Genome-Wide Dosage-Dependent and -Independent Regulation Contributes to Gene Expression and Evolutionary Novelty in Plant Polyploids

Xiaoli Shi,[†,1] Changqing Zhang,[‡,1] Dae Kwan Ko,[1] and Z. Jeffrey Chen[*,1,2]

[1]Department of Molecular Biosciences, Center for Computational Biology and Bioinformatics, and Institute for Cellular and Molecular Biology, The University of Texas at Austin

[2]State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China

[†]Present address: Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing, China

[‡]Present address: Plants for Human Health Institute, North Carolina State University, Kannapolis

***Corresponding author:** E-mail: zjchen@austin.utexas.edu.

Associate editor: Aoife McLysaght

The RNA-seq data have been deposited in NCBI's Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) (accession nos. GSE29687 and GSE50715).

Abstract

Polyploidy provides evolutionary and morphological novelties in many plants and some animals. However, the role of genome dosage and composition in gene expression changes remains poorly understood. Here, we generated a series of resynthesized *Arabidopsis* tetraploids that contain 0–4 copies of *Arabidopsis thaliana* and *Arabidopsis arenosa* genomes and investigated ploidy and hybridity effects on gene expression. Allelic expression can be defined as dosage dependent (expression levels correlate with genome dosages) or otherwise as dosage independent. Here, we show that many dosage-dependent genes contribute to cell cycle, photosynthesis, and metabolism, whereas dosage-independent genes are enriched in biotic and abiotic stress responses. Interestingly, dosage-dependent genes tend to be preserved in ancient biochemical pathways present in both plant and nonplant species, whereas many dosage-independent genes belong to plant-specific pathways. This is confirmed by an independent analysis using *Arabidopsis* phylostratigraphic map. For *A. thaliana* loci, the dosage-dependent alleles are devoid of TEs and tend to correlate with H3K9ac, H3K4me3, and CG methylation, whereas the majority of dosage-independent alleles are enriched with TEs and correspond to H3K27me1, H3K27me3, and CHG (H = A, T, or C) methylation. Furthermore, there is a parent-of-origin effect on nonadditively expressed genes in the reciprocal allotetraploids especially when *A. arenosa* is used as the pollen donor, leading to metabolic and morphological changes. Thus, ploidy, epigenetic modifications, and cytoplasmic-nuclear interactions shape gene expression diversity in polyploids. Dosage-dependent expression can maintain growth and developmental stability, whereas dosage-independent expression can facilitate functional divergence between homeologs (subfunctionalization and/or neofunctionalization) during polyploid evolution.

Key words: **polyploidy, gene expression, epigenetics, evolution, *Arabidopsis*.**

Introduction

Whole-genome duplication (WGD) or polyploidy plays prominent roles in shaping genome evolution and genetic diversity of eukaryotes, including most plants and some animals (Wendel 2000; Comai 2005; Leitch and Leitch 2008; Soltis PS and Soltis DE 2009). Polyploidy has been traditionally considered an evolutionary dead end (Mayrose et al. 2011). Polyploidy and aneuploidy often lead to carcinogenesis or birth defects in humans (Storchova and Pellman 2004), and aneuploidy impairs proliferation and alters metabolic properties in mouse cell lines (Williams et al. 2008) and induces proteomic changes and phenotypic variation in yeast (Pavelka et al. 2010). Aneuploids generally have larger changes than polyploids probably because dosage imbalance could affect the stability and interaction of a protein in a regulatory complex (Birchler et al. 2005; Veitia et al. 2008). However, recent studies suggest that polyploidy is a key driver of macroevolutionary success (Wood et al. 2009; Mayrose et al. 2011). Estimates indicate two rounds of ancestral WGDs occurred before the divergence of extant seed plants and angiosperms, giving rise to the diversification of genes and pathways important to seed and flower development and eventually the dominance of angiosperms on the earth (Van de Peer et al. 2009; Jiao et al. 2011).

The evolutionary success of polyploids is thought to associate with novel phenotypes that exceed the range of diploid progenitors or are absent in the progenitors (Ramsey and Schemske 2002). This increased range of phenotypic variation could be caused by dosage regulation (Birchler et al. 2005). Allelic expression can be defined as dosage-dependent expression, in which the expression level of an allele is correlated with the dosage, and dosage-independent expression

(expression levels do not correlate with the dosage). In some literature, dosage-dependent expression is used interchangeably with dosage-sensitive expression (Thomas et al. 2006) or dosage-insensitive expression (Zhou et al. 2011). In yeast polyploids, cell cycle regulation is repressed to promote larger cell size (Galitski et al. 1999), and Arabidopsis polyploids show higher potassium uptake and salinity tolerance (Chao et al. 2013). In addition, dosage-independent regulation leads to activation or silencing of homeologous rRNA and protein-coding genes in different developmental stages, as observed in allopolyploid *Arabidopsis* (Chen and Pikaard 1997), *Tragopogon* (Buggs et al. 2011), *Senecio* (Hegarty et al. 2006), cotton (Adams et al. 2003), soybean (Coate and Doyle 2010), and wheat (Qi et al. 2012). Expression divergence between homeologous genes could lead to neo- and subfunctionalization (Lynch and Force 2000; Lynch et al. 2001). Moreover, epigenetic factors such as histone marks (Ha et al. 2011) and small RNAs mediate genetic buffering and defense response in allotetraploid *Arabidopsis* (Ha et al. 2009; Ng et al. 2011) and fiber development in allotetraploid cotton (Guan et al. 2014). Epigenetic reprogramming of circadian clock regulators results in starch and biomass heterosis in *Arabidopsis* allotetraploids and hybrids (Ni et al. 2009; Chen 2013).

In spite of extensive studies, biological consequences of gene expression changes in polyploids with variable genome dosages and compositions remain elusive partly because suitable genetic materials and/or technologies such as allelic expression had not been available until recently. Here, we created a series of *Arabidopsis* tetraploids that contain variable dosages of *Arabidopsis thaliana* and *Arabidopsis arenosa* genomes and systematically investigated consequences and mechanisms for genome-wide dosage regulation of allelic expression and evolution. We studied genome-wide correlation between allelic expression and dosage variation and classified genes into dosage-dependent and -independent groups. We further examined roles of dosage-dependent and -independent genes in biochemical pathways and evolution (polyploidization). Finally, genetic and epigenetic mechanisms for gene expression and metabolic changes were investigated in allotetraploids.

Results

Dosage-Dependent and -Independent Allelic Expression in Arabidopsis Allopolyploids

Arabidopsis allotetraploids (TTAA) were resynthesized by pollinating *A. thaliana* (TTTT) with *A. arenosa* (AAAA) pollen (F1). The synthetic F1 allotetraploids were highly sterile (Comai et al. 2000) and self-pollinated manually for eight generations to improve fertility and to reduce heterozygosity originating from the outcrossing tetraploid *A. arenosa* (fig. 1*A*). These resynthesized F8 allotetraploids were genetically stable and morphologically similar to *Arabidopsis suecica*, a natural allotetraploid (Comai et al. 2000; Wang et al. 2006; Shi et al. 2012). Reciprocal crosses between allotetraploids (TTAA, F8) and *A. arenosa* (AAAA) produced two allopolyploids TAAA (TTAA x AAAA, by convention the maternal parent is listed first in a genetic cross) and AAAT (AAAA x TTAA) that contained the same nuclear genomes but different cytoplasms (*A. thaliana* vs. *A. arenosa*) (fig. 1*A*). Likewise, reciprocal crosses between F8 allotetraploids (TTAA) and *A. thaliana* (TTTT) produced the allopolyploids TATT (TTAA x TTTT) and TTTA (TTTT x TTAA) with the same nuclear genomes and similar cytoplasms. These materials are suitable for studying expression of *A. thaliana* and *A. arenosa* alleles in response to genome-dosage changes within the same tetraploid genome composition. A wide range of morphological diversity, including leaf and flower shape and size, was observed in these lines (fig. 1*B*). Morphological differences between the polyploids were probably caused by varied genome dosages of *A. thaliana* and *A. arenosa*, whereas those between reciprocal polyploids (TAAA vs. AAAT and TATT vs. TTTA) were likely caused by the parent-of-origin effects. These differences are unlikely to result from aneuploids because all lines contained expected chromosome numbers based on fluorescence in situ hybridization data (fig. 1*C*).

Genome-wide allelic expression was estimated in the polyploid series of TTTT, TTTA (and TATT), TTAA (F1, F8, and *A. suecica*), TAAA (and AAAT), and AAAA. Among these tetraploids, T-genome dosages varied 4:3:2:1:0, whereas A-genome dosages varied in the opposite direction: 0:1:2:3:4. Expression levels of the alleles within the same dosage (e.g., F1, F8, and *A. suecica*) were calculated separately but for simplicity were shown at one level (2:2, see Materials and Methods). We mapped the RNA-seq reads and computed allelic transcript levels as described previously (Mortazavi et al. 2008; Shi et al. 2012). Considering that expression levels with lower reads per kilobase per million (RPKM) have greater variation, we chose an RPKM cutoff value (>2) to reduce false positives in a correlation test (Ramskold et al. 2009) (supplementary fig. S1, Supplementary Material online). In addition, biological and technical noise could cause stochastic effects on dosage and expression correlation between replicates. To test this, we calculated the *R* values between allelic and dosage expression in each replicate independently and further compared the *R* value correlation between biological replicates. The correlation was significantly high ($R = 0.92$, $P < 2.2e\text{-}16$) (supplementary fig. S2*A*, Supplementary Material online). Because the noise level is relatively low, we averaged expression levels of three biological replicates to reduce the stochastic effects. Surprisingly, we found that expression levels of most alleles were positively correlated with dosage changes. Only 158 T alleles (~1%) and 171 A alleles (~1%) exhibited negative correlation between their expression levels and dosages (supplementary fig. 2*B* and *C*, Supplementary Material online). Moreover, the negative correlation except for seven T alleles and four A alleles, was not statistically significant (supplementary table S1, in red, Supplementary Material online). This small group of negatively correlated alleles is excluded from further analysis, although it could be potentially interesting and remains to be investigated.

Because the majority of genes show positive correlation, we used coefficients of determination (R^2) between the allele dosage and expression level to quantify the strength of dosage

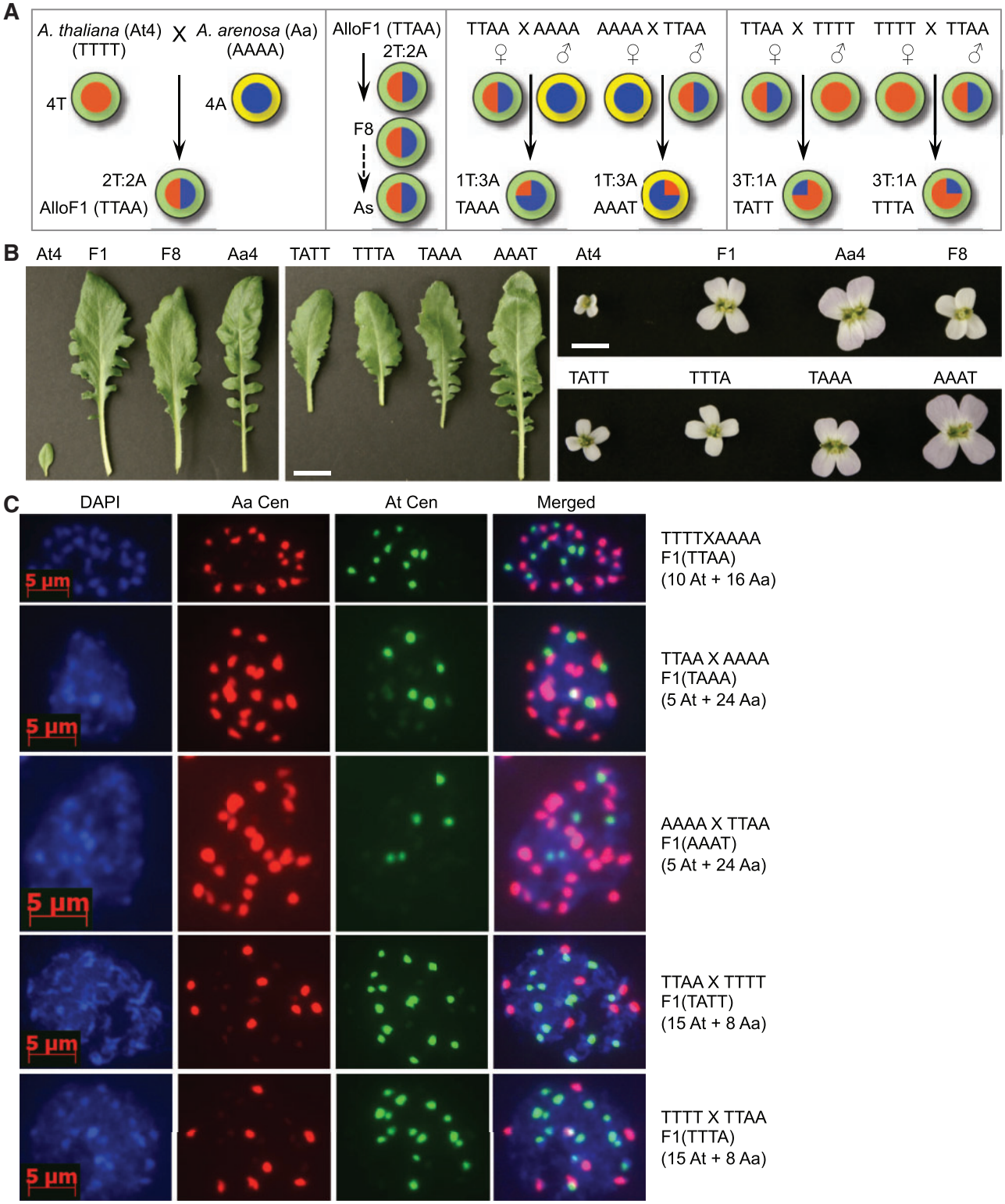

Fig. 1. Genetic materials used to study dosage regulation of gene expression changes in *Arabidopsis* tetraploids. (*A*) An F1 allotetraploid (AlloF1, TTAA) was made between tetraploid *Arabidopsis thaliana* (At, TTTT) and *Arabidopsis arenosa* (Aa, AAAA) and self-pollinated to produce F8 generation, which resembles natural *Arabidopsis suecica* (As). Reciprocal crosses between TTAA(F8) and TTTT resulted in TTTA and ATTT allopolyploids, whereas reciprocal crosses between TTAA(F8) and AAAA gave rise to AAAT and TAAA allopolyploids. Large green and yellow circles indicate *A. thaliana* and *A. arenosa* cytoplasms, respectively; small red and blue circles indicate *A. thaliana* and *A. arenosa* nuclear genomes, respectively. At:Aa genome dosage is also shown for these allotetraploids. (*B*) Leaf (left) and flower (right) morphologies of these materials as shown in (*A*). Scale bars = 10 mm. (*C*) Validation of chromosome compositions in TTAA, TATT, TTTA, TAAA, and AAAT plants. Chromosomes were painted with *A. thaliana* (green) and *A. arenosa* (red) centromeric DNA probes, respectively. Interphase chromosomes (left most in each panel) were stained with 4′,6-diamidino-2-phenylindole (DAPI). Scale bars = 5 mm.

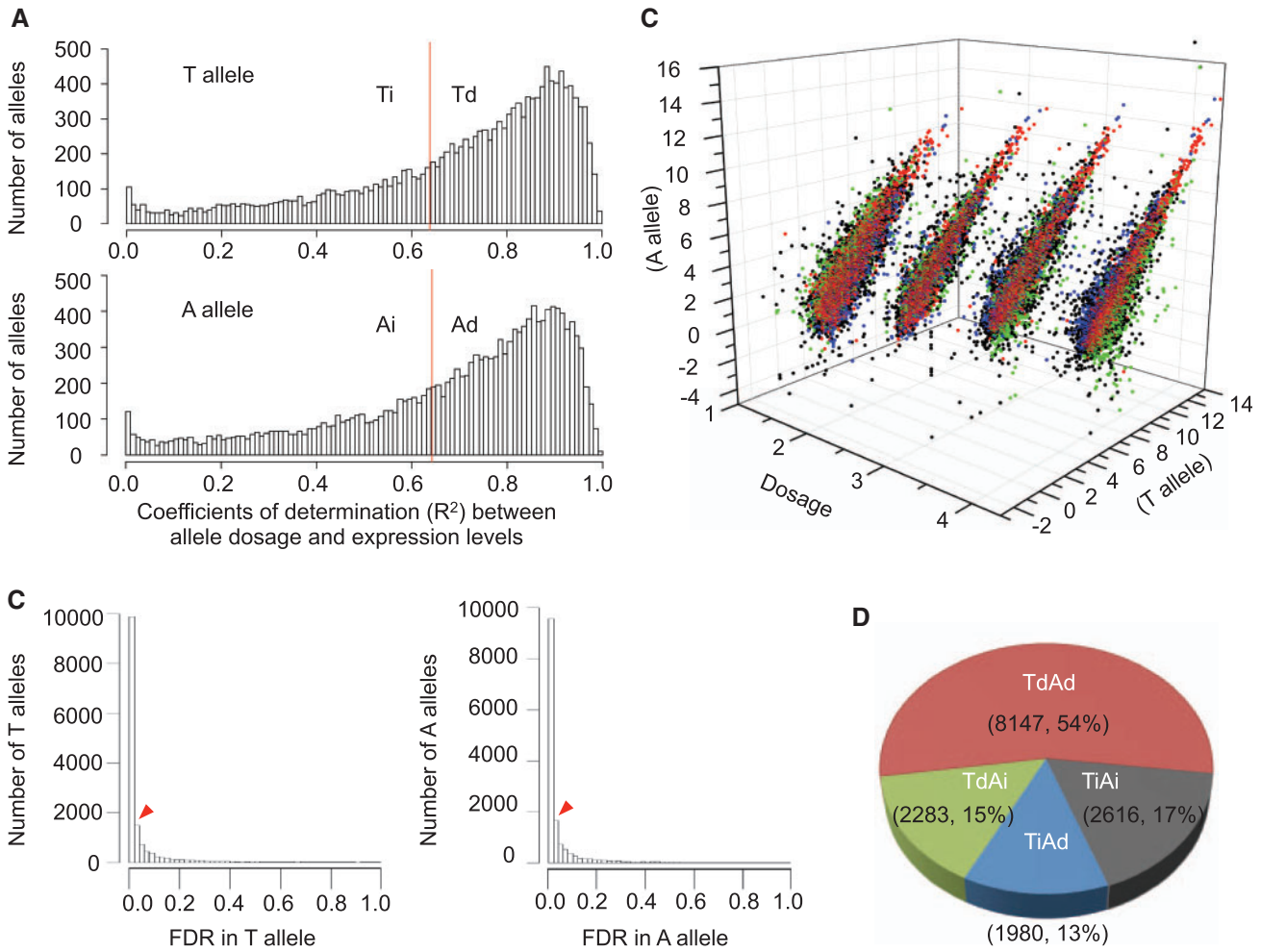

Fig. 2. Genome-wide dosage regulation of homeologous allelic expression in *Arabidopsis* tetraploids. (*A*) Histograms displaying the distribution of coefficient of determination (R^2) in T and A alleles. *X* axis: R^2 value bins that are evenly divided; *Y* axis: number of alleles with the R^2 value in each bin. Red lines separate alleles with dosage-independent and dosage-dependent expression. (*B*) Histograms displaying the distribution of FDR in T (left) and A (right) alleles. *X* axis: FDR-value bins that are evenly divided; *Y* axis: number of alleles with the FDR in each bin. FDR = 0.025 (red arrows) was cutoff value between dosage-dependent and -independent genes. (*C*) Three-dimension display of *Arabidopsis arenosa* (A) allele (*Y* axis) and *Arabidopsis thaliana* (T) allele (*Z* axis) expression (log2 RPKM) in response to dosage changes (*X* axis). Red, green, blue, and black dots indicate TdAd, TdAi, TiAd, and TiAi expression patterns, respectively. Td and Ad represent dosage-dependent expression of T (Td) or A (Ad) alleles, respectively; Ti and Ai represent dosage-independent expression of T (Ti) and A (Ai) alleles, respectively. (*D*) Number and percentage of TdAd, TdAi, TiAd, and TiAi genes.

effects on expression of T and A alleles, respectively (fig. 2*A*). For R^2 in the range from 0 to 0.9, the number of corresponding T and A alleles increased slowly and monotonically and decreased sharply after $R^2 \approx 0.9$. High R^2 values indicate that allelic expression is strongly correlated with dosage changes. The corresponding alleles are defined as dosage dependent, in which the allelic expression level correlates proportionally with the dosage variation. Low R^2 values suggest that the expression level is weakly correlated to the dosage variation, and the corresponding alleles are called dosage-independent. For instance, expression ratios of T alleles of a fructose-1,6-bisphosphatase gene (*HCEF1*) involved in gluconeogenesis and photosystem II assembly were 4.9:3.8:2.6:1, which was close to 4:3:2:1 and dosage dependent ($R^2 = 0.9774$) (supplementary fig. S2*A*, Supplementary Material online). In contrast, T-allele expression ratios of a lipoxygenase gene (*LOX3*) involved in the ethylene biosynthetic process were 0.4:1.7:1.2:1, which deviated substantially from 4:3:2:1, and its expression is dosage independent ($R^2 = 0.0014$) (supplementary fig. S2*B*, Supplementary Material online). Similarly, there were examples of A-alleles that displayed dosage-dependent and -independent expression (supplementary fig. S2*C* and *D*, Supplementary Material online).

Biological Features of Dosage-Dependent and -Independent Genes

To compare biological features of dosage-dependent and -independent alleles, we empirically categorized all alleles into two groups based on coefficient of determination (R^2) using the Pearson correlation test (Benjamini and Hochberg false-discovery rate [FDR]) (fig. 2*A*). The distribution of T and A alleles was clustered around the FDR ranging from 0 to 0.025 and declined sharply at the FDR higher than 0.025

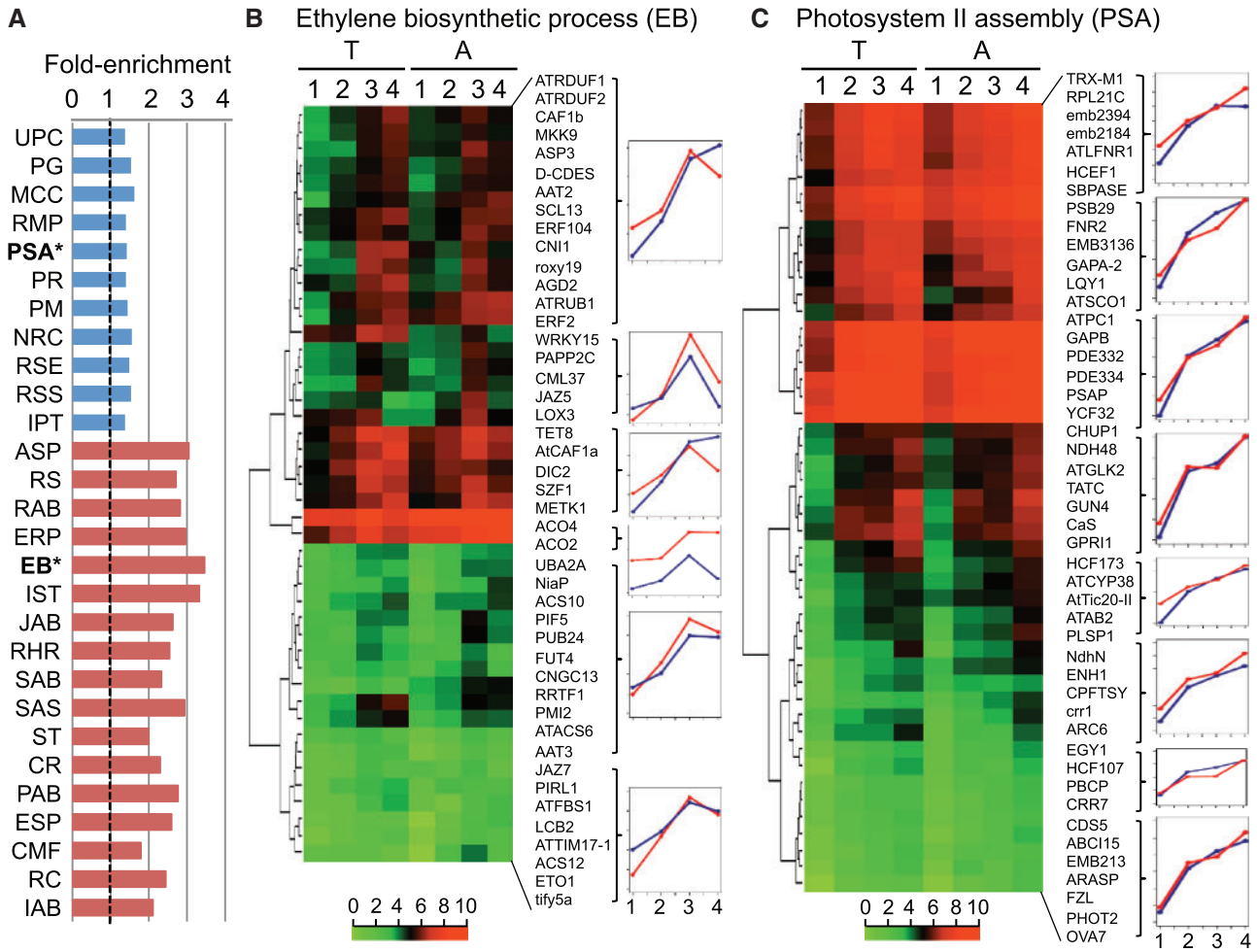

Fig. 3. Dosage-dependent and dosage-independent expression of homeologous loci in *Arabidopsis* tetraploids. (*A*) Fold enrichment of TdAd (blue) and TiAi (red) homeologous genes in GO groups. PSA, photosystem II assembly; EB, ethylene biosynthetic process; see Supplementary figure S4, Supplementary Material online, for names of GO classifications. The dashed line indicates the ratio (=1) of fold changes in each GO group relative to all genes. (*B*, *C*) Heatmaps showing T- and A-allele expression levels (log2 RPKM) of the genes clustered within EB (*B*) and PSA (*C*). Dosage of T or A allele is indicated. Representative expression levels of A allele (red) and T allele (blue) in subgroups of EB (*B*) and PSA (*C*) are shown in right side. *X* axis: dosage; *Y* axis: log 2-expression values.

(fig. 2*B*), which was used as the cutoff value for dosage-dependent and -independent alleles (fig. 2*A*). Those at the statistically significant level were called dosage-dependent T (Td) and A (Ad) alleles, whereas those at the insignificant level are called dosage-independent T (Ti) or A (Ai) alleles. A 3D display of the relationships among dosage (*X* axis), A alleles (*Y* axis), and T alleles (*Z* axis) is shown in figure 2*C*. Note that the distribution is continuous; therefore, under any given cutoff value, some dosage-independent genes could show at least some levels of dosage dependency or vice versa.

Using the cutoff value, a total of 15,026 transcripts analyzed were classified into four groups: TdAd (54%), TdAi (15%), TiAd (13%), and TiAi (17%) (fig. 2*D* and supplementary table S1, Supplementary Material online). At the single-allele level, the majority of T allele (69%) and A allele (67%) displayed dosage-dependent expression, whereas approximately 1/3 showed dosage-independent expression. At the dual-allele level, 54% and 17% of both T and A alleles were dosage-dependent and -independent, respectively (fig. 2*D*). Only 15% (TdAi) and 13% (TiAd) of T and A alleles were in the opposite groups. Based on single-allele numbers, the expected percentages were TdAd (46%), TdAi (23%) and TiAd (21%), and TiAi (10%). The observed proportion of the dual alleles TdAd and TiAi was approximately 7–8% more than expected (Fisher's exact test, $P < 2.2e\text{-}16$), whereas that of TdAi and TiAd alleles was approximately 7–8% less than expected ($P < 2.2e\text{-}16$), indicating that expression of both A and T alleles are likely to have the similar dependency on the dosage.

Gene ontology (GO) analysis indicated that TdAd genes were significantly enriched (FDR < 0.05) in functional categories such as photosynthesis, photorespiration, cell cycle regulation, and primary metabolite production (fig. 3*A*), whereas TiAi genes were enriched in the groups including plant defense and stress responses, cell signaling, and secondary metabolite production (fig. 3*A* and supplementary fig. S4, Supplementary Material online). Examples of overrepresented TiAi genes are given in ethylene biosynthesis process (fig. 3*B*)

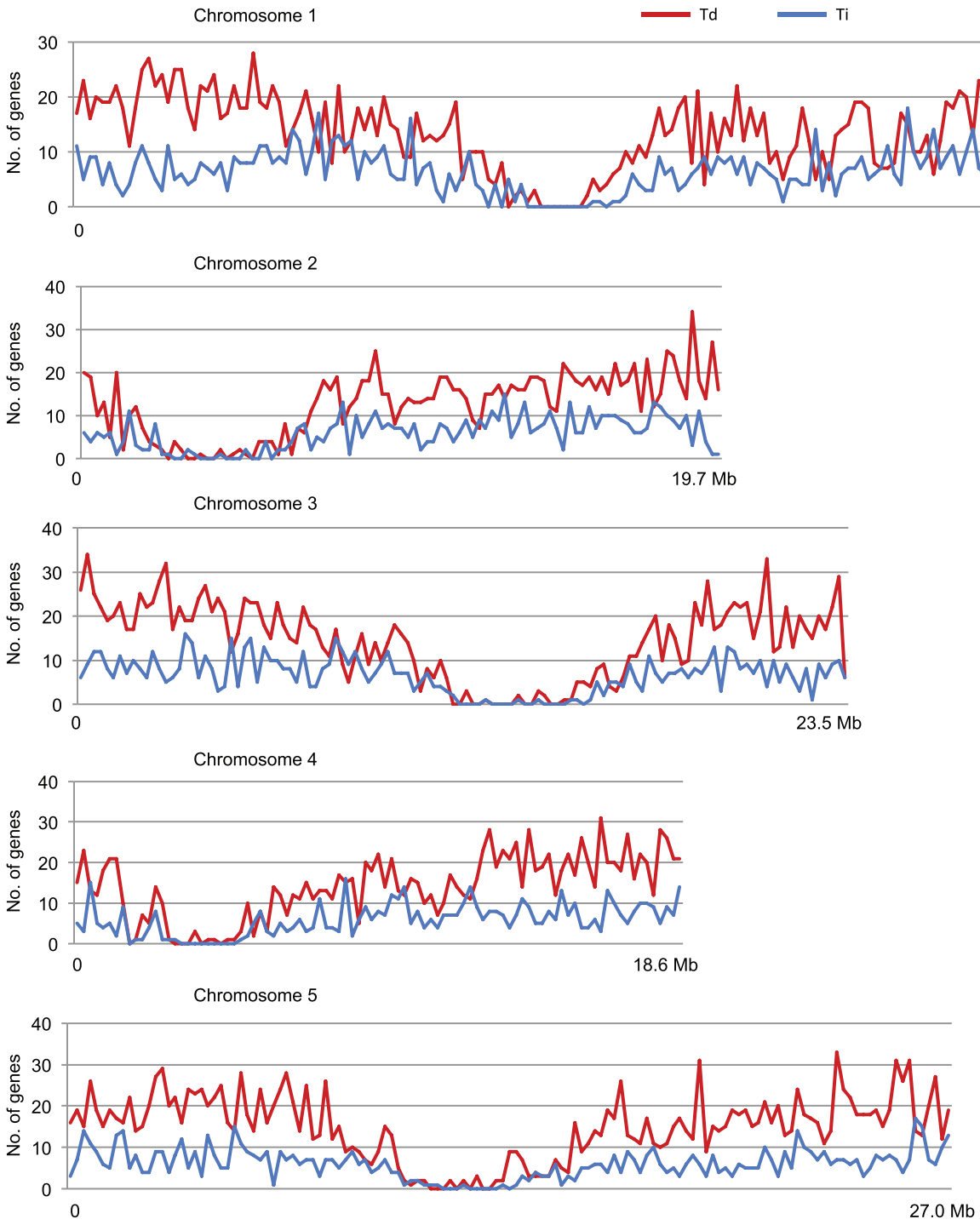

Fig. 4. Distribution of the genes with T allele dosage-dependent (Td, red) and dosage-independent (Td, blue) expression patterns on five chromosomes in *Arabidopsis thaliana*.

and circadian regulation (supplementary fig. S4C, Supplementary Material online). Examples of enriched TdAd genes are shown in photosystem II assembly (fig. 3C) and mitotic cell cycles (supplementary fig. S4D, Supplementary Material online). TdAi and TiAd genes, however, were not enriched in any GO classes. The enrichment of distinct functional classifications suggests that these genes are not randomly grouped. The genes involved in cell cycle, metabolic, and developmental processes tend to display dosage-dependent expression, whereas the genes involved in stress responses are likely subjected to dosage-independent expression.

Notably, the correlation between allelic expression and dosage could be nonlinear for some dosage-independent and -dependent genes. For example, ethylene pathway plays many roles in plant growth and development, including stress response, cell elongation, and fruit ripening (Guo and Ecker 2004). Allelic expression levels increased from one to three copies, which is likely associated with changes in growth and development in the corresponding

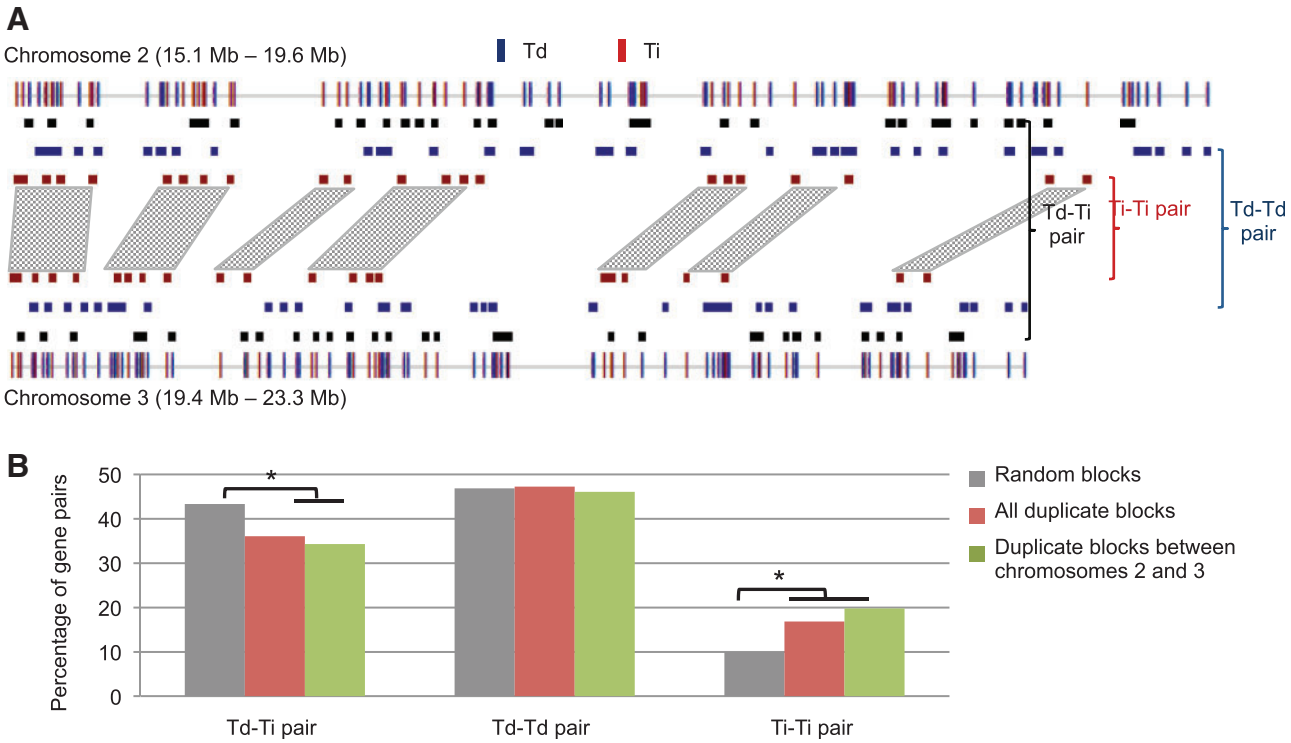

FIG. 5. Dosage-dependent and dosage-independent allelic expression in two segmental blocks between chromosomes 2 and 3. (*A*) Distribution of Td and Ti genes in two duplicate blocks on chromosomes 2 (upper) and 3 (lower) in *Arabidopsis thaliana*. Dark blue and red indicate Td genes and Ti genes, respectively. Gene pairs between two chromosomal regions are indicted by Td–Ti or Ti–Td (black boxes), Ti–Ti (red boxes), and Td–Td (blue boxes), respectively. (*B*) Percentage of Td–Ti, Td–Td, and Ti–Ti gene pairs in random blocks (gray), all duplicate blocks in the genome (red), and duplicate regions between chromosome 2 and 3 (green). Asterisks indicate the difference at a statistically significant level ($P < 0.05$).

allopolyploids (e.g., TAAA, TTAA, and TTTA) (fig. 3*B*). However, the expression level at four copies in the autotetrpaloid parent (either TTTT or AAAA) was decreased. This is consistent with the previous data that many genes in autotetraploids are expressed at similar levels to that of the diploid parent (Wang et al. 2006), probably via a dosage-compensation mechanism (Birchler et al. 2005). The nonlinear correlation between expression and dosage levels may represent complex allelic interactions that could lead to potential advantages or constraints of dosage regulation in polyploids.

Genomic Distribution of Dosage-Dependent and -Independent Genes

At the genome-wide level, neither Td nor Ti genes were clustered along chromosomes (χ^2 test, $P = 1$ [Td], $P = 0.52$ [Ti]), and there were more Td alleles than Ti alleles on most chromosomal segments (*t*-test, $P < 2.2e\text{-}16$) (fig. 4). In the *A. thaliana* genome, there are several rounds of WGD (Vision et al. 2000; Bowers et al. 2003). To test ploidy effects on those ancient duplicate genes, we compared dosage and expression correlations between 1,647 duplicate gene pairs at the whole genome level in *Arabidopsis* diploid (http://chibba.agtec.uga.edu/duplication/), including 168 pairs of paralogs in two large WGD duplicate blocks on chromosomes 2 and 3 (fig. 5*A*). The observed and expected probabilities of duplicate gene pairs with the same (Td–Td or Ti–Ti) or opposite dosage/expression relationship (Td–Ti) were comparatively analyzed (fig. 5*B*). The majority of Td–Td (46–47%) and Td–Ti (34–36%) pairs retained among all duplicate blocks, although they were not statistically different from the genome-wide average. The total number of Ti–Ti pair genes (17–20%) was relatively small but higher than expected. These data suggest that copy number variation of the Ti genes is tolerable in the diploidization process than that of the Td genes. Loss of a Td gene copy could induce a cascade effect on a series of closely related Td gene duplicates to maintain a biosynthetic balance as in the original state prior to duplication. The overrepresentation of dosage-independent genes also suggests an evolutionary potential for functional diversification between homeologous loci in allopolyploids.

Distribution of Dosage-Dependent and -Independent Genes in Biochemical Pathways and Phylostratigraphic Profiles

During evolution, dosage-dependent and -independent allelic expression could lead to functional consequences, including the emergence and maintenance of biochemical pathways that are commonly and/or specifically present among organisms (Fani and Fondi 2009). Living organisms emerged at least 3.8 Ga (Cooper 2000), and green plants appeared approximately 1,400 Ma and approximately 2,400 My after cellular origins. Based on the expected taxonomic range (ETR), biochemical pathways in *Arabidopsis* could be classified into

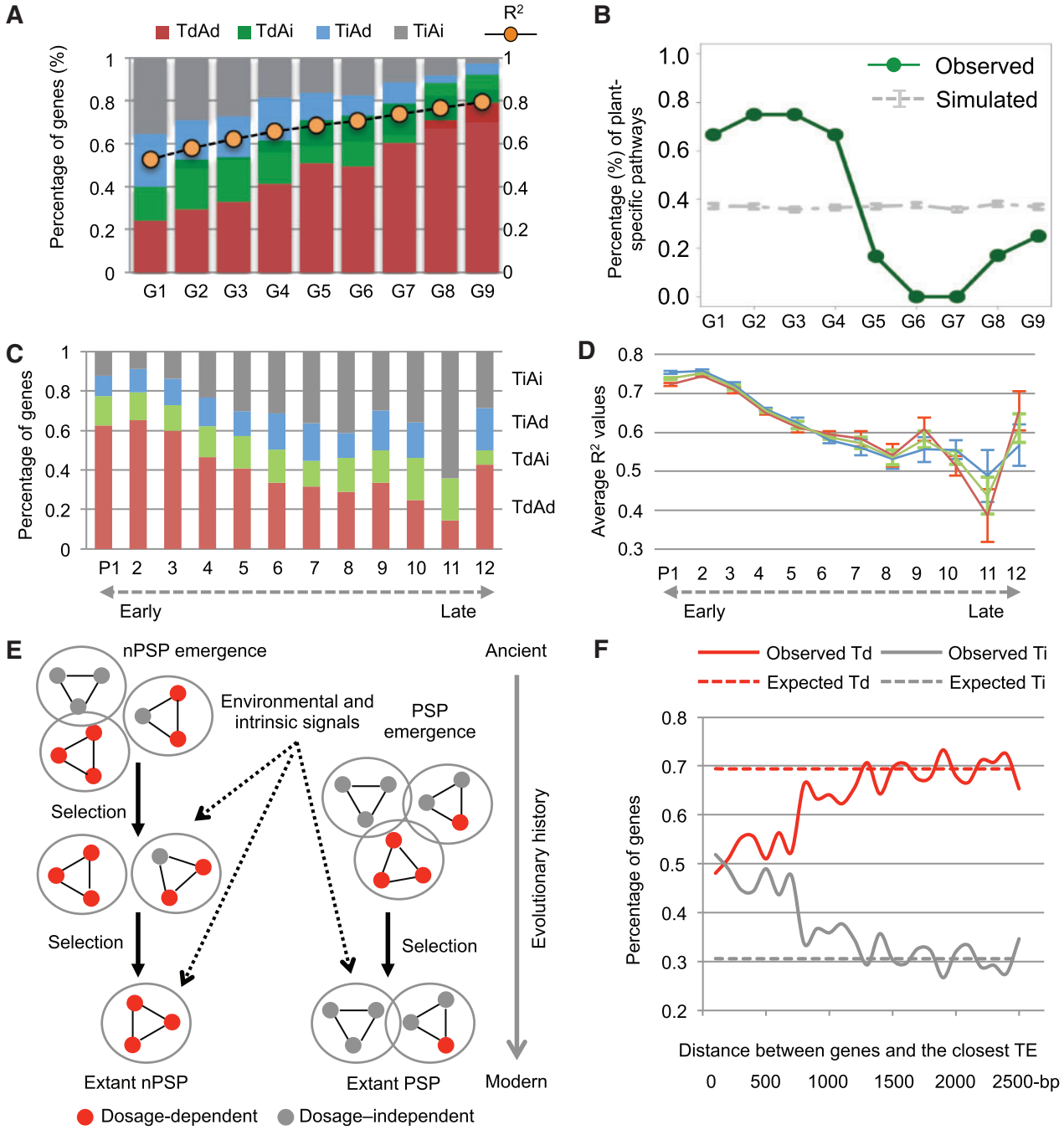

Fig. 6. Evolution of biological pathways in response to dosage-dependent and -independent expression. (*A*) Percentages of dosage-dependent and -independent genes within 108 biochemical pathways, which were divided into nine groups (from G1 to G9) each containing 12 pathways. Barplot displaying percentage of TdAd, TdAi, TiAd, and TiAi genes in each group. Orange circles denote the mean of coefficients of determination (R^2) between dosage and expression levels in each biochemical pathways. (*B*) Percentage of observed (green) and expected (gray) plant-specific biochemical pathways. Expected values were obtained through 1,000 random simulations. (*C*) Percentages of TdAd, TdAi, TiAd, and TiAi genes in 12 phylostrata. (*D*) Correlation of dosage and expression levels of the genes in each phylostratum. Average R^2 values of T (blue) and A (red) alleles were higher in early emerged genes than in late emerged genes. Green denotes mean of A and T alleles. (*E*) An evolutionary model for the dosage-dependent (red balls) and dosage-independent (gray balls) genes in the two types of biochemical pathways. Solid lines connecting balls represent genes in the same pathway. Gray circles surround different individuals carrying the homologous pathway within a population. Environmental cues and intrinsic signals could facilitate process of natural selection, leading to the maintenance of dosage-dependent genes in nPSP (left) and giving rise to dosage-independent genes in PSP (right). (*F*) Percentage of Td (red) and Ti (gray) genes (*Y* axis) that contain TEs close to the genes within a range of distances (base pair, bp; *X* axis).

plant-specific pathway (PSP), and nonplant-specific pathway (nPSP) (Mueller et al. 2003). nPSPs, such as aerobic respiration and tRNA charging, are present in ETR from Archaea, Bacteria to Eukaryote. These nPSP pathways were formed in the remote ancient time before plants diverged from other species and were conserved during evolution. Some pathways such as photosynthesis are conserved in plants, but they are widely spread from Bacteria to Eukaryota. These pathways are

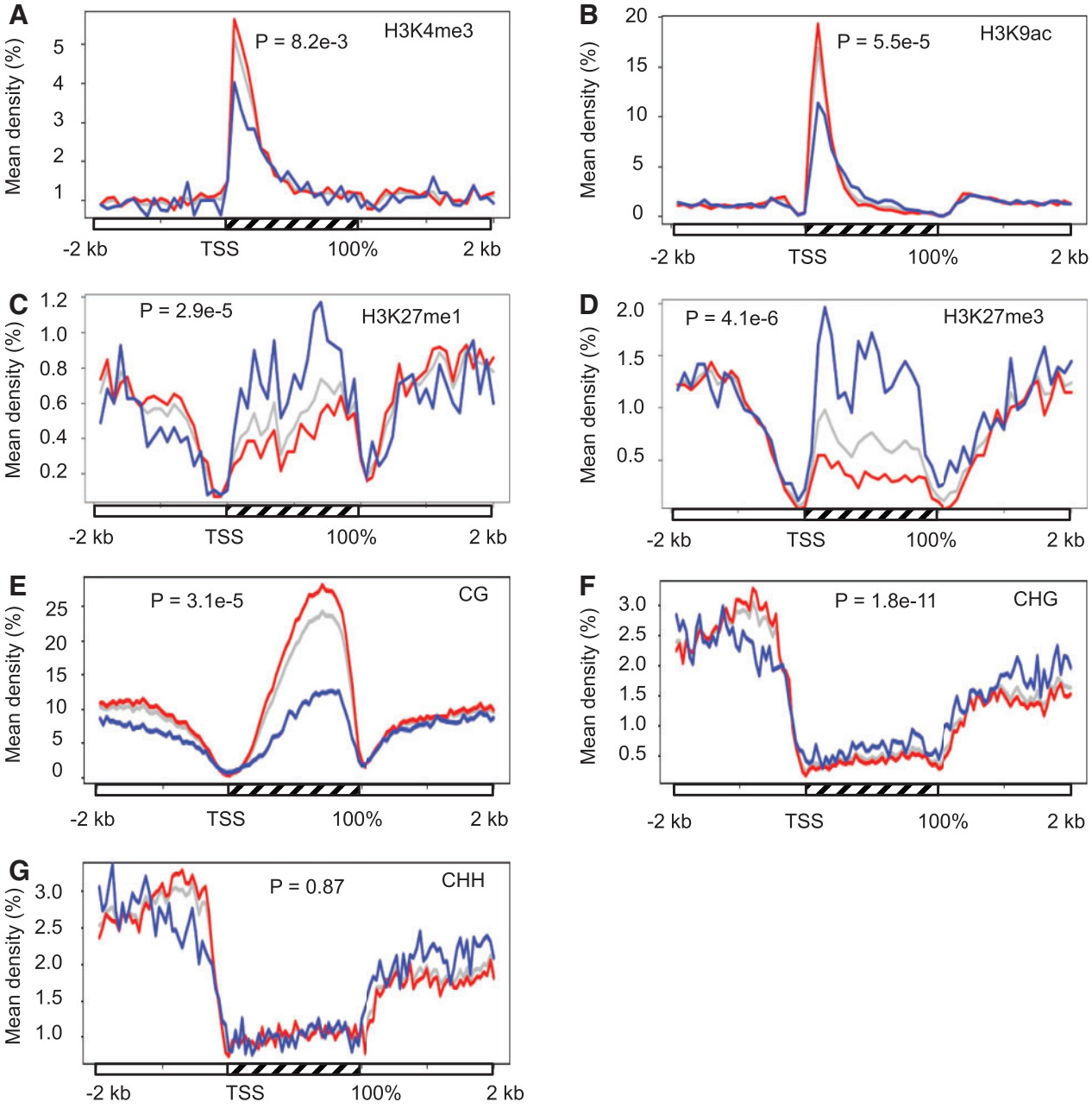

FIG. 7. Epigenetic regulation of dosage-dependent and -independent gene expression. (*A*, *B*) Mean density (50-kb window) of H3K4me3 (*A*) and H3K9ac (*B*) in Td (red), Ti (blue), and all (gray) genes, which were enriched more around TSS in Td genes than in Ti genes. TSS, transcription start site. (*C*, *D*) Mean density (50-kb window) of H3K27me1 (*C*) and H3K27me3 (*D*), which were enriched more near the 5′ and transcribed region in Ti genes than in Td genes. (*E*–*G*) Mean density (50-kb window) of CG (*E*), CHG (*F*), and CHH (*G*) methylation sites. CG methylation was enriched more in transcribed regions in Td than in Ti genes (*E*), whereas CHG methylation was enriched more around transcribed regions in Ti genes than Td genes (*F*). No significant difference was found in the CHH methylation between Td and Ti genes (*G*).

also classified into nPSP. PSPs, such as jasmonic acid biosynthesis and anthocyanin biosynthesis, exist in ETR of the plant kingdom including Viridiplantae, Brassicales, and Embryophyta. These PSP pathways either emerged in plants after their split from other species or lost in other clades during evolution.

We classified 108 biochemical pathways into PSPs and a wider range of taxonomic nPSPs, each consisting of at least ten genes with known dosage expression correlation. Based on the ranked average correlations between dosage and gene expression, 108 pathways were evenly divided into nine groups. Each group consists of 12 pathways, and the average dosage expression correlation increases monotonically from group one to nine (fig. 6*A*, supplementary fig. S5 and table S2, Supplementary Material online). The proportion of PSPs decreased as the correlation between allelic dosage and expression increased (fig. 6*A* and *B*), indicating that most pathways consisting of dosage-independent genes appeared after those involving dosage-dependent genes. The dosage-dependent (Td and Ad) genes represent the majority of common pathways. In contrast, dosage-independent (Ti and Ai) genes appear more in PSPs. For example, aerobic respiration is

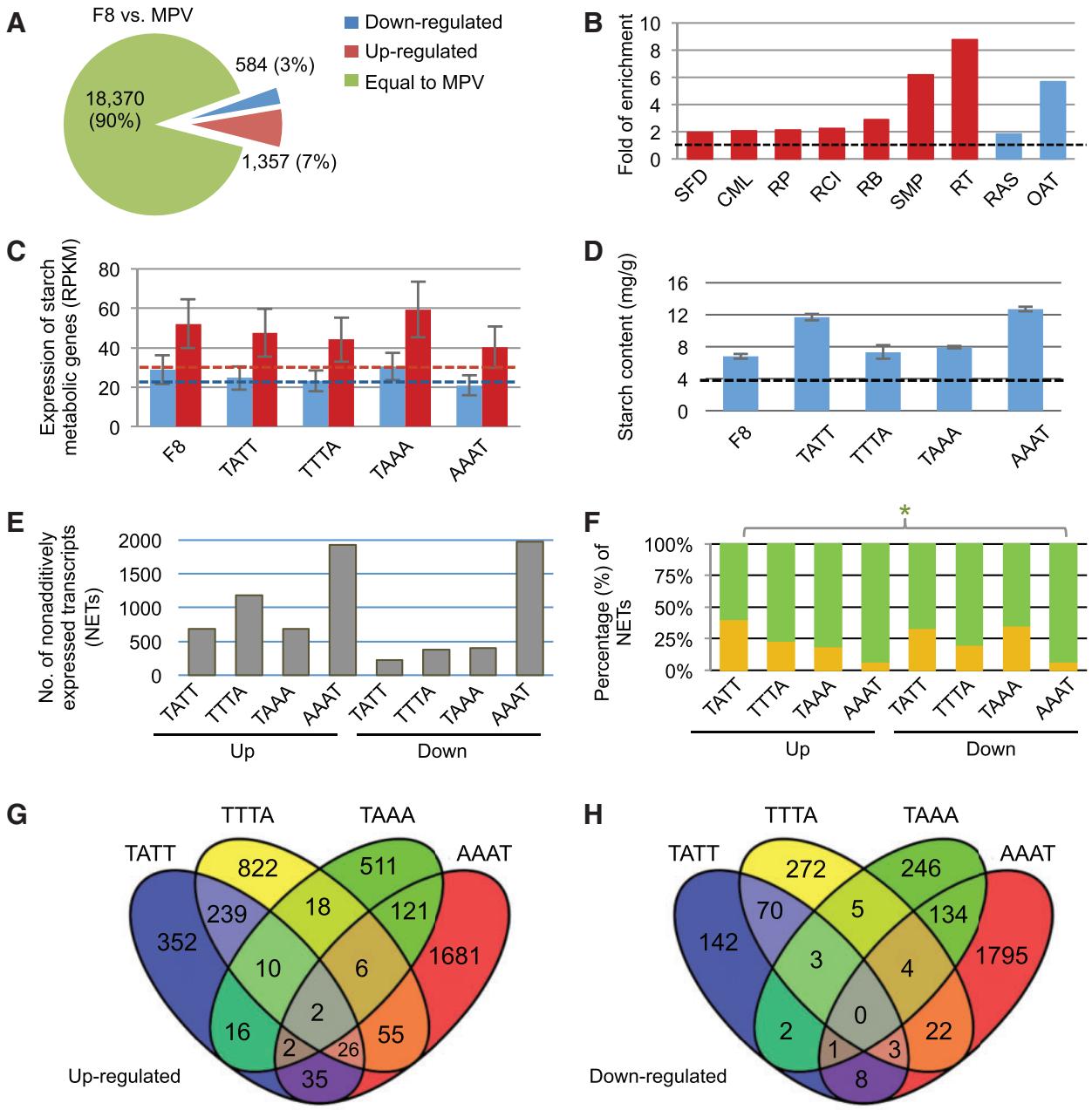

FIG. 8. Gene expression differences in polyploids with the same/different T and A dosages. (*A*) Number of genes that show additive and nonadditive (up- or downexpression) patterns in TTAA F8 relative to the MPV. (*B*) Fold enrichment of GO groups of nonadditively expressed genes in stable F8 allotetraploids (TTAA). SFD, seed and fruit development; CML, cellular macromolecule localization; RP, RNA processing; RCI, response to cadmium ion; RB, ribosome biogenesis; SMP, starch metabolic process; RT, RNA transport; RAS, response to abiotic stimulus; OAT, organic acid transport. (*C*) Expression of starch synthesis (blue) and degradation (red) genes in the parents (MPV, dot lines), F8, TATT, TTTA, AAAT, and TAAA. Averages and standard errors for all studied transcripts are shown. (*D*) Starch content (mean and standard error of three replicates) in the parents (MPV, dot line), F8, TATT, TTTA, AAAT, and TAAA. (*E*) Number of nonadditively expressed transcripts (NETs) in TATT, TTTA, TAAA, and AAAT allotetraploids. (*F*) Percentage of NETs whose expression changes were shared between reciprocal crosses (orange) or unshared (green). Asterisks indicate significant difference between unshared (green) and shared (orange) transcripts ($P < 0.05$). (*G*, *H*) Venn-diagram analysis for the number of upregulated (*G*) or downregulated (*H*) transcripts in TATT, TTTA, TAAA, and AAAT allopolyploids.

required for almost all organisms living in an aerobic environment to synthesize ATP and exists in Bacteria, Fungi, and Viridiplantae. In contrast, jasmonates, the derivatives of jasmonic acid, are important regulators of diverse developmental processes in plants. The jasmonic acid biosynthesis pathway is specifically identified in Viridiplantae. The average correlation values between dosage and expression levels of the genes in these two pathways were 0.82 and 0.45, respectively.

In an independent test, the genes were separated into more taxonomic units before and after plant origins according to phylogenetic hierarchy phylostratigraphic profiles

(Tautz and Domazet-Loso 2011; Quint et al. 2012). Each gene was classified into its phylostratum (PS1–PS12). The profiles of dosage dependency of T and A allele across 12 phylostrata are shown in supplementary fig. S6A, Supplementary Material online. PS3 denotes the divergence of plants from other organisms. Correlation between the dosage and allelic expression levels monotonically decreased from PS1 to PS8 but slightly rose from PS9 to PS12. The fluctuations could be related to the small number of genes in the last two groups (PS11 and PS12). Together with the data of pathway analysis, the results collectively indicate that expression of evolutionarily older genes is strongly correlated with the dosage. This trend is applied not only to the genes involved in common PSPs but also to the genes successively arising at different levels of plant phylogenetic hierarchy. The dosage-dependent genes might maintain and conserve common pathways through natural selection and regulation of cellular functions (fig. 6*E*). On the other hand, the dosage-independent genes might have contributed to the optimization of PSPs through the combination of gene expression regulation and natural selection.

Distribution of TEs Near the Dosage-Dependent and -Independent Genes

Plant and animal genomes contain 50–80% or more of transposable elements (TEs) (Feschotte et al. 2002), which could affect expression of inserted and neighboring genes. It is shown that the distance of TE insertion is associated with the expression of neighboring genes (Zemach et al. 2013).

Here, we found that dosage-dependent genes tend to be farther away from TEs than dosage-independent genes for T allele ($P < 4.0e\text{-}7$, fig. 6*F*). It is reasonable to assume that there is generally equal chance for a TE to insert into or be close to Td and Ti genes. Fewer TEs present near the dosage-dependent genes suggest that such TEs are deleterious and have been eliminated by purifying selection. On the other hand, more TEs near the dosage-independent genes indicate that such TEs are likely to be advantageous and preserved by positive selection or genetic drift. The enrichment of TEs near the dosage-independent genes including stress-responsive genes is consistent with the finding that the dynamic DNA methylation changes in transposons and repeats can regulate expression of neighboring genes in response to biotic stress (Dowen et al. 2012). The data suggest TEs could serve as a sensor for dosage-independent genes to change their expression in response to internal (dosage change) and external (environmental) signals.

Association of Allelic Expression with Histone Modifications in *A. Thaliana*

Allelic expression in polyploids could be regulated by various epigenetic factors, including DNA methylation and histone modifications (Chen 2007; Song et al. 2015). Histone modifications mediate epigenetic regulation of gene expression and development in plants and animals (Li et al. 2007; Zhang 2008). Some modifications, such as acetylation of histones H3 and H4 and trimethylation of H3 lysine 4 (H3K4me3), are known as euchromatic marks and often associated with gene activation, whereas other modifications, such as methylation of H3K9 and H3K27, are known as heterochromatic marks and related to gene repression (Jenuwein and Allis 2001; Li et al. 2007).

Two active histone marks, H3K9ac and H3K4me3, were correlated and enriched near transcriptional start sites in *A. thaliana* and its related species (Ha et al. 2011). *Arabidopsis arenosa* genome has yet to be sequenced, so we used the data in *A. thaliana* to infer the association of allelic expression with epigenetic marks. The number of Td alleles associated with the density of H3K4me3 (Wilcoxon signed-rank test, $P = 8.2e\text{-}3$, fig. 7*A*) and H3K9ac ($P = 5.5e\text{-}5$, fig. 7*B*) was clearly higher than that of Ti alleles, suggesting a role for these active histone marks in maintaining dosage-dependent expression. On the other hand, the number of Ti alleles associated with the density of H3K27me1 ($P = 2.9e\text{-}5$, fig. 7*C*) and H3K27me3 ($P = 4.1e\text{-}6$, fig. 7*D*) histone marks was higher than Td alleles. A large number of genes (~4,400) are associated with H3K27me3 in *A. thaliana* (Zhang et al. 2007). Together, the enrichment of H3K27me1 and H3K27me3 in Ti genes and the large number of H3K27 methylation-associated genes imply an important role for H3K27 methylation in allelic expression variation in a dosage-independent manner.

In *A. thaliana*, DNA methylation of CG, CHG, and CHH (H = A, T, or C) sites directly or indirectly affects gene expression (Zhang et al. 2006; Lister et al. 2008). We found that CG methylation was significantly enriched in the 5′ and coding regions of the dosage-dependent alleles but underrepresented in the dosage-independent alleles ($P = 3.1e\text{-}5$; Fig. 7*E*). CHG methylation distributions were higher in genic regions of dosage-independent alleles than that of dosage-dependent alleles but higher in the 5′ regions of dosage-dependent alleles than that of dosage-independent alleles ($P = 1.8e\text{-}11$; fig. 7*F*). CHH methylation distributions were similar between the genes with dosage-dependent and -independent expression (fig. 7*G*). The CG methylation maintains the stability of dosage-dependent expression of house-keeping genes, which is essential for cellular biosynthesis and metabolism (Zhang et al. 2006). CG methylation is enriched at TEs and gene bodies, whereas CHG methylation is predominantly restricted to TEs (Zhang et al. 2006; Cokus et al. 2008). Only a small fraction of genes with dosage-independent expression is associated with CHG methylation within the gene body.

Nonadditive Expression and Starch Metabolic Activities in Polyploids

Growth vigor is commonly observed in allopolyploids and hybrids at different ploidy levels (Ni et al. 2009; Miller et al. 2012), which correlates with increased metabolic activities. Because *Arabidopsis* allotetraploids in early generations (F1–F4) are unstable with low fertility (Comai et al. 2000), we investigated allelic expression in stable allotetraploids (TTAA, F8), when fertility is improved and heterozygosity within genotypes is reduced by self-pollination. As predicted, in F8 allotetraploids, the majority of genes (~90%) were expressed at the midparent value (MPV), whereas

approximately 7% of genes were upregulated and approximately 3% of genes were downregulated relative to the MPV (fig. 8A, supplementary fig. S6B–E and table S3, Supplementary Material online). The upregulated genes were enriched in the GO groups of ribosome biogenesis, RNA processing, and starch metabolic process (fig. 8B and supplementary fig. S6C, Supplementary Material online), whereas the downregulated genes belonged to GO groups of abiotic stimuli and organic acid transport (fig. 8B and supplementary fig. S6E, Supplementary Material online). Interestingly, expression levels of starch degradation genes in five allotetraploids, including F8, TATT, TTTA, ATAA, and TAAA, were higher than the MPV (fig. 8C and supplementary fig. S7, Supplementary Material online), whereas the starch biosynthetic genes were expressed at slightly higher than or similar to the MPV. The starch content of these five allotetraploids was significantly higher than the MPV (fig. 8D). The data indicated that allotetraploids accumulated more starch than their parents at noon (ZT6; Zeitgeber time 0 [ZT0] = dawn, when the samples were harvested). In *A. thaliana*, starch accumulates during the day and is degraded at night to promote carbohydrate metabolism, and almost all starch is used by dawn (Graf et al. 2010). The more starch accumulates during the day; the more starch can be degraded at night to promote growth (Chen 2013). These data suggest an important role for starch biosynthesis and utilization in stimulating growth vigor in allopolyploids.

Parent-of-Origin Effects on Nonadditive Gene Expression in Reciprocal Tetraploids

The polyploids from the reciprocal crosses have the same nuclear genome composition but different cytoplasms. TAAA is derived from pollinating Allo (TTAA) with *A. arenosa* (AAAA) pollen, and AAAT is derived from pollinating *A. arenosa* (AAAA) with Allo (TTAA) pollen. Likewise, pollinating Allo (TTAA) with *A. thaliana* (TTTT) pollen produced TATT and pollinating *A. thaliana* (TTTT) with Allo (TTAA) pollen produced TTTA. The parent-of-origin effect on starch content (supplementary table S4, Supplementary Material online) and morphological differences (fig. 1B) is consistent with gene expression changes in these reciprocal polyploids. Nonadditive expression of a gene in the F1 cross denotes up- or downregulation of the gene relative to the MPV. Overall, more transcripts were nonadditively expressed in TAAA (686 upregulated and 395 downregulated) than in TATT (682 upregulated and 229 downregulated) and a lot more in AAAT (1,928 upregulated and 1,967 downregulated) than in TTTA (1,178 upregulated and 379 downregulated) (fig. 8E). The data suggest that *A. arenosa* has larger effects on gene expression in the allopolyploids than *A. thaliana*, which are consistent with overall gene expression dominance of *A. arenosa* over *A. thaliana* in resynthesized allotetraploids (Wang et al. 2006). Furthermore, over 93% of upregulated and 93% of downregulated transcripts were found only in TAAA (TTAA X AAAA), and over 80% of upregulated and 64% of downregulated transcripts were specific to AAAT (AAAA X TTAA) (fig. 8F). The data suggest that in the reciprocal crosses involving *A. arenosa*, more genes displayed expression changes when *A. arenosa* is used as the paternal parent than as the maternal parent.

Consistent with the above notion, the unshared nonadditively expressed transcripts between each pair of reciprocal crosses were significantly higher than the shared ones (fig. 8F, $P < 0.05$, χ^2 test), demonstrating a parent-of-origin effect on nonadditive gene expression. In each pair of reciprocal polyploids, more transcripts were upregulated than downregulated, except for TAAA that had similar number of up- and downregulated genes (fig. 8E–G). Among upregulated transcripts, 277 were shared between TATT and TTTA, whereas 405 (59.4%) and 901 (76.5%) were unique to TATT and TTTA crosses, respectively (fig. 8G). Among downregulated transcripts, 76 were shared, whereas 153 (66.8%) and 303 (79.9%) were found only in TATT and TTTA crosses, respectively (fig. 8H). A similar trend of nonadditive expression was also found in AAAT and TAAA crosses but with a larger number of genes. These nonadditively expressed genes specific to each reciprocal cross may contribute to morphological differences between individual allopolyploid lines (fig. 1B). Notably, the crosses between TTAA and AAAA showed the parent-of-origin effect on the largest number of genes (fig. 8G and H). This could be attributed to larger cytoplasmic effects of *A. arenosa* and/or high levels of heterozygosity derived from outcrossing *A. arenosa*.

Discussion

After neopolyploids overcome a bottleneck of sterility and instability (Mayer and Aguilera 1990; Comai et al. 2000), they become genetically stable with respect to meiotic chromosome pairing and segregation (Wang et al. 2006). In the allotetraploids with different combinations of genome dosages, gene expression is dependent or independent of the genome dosage changes. Dosage-dependent genes are mainly associated with primary metabolic and biosynthetic activities, which are in anciently formed biochemical pathways and evolutionarily older phylostrata in an *Arabidopsis* phylostratigraphic profiles. These dosage-dependent genes are largely devoid of TE insertions, suggesting an evolutionary importance of these genes for maintaining growth and developmental stability. In contrast, dosage-independent genes are involved in processing internal (dosage and hybridization) and external (environmental stress) cues, which are present in PSPs and late-originating phylostrata. Dosage-independent genes are prone to TE insertions, which may facilitate expression modulation in response to external and internal changes, leading to improved fitness and adaptive evolution. Indeed, TEs and small RNAs contribute to gene expression divergence between *A. thaliana* and its related species *Arabidopsis lyrata* (Hollister et al. 2011).

In the allopolyploid series, both *A. thaliana* (T) and *A. arenosa* (A) alleles are subjected to dosage-dependent and -independent regulation. Approximately 56% of T and A alleles are in the same direction in terms of dosage-dependent expression (46% TdAd) or dosage-independent expression (10% TiAi), whereas 44% of T and A alleles (23% TdAi and 21% TiAd) are in different directions. For those genes with allelic dosage-dependent expression, their

expression levels are dependent on allele dosage, and their functions are likely to maintain and stabilize ancestral functions including cell cycle regulation, photosynthesis, and metabolism. In polyploids, mutation on a given allele would be compensated by other copies to buffer deleterious effects on important biological functions. This alleviates negative selection and allows the survival of individuals bearing such mutations. Thus, the individuals with a mutation in the dosage-dependent genes would have a higher chance of survival in polyploids.

For those genes with different allelic expression directions or dosage-independent expression, they can diverge their expression from constraints of dosage regulation, which would allow them to gain new functions (neofunctionalization) (Lynch et al. 2001) or diverge from their ancestral functions (subfunctionalization) (Lynch and Force 2000). These genes are largely responsible for the response to external signals and environmental stresses, and their expression is likely to rapidly and stochastically change during evolution. These duplicate loci can diverge their expression in response to internal (intergenomic interactions) and external (new environmental signals) changes. These allelic expression changes likely provide a molecular basis for the polyploids to prevent from entering an evolutionary dead end, as previously predicted (Stebbins 1971). This "dead-end" prediction has been revisited (Soltis et al. 2014), and our results have provided a new genomic basis for polyploidy to become a pervasive force in evolution and speciation. The overall genomic structure and dosage-dependent expression can remain relatively stable in polyploids, but allelic expression divergence between the homeologous loci can promote genetic diversity within and between polyploid populations. The dosage-dependent and -independent regulation leads to dual consequences of allelic expression stability and divergence between homeologous loci, promoting gene expression basis for polyploid evolution and speciation.

At the mechanistic level, dosage-dependent genes tend to be associated with H3K9ac, H3K4me3, and CG methylation, whereas dosage-independent genes tend to correlate with H3K27me3 and CHG methylation. H3K4me3 in coordination with CG methylation may maintain stability of dosage-dependent gene expression, as gene-body methylation is associated with constitutively expressed genes (Zhang et al. 2006). Enrichment of H3K27me3 and CHG methylation in TEs may suggest a role for these modifications to regulate dosage-independent gene expression through epigenetic modifications of TEs adjacent to these genes (Hollister et al. 2011). TEs could serve as a sensor for dosage-independent genes to change their expression in response to internal and external signals. Together, both genetic and epigenetic mechanisms could provide ploidy and parent-of-origin effects on gene expression during formation and evolution of allopolyploids.

Materials and Methods

Plant Materials

Plant materials included *A. thaliana* autotetraploids (At4; CS3900), *A. arenosa* (Aa; CD3901), F1 resynthesized allotetraploids (AlloF1), F8 allotetraploids (AlloF8; CS3895), a natural allotetraploid *A. suecica*, reciprocal crosses between AlloF8 and At4, generating TATT and TTTA tetraploids, and reciprocal crosses between AlloF8 and Aa, producing TAAA and AAAT tetraploids. Plant care and growth followed the protocols as described previously (Wang et al. 2006). Rosette leaves before bolting (flowering) were collected for DNA and RNA analysis.

Fluorescence In Situ Hybridization

The chromosome compositions in polyploids were confirmed using fluorescent in situ hybridization (FISH). General FISH procedures follow those described in Lysak et al. (2006) and Kato et al. (2004). Flower buds (<1 mm) were fixed in ice-cold Carnoy's Fixative for at least 1.5 h. Chromosomal DNA was cross linked to slides by exposure to UV light (120 mJ/cm^2). Centromere repetitive sequences of At (AtCEN) and Aa (AaCEN) were polymerase chain reaction amplified from genomic DNA using primers as described in Comai et al. (2003) and used as hybridization probes.

Hybridization of chromosome mounts was performed at 55 °C for 4–6 h in the presence of the labeled probes. DAPI (Vector Laboratories, Inc., Burlingame, CA) was applied to hybridized slides before observation. Images were taken using a Zeiss Axiovert 200 M microscope equipped with a Zeiss Axiocam fluorescence camera. Images were processed using Adobe Photoshop 7.0.

Starch Assays

Starch content was measured from rosette leaves of three plants before bolting as described previously (Ni et al. 2009).

RNA Preparation and RNA-Seq Data Analysis

Total RNA was isolated from rosette leaves before bolting (3–4-week old in A. *thaliana* and 6–7-week old in *A. arenosa* or allotetraploids). mRNA-seq libraries were prepared using Illumina mRNA-seq sample preparation kit (Illumina, San Diego, CA) and sequenced by HiSeq 2000 (40–100-bp read length). RNA-seq reads from 27 libraries (three biological replicates in each of nine genotypes) were mapped to TAIR9 genome, and transcript levels were quantified by counting RPKM as described previously (Mortazavi et al. 2008; Shi et al. 2012). Read assignment and transcript quantification of At and Aa alleles in an allotetraploid using two or more distinct single-nucleotide polymorphisms (SNPs) between At and Aa were performed using a published method (Shi et al. 2012). Genes whose two alleles have too few SNPs were not considered. Alleles with small SNP numbers produce the same or similar transcripts and proteins. Therefore, it is reasonable to neglect effect of dosage change to expression of such alleles. We selected 15,026 transcripts, each of which averagely had two or more RPKM in nine genotypes, for further analyses.

Pearson correlation tests between allelic expression and genotype dosage (4:3:2:1) for each T and A alleles were performed, respectively. Expression levels of the alleles with the same dosage (e.g. F1, F8, and *A. suecica*) were computed

separately. For example, for T allele of AT1G01040, a RNA helicase involved in microRNA processing, expression level of the T allele at two 1 T genomes (AAAT and TAAA), three 2 T genomes (F1, F8, and *A. suecica*), two 3 T genomes (TTTA and TATT), and one 4 T genome (At) are 1.64, 1.77, 3.37, 3.59, 5.08, 4.53, 5.06, and 6.36, respectively. Pearson correlation between the dosage (1, 1, 2, 2, 2, 3, 3, and 4) and expression (1.64, 1.77, 3.37, 3.59, 5.08, 4.53, 5.06, and 6.36) is 0.92, and R^2 is 0.8476 (supplementary table S1, Supplementary Material online). FDR multiple test correction was calculated using adjustment method in *R*. Alleles with significant dosage and expression correlation (FDR < 0.025) were defined as dosage dependent, whereas alleles with no significant dosage and expression correlation were defined as dosage independent. We calculated Pearson correlation coefficient of dosage-associated allelic expression between genes. High correlation coefficient indicated that the two genes had similar expression profiles and were assigned to the same cluster by implementing Kmeans clustering. All genes tested were classified into ten clusters with several combinations of dosage-dependent and -independent regulation (supplementary fig. S2, Supplementary Material online). The transcripts were assigned to different biochemical pathways according to the *Arabidopsis* information resource database AraCyc.

GO analysis was performed using the DAVID Functional Annotation Tool (http://david.abcc.ncifcrf.gov/tools.jsp). TdAd, TdAi, TiAd, and TiAi genes were compared with the whole *Arabidopsis* genome to display whether the genes in a specific dosage group were enriched in certain GO categories (FDR < 0.05).

For distribution of Td and Ti genes on each chromosome, we divided the chromosome into "bins" each containing 25 genes and counted the number of Td and Ti genes in each bin. Chi-square goodness of fit test was used to test the null hypothesis that Td and Ti genes are equally represented among bins. The Student's *t*-test was used to test whether number of Td and Ti genes is significantly different.

Epigenomic Data Analysis

Data of two active histone marks, H3K9ac and H3K4me3, were from a previous publication of our laboratory (Ha et al. 2011). ChIP-Seq data of H3K27me1 (Jacob et al. 2010) and H3K27me3 (Luo et al. 2013) were downloaded from NCBI GEO. Peak identification was implemented by ERANGE3.3 (Mortazavi et al. 2008) using H3 ChIP-seq as the control sample. Peak number within the sliding window between 2 kb upstream and 2 kb downstream of the transcribed region was computed for frequencies of Td and Ti genes, respectively. In flanking regions, sliding window size is 100 bp, and varied window size was applied in coding region according to coding sequence length. Methylome (bisulfite) sequencing data were downloaded from http://www.ebi.ac.uk/ena/data/view/ERP000902 (Becker et al. 2011). BSMAP 2.74 was implemented to perform the bisulfite mapping and compute ratio of methylated cytosines_ENREF_35 (Xi and Li 2009). Statistical analyses were subjected to *t*-test and multiple test correction using SGoF + v7.2 (Carvajal-Rodriguez et al. 2009).

Supplementary Material

Supplementary figures S1–S7 and tables S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Prof. Luca Comai (University of California, Davis) for stimulating discussions in the polyploid genome project, Dr Danny Ng and Dr Lu Tian, former members in the Chen Laboratory for experimental assistance, and Angela Bardo (Core Facility at The University of Texas at Austin) for assistance in image acquisition. Special thanks are given to the associate editor and two anonymous reviewers for their insightful and constructive comments on improving data analysis and paper presentation. This work was supported by grants (ISO1238048 and ISO1025947 to Z.J.C.) from the NSF Plant Genome Research Program.

References

Adams KL, Cronn R, Percifield R, Wendel JF. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Natl Acad Sci U S A.* 100:4649–4654.

Becker C, Hagmann J, Muller J, Koenig D, Stegle O, Borgwardt K, Weigel D. 2011. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* 480:245–249.

Birchler JA, Riddle NC, Auger DL, Veitia RA. 2005. Dosage balance in gene regulation: biological implications. *Trends Genet.* 21:219–226.

Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unraveling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422:433–438.

Buggs RJ, Zhang L, Miles N, Tate JA, Gao L, Wei W, Schnable PS, Barbazuk WB, Soltis PS, Soltis DE. 2011. Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr Biol.* 21:551–556.

Carvajal-Rodriguez A, de Una-Alvarez J, Rolan-Alvarez E. 2009. A new multitest correction (SGoF) that increases its statistical power when increasing the number of tests. *BMC Bioinformatics* 10:209.

Chao DY, Dilkes B, Luo H, Douglas A, Yakubova E, Lahner B, Salt DE. 2013. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science* 341:658–659.

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

Chen ZJ. 2013. Genomic and epigenetic insights into the molecular bases of heterosis. *Nat Rev Genet.* 14:471–482.

Chen ZJ, Pikaard CS. 1997. Transcriptional analysis of nucleolar dominance in polyploid plants: biased expression/silencing of progenitor rRNA genes is developmentally regulated in Brassica. *Proc Natl Acad Sci U S A.* 94:3442–3447.

Coate JE, Doyle JJ. 2010. Quantifying whole transcriptome size, a prerequisite for understanding transcriptome evolution across species: an example from a plant allopolyploid. *Genome Biol Evol.* 2:534–546.

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452:215–219.

Comai L. 2005. The advantages and disadvantages of being polyploid. *Nat Rev Genet.* 6:836–846.

Comai L, Tyagi AP, Lysak MA. 2003. FISH analysis of meiosis in *Arabidopsis* allopolyploids. *Chromosome Res.* 11:217–226.

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B. 2000. Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. *Plant Cell* 12:1551–1568.
Cooper GM. 2000. The cell: a molecular approach. Sunderland (MA): Sinauer Associates.
Dowen RH, Pelizzola M, Schmitz RJ, Lister R, Dowen JM, Nery JR, Dixon JE, Ecker JR. 2012. Widespread dynamic DNA methylation in response to biotic stress. *Proc Natl Acad Sci U S A.* 109:E2183–E2191.
Fani R, Fondi M. 2009. Origin and evolution of metabolic pathways. *Phys Life Rev.* 6:23–52.
Feschotte C, Jiang N, Wessler SR. 2002. Plant transposable elements: where genetics meets genomics. *Nat Rev Genet.* 3:329–341.
Galitski T, Saldanha AJ, Styles CA, Lander ES, Fink GR. 1999. Ploidy regulation of gene expression. *Science* 285:251–254.
Graf A, Schlereth A, Stitt M, Smith AM. 2010. Circadian control of carbohydrate availability for growth in *Arabidopsis* plants at night. *Proc Natl Acad Sci U S A.* 107:9458–9463.
Guan X, Pang M, Nah G, Shi X, Ye W, Stelly DM, Chen ZJ. 2014. miR828 and miR858 regulate homoeologous MYB2 gene functions in *Arabidopsis* trichome and cotton fibre development. *Nat Commun.* 5:3050.
Guo H, Ecker JR. 2004. The ethylene signaling pathway: new insights. *Curr Opin Plant Biol.* 7:40–49.
Ha M, Lu J, Tian L, Ramachandran V, Kasschau KD, Chapman EJ, Carrington JC, Chen X, Wang XJ, Chen ZJ. 2009. Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids. *Proc Natl Acad Sci U S A.* 106:17835–17840.
Ha M, Ng DW, Li WH, Chen ZJ. 2011. Coordinated histone modifications are associated with gene expression variation within and between species. *Genome Res.* 21:590–598.
Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ. 2006. Transcriptome shock after interspecific hybridization in senecio is ameliorated by genome duplication. *Curr Biol.* 16:1652–1659.
Hollister JD, Smith LM, Guo YL, Ott F, Weigel D, Gaut BS. 2011. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proc Natl Acad Sci U S A.* 108:2322–2327.
Jacob Y, Stroud H, Leblanc C, Feng S, Zhuo L, Caro E, Hassel C, Gutierrez C, Michaels SD, Jacobsen SE. 2010. Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* 466:987–991.
Jenuwein T, Allis CD. 2001. Translating the histone code. *Science* 293:1074–1080.
Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* 473:97–100.
Kato A, Lamb JC, Birchler JA. 2004. Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc Natl Acad Sci U S A.* 101:13554–13559.
Leitch AR, Leitch IJ. 2008. Genomic plasticity and the diversity of polyploid plants. *Science* 320:481–483.
Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. *Cell* 128:707–719.
Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133:523–536.
Luo C, Sidote DJ, Zhang Y, Kerstetter RA, Michael TP, Lam E. 2013. Integrative analysis of chromatin states in *Arabidopsis* identified potential regulatory mechanisms for natural antisense transcript production. *Plant J.* 73:77–90.
Lynch M, Force A. 2000. The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154:459–473.
Lynch M, O'Hely M, Walsh B, Force A. 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics* 159:1789–1804.
Lysak M, Fransz P, Schubert I. 2006. Cytogenetic analyses of *Arabidopsis*. *Methods Mol Biol.* 323:173–186.
Mayer VW, Aguilera A. 1990. High levels of chromosome instability in polyploids of *Saccharomyces cerevisiae*. *Mutat Res.* 231:177–186.
Mayrose I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP. 2011. Recently formed polyploid plants diversify at lower rates. *Science* 333:1257.
Miller M, Zhang C, Chen ZJ. 2012. Ploidy and hybridity effects on growth vigor and gene expression in *Arabidopsis thaliana* hybrids and their parents. *G3 (Bethesda)* 2:505–513.
Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 5:621–628.
Mueller LA, Zhang P, Rhee SY. 2003. AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.* 132:453–460.
Ng DW, Zhang C, Miller M, Palmer G, Whiteley M, Tholl D, Chen ZJ. 2011. cis- and trans-regulation of miR163 and target genes confers natural variation of secondary metabolites in two *Arabidopsis* species and their allopolyploids. *Plant Cell* 23:1729–1740.
Ni Z, Kim ED, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ. 2009. Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* 457:327–331.
Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattem GL, Li R. 2010. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* 468:321–325.
Qi B, Huang W, Zhu B, Zhong X, Guo J, Zhao N, Xu C, Zhang H, Pang J, Han F, Liu B. 2012. Global transgenerational gene expression dynamics in two newly synthesized allohexaploid wheat (*Triticum aestivum*) lines. *BMC Biol.* 10:3.
Quint M, Drost HG, Gabel A, Ullrich KK, Bonn M, Grosse I. 2012. A transcriptomic hourglass in plant embryogenesis. *Nature* 490:98–101.
Ramsey J, Schemske DW. 2002. Neopolyploidy in flowering plants. *Annu Rev Ecol Syst.* 33:589–639.
Ramskold D, Wang ET, Burge CB, Sandberg R. 2009. An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput Biol.* 5:e1000598.
Shi X, Ng DW, Zhang C, Comai L, Ye W, Jeffrey Chen Z. 2012. Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nat Commun.* 3:950.
Soltis DE, Visger CJ, Soltis PS. 2014. The polyploidy revolution then . . . and now: Stebbins revisited. *Am J Bot.* 101:1057–1078.
Soltis PS, Soltis DE. 2009. The role of hybridization in plant speciation. *Annu Rev Plant Biol.* 60:561–588.
Song Q, Chen ZJ. 2015. Epigenetic and developmental regulation in plant polyploids. *Curr Opin Plant Biol.* 24:101–109.
Stebbins GL. 1971. Chromosomal evolution in higher plants. London: Edward Arnold.
Storchova Z, Pellman D. 2004. From polyploidy to aneuploidy, genome instability and cancer. *Nat Rev Mol Cell Biol.* 5:45–54.
Tautz D, Domazet-Loso T. 2011. The evolutionary origin of orphan genes. *Nat Rev Genet.* 12:692–702.
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.
Van de Peer Y, Maere S, Meyer A. 2009. The evolutionary significance of ancient genome duplications. *Nat Rev Genet.* 10:725–732.
Veitia RA, Bottani S, Birchler JA. 2008. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet.* 24:390–397.
Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in *Arabidopsis*. *Science* 290:2114–2117.
Wang J, Tian L, Lee HS, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, et al. 2006. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. *Genetics* 172:507–517.
Wendel JF. 2000. Genome evolution in polyploids. *Plant Mol Biol.* 42:225–249.
Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE, Amon A. 2008. Aneuploidy affects proliferation and

spontaneous immortalization in mammalian cells. *Science* 322: 703–709.
Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci U S A.* 106:13875–13879.
Xi Y, Li W. 2009. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* 10:232.
Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. 2013. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153:193–205.
Zhang X. 2008. The epigenetic landscape of plants. *Science* 320:489–492.
Zhang X, Clarenz O, Cokus S, Bernatavichute YV, Pellegrini M, Goodrich J, Jacobsen SE. 2007. Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol.* 5:e129.
Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al. 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 126:1189–1201.
Zhou J, Lemos B, Dopman EB, Hartl DL. 2011. Copy-number variation: the balance between gene dosage and expression in *Drosophila melanogaster*. *Genome Biol Evol.* 3:1014–1024.