA biogenesis genes and preference of miRNAs for repressing *A. thaliana* or *A. arenosa* targets suggest a miRNA-dependent mechanism for repression of some *A. thaliana* genes observed in interspecific hybrids and allopolyploids (7). The miRNA preference for targets is probably resulted from divergence between target sequences or secondary structures (27). Consistent with the high sensitivity of the *A. thaliana* genes to expression changes in the allotetraploids (3, 7), there is a tendency for miRNA-triggered degradation of *A. thaliana* targets.

Many miRNA targets encode transcription factors or proteins that are important to growth and development in plants and animals (13). For example, miR164 is associated with age-dependent cell death (35); tasiRNAs affect lateral organ development and the transition from juvenile to adult stages (16); and miR168 and miR403 provide feedback regulation in miRNA biogenesis (17, 25). Nonadditive regulation of these miRNAs and their targets may lead to nonadditive or novel phenotypes in the allopolyploids. The data collectively suggest that allopolyploids take advantages from the control of progenitors' siRNAs and miRNAs as a buffer against genetic clashes resulting from genome merger and hybridization between the related species, leading to siRNA-dependent genome stability and miRNA-mediated gene expression variation and developmental changes in new allopolyploid species.

Materials and Methods

Plant materials and experimental designs were described in *SI Materials and Methods*. Small RNAs were prepared using a protocol published in ref. 18 and subjected to pyrosequencing. Library construction and analyses of small RNA sequences, miRNA microarrays, small RNA blots, and target cleavages were described in *SI Materials and Methods*. The small RNA sequence and microarray data are deposited in GEO with accession numbers GSM387513-387522, GSM425081-085, and 425098-100, respectively.

ACKNOWLEDGMENTS. We thank Wen-Hsiung Li and two anonymous reviewers for providing constructive comments and valuable advice to improve the manuscript, Zhixin Xie for helpful assistance during small RNA library construction, Chris Sullivan and Guanglin Li for initial parsing of the small RNA sequence data, and members in the Z.J.C. laboratory for valuable discussions. The work is supported by National Institutes of Health Grant GM067015 and National Science Foundation Plant Genome Research Program Grant DBI0733857 (to Z.J.C.).

1. Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol* 58:377–406.
2. Leitch AR, Leitch IJ (2008) Genomic plasticity and the diversity of polyploid plants. *Science* 320:481–483.
3. Ni Z, et al. (2009) Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457:327–331.
4. Jakobsson M, et al. (2006) A unique recent origin of the allotetraploid species Arabidopsis suecica: Evidence from nuclear DNA markers. *Mol Biol Evol* 23:1217–1231.
5. Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in Arabidopsis, Arabis, and related genera (Brassicaceae) *Mol Biol Evol* 17:1483–1498.
6. Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D (2006) Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 441:1103–1108.
7. Wang J, et al. (2006) Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. *Genetics* 172:507–517.
8. Comai L, et al. (2000) Phenotypic instability and rapid gene silencing in newly formed Arabidopsis allotetraploids. *Plant Cell* 12:1551–1568.
9. McClintock B (1984) The significance of responses of the genome to challenge. *Science* 226:792–801.
10. O'Neill RJ, O'Neill MJ, Graves JA (1998) Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393:68–72.
11. Vrana PB, et al. (2000) Genetic and epigenetic incompatibilities underlie hybrid dysgenesis in Peromyscus. *Nat Genet* 25:120–124.
12. Ha M, Pang M, Agarwal V, Chen ZJ (2008) Interspecies regulation of microRNAs and their targets. *Biochim Biophys Acta* 1779:735–742.
13. Bartel DP (2004) MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell* 116:281–297.
14. Baulcombe D (2004) RNA silencing in plants. *Nature* 431:356–363.
15. Vazquez F, et al. (2004) Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. *Mol Cell* 16:69–79.
16. Peragine A, Yoshikawa M, Wu G, Albrecht HL, Poethig RS (2004) SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. *Genes Dev* 18:2368–2379.
17. 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.
18. Kasschau KD, et al. (2007) Genome-wide profiling and analysis of arabidopsis siRNAs. *PLoS Biol* 5:e57.
19. Lu C, et al. (2005) Elucidation of the small RNA component of the transcriptome. *Science* 309:1567–1569.
20. Schmid M, et al. (2005) A gene expression map of Arabidopsis thaliana development. *Nat Genet* 37:501–506.
21. Lippman Z, Martienssen R (2004) The role of RNA interference in heterochromatic silencing. *Nature* 431:364–370.
22. Griffiths-Jones S, Saini HK, Dongen SV, Enright AJ (2008) miRBase: Tools for microRNA genomics. *Nucleic Acids Res* 36:D154–D158.
23. Allen E, et al. (2004) Evolution of microRNA genes by inverted duplication of target gene sequences in Arabidopsis thaliana. *Nat Genet* 36:1282–1290.
24. Rhoades MW, et al. (2002) Prediction of plant microRNA targets. *Cell* 110:513–520.
25. Vaucheret H, Mallory AC, Bartel DP (2006) AGO1 homeostasis entails coexpression of MIR168 and AGO1 and preferential stabilization of miR168 by AGO1. *Mol Cell* 22:129–136.
26. Brown BD, et al. (2007) Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. *Nat Biotechnol* 25:1457–1467.
27. Long D, et al. (2007) Potent effect of target structure on microRNA function. *Nat Struct Mol Biol* 14:287–294.
28. Chen M, Ha M, Lackey E, Wang J, Chen ZJ (2008) RNAi of met1 reduces DNA methylation and induces genome-specific changes in gene expression and centromeric small RNA accumulation in Arabidopsis allopolyploids. *Genetics* 178:1845–1858.
29. Wang J, et al. (2004) Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. *Genetics* 167:1961–1973.
30. Wierzbicki AT, Haag JR, Pikaard CS (2008) Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. *Cell* 135:635–648.
31. Mosher RA, et al. (2009) Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. *Nature*.
32. Slotkin RK, et al. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136:461–472.
33. Swiezewski S, et al. (2007) Small RNA-mediated chromatin silencing directed to the 3′ region of the Arabidopsis gene encoding the developmental regulator, FLC. *Proc Natl Acad Sci USA* 104:3633–3638.
34. Brennecke J, et al. (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322:1387–1392.
35. Kim JH, et al. (2009) Trifurcate feed-forward regulation of age-dependent cell death involving miR164 in Arabidopsis. *Science* 323:1053–1057.