



# Coordinated Functional Divergence of Genes after Genome Duplication in *Arabidopsis thaliana*<sup>OPEN</sup>

Riet De Smet,<sup>a,b,1</sup> Ehsan Sabaghian,<sup>a,b,1</sup> Zhen Li,<sup>a,b,1</sup> Yvan Saeys,<sup>c,d</sup> and Yves Van de Peer<sup>a,b,e,f,2</sup>

<sup>a</sup>Department of Plant Biotechnology and Bioinformatics, Ghent University, B-9052 Ghent, Belgium

<sup>b</sup>Center for Plant Systems Biology, VIB, B-9052 Ghent, Belgium

<sup>c</sup>Center for Inflammation Research, VIB, B-9052 Ghent, Belgium

<sup>d</sup>Department of Applied Mathematics, Computer Science and Statistics, Ghent University, B-9052 Ghent, Belgium

<sup>e</sup>Bioinformatics Institute Ghent, Ghent University, B-9052 Ghent, Belgium

<sup>f</sup>Genomics Research Institute, University of Pretoria, Pretoria 0028, South Africa

ORCID IDs: 0000-0001-8920-9270 (Z.L.); 0000-0003-4327-3730 (Y.V.d.P.)

**Gene and genome duplications have been rampant during the evolution of flowering plants. Unlike small-scale gene duplications, whole-genome duplications (WGDs) copy entire pathways or networks, and as such create the unique situation in which such duplicated pathways or networks could evolve novel functionality through the coordinated sub- or neofunctionalization of its constituent genes. Here, we describe a remarkable case of coordinated gene expression divergence following WGDs in *Arabidopsis thaliana*. We identified a set of 92 homoeologous gene pairs that all show a similar pattern of tissue-specific gene expression divergence following WGD, with one homoeolog showing predominant expression in aerial tissues and the other homoeolog showing biased expression in tip-growth tissues. We provide evidence that this pattern of gene expression divergence seems to involve genes with a role in cell polarity and that likely function in the maintenance of cell wall integrity. Following WGD, many of these duplicated genes evolved separate functions through subfunctionalization in growth/development and stress response. Uncoupling these processes through genome duplications likely provided important adaptations with respect to growth and morphogenesis and defense against biotic and abiotic stress.**

## INTRODUCTION

Whole-genome duplication (WGD) is widespread within the angiosperms, and although the exact causes and consequences are still under debate, it has been suggested that WGD plays an important role in the evolution of novel traits and increased biological complexity (Van de Peer et al., 2009; Vanneste et al., 2014; Soltis and Soltis, 2016). A range of phenotypic and genotypic effects might explain the relative success of polyploids (Van de Peer et al., 2017). It has been observed that polyploidy can have a direct influence on the plant phenotype, leading, for instance, to bigger flowers and/or increased stature (Chen, 2010; Gross and Schiestl, 2015), as well as on the plant genotype, leading to genomic rearrangements and gene expression changes (Adams and Wendel, 2005; Chen, 2007; Buggs et al., 2011; Chester et al., 2012; Shi et al., 2015).

Besides the direct effects associated with polyploidy, the evolutionary success of polyploids might also in part be explained by the increased diversification potential that comes with WGD. Indeed, gene duplication is undoubtedly the most important source of novel genes (Taylor and Raes, 2004; Flagel and Wendel, 2009). These “extra” genes can serve as substrates for the

evolution of novel functions through sub- or neofunctionalization, or a combination of both (Prince and Pickett, 2002; Conant and Wolfe, 2008). In contrast to small-scale duplications, duplicates created by WGD (also called homoeologs) tend to be retained at much higher fractions and for longer periods of time, likely due to dosage-balance constraints opposing their loss (Maere et al., 2005a; Aury et al., 2006; Birchler and Veitia, 2012; Li et al., 2016), thereby extending the temporal window for sub- and neofunctionalization. Hence, studies have focused on identifying duplicated genes that have been retained following WGD and subsequently diverged and that can explain the evolution of novel traits or increase in biological complexity. Examples of such studies include enhanced root nodule symbiosis in Papilionoidae (Li et al., 2013), the glucosinolate pathway in the Brassicales (Hofberger et al., 2013; Edger et al., 2015), the ethylene biosynthesis pathway in banana (*Musa acuminata*; Jourda et al., 2014), and response to wounding in *Nicotiana* (Zhou et al., 2016).

However, an often-overlooked aspect of WGD in these studies is that WGD duplicates not only individual genes, but also entire pathways. The coordinated evolution of all genes in a pathway is of particular interest when studying the evolution of novel traits following WGD (De Smet and Van de Peer, 2012). For instance, Freeling and Thomas (2006) argued that gene balance maintains duplicated functional modules as the spandrels of purifying selection and that these modules are likely precursors of coadapted gene complexes. These authors also postulated that such duplicated functional networks or modules—representing functional pathways or parts thereof—would lead to an increase in biological complexity and/or morphology. Another argument that coordinated

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Address correspondence to [yves.vandeppeer@psb.ugent.be](mailto:yves.vandeppeer@psb.ugent.be).

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors ([www.plantcell.org](http://www.plantcell.org)) is: Yves Van de Peer ([yves.vandeppeer@psb.ugent.be](mailto:yves.vandeppeer@psb.ugent.be)).

<sup>OPEN</sup>Articles can be viewed without a subscription.

[www.plantcell.org/cgi/doi/10.1105/tpc.17.00531](http://www.plantcell.org/cgi/doi/10.1105/tpc.17.00531)

evolution might follow WGD is the observation that the evolution of traits often involves mutations in multiple genes (Bullard et al., 2010; Fraser et al., 2011; He et al., 2016; Roop et al., 2016), which makes sense given that many traits are polygenic in nature. Inquiries into the existence of such patterns of coordinated divergence in *Arabidopsis thaliana* (Blanc and Wolfe, 2004a; Ruprecht et al., 2016) and *Saccharomyces cerevisiae* (Pereira-Leal and Teichmann, 2005; Conant and Wolfe, 2006; Wapinski et al., 2007), which were mainly assessed at the expression level, have reported such cases to be rare. One exception was provided by Ihmels et al. (2005), who identified a link between the evolution of anaerobic growth and transcriptional rewiring of dozens of genes following WGD in *S. cerevisiae*. The paucity of well-described cases might indicate that coordinated evolution of duplicated genes following WGD is a rare evolutionary event.

Here, we aimed to identify cases of polygenic evolution following WGD for Arabidopsis. To this end, we used a gene-centric approach to interrogate large Arabidopsis gene expression compendia for coordinated gene expression divergence following WGD. We identified a clear case of 92 homoeologous gene pairs that show the same pattern of gene expression divergence, representing a shift in expression between aerial tissues and polarized cell types/tissues (pollen tube and root tip). We provide evidence that this tissue-specific gene expression divergence pattern parallels functional divergence between defense and growth/development, and hypothesize that many of the identified duplicated gene pairs might function in cell wall integrity pathways that evolved after WGD to have separate or at least partly separate functions in growth and stress.

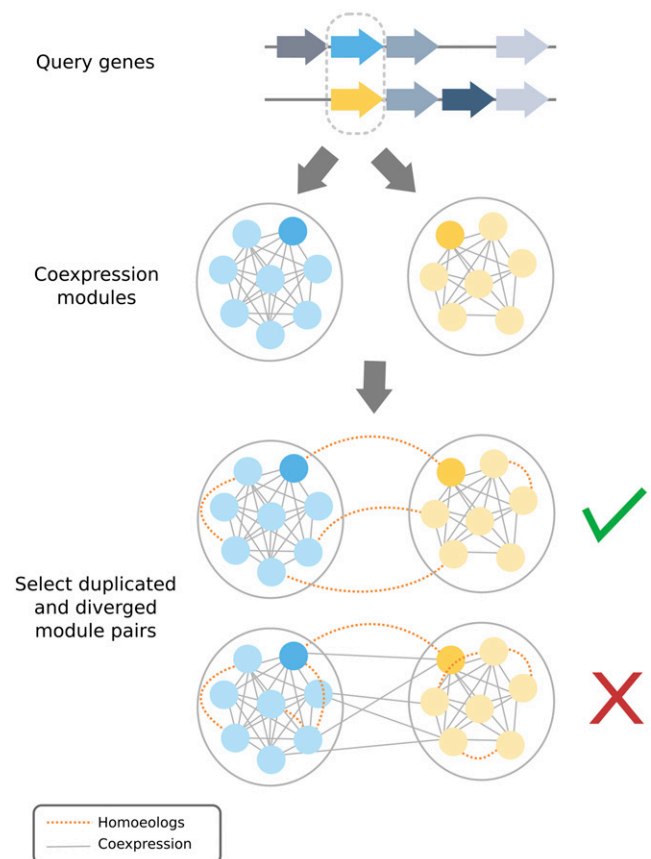
## RESULTS

### Identifying Homoeologous Gene Pairs with Coordinated Gene Expression Divergence

To identify WGD-derived gene pairs that show coordinated gene expression divergence, we used biclustering, a clustering technique that allows the simultaneous clustering of rows and columns of a matrix and the identification of subclusters. Biclustering analysis of a large gene expression compendium allowed us to construct modules of genes with similar expression profiles across a range of experimental conditions (i.e., coexpression modules). Based on these coexpression modules, we defined gene expression divergence of duplicated gene pairs as cases in which the homoeologs did not belong to the same coexpression module and that hence were coexpressed with distinct sets of genes. To identify cases of coordinated gene expression divergence, we searched for pairs of coexpression modules that differed in their gene expression patterns and that shared a large number of homoeologous gene pairs (Figure 1) (see Methods). We used a module-based approach to study coordinated gene expression divergence for the following reasons: (1) coordinated gene expression divergence likely involves genes that belong to the same pathway, signaling network, or protein complex and therefore are expected to be mutually coexpressed (Wu et al., 2002); (2) patterns of gene expression divergence can be more robustly inferred based on modules, since gene expression

divergence of a pair of genes is supported by these genes being coexpressed with distinct gene sets; and (3) the module concept can be used to improve our understanding of gene function by taking advantage of the guilt-by-association principle and can as such contribute to an increased understanding of how gene function might have changed following WGD.

To obtain coexpression modules, we used a gene-centric biclustering approach to interrogate a gene expression compendium for genes that were specifically coexpressed with the homoeologous genes retained in duplicate following WGD in Arabidopsis (i.e., the query genes for module retrieval). As the biclustering method, we used a modified version of the Signature Algorithm (Ihmels et al., 2002), which can be applied in a query-based mode (De Smet and Marchal, 2011) (see Methods). As a gene expression data set for module retrieval, we used a compendium that contains 134 different conditions from different



**Figure 1.** Procedure for Selecting Modules with Divergent Gene Expression.

To identify coordinated gene expression divergence following WGD, we considered all gene pairs in conserved syntenic blocks as a starting point for our analyses. A gene-centric biclustering approach was applied to retrieve a coexpression module for each query gene (i.e., homoeologs). Candidate module pairs are module pairs for which the genes show divergent gene expression patterns (for all or a large majority of genes) and that share a maximum number of homoeologous gene pairs. See Methods and Supplemental Figure 6 for details.

experiments, describing the expression of 19,825 genes in different tissues and developmental stages in *Arabidopsis* (Schmid et al., 2005; De Bodt et al., 2012). The query genes for module retrieval came from 2699 putative homoeologous gene pairs from consecutive *Arabidopsis* WGDs, taken from PLAZA 2.5 (Van Bel et al., 2012), where conservation of intragenomic synteny was used as evidence for retention following WGD. In total, these pairs contained 4759 genes that were used as query genes to obtain 4759 coexpression modules. We had 4759 genes (i.e., less than  $2 \times 2699$  genes) in these pairs because multiple WGDs have occurred during the evolutionary past of *Arabidopsis*; hence, the duplication history of genes is often nested such that certain genes can be part of multiple syntenic gene pairs. The average size of these modules was 2207 genes (Supplemental Figure 1). Since the coexpression modules were obtained independently from each other, the modules could overlap: On average, each homoeolog belonged to 555 different coexpression modules.

We used this set of 4759 modules to identify candidate module pairs that shared a large number of homoeologous gene pairs with coordinated gene expression divergence and obtained 22 such module pairs (Figure 1; see Methods for further details). Below, we will discuss in detail two module pairs that showed clear gene expression divergence. Module Pair 1 consisted of a module of 2033 genes (module 1-A) and a module of 1934 genes (module 1-B). Module Pair 2 comprised a module of 2113 genes (module 2-A) and a module of 2230 genes (module 2-B). The modules within each pair showed substantial overlap; for example, modules 1-A and 2-A shared 19% of their genes, while modules 1-B and 2-B shared 32% of their genes. In total, these two module pairs contained 92 unique homoeologous pairs, consisting of 173 unique genes (Supplemental Data Set 1). Again, these 92 pairs contained fewer than  $2 \times 92$  unique genes because these homoeologous gene pairs originated from consecutive independent WGDs and a small subset of them were repeatedly retained in duplicate. Module Pair 1 contained 53 homoeologous pairs, Module Pair 2 contained 54 homoeologous pairs, and the overlap of both module pairs comprised 15 homoeologous pairs (Figure 2).

We also identified the experimental conditions for which the genes in the two module pairs were most divergent in gene expression. We used Random Forest classifiers to test how well the gene expression levels of the genes in module A could be differentiated from the gene expression levels of their homoeologs in module B under a certain condition (see Methods for details). Using this approach, we found that Module Pair 1 homoeologous gene pairs had opposite gene expression patterns in pollen (male gametophyte) and aerial tissues (leaf, shoot, cotyledon, etc.), with the genes in module 1-A upregulated in the pollen tissues and those in module 1-B upregulated in the aerial tissues (Figure 2). Therefore, we will further refer to module 1-A as the Pollen module and module 1-B as the Leaf\_P (i.e., leaf as opposed to pollen) module. We identified a similar pattern for the homoeologs in Module Pair 2, with the gene pairs having opposite expression patterns in root (primarily root tip) (module 2-A) and aerial (module 2-B) tissues (Figure 2). We will further refer to these modules as the Root and Leaf\_R (i.e., leaf as opposed to root) modules. The gene expression patterns for genes in the Root and Pollen modules were quite similar, since they were both downregulated in the aerial tissues, where the homoeologs in

both Leaf\_P and Leaf\_R modules were upregulated (Figure 2). These observed tissue-specific expression biases together with the high overlap in genes between both modules might be explained by both Root and Pollen representing tissue types with a polarized growth signature (Rounds and Bezanilla, 2013). Interestingly, the selected conditions for the Root module also included apical and basal embryonic cells (Figure 2), for which it is well established that cell polarity maintenance is important for correct plant development (Lau et al., 2012). Consequently, the top-divergent conditions for the Root and Pollen modules indicate that genes within these modules function in cell or tissue polarization; hence, we hereafter considered them as one Root/Pollen module. Similarly, we could not identify any distinction between the genes in the Leaf\_R and Leaf\_P modules based on their expression patterns; therefore, we further considered them as one Leaf module.

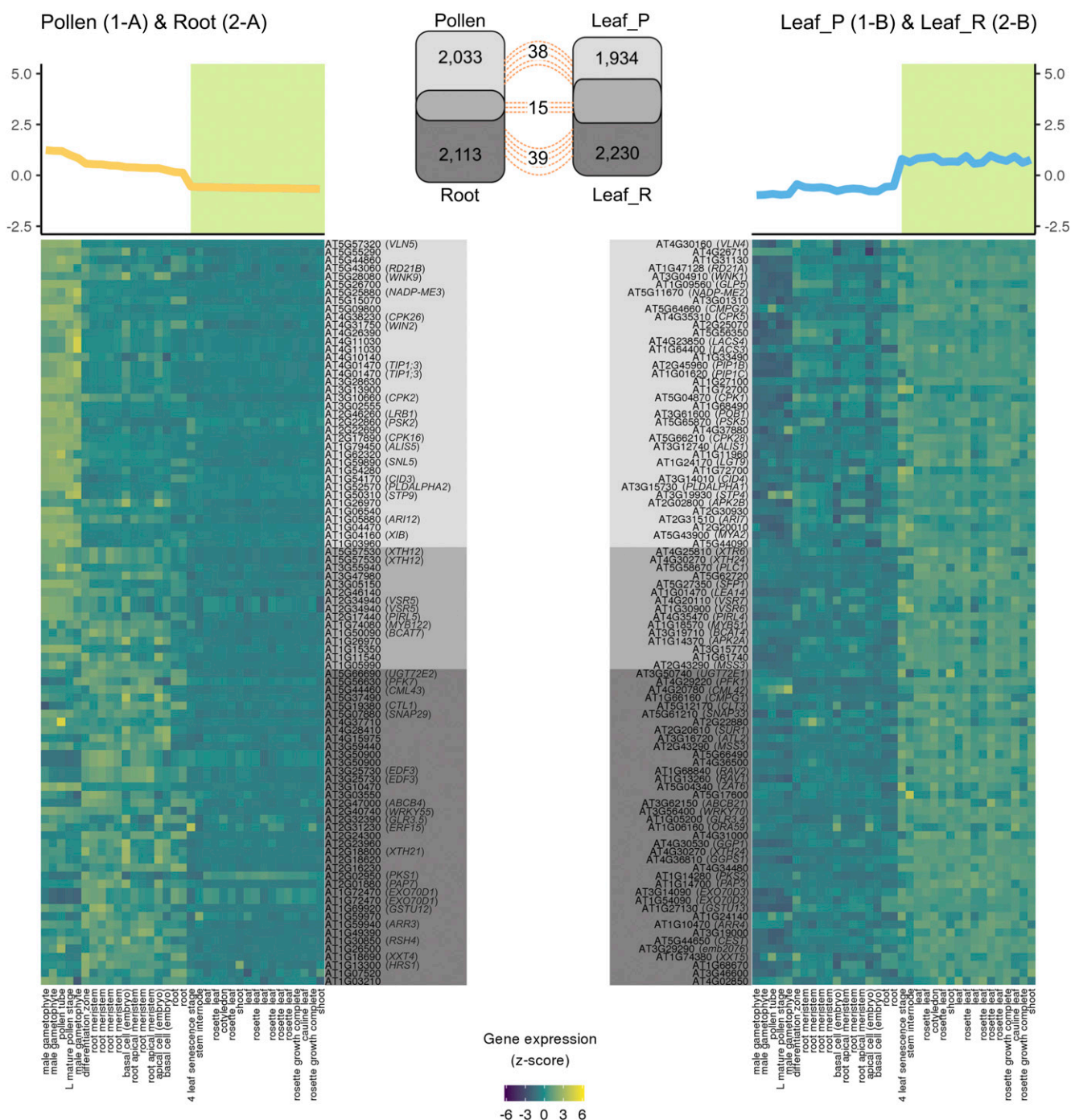
We next assessed whether the above-described gene expression divergence patterns were robust outside the chosen gene expression data set, by investigating whether similar gene expression divergence patterns could be observed for other data sets. As a first alternative data set, we used all 946 conditions remaining in CORNET (De Bodt et al., 2012) after removing the ones in the Development compendium, which was used for the above analysis. We again used the Random Forest approach to identify conditions of divergent expression for genes in module A and module B in both pairs. The top conditions were consistent with the observations made for the Development compendium: Genes within the modules 1-A (Pollen) and 2-A (Root) seemed to be biased toward being expressed in flower/male gametophyte (stamen and inflorescence) and root tissues, respectively. Genes within the modules 1-B (Leaf\_P) and 2-B (Leaf\_R) seemed to show biased expression in aerial tissues, such as leaves and the shoot (Supplemental Figure 2). Hence, this data set confirmed the gene expression dichotomy observed in the Development data set.

As a second alternative data set, we profiled the Rootmap data from Brady et al. (2007) to identify the specific regions in the *Arabidopsis* root where the homoeologs with a Root-Leaf biased expression showed gene expression divergence. This analysis supported the observation that genes in the Root module were specifically biased toward being expressed in the root tip and the meristematic tissues, whereas gene expression levels for the genes in the Leaf\_R module were comparatively low for these particular tissues. By contrast, the genes with a biased expression toward the leaves (Leaf\_R module) appeared to be primarily expressed in the maturation zone of the root, phloem, and the lateral root (Supplemental Figure 3).

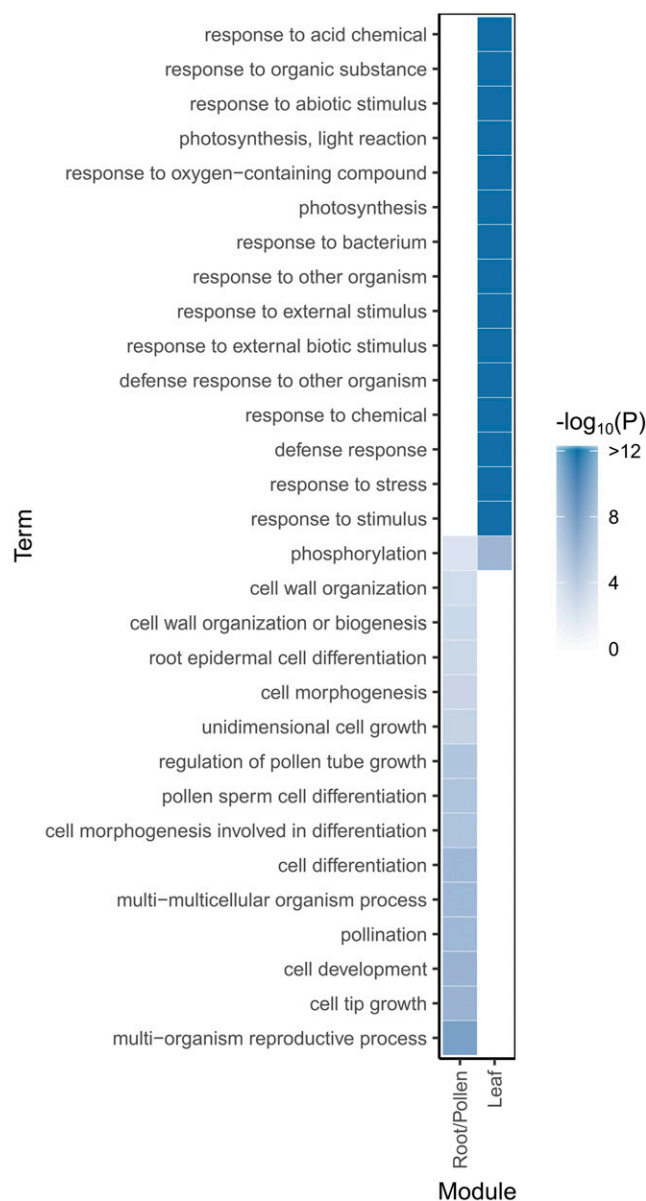
In summary, using an approach to identify modules that contained gene pairs that duplicated and diversified following WGD, we identified two module pairs with clear gene expression divergence signatures. The first module pair corresponded with a gene expression shift of 53 gene pairs between primarily leaf-biased expression and expression in the pollen tube. A second module pair reflected a gene expression shift of 54 gene pairs between primarily leaf-biased expression and expression in the root tip.

### Functional Divergence of Homoeologs and Their Respective Coexpressed Genes

To study how the functions of the 92 homoeologous gene pairs might have changed following WGD, we analyzed their functions



**Figure 2.** Gene Expression Patterns for the 92 Homoeologous Gene Pairs under Divergent Conditions as Selected by Random Forest Classification. The left panel corresponds to the Pollen and Root modules, while the right panel corresponds to the Leaf\_P and Leaf\_R modules. The top line plots represent the average gene expression profiles under selected conditions for all genes in the Root and Pollen modules (left) and the Leaf\_P and Leaf\_R modules (right). The diagram in the top middle schematically represents the module sizes in terms of the total number of genes they contain and the number of homoeologous gene pairs (orange dashed lines). Heat maps represent gene expression patterns for the 92 divergent homoeologous gene pairs (rows); hence, one row in the heat map panels corresponds to one homoeologous gene pair. Conditions (columns) are the same in the left and the right panels. Colored boxes around the gene labels indicate the module pair they come from: Pollen-Leaf\_P (light gray), Root-Leaf\_R (dark gray), and the overlap of both modules pairs (intermediate shade) (see also diagram, top middle).



**Figure 3.** The 15 Most Significant GO Terms Enriched for the Aggregated Root/Pollen and Aggregated Leaf Modules.

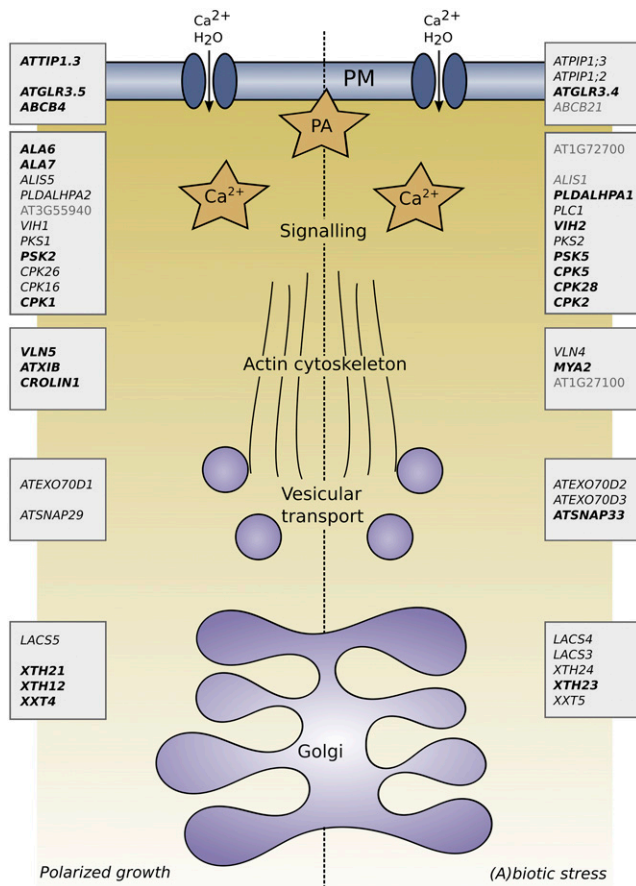
Shades of blue represent the logarithm of the enrichment P value following Benjamini-Hochberg multiple testing correction of the indicated GO term in both modules.

based on Gene Ontology (GO) terms (Supplemental Data Set 1). Because of the limited number of genes with an annotated function, GO term overrepresentation analyses resulted in a short list of enriched GO terms. For instance, homoeologs in the Root/Pollen module appear to be overrepresented for terms associated with calcium-dependent signaling (Supplemental Table 1). For the homoeologs with a leaf-biased gene expression, we found overrepresentation for responses to multiple exogenous and endogenous stimuli. This overrepresentation was maintained if we

used all homoeologs in Arabidopsis as a reference set, suggesting that especially homoeologs responding to stimuli belonged to our coexpression modules (Supplemental Table 2). Cross-checking the GO terms with experimental evidence for the homoeologous gene pairs identified a large number of regulatory genes among the 92 homoeologous gene pairs, with 36 genes involved in “signal transduction.” Furthermore, the large number of genes involved in “ubiquitination” (11 genes) and “DNA-templated transcription” (16 genes) suggested that many of these homoeologs are involved in gene regulation at different levels (signaling, transcription regulation, and posttranslational regulation/protein turnover).

We also performed a GO enrichment analysis of all genes in the coexpression modules to which the 92 genes belong. These GO enrichment results supported the tissue-specific gene expression patterns described above: For instance, genes in the Root module were enriched for GO terms associated with root development, such as “trichoblast maturation” and “root hair cell development and differentiation” (Supplemental Table 3). The Pollen module was enriched for GO terms associated with the male gametophyte, for instance, “pollen tube,” “pollen tube growth and development,” “pollen germination,” and “male gamete generation” (Supplemental Table 3). The two Leaf modules were enriched for GO terms that are expected to be associated with aerial tissues, such as photosynthesis-related terms (Supplemental Table 3). In agreement with the relatively high fractions of signal transduction genes among the 92 homoeologous gene pairs (Supplemental Figure 4), the only GO term that was most enriched for both the Root/Pollen and Leaf modules was “phosphorylation” (Figure 3). More remarkable, the GO analysis revealed a difference in functional bias between genes in the Root and Pollen modules on the one hand and the two Leaf modules on the other hand. Enriched GO terms were different for both sets of modules, with top-enriched GO terms for the Root/Pollen modules biased toward functions related to cell growth and differentiation, whereas enriched GO terms for the Leaf modules were biased to response to stress and more specifically defense (Figure 3).

Detailed literature searches for the 173 genes in the 92 homoeologous pairs confirmed that many of the homoeologous genes with described functions have a role in growth and development, and more specifically in tip growth and defense (Supplemental Data Set 1). The link between genes involved in growth and development and their expression in pollen and root tissue seems reasonable, considering that the pollen tube represents a prime example of directional cellular growth (Cheung and Wu, 2008) and that the root tip is also involved in directional responses to developmental and environmental cues (Abas et al., 2006; Galvan-Ampudia et al., 2013). Maybe less clear is the link between genes involved in defense and their expression in aerial tissues, especially since the expression compendium we used does not contain any stress conditions. However, it has been observed that subtle variations in growth conditions are difficult to control under laboratory conditions and affect gene expression (Bhosale et al., 2013). In an analysis of functional processes that showed gene expression variation in individuals grown under tightly controlled conditions, “defense”-related functions and “hormone signaling” outperformed other functional categories (Bhosale et al., 2013). This finding suggests that these specific functional processes are especially difficult to control in lab



**Figure 4.** Schematic Representation of the Processes Involved in Cell Wall Maintenance and the Homoeologous Genes That Participate in These Processes.

From bottom to top: Cell wall enzymes are usually synthesized in the Golgi apparatus and transported by vesicular transport (involving exocytosis and the actin cytoskeleton) to the plasma membrane. Different signaling proteins are activated by extracellular or plasma membrane-associated signals and (in)directly interact with the cytoskeleton and vesicular transport. The main signaling molecules involved in cell wall maintenance are signaling molecules derived from phospholipids (e.g., PA) and calcium. At the level of the plasma membrane, transporter proteins are responsible for maintaining ion homeostasis (e.g., Ca<sup>2+</sup> concentrations) and for controlling cellular turgor. Homoeologous genes involved in each of these processes are indicated in boxes. Genes from the Root/Pollen modules are listed on the left, whereas homoeologs from the Leaf modules are on the right. Gene names in bold are genes for which experimental evidence exists that they are involved in polarized growth or (a)biotic stress. Gene names not in bold are genes for which there exists indirect evidence that they are involved in either of these processes, for instance, because they belong to a gene family containing other members that function in these processes and/or with experimental evidence that they are expressed in the relevant tissues. For other genes, indicated in light gray, no relation to any of these processes could be found in literature. References to relevant literature are in Supplemental Data Set 1.

conditions and that even in the absence of specific stressors, individuals show variation in their responses for these specific functions.

In summary, the observed gene expression divergence of the 92 homoeologs and their respective coexpressed genes seems to reflect true functional divergence, with genes in the Root/Pollen module(s) functioning in (tip) growth and development, and those in the Leaf module(s) having a stress-related function. It is worth noting that the GO enrichment analysis suggests a statistically significant functional profile for a set of genes, so the enriched terms do not necessarily apply to all genes (Supplemental Data Set 1). We noticed that some genes in the Leaf module are also involved in (tip) growth and development, based on evidence from the literature. For instance, *VLN4* (AT4G30160) has a described function in root hair growth (Zhang et al., 2011); mutants of *XXT5* (AT1G74380) have a root hair morphology phenotype (Zabotina et al., 2008); *ARR4* (AT1G10470) functions in growth and development (Chi et al., 2016); *PKS2* (AT1G14280) functions in leaf flattening and leaf positioning (de Carbonnel et al., 2010); and *GLR3.4* (AT1G05200) functions in root development (Vincill et al., 2013). GO terms for these genes also reflect the above described functions (Supplemental Data Set 1).

#### Homoeologous Gene Pairs Show Associated Functions in Cell Wall Integrity

While the genes in the Root/Pollen modules and the Leaf modules tended to have evolved divergent functions in (tip) growth and defense, GO analysis did not reveal a single underlying pathway that has been duplicated. Therefore, it was not clear what connects the homoeologs mutually and how these duplicated genes could be involved in such diverse functions as growth and defense. However, exhaustive literature searches (Supplemental Data Set 1) revealed that a considerable number of homoeologs seemed to be (in)directly involved in cell wall maintenance, either under growth/morphogenesis (Root/Pollen module) or (a)biotic stress (Leaf module). The most obvious examples of homoeologs that may be involved in cell wall maintenance were a set of genes in our modules that function in cell wall biogenesis and metabolism, such as xyloglucan endotransglucosylases (Figure 4). In addition, the set of homoeologs included genes involved in the establishment and maintenance of cell polarity, needed to transport cell wall material to the site of growth/cell wall damage, such as cytoskeletal genes, genes involved in exocytosis, and signaling genes involved in the polarization of the plasma membrane. Specifically, genes encoding components of the acto-myosin cytoskeleton and vesicular transport (exocyst and SNAREs) were found among our homoeologs (Figure 4). For instance, targeted vesicular transport of cell wall components through the actin cytoskeleton plays a pivotal role in cell wall maintenance (Hepler et al., 2013; Sampathkumar et al., 2013; Thomas and Staiger, 2014). Notably, the roles of the actin cytoskeleton and vesicular trafficking in both cell elongation and defense are well described (Cheung and Wu, 2008; Day et al., 2011). Besides these structural components that are involved in cell wall maintenance, as stated previously, the homoeologs also included many signaling genes (Supplemental Figure 4). Of note were different genes that function in phospholipid signaling, such as *PHOSPHOLIPASE D* (*PLD*) and *PLC*, which modulate levels of intracellular phosphatidic acid (PA). PA is a signaling phospholipid with diverse functions in biotic and abiotic stress (Testerink and Munnik, 2005) but also with a crucial

role in the regulation of cell expansion in pollen tubes (Pleskot et al., 2013) and in root tropism (Galvan-Ampudia et al., 2013) (Figure 4). Interestingly, PA directly interacts with the actin cytoskeleton and as such influences vesicular trafficking to the cellular site of PA signaling (Pleskot et al., 2013; Hong et al., 2016). In addition, PLCs are possibly involved in plasma membrane cell wall interactions and are thus likely involved in cell wall remodeling and transferring signals from the outside of the cell to the inside (Borner et al., 2003; Liu et al., 2015). Besides PA,  $\text{Ca}^{2+}$  also functions as an important intracellular signaling molecule involved in cell wall maintenance, for instance, by interacting with the actin cytoskeleton (Hepler, 2016), while it is also an important component of the cell wall itself (Hepler et al., 2013). Transporter genes among our homoeologs (such as GLRs) have a role in regulating  $\text{Ca}^{2+}$ -gradients during pollen tube growth and possibly also root hair growth (Michard et al., 2017), but also function in stress (Meyerhoff et al., 2005). In addition, our homoeologous gene pairs included  $\text{Ca}^{2+}$ -sensitive signaling genes, such as multiple CDPKs, that act as  $\text{Ca}^{2+}$  sensors and trigger downstream responses, such as polarity maintenance in the growing pollen tube (Myers et al., 2009) and defense response (Boudsocq and Sheen, 2013).

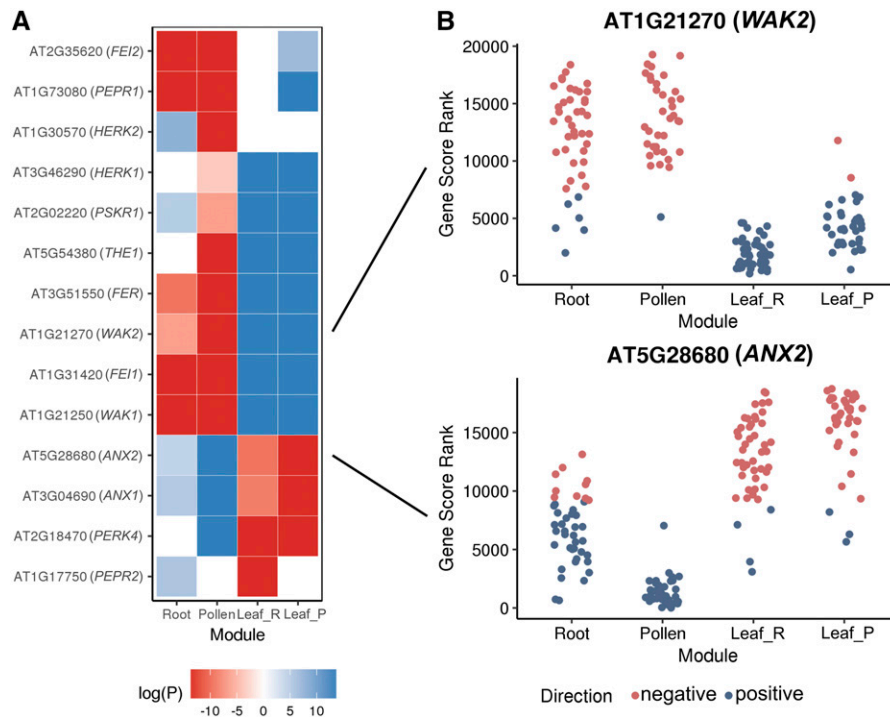
In agreement with our observation that these duplicated genes might function in cell wall maintenance are observations that the plant cell wall is intimately involved in both growth/morphogenesis and (abiotic stress (Wolf et al., 2012; Engelsdorf and Hamann, 2014), explaining the functional bias of homoeologs to either growth or defense. The opposite expression patterns we observed for the 92 homoeologous gene pairs for growth-related genes versus defense-related genes could be explained by the antagonistic requirements for growth and stress on the plant cell wall. For instance, growth requires controlled loosening of the cell wall (Cosgrove, 2005; Braidwood et al., 2014), whereas in case of herbivore or pathogen attack, further loosening of the cell wall (as caused by pathogen enzymes) will be halted as to prevent further damage (Hématy et al., 2007).

To further validate our hypothesis that many of these homoeologs are potentially involved in controlling and maintaining cell wall integrity (CWI) under both growth and stress, we determined whether probable marker genes of CWI maintenance are coregulated with the 92 homoeologous gene pairs. As marker genes for CWI maintenance, we considered 14 genes encoding receptor-like kinases (RLKs) that have been implicated in cell wall damage signaling and that likely function upstream of pathways involved in the maintenance of CWI (Engelsdorf and Hamann, 2014). These RLK genes include those that were identified as primarily functioning in tip growth (e.g., *ANX1* and *ANX2*) as well as genes that primarily function in defense (e.g., *WAK1* and *WAK2*), although functions of RLKs are often pleiotropic (Nissen et al., 2016). For each of these 14 RLK genes, we tested whether they show condition-specific coexpression with any of the 173 homoeologs in the 92 pairs. Specifically, we used each of the RLKs as query genes for the modified Signature Algorithm (SA), applied to the Development expression compendium. Next, we ranked all 19,825 genes in the expression compendium according to the Gene Scores generated by the SA (see Methods). Briefly, a Gene Score of a certain gene is a proxy for the extent to which it shows condition-dependent coexpression with the query gene. Using this approach, for each of the 14 RLK genes, we can thus rank the

19,825 genes in the expression compendium in decreasing order of coexpression based on the calculated Gene Scores (see Methods). We then used Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005; Våremo et al., 2013) to test whether the 173 homoeologous genes were significantly coexpressed (meaning they had collectively high ranks in the list) or showed significant anti-coexpression (meaning they had collectively low ranks in the list) with each of the 14 RLKs (see Methods). The outcome of this guilt-by-association analysis was again supportive of our hypothesis that the homoeologs function in CWI maintenance in either growth (Root/Pollen module) or defense (Leaf modules) (Figure 5A). In particular, we found that for all 14 RLKs, homoeologs within either the Root/Pollen module or the Leaf module were significantly associated with either top-coexpressed or top-anti-coexpressed genes. In addition, homoeologs in the Root/Pollen and Leaf modules showed opposite patterns of statistical association, with for the majority of the cases the Root/Pollen homoeologs showing statistically significant anti-coexpression to the RLKs, whereas the Leaf homoeologs showed statistically significant coexpression. This result was consistent with the opposite tissue-specific expression patterns we observed earlier for the homoeologs within the pairs. In addition, the observed patterns agreed with known roles for the RLKs. For instance, *FERONIA* arrests pollen tube growth (Escobar-Restrepo et al., 2007) and was indeed anti-coexpressed to the Pollen homoeologs. Conversely, *ANXUR1* (*ANX1*) and *ANX2* maintain pollen tube integrity during growth (Boisson-Dernier et al., 2009) and were coexpressed with the pollen homoeologs (Figure 5). Similarly, *WALL-ASSOCIATED KINASE1* (*WAK1*) and *WAK2* play a role in defense (Kohorn and Kohorn, 2012), consistent with their coexpression with the Leaf homoeologs (Figure 5).

### Recurrent WGDs Have Contributed to the Functional Divergence of Duplicated Genes

Besides the 92 homoeologous gene pairs identified through synteny analysis (see Methods), we wondered whether the other genes in the modules could also have been created through WGD and hence represent false negatives of our synteny-based filtering method. Therefore, we estimated their age of duplication. First, we identified all gene families with gene members shared between both modules and that showed a similar gene expression divergence pattern as determined above for the 92 homoeologous gene pairs. Because of the potentially complex many-to-many relationship within plant gene families, we used gene tree-species tree reconciliation to identify the specific duplication nodes in the gene trees associated with gene expression divergence, i.e., those nodes for which the duplicated genes belonged to different coexpression modules (see Methods). As such, we identified 202 additional duplication nodes for which the paralogous genes showed a similar gene expression divergence as the 92 homoeologous gene pairs identified earlier in this article (i.e., one gene belongs to the Root/Pollen module and its paralog to one of the Leaf modules). We analyzed the phylogenetic distribution of these duplication nodes by evaluating their position on the species trees and found that 64.9% (131/202) of the identified duplication nodes mapped to three nodes associated with WGD events in *Arabidopsis* (Van de Peer et al., 2017) (Figure 6). If we ignored



**Figure 5.** Coexpression Association of the Homoeologous Genes with 14 Cell Wall Integrity Marker Genes.

**(A)** Outcome of GSEA that tests coexpression association between divergent homoeologs in the Root, Pollen, Leaf\_R, and Leaf\_P modules and 14 known CWI marker genes (y axis). Boxes are colored if the outcome of GSEA was significantly up (coexpression, blue) or down (anti-coexpression, red), while white boxes indicate that no significant association was found.

**(B)** Gene Score ranks of homoeologs in the four different modules when using *WAK2* and *ANX2* as query genes for module detection by the SA. Low ranks indicate high expression similarity to the query gene. The Gene Scores are signed, with positive signs (blue) indicating coexpression and negative signs (red) indicating anti-coexpression under selected conditions. These Gene Score ranks are used by GSEA to assess statistical significant association of CWI marker gene expression to the homoeologs (see Methods for details).

53 duplication events that mapped to the Angiosperm node, representing putative ancient duplication nodes that predate the divergence of basal angiosperms from the other angiosperm species, this fraction increased to 87.9% (131/149). In comparison, 93.4% (2005/2147) of all syntenic gene pairs used in this study and for which phylogenetic trees could be constructed, not limited to the ones that show divergence in the module pairs, mapped to species tree nodes associated with WGD events. Therefore, we conclude that a large fraction of these 149 pairs that do not belong to the filtered set of syntenic gene pairs likely also originated through WGD events. We also studied the  $K_s$  values (number of synonymous substitutions per synonymous site) of the duplicated gene pairs and used these values as another proxy to date duplication nodes (see Methods). With WGD, all duplicates originated at the same time and therefore should have similar  $K_s$  values (Blanc and Wolfe, 2004b; Vanneste et al., 2013). We compared  $K_s$  distributions for these 149 duplication nodes to the  $K_s$  distributions of the duplication nodes for all syntenic gene pairs in our data set. This analysis showed that on average the  $K_s$  values associated with the 149 duplication nodes were larger than those of all syntenic pairs in the data ( $P = 5.781 \times 10^{-6}$ , Wilcoxon rank sum test), whereas they were similar to the 92 divergent homoeologous gene pairs in our data set ( $P = 0.05683$ , Wilcoxon rank sum test).

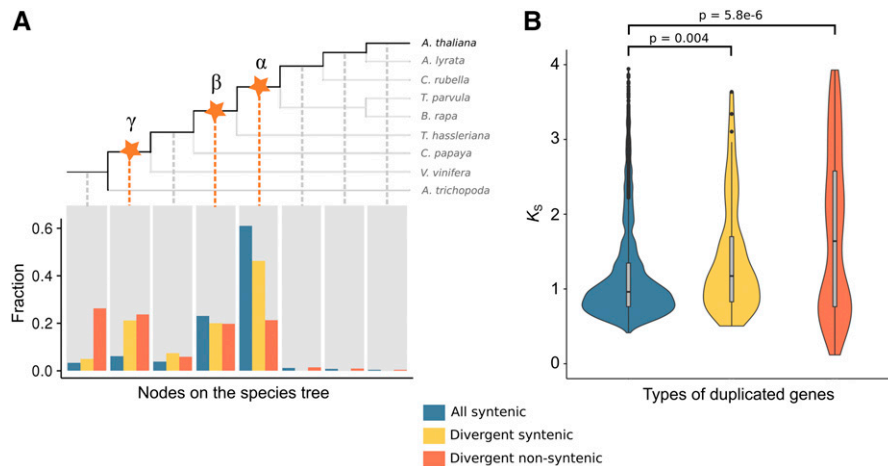
This result would suggest that these 149 duplication nodes are of ancient origin, which is in line with the observations of their duplication nodes mapping to WGD events (Figure 6).

We also looked at the phylogenetic distribution for the 92 divergent homoeologous (and syntenic) gene pairs and observed that they originated from different WGD events and hence do not trace back to a single WGD event (Figure 6). Specifically, they had a biased origin toward more ancient WGD events in comparison to the distribution of all syntenic pairs.

## DISCUSSION

In this study, we analyzed an Arabidopsis development gene expression compendium for patterns of coordinated gene expression divergence of homoeologous gene pairs. Using a gene-centric biclustering approach, we found a striking case of coordinated tissue-specific gene expression divergence of 92 homoeologous gene pairs, with one homoeolog showing biased expression toward tip growth tissue (Root and/or Pollen modules) and the other homoeolog showing biased expression toward aerial tissues (Leaf module). We provide evidence that this pattern of tissue-specific gene expression divergence potentially reflects a pattern of functional divergence where one gene functions in polarized growth,





**Figure 6.** Phylogenetic Distribution and  $K_s$  Distribution of Homoeologous Gene Pairs and Nonsynthetic Duplicated Pairs.

**(A)** We assessed the phylogenetic distribution for three different types of duplicated gene pairs: all syntenic gene pairs (All syntenic), nonsynthetic duplicated pairs showing Root/Pollen versus Leaf gene expression divergence (Divergent nonsyntenic), and the 92 functionally divergent homoeologous gene pairs (Divergent syntenic). Bars represent the fraction of duplicated pairs observed for each node in the species tree. Stars denote ancient WGDs as described in literature.

**(B)**  $K_s$  distributions corresponding to the duplicates in **(A)**. Significant differences between  $K_s$  distributions were calculated by Wilcoxon rank sum test.

while its homoeolog tends to be involved in defense response/stress. By investigating the functions of the 173 genes in these 92 pairs, we hypothesize that many of these genes may have a function with respect to maintaining cell wall integrity and include genes involved in plasma membrane-associated and  $Ca^{2+}$ -dependent signaling, vesicular transport, actin cytoskeleton, and cell wall metabolism. We corroborated this hypothesis by showing that 14 RLK genes that function upstream in cell wall integrity pathways in both tip growth and defense have gene expression patterns that are very similar to the 92 homoeologous gene pairs, suggesting that following WGD, gene pairs that are involved in the maintenance of cell wall integrity have been maintained in duplicate but subsequently diversified to function in defense or growth.

Both tip growth and defense are processes where the maintenance of the cell wall is particularly challenged (Engelsdorf and Hamann, 2014). It is becoming clear that similar cellular components that are used for polarized growth and morphogenesis are also used for defense responses, including the acto-myosin cytoskeleton, targeted exocytosis (of cell wall material), and signaling molecules that redirect the cytoskeleton to the site of cell wall damage (Dettmer and Friml, 2011; Pleskot et al., 2013). Indeed, both processes require a focal distribution of cellular material, either at the site of growth or at the site of cell wall damage. Interestingly, the usage of similar cellular components for polarized growth and defense response appears to be a common theme in eukaryotic cells that is not unique to plants because in budding yeast, but also in higher organisms, a similar cellular machinery is used for polarized growth and cellular wound healing (Sonnemann and Bement, 2011; Kono et al., 2012). In budding yeast, it was even shown that the gene sets involved in polarized growth and wound healing are largely overlapping (Kono et al., 2012).

We observed that the homoeologous gene pairs originated from different WGD events. These observations suggest that this gene

expression divergence pattern is a recurrent event that is tightly associated with WGD. We hypothesize that genes that are involved in the cell wall might be specifically maintained in duplicate following WGD because at the cellular level, polyploidy is often associated with increase in cell size (Kondorosi et al., 2000) and, consequently, polyploid cells require extra cell wall material to accommodate for the increase in cell volume. In support of this hypothesis, Wu et al. (2010) observed that gene expression changes in tetraploid yeast cells involved genes encoding cell surface proteins. They showed that these gene expression changes are likely due to the increased cell size associated with polyploidy, with in particular genes annotated with the “cell wall” GO term being differentially expressed in tetraploid versus diploid cells. This increased demand in cell wall material of polyploid cells might be compensated for following WGD by preserving genes in duplicate that are involved in the production and transport of cell wall material, possibly affecting genes with a wide range of functions, including cell wall biogenesis enzymes, genes involved in cellular trafficking of cell wall material to the site of growth, and signaling molecules that coordinate controlled cellular expansion. In support of this hypothesis, we indeed observe that genes with the “cell wall” GO term are overrepresented among genes that were repeatedly retained in duplicate following multiple independent WGD events in the angiosperms (Li et al., 2016) (Supplemental Figure 5). Prolonged retention of these specific genes following WGD because of the increased demand for cell wall material extends the temporal window during which sub- or neofunctionalization can occur.

How did this pattern of functional divergence evolve in Arabidopsis? There are two possible scenarios: In the first scenario, we assume that the original unduplicated genes had multiple functions and were involved in both polarized growth and defense, while a partitioning of functions (i.e., subfunctionalization)

occurred after duplication. Alternatively, the ancestral genes only possessed one of the functions (i.e., polarized growth or defense response) and following duplication the entire duplicated pathway was co-opted to function in either polarized growth or defense. Hence, whereas the former case postulates partitioning of ancestral functions, the latter involves the re-deployment of entire sets of genes in a novel context. In favor of the first scenario, we hypothesize that partitioning of functions following duplication would represent an improved way to balance growth and stress responses and thus would have an immediate effect on plant fitness. Growth and stress/defense are biological processes that can compete for nutrients and energy (Karasov et al., 2017). Therefore, it is crucial for plants to properly allocate the limited resources between such energy-intensive processes to gain better fitness. Since plants grow in environments in which they are constantly confronted with (a)biotic stresses, they have evolved sophisticated mechanisms to balance growth and defense (Huot et al., 2014). Indeed, there appears to be abundant crosstalk between growth and defense pathways (Ballaré, 2014; Chaiwanon et al., 2016). Under a scenario where, before duplication, the ancestral genes functioned in both growth and stress, we can expect that growth and defense responses compete for the same cellular resources. An elegant study in budding yeast indeed showed that wound healing and polarized cell growth use the same genes and that stabilization of polarity compromises wound healing (Kono et al., 2012). Consequently, duplication of components of these pathways and their subsequent divergence in functioning in growth and defense would at least partially resolve this competition since distinct gene sets would be available to function in polarized growth and defense.

It should be noted that several of the genes in our modules have described functions in both growth and development as well as (a)biotic stress (Supplemental Data Set 1). If these functions actually reflect the ancestral functions of the genes before duplication, the fact that such genes have retained (part of) their ancestral function(s) in growth and development and stress response is not necessarily in conflict with a scenario of subfunctionalization as proposed here. For instance, subfunctionalization is a process that takes time and some genes might still show some residual functionality in stress/defense versus a primary role in growth/stress (or vice versa) and hence, depending on the experimental conditions, one might observe one over the other.

Concerning the alternative scenario, i.e., that following duplication, the existing cellular machinery involved in cell wall maintenance was redeployed for a different function (i.e., either growth or defense), there is a growing body of evidence that existing pathways or regulatory networks can be reused in new settings. Famous examples of co-option of pathways are butterfly wing spot variety, tetrapod limb evolution, and the evolution of complex leaves in plants (True and Carroll, 2002). While it is not known what evolutionary events underlie such network co-option events, it is tempting to speculate that novel developmental or stress-related functions could be created by for instance the evolution of novel expression patterns of just one or few upstream regulatory genes and hence require a minimal number of evolutionary changes. Previously reported cases of network or pathway co-option (e.g., Shimeld et al., 2005; Werner et al., 2010) usually relate to existing gene sets being used in a new setting, but some examples exist of

pathway co-option following gene duplication (Rosin and Kramer, 2009; Hoffmann et al., 2010). The latter has the advantage of producing template pathways that can be molded by evolution while simultaneously limiting possible interference with the ancestral function.

Here, we described two diverged coexpression modules connected by 92 homoeologous gene pairs with predominant expression in either tip growth tissues or aerial tissues following WGDs. The expression patterns seem to be in agreement with their different roles in growth and defense and likely related with the maintenance of cell wall integrity. Hence, our data might describe a case in which a template pathway created by WGD can be subsequently redeployed for other functions by upstream regulatory changes. However, to distinguish between the subfunctionalization and co-option scenarios, we will need additional insight into the ancestral functions of these genes, likely obtained from phenotyping of mutants of orthologs in early branching angiosperms.

## METHODS

### Homoeologous Gene Pairs

Homoeologous gene pairs for *Arabidopsis thaliana* were obtained from PLAZA 2.5 (Van Bel et al., 2012). These contained 5101 gene pairs that reside within genomic blocks that show intraspecies synteny, encompassing 7848 unique Arabidopsis genes. There were 7848 genes and thus fewer than  $2 \times 5101$  genes in these pairs because Arabidopsis has undergone multiple WGDs; hence, the duplication history of genes is often nested such that certain genes can be part of multiple syntenic gene pairs. We also retrieved the  $K_s$  values for these homoeologous gene pairs from PLAZA 2.5 and further filtered these gene pairs based on their  $K_s$  values. We assumed that all gene pairs in the same syntenic block originated from the same duplication event; therefore, we calculated a median  $K_s$  value for each block to represent its age. We plotted the  $K_s$  distribution of these median  $K_s$  values and observed a bimodal distribution, as has been previously documented (Blanc and Wolfe, 2004b; Vanneste et al., 2013), with the first mode likely corresponding to the most recent alpha event and the second mode representing a mixture of pairs from the beta and gamma events (Vanneste et al., 2013). We removed all homoeologous gene pairs on syntenic blocks with a median  $K_s$  value lower than 0.1, as their origin is too recent to come from any of the ancient polyploidization events and likely represent segmental duplications; we also removed all homoeologous gene pairs corresponding to syntenic blocks with a median  $K_s$  value that exceeds 4 to account for  $K_s$  saturation effects (Vanneste et al., 2013). As such, we retained a set of 3962 high-confidence homoeologous gene pairs that were used in subsequent analyses. We also removed the gene pairs in which one of the homoeologs had no matching probe set for the ATH1 array (see below), ending up with a set of 2699 gene pairs (encompassing 4759 unique genes) used for all analyses in this article.

### Gene Expression Data

The main analyses of this article were conducted on the CORNET 2.0 Development compendium (De Bodt et al., 2012), which contains data for 134 different experimental conditions and compiles data from different experiments describing gene expression in different tissues and developmental stages in Arabidopsis. All experiments in this compendium used the ATH1 array and the probe sets could be mapped to 21,428 Arabidopsis genes when using TAIR10 gene annotation. From this compendium, we filtered out all genes that showed little variation in their expression across conditions, by only maintaining 90% of the genes with the

largest gene expression variation (as assessed by  $sd$ ). The final expression matrix contained 19,825 genes and 134 experimental conditions.

Furthermore, to verify that the observed expression patterns were valid in other expression data sets, we also considered the 946 remaining conditions in CORNET 2.0 that were not in the Development compendium and we also used the Rootmap data from Brady et al. (2007). ATH1 probe sets were always mapped to TAIR10 gene annotations.

### Inferring Coexpression Modules

To detect coexpression modules, we used the Signature Algorithm (SA; Ihmels et al., 2002). The SA is a query-driven biclustering approach to identify the genes that show condition-dependent coexpression with a certain seed gene (also called the query gene). Here, we used all 4759 genes from the selected homoeologous gene pairs as seeds for the SA to identify the coexpression modules associated with each of these seed genes. Starting from a certain seed gene, the SA defines two vectors: the condition score vector and the gene score vector that define the coexpression module and are adjusted in subsequent iterations of the algorithm. For each condition, the condition score vector represents the extent to which the genes in the module are up- or downregulated, i.e., reflecting an average gene expression pattern for all genes in the module. In the first iteration of the algorithm, the condition score vector corresponds to the normalized gene expression profile of the seed gene. For each gene, the gene score vector represents the extent to which a gene is coexpressed with the seed gene, giving higher weight to conditions with high values in the condition score vector. In the original SA, coexpression modules, or biclusters, are obtained by putting thresholds on the gene and condition score vectors with threshold values or a range of threshold values chosen by the user, i.e., only genes for which the gene score exceeds a certain threshold value will be assigned to the coexpression module. Here, we implemented a more robust and statistically sound approach to define the threshold on the gene scores. We permuted the original gene expression matrix in a gene-wise manner, i.e., permuting for each gene the expression values across all conditions, keeping only the gene expression profile of the seed gene fixed. We then recalculated the gene and condition scores for this scrambled matrix. We repeated this procedure 1000 times for each seed gene and constructed as such a background distribution of gene scores. We finally selected the 90th, 95th, and 99th percentile of the background distribution as thresholds for the gene score vector and thus generated three sets of coexpression module pairs for further identification of diverged module pairs (Supplemental Figure 1). Since we were mainly interested in the genes that belonged to the same coexpression module, we did not put a threshold on the condition score vectors.

### Identifying Diverged Module Pairs

To identify module pairs that have putatively diverged following WGD, we considered all relevant module pairs associated to homoeologous gene pairs (i.e., 2699 module pairs, corresponding to 2699 homoeologous gene pairs). We defined divergent module pairs as those module pairs that showed gene expression divergence and that shared a large number of homoeologous gene pairs (see Figure 1). To this end, we introduced two different scores: (1) the Divergence score to assess to what extent two modules have diverged in their coexpression profile and (2) the Paralogy score to assess to what extent the genes in two modules are evolutionary related by WGD (i.e., are each other's homoeologs) (Supplemental Figure 6).

Under the assumption that modules that have divergent gene expression patterns will largely contain different genes, we measured the divergence of genes in two modules based on the Jaccard Index, which calculates the similarity between finite sample sets by using the size of the

intersection divided by the size of the union of the sample sets. We defined the Divergence scores as:

$$\text{Divergence score} = 1 - \frac{\text{Number of shared genes between the two modules}}{\text{Total number of genes in the two modules}}$$

Hence, a Divergence score of 1 represents module pairs that have no genes in common. In this work, we considered only module pairs with Divergence scores in the 95th percentile.

The Paralogy score reflects the fraction of homoeologous gene pairs that are shared by two modules in a module pair. To assess the significance of the paralogy score, we sampled two random gene sets from the total number of genes measured on the microarray, keeping the module sizes, the Divergence score, and the total number of homoeologous genes (i.e., the sum of the intermodule and intramodule homoeologous pairs) fixed and compared the number of observed homoeologous gene pairs across both modules to that observed for the randomly sampled gene sets. We repeated this procedure 1000 times for each module pair and selected only module pairs for which the Divergence score belonged to the 95th percentile of the background distribution.

Using the above strategy, we identified diverged module pairs from each of the three sets of coexpression modules (see one example in Supplemental Figure 6). Among these, 22 diverged module pairs existed in all the three sets, suggesting they were independent of the threshold scores used in the SA. These 22 module pairs were partially overlapping module pairs that contained a large number of homoeologous gene pairs that show coordinated gene expression divergence. For these 22 module pairs, we manually checked gene expression divergence by inspecting their gene expression levels in heat map representations and ended up with two module pairs for which the genes showed clear gene expression divergence.

### Assessing Module Stability

Since SA is a query-driven module detection approach, each seed gene (i.e., all 4759 homoeologs) has its own coexpression module. A stable module in the query-driven approach is a module in which each gene from the module can retrieve its own module: If this gene is used as a query gene for module detection, the outcome is again the same module. Hence, in the optimal case there is a reciprocal relationship between all gene members of a module. We used this knowledge to assess the robustness of the four coexpression modules reported in this article, i.e., the Root, Pollen, Leaf\_R, and Leaf\_P modules, specifically focusing on the homoeologs within these modules. In particular, for each homoeolog  $q$  in a certain module, we assessed the success rate of its retrieval by the remaining  $Q-q$  homoeologous genes in the same module. We call this success rate the "retrieval score." The average retrieval score for all 173 homoeologs was 0.57. We compared this number to random expectations by randomly selecting the same number of homoeologs as in the module that we are interested in, without the limitation that these should belong to the same module, and then calculated the median of their individual retrieval scores. We repeated this procedure 1000 times to construct a background distribution of retrieval scores and found the retrieval scores of the 173 homoeologous genes to be significantly larger than that of the background ( $P$  value <  $2.2e-16$ , Welch's two-sample  $t$  test) (Supplemental Figure 7).

### Selecting Conditions Underlying Gene Expression Divergence

We identified the conditions for which all homoeologous gene pairs had divergent gene expression by defining a two-class classification problem in which we assigned all homoeologs in the same coexpression module to one class and the homoeologs in the other coexpression module to the other class. We then used Random Forests (RFs) (Breiman, 2001) to identify those conditions for which the classifier could, based on the gene expression levels, discriminate well between genes in the first class and those

in the second class. To select the conditions, we used the RF approach implemented in the “varSelRF” R package (Diaz-Uriarte, 2007), which combines RF classification with a variable selection method based on backward elimination. We applied this package to our data with 5000 trees and 200 bootstrap samples.

### Functional Enrichment Analysis

We identified GO terms that are significantly overrepresented in the modules using the Bingo package (Maere et al., 2005b). Gene-GO term associations were obtained from geneontology.org, and the file used in this work was submitted by TAIR to the Gene Ontology Consortium on October 5, 2015. Reported P values are obtained after multiple-testing correction using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995).

We used GSEA to test whether module genes showed a statistically significant coexpression association to 14 cell wall integrity marker genes obtained from Engelsdorf and Hamann (2014). In this GSEA approach, we used the Gene Scores obtained by the SA to rank all genes in the gene expression compendium in decreasing order of gene expression similarity to each of the 14 cell wall integrity marker genes (see Figure 5B for examples). We then used this ranked vector of genes to test whether homoeologs in each of the four modules had gene expression patterns that were statistically significantly associated with those of the cell wall integrity marker genes. Specifically, we used the “mean” function from the “piano” R package (Våremo et al., 2013) to test the gene expression association between the cell wall integrity marker genes and the homoeologs. To test significance of the association, the gene rankings were permuted 10,000 times. The Benjamini-Hochberg (Benjamini and Hochberg, 1995) method was used to control the false discovery rate associated with multiple testing.

### Inferring Gene Families and Gene Trees

To infer gene families, we retrieved protein coding sequences from the following sequenced Brassicaceae/Brassicales species: *A. thaliana*, *Arabidopsis lyrata*, *Capsella rubella*, *Brassica rapa*, *Thellungiella parvula*, and *Tarenaya hassleriana*. To delineate the three polyploidy events, we also included three other species, *Carica papaya*, *Vitis vinifera*, and *Amborella trichopoda*, as outgroup species. All protein coding sequences come from Li et al. (2016), except for *T. hassleriana* (Cheng et al., 2013), which was downloaded from NCBI with RefSeq assembly accession GCF\_000463585.1. To identify gene families, we used OrthoMCL (Li et al., 2003) with the inflation parameter defined as  $I = 1.5$  to cluster all-against-all BLASTP results ( $E$ -value  $< 1 \times 10^{-5}$ ). As some of the 2699 homoeologous gene pairs belonged to different orthologous groups identified by OrthoMCL, we further combined those gene families into one gene family for further analysis. Therefore, we ended up with 4139 multigene families that contained each at least one Arabidopsis paralogous gene pair.

A species tree was constructed based on a concatenated multiple sequence alignment of 1311 gene families with exactly one copy in all species and with orthologs in all species (i.e., single-copy gene families) (Supplemental Data Set 2). MUSCLE (3.8.31) (Edgar, 2004) was used to perform multiple sequence alignment on amino acid sequences for each single-copy gene family. Low-quality aligned regions were removed by trim1 in a heuristic mode (“-automated1”), followed by back-translation from amino acid alignments into nucleotide sequence alignments (Capella-Gutiérrez et al., 2009) (Supplemental Data Set 3). We used RAxML (8.2) with the GTR+GAMMA model to infer a maximum likelihood tree that was obtained by optimizing every 5th bootstrap tree in 100 rapid bootstraps; the remaining trees were used to calculate bootstrap-based branch support (Stamatakis, 2014).

For gene tree construction and gene tree–species tree reconciliation, we used the pipeline described by Li et al. (2016). The 4139 gene families with at least a pair of Arabidopsis paralogs were given as input to the reconciliation pipeline. Six gene families with more than 300 members were

removed due to the large computational resources required by large gene families. Gene tree–species tree reconciliation defines nodes in a gene tree as either “duplication” nodes or “speciation” nodes, depending on whether children nodes represent orthologs or paralogs. In addition, they report a relative age of gene duplication based on predictions of where duplications likely map on the species tree (Stolzer et al., 2012). After filtering out low-quality duplication nodes (Li et al., 2016), we obtained duplication nodes in 3981 gene families, including 2350 of 2699 homoeologous gene pairs (Supplemental Data Set 4).  $K_s$  values for the predicted duplication nodes inferred above were estimated in the same way as described by Li et al. (2016) (Supplemental Data Set 4). For each paralogous pair, protein sequences were aligned by ClustalW (Oliver et al., 2005) using parameters recommended by Hall (2004). PAL2NAL (Suyama et al., 2006) was used to back-translate the aligned amino acids into corresponding codon sequences followed by removing gaps.  $K_s$  values were estimated by codeml from PAML using the GY model with stationary codon frequencies empirically estimated by the F3×4 model (Yang, 2007). To remove redundant  $K_s$  values for duplication nodes with multiple paralogous pairs, the minimum  $K_s$  value was selected from all possible gene pairs to represent the timing of the duplication event.

### Accession Numbers

Sequence data from this article can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the accession numbers listed in Supplemental Data Set 1.

### Supplemental Data

**Supplemental Figure 1.** Distribution of module sizes for different threshold parameters.

**Supplemental Figure 2.** Gene expression divergence of duplicated gene pairs in other CORNET conditions.

**Supplemental Figure 3.** Gene expression divergence of homoeologous gene pairs in the Root and Leaf\_R modules for the Rootmap data set (Brady et al., 2007).

**Supplemental Figure 4.** Functional annotation of the genes in the 92 homoeologous gene pairs.

**Supplemental Figure 5.** Overretention of cell wall genes following repeated, independent whole-genome duplication events.

**Supplemental Figure 6.** Criteria used to choose module pairs.

**Supplemental Figure 7.** Module stability for the Pollen, Root, Leaf\_P, and Leaf\_R modules.

**Supplemental Table 1.** Gene Ontology overrepresentation of the homoeologs in the Root/Pollen modules.

**Supplemental Table 2.** Gene Ontology overrepresentation of the homoeologs in the Leaf modules.

**Supplemental Table 3.** Gene Ontology overrepresentation of all module genes in the Root/Pollen and Leaf modules.

**Supplemental Data Set 1.** Overview of the 92 homoeologous gene pairs with literature references and Gene Ontology terms related to their functions in either growth/development or stress.

**Supplemental Data Set 2.** Accession numbers of genes from 1311 single-copy gene families used for building the species tree.

**Supplemental Data Set 3.** Concatenated multiple sequence alignment of 1311 single-copy gene families to build the species tree.

**Supplemental Data Set 4.** Duplication events identified by gene tree–species tree reconciliation and  $K_s$  values for paralogous pairs in Arabidopsis.

## ACKNOWLEDGMENTS

R.D.S. is a postdoctoral fellow of The Research Foundation-Flanders (FWO). Y.V.d.P. acknowledges the Multidisciplinary Research Partnership “Bioinformatics: from nucleotides to networks” Project of Ghent University (no. 01MR0310W) and the European Union Seventh Framework Programme (FP7/2007-2013) under European Research Council Advanced Grant Agreement 322739-DOUBLE-UP. This project is supported by The Research Foundation-Flanders (FWO) (G008812N).

## AUTHOR CONTRIBUTIONS

R.D.S. and Y.V.d.P. designed the research. E.S., Y.S., and R.D.S. developed and implemented the computational approach to identify diverged coexpressed modules. Z.L. and R.D.S. designed and performed analyses on gene family evolution. R.D.S. wrote the manuscript with the assistance of the other coauthors.

Received July 6, 2017; revised October 10, 2017; accepted October 25, 2017; published October 26, 2017.

## REFERENCES

- Abas, L., Benjamins, R., Malenica, N., Paciorek, T., Wiśniewska, J., Moulinier-Anzola, J.C., Sieberer, T., Friml, J., and Luschnig, C. (2006). Intracellular trafficking and proteolysis of the Arabidopsis auxin-efflux facilitator PIN2 are involved in root gravitropism. *Nat. Cell Biol.* **8**: 249–256. Erratum. *Nat. Cell Biol.* **8**: 424.
- Adams, K.L., and Wendel, J.F. (2005). Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* **8**: 135–141.
- Aury, J.M., et al. (2006). Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**: 171–178.
- Ballaré, C.L. (2014). Light regulation of plant defense. *Annu. Rev. Plant Biol.* **65**: 335–363.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**: 289–300.
- Bhosale, R., Jewell, J.B., Hollunder, J., Koo, A.J.K., Vuylsteke, M., Michoel, T., Hilson, P., Goossens, A., Howe, G.A., Browse, J., and Maere, S. (2013). Predicting gene function from uncontrolled expression variation among individual wild-type Arabidopsis plants. *Plant Cell* **25**: 2865–2877.
- Birchler, J.A., and Veitia, R.A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proc. Natl. Acad. Sci. USA* **109**: 14746–14753.
- Blanc, G., and Wolfe, K.H. (2004a). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell* **16**: 1679–1691.
- Blanc, G., and Wolfe, K.H. (2004b). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **16**: 1667–1678.
- Boisson-Dernier, A., Roy, S., Kritsas, K., Grobei, M.A., Jaciubek, M., Schroeder, J.I., and Grossniklaus, U. (2009). Disruption of the pollen-expressed FERONIA homologs ANXUR1 and ANXUR2 triggers pollen tube discharge. *Development* **136**: 3279–3288.
- Borner, G.H.H., Lilley, K.S., Stevens, T.J., and Dupree, P. (2003). Identification of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A proteomic and genomic analysis. *Plant Physiol.* **132**: 568–577.
- Boudsocq, M., and Sheen, J. (2013). CDPKs in immune and stress signaling. *Trends Plant Sci.* **18**: 30–40.
- Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318**: 801–806.
- Braidwood, L., Breuer, C., and Sugimoto, K. (2014). My body is a cage: mechanisms and modulation of plant cell growth. *New Phytol.* **201**: 388–402.
- Breiman, L. (2001). Random forests. *Mach. Learn.* **45**: 5–32.
- Buggs, R.J.A., Zhang, L., Miles, N., Tate, J.A., Gao, L., Wei, W., Schnable, P.S., Barbazuk, W.B., Soltis, P.S., and Soltis, D.E. (2011). Transcriptomic shock generates evolutionary novelty in a newly formed, natural allopolyploid plant. *Curr. Biol.* **21**: 551–556.
- Bullard, J.H., Mostovoy, Y., Dudoit, S., and Brem, R.B. (2010). Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proc. Natl. Acad. Sci. USA* **107**: 5058–5063.
- Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.
- Chaiwanon, J., Wang, W., Zhu, J.Y., Oh, E., and Wang, Z.Y. (2016). Information integration and communication in plant growth regulation. *Cell* **164**: 1257–1268.
- Chen, Z.J. (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.* **58**: 377–406.
- Chen, Z.J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* **15**: 57–71.
- Cheng, S., et al. (2013). The *Tarenaya hassleriana* genome provides insight into reproductive trait and genome evolution of crucifers. *Plant Cell* **25**: 2813–2830.
- Chester, M., Gallagher, J.P., Symonds, V.V., Cruz da Silva, A.V., Mavrodiev, E.V., Leitch, A.R., Soltis, P.S., and Soltis, D.E. (2012). Extensive chromosomal variation in a recently formed natural allopolyploid species, *Tragopogon miscellus* (Asteraceae). *Proc. Natl. Acad. Sci. USA* **109**: 1176–1181.
- Cheung, A.Y., and Wu, H.M. (2008). Structural and signaling networks for the polar cell growth machinery in pollen tubes. *Annu. Rev. Plant Biol.* **59**: 547–572.
- Chi, W., Li, J., He, B., Chai, X., Xu, X., Sun, X., Jiang, J., Feng, P., Zuo, J., Lin, R., Rochaix, J.D., and Zhang, L. (2016). DEG9, a serine protease, modulates cytokinin and light signaling by regulating the level of ARABIDOPSIS RESPONSE REGULATOR 4. *Proc. Natl. Acad. Sci. USA* **113**: E3568–E3576.
- Conant, G.C., and Wolfe, K.H. (2006). Functional partitioning of yeast co-expression networks after genome duplication. *PLoS Biol.* **4**: e109.
- Conant, G.C., and Wolfe, K.H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* **9**: 938–950.
- Cosgrove, D.J. (2005). Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* **6**: 850–861.
- Day, B., Henty, J.L., Porter, K.J., and Staiger, C.J. (2011). The pathogen-actin connection: a platform for defense signaling in plants. *Annu. Rev. Phytopathol.* **49**: 483–506.
- De Bodt, S., Hollunder, J., Nelissen, H., Meulemeester, N., and Inzé, D. (2012). CORNET 2.0: integrating plant coexpression, protein-protein interactions, regulatory interactions, gene associations and functional annotations. *New Phytol.* **195**: 707–720.
- de Carbonnel, M., Davis, P., Roelfsema, M.R.G., Inoue, S., Schepens, I., Lariguet, P., Geisler, M., Shimazaki, K., Hangarter, R., and Fankhauser, C. (2010). The Arabidopsis PHYTOCHROME KINASE SUBSTRATE2 protein is a phototropin signaling element that regulates leaf flattening and leaf positioning. *Plant Physiol.* **152**: 1391–1405.
- De Smet, R., and Marchal, K. (2011). An ensemble biclustering approach for querying gene expression compendia with experimental lists. *Bioinformatics* **27**: 1948–1956.
- De Smet, R., and Van de Peer, Y. (2012). Redundancy and rewiring of genetic networks following genome-wide duplication events. *Curr. Opin. Plant Biol.* **15**: 168–176.

- Dettmer, J., and Friml, J.** (2011). Cell polarity in plants: when two do the same, it is not the same.... *Curr. Opin. Cell Biol.* **23**: 686–696.
- Diaz-Uriarte, R.** (2007). GeneSrf and varSelRF: a web-based tool and R package for gene selection and classification using random forest. *BMC Bioinformatics* **8**: 328.
- Edgar, R.C.** (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797.
- Edger, P.P., et al.** (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. USA* **112**: 8362–8366.
- Engelsdorf, T., and Hamann, T.** (2014). An update on receptor-like kinase involvement in the maintenance of plant cell wall integrity. *Ann. Bot.* **114**: 1339–1347.
- Escobar-Restrepo, J.M., Huck, N., Kessler, S., Gagliardini, V., Gheyselinck, J., Yang, W.C., and Grossniklaus, U.** (2007). The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science* **317**: 656–660.
- Flagel, L.E., and Wendel, J.F.** (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* **183**: 557–564.
- Fraser, H.B., Babak, T., Tsang, J., Zhou, Y., Zhang, B., Mehrabian, M., and Schadt, E.E.** (2011). Systematic detection of polygenic cis-regulatory evolution. *PLoS Genet.* **7**: e1002023.
- Freeling, M., and Thomas, B.C.** (2006). Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* **16**: 805–814.
- Galvan-Ampudia, C.S., Julkowska, M.M., Darwish, E., Gandullo, J., Korver, R.A., Brunoud, G., Haring, M.A., Munnik, T., Vernoux, T., and Testerink, C.** (2013). Halotropism is a response of plant roots to avoid a saline environment. *Curr. Biol.* **23**: 2044–2050.
- Gross, K., and Schiestl, F.P.** (2015). Are tetraploids more successful? Floral signals, reproductive success and floral isolation in mixed-ploidy populations of a terrestrial orchid. *Ann. Bot.* **115**: 263–273.
- Hall, B.G.** (2004). *Phylogenetic Trees Made Easy*. (Sunderland, MA: Sinauer Associates).
- He, F., Arce, A.L., Schmitz, G., Koornneef, M., Novikova, P., Beyer, A., and de Meaux, J.** (2016). The footprint of polygenic adaptation on stress-responsive cis-regulatory divergence in the Arabidopsis genus. *Mol. Biol. Evol.* **33**: 2088–2101.
- Hématy, K., Sado, P.E., Van Tuinen, A., Rochange, S., Desnos, T., Balzergue, S., Pelletier, S., Renou, J.P., and Höfte, H.** (2007). A receptor-like kinase mediates the response of Arabidopsis cells to the inhibition of cellulose synthesis. *Curr. Biol.* **17**: 922–931.
- Hepler, P.K.** (2016). The cytoskeleton and its regulation by calcium and protons. *Plant Physiol.* **170**: 3–22.
- Hepler, P.K., Rounds, C.M., and Winship, L.J.** (2013). Control of cell wall extensibility during pollen tube growth. *Mol. Plant* **6**: 998–1017.
- Hofberger, J.A., Lyons, E., Edger, P.P., Chris Pires, J., and Eric Schranz, M.** (2013). Whole genome and tandem duplicate retention facilitated glucosinolate pathway diversification in the mustard family. *Genome Biol. Evol.* **5**: 2155–2173.
- Hoffmann, F.G., Opazo, J.C., and Storz, J.F.** (2010). Gene cooption and convergent evolution of oxygen transport hemoglobins in jawed and jawless vertebrates. *Proc. Natl. Acad. Sci. USA* **107**: 14274–14279.
- Hong, Y., Zhao, J., Guo, L., Kim, S.C., Deng, X., Wang, G., Zhang, G., Li, M., and Wang, X.** (2016). Plant phospholipases D and C and their diverse functions in stress responses. *Prog. Lipid Res.* **62**: 55–74.
- Huot, B., Yao, J., Montgomery, B.L., and He, S.Y.** (2014). Growth-defense tradeoffs in plants: a balancing act to optimize fitness. *Mol. Plant* **7**: 1267–1287.
- Ihmels, J., Bergmann, S., Gerami-Nejad, M., Yanai, I., McClellan, M., Berman, J., and Barkai, N.** (2005). Rewiring of the yeast transcriptional network through the evolution of motif usage. *Science* **309**: 938–940.
- Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y., and Barkai, N.** (2002). Revealing modular organization in the yeast transcriptional network. *Nat. Genet.* **31**: 370–377.
- Jourda, C., Cardí, C., Mbéguié-A-Mbéguié, D., Bocs, S., Garsmeur, O., D'Hont, A., and Yahiaoui, N.** (2014). Expansion of banana (*Musa acuminata*) gene families involved in ethylene biosynthesis and signalling after lineage-specific whole-genome duplications. *New Phytol.* **202**: 986–1000.
- Karasov, T.L., Chae, E., Herman, J.J., and Bergelson, J.** (2017). Mechanisms to mitigate the trade-off between growth and defense. *Plant Cell* **29**: 666–680.
- Kohorn, B.D., and Kohorn, S.L.** (2012). The cell wall-associated kinases, WAKs, as pectin receptors. *Front. Plant Sci.* **3**: 88.
- Kondrosi, E., Roudier, F., and Gendreau, E.** (2000). Plant cell-size control: growing by ploidy? *Curr. Opin. Plant Biol.* **3**: 488–492.
- Kono, K., Saeki, Y., Yoshida, S., Tanaka, K., and Pellman, D.** (2012). Proteasomal degradation resolves competition between cell polarization and cellular wound healing. *Cell* **150**: 151–164.
- Lau, S., Slane, D., Herud, O., Kong, J., and Jürgens, G.** (2012). Early embryogenesis in flowering plants: setting up the basic body pattern. *Annu. Rev. Plant Biol.* **63**: 483–506.
- Li, L., Stoekert, C.J., Jr., and Roos, D.S.** (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**: 2178–2189.
- Li, Q.G., Zhang, L., Li, C., Dunwell, J.M., and Zhang, Y.M.** (2013). Comparative genomics suggests that an ancestral polyploidy event leads to enhanced root nodule symbiosis in the Papilionoideae. *Mol. Biol. Evol.* **30**: 2602–2611.
- Li, Z., Defoort, J., Tasdighian, S., Maere, S., Van de Peer, Y., and De Smet, R.** (2016). Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell* **28**: 326–344.
- Liu, Z., Persson, S., and Sánchez-Rodríguez, C.** (2015). At the border: the plasma membrane-cell wall continuum. *J. Exp. Bot.* **66**: 1553–1563.
- Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., and Van de Peer, Y.** (2005a). Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. USA* **102**: 5454–5459.
- Maere, S., Heymans, K., and Kuiper, M.** (2005b). BINGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**: 3448–3449.
- Meyerhoff, O., Müller, K., Roelfsema, M.R., Latz, A., Lacombe, B., Hedrich, R., Dietrich, P., and Becker, D.** (2005). AtGLR3.4, a glutamate receptor channel-like gene is sensitive to touch and cold. *Planta* **222**: 418–427.
- Michard, E., Simon, A.A., Tavares, B., Wudick, M.M., and Feijó, J.A.** (2017). Signaling with ions: the keystone for apical cell growth and morphogenesis in pollen tubes. *Plant Physiol.* **173**: 91–111.
- Myers, C., Romanowsky, S.M., Barron, Y.D., Garg, S., Azuse, C.L., Curran, A., Davis, R.M., Hatton, J., Harmon, A.C., and Harper, J.F.** (2009). Calcium-dependent protein kinases regulate polarized tip growth in pollen tubes. *Plant J.* **59**: 528–539.
- Nissen, K.S., Willats, W.G., and Malinovsky, F.G.** (2016). Understanding CrRLK1L function: cell walls and growth control. *Trends Plant Sci.* **21**: 516–527.
- Oliver, T., Schmidt, B., Nathan, D., Clemens, R., and Maskell, D.** (2005). Using reconfigurable hardware to accelerate multiple sequence alignment with ClustalW. *Bioinformatics* **21**: 3431–3432.
- Pereira-Leal, J.B., and Teichmann, S.A.** (2005). Novel specificities emerge by stepwise duplication of functional modules. *Genome Res.* **15**: 552–559.
- Pleskot, R., Li, J., Zárský, V., Potocký, M., and Staiger, C.J.** (2013). Regulation of cytoskeletal dynamics by phospholipase D and phosphatidic acid. *Trends Plant Sci.* **18**: 496–504.

- Prince, V.E., and Pickett, F.B.** (2002). Splitting pairs: the diverging fates of duplicated genes. *Nat. Rev. Genet.* **3**: 827–837.
- Roop, J.I., Chang, K.C., and Brem, R.B.** (2016). Polygenic evolution of a sugar specialization trade-off in yeast. *Nature* **530**: 336–339.
- Rosin, F.M., and Kramer, E.M.** (2009). Old dogs, new tricks: regulatory evolution in conserved genetic modules leads to novel morphologies in plants. *Dev. Biol.* **332**: 25–35.
- Rounds, C.M., and Bezanilla, M.** (2013). Growth mechanisms in tip-growing plant cells. *Annu. Rev. Plant Biol.* **64**: 243–265.
- Ruprecht, C., Mendrinna, A., Tohge, T., Sampathkumar, A., Klie, S., Fernie, A.R., Nikoloski, Z., Persson, S., and Mutwil, M.** (2016). FamNet: a framework to identify multiplied modules driving pathway expansion in plants. *Plant Physiol.* **170**: 1878–1894.
- Sampathkumar, A., Gutierrez, R., McFarlane, H.E., Bringmann, M., Lindeboom, J., Emons, A.M., Samuels, L., Ketelaar, T., Ehrhardt, D.W., and Persson, S.** (2013). Patterning and lifetime of plasma membrane-localized cellulose synthase is dependent on actin organization in *Arabidopsis* interphase cells. *Plant Physiol.* **162**: 675–688.
- Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B., Weigel, D., and Lohmann, J.U.** (2005). A gene expression map of *Arabidopsis thaliana* development. *Nat. Genet.* **37**: 501–506.
- Shi, X., Zhang, C., Ko, D.K., and Chen, Z.J.** (2015). Genome-wide dosage-dependent and -independent regulation contributes to gene expression and evolutionary novelty in plant polyploids. *Mol. Biol. Evol.* **32**: 2351–2366.
- Shimeld, S.M., Purkiss, A.G., Dirks, R.P., Bateman, O.A., Slingsby, C., and Lubsen, N.H.** (2005). Urochordate betagamma-crystallin and the evolutionary origin of the vertebrate eye lens. *Curr. Biol.* **15**: 1684–1689.
- Soltis, P.S., and Soltis, D.E.** (2016). Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* **30**: 159–165.
- Sonnemann, K.J., and Bement, W.M.** (2011). Wound repair: toward understanding and integration of single-cell and multicellular wound responses. *Annu. Rev. Cell Dev. Biol.* **27**: 237–263.
- Stamatakis, A.** (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
- Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., and Durand, D.** (2012). Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* **28**: i409–i415.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P.** (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**: 15545–15550.
- Suyama, M., Torrents, D., and Bork, P.** (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**: W609–W612.
- Taylor, J.S., and Raes, J.** (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* **38**: 615–643.
- Testerink, C., and Munnik, T.** (2005). Phosphatidic acid: a multifunctional stress signaling lipid in plants. *Trends Plant Sci.* **10**: 368–375.
- Thomas, C., and Staiger, C.J.** (2014). A dynamic interplay between membranes and the cytoskeleton critical for cell development and signaling. *Front. Plant Sci.* **5**: 335.
- True, J.R., and Carroll, S.B.** (2002). Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* **18**: 53–80.
- Van Bel, M., Proost, S., Wischnitzki, E., Movahedi, S., Scheerlinck, C., Van de Peer, Y., and Vandepoele, K.** (2012). Dissecting plant genomes with the PLAZA comparative genomics platform. *Plant Physiol.* **158**: 590–600.
- Van de Peer, Y., Maere, S., and Meyer, A.** (2009). The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**: 725–732.
- Van de Peer, Y., Mizrachi, E., and Marchal, K.** (2017). The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**: 411–424.
- Vanneste, K., Maere, S., and Van de Peer, Y.** (2014). Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philos. T. R. Soc. B* **369**: 20130353.
- Vanneste, K., Van de Peer, Y., and Maere, S.** (2013). Inference of genome duplications from age distributions revisited. *Mol. Biol. Evol.* **30**: 177–190.
- Väremo, L., Nielsen, J., and Nookaew, I.** (2013). Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res.* **41**: 4378–4391.
- Vincill, E.D., Clarin, A.E., Molenda, J.N., and Spalding, E.P.** (2013). Interacting glutamate receptor-like proteins in phloem regulate lateral root initiation in *Arabidopsis*. *Plant Cell* **25**: 1304–1313.
- Wapinski, I., Pfeffer, A., Friedman, N., and Regev, A.** (2007). Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**: 54–61.
- Werner, T., Koshikawa, S., Williams, T.M., and Carroll, S.B.** (2010). Generation of a novel wing colour pattern by the *Wingless* morphogen. *Nature* **464**: 1143–1148.
- Wolf, S., Hématy, K., and Höfte, H.** (2012). Growth control and cell wall signaling in plants. *Annu. Rev. Plant Biol.* **63**: 381–407.
- Wu, C.Y., Rolfe, P.A., Gifford, D.K., and Fink, G.R.** (2010). Control of transcription by cell size. *PLoS Biol.* **8**: e1000523.
- Wu, L.F., Hughes, T.R., Davierwala, A.P., Robinson, M.D., Stoughton, R., and Altschuler, S.J.** (2002). Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters. *Nat. Genet.* **31**: 255–265.
- Yang, Z.** (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**: 1586–1591.
- Zabotina, O.A., van de Ven, W.T.G., Freshour, G., Drakakaki, G., Cavalier, D., Mouille, G., Hahn, M.G., Keegstra, K., and Raikhel, N.V.** (2008). *Arabidopsis* XXT5 gene encodes a putative alpha-1,6-xylosyltransferase that is involved in xyloglucan biosynthesis. *Plant J.* **56**: 101–115.
- Zhang, Y., Xiao, Y., Du, F., Cao, L., Dong, H., and Ren, H.** (2011). *Arabidopsis* VILLIN4 is involved in root hair growth through regulating actin organization in a Ca<sup>2+</sup>-dependent manner. *New Phytol.* **190**: 667–682.
- Zhou, W., Brockmüller, T., Ling, Z., Omdahl, A., Baldwin, I.T., and Xu, S.** (2016). Evolution of herbivore-induced early defense signaling was shaped by genome-wide duplications in *Nicotiana*. *eLife* pii: e19531.

## Coordinated Functional Divergence of Genes after Genome Duplication in *Arabidopsis thaliana*

Riet De Smet, Ehsan Sabaghian, Zhen Li, Yvan Saeys and Yves Van de Peer  
*Plant Cell* 2017;29;2786-2800; originally published online October 25, 2017;  
DOI 10.1105/tpc.17.00531

This information is current as of January 2, 2018

<b>Supplemental Data</b>	<a href="/content/suppl/2017/10/25/tpc.17.00531.DC1.html">/content/suppl/2017/10/25/tpc.17.00531.DC1.html</a>
<b>References</b>	This article cites 106 articles, 29 of which can be accessed free at: <a href="/content/29/11/2786.full.html#ref-list-1">/content/29/11/2786.full.html#ref-list-1</a>
<b>Permissions</b>	<a href="https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&amp;issn=1532298X&amp;WT.mc_id=pd_hw1532298X">https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&amp;issn=1532298X&amp;WT.mc_id=pd_hw1532298X</a>
<b>eTOCs</b>	Sign up for eTOCs at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a>
<b>CiteTrack Alerts</b>	Sign up for CiteTrack Alerts at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a>
<b>Subscription Information</b>	Subscription Information for <i>The Plant Cell</i> and <i>Plant Physiology</i> is available at: <a href="http://www.aspb.org/publications/subscriptions.cfm">http://www.aspb.org/publications/subscriptions.cfm</a>