Plant Biotechnology Journal
Open Access

Plant Biotechnology Journal (2020) **18**, pp. 553–567 doi: 10.1111/pbi.13223

Comparative transcriptomics enables the identification of functional orthologous genes involved in early leaf growth

Jasmien Vercruysse[1,2], Michiel Van Bel[1,2], Cristina M. Osuna-Cruz[1,2], Shubhada R. Kulkarni[1,2], Véronique Storme[1,2], Hilde Nelissen[1,2], Nathalie Gonzalez[1,2,3,†], Dirk Inzé[1,2,*,†] and Klaas Vandepoele[1,2,†]

[1]*Department of Plant Biotechnology and Bioinformatics, Ghent University, Gent, Belgium*
[2]*Center for Plant Systems Biology, VIB, Gent, Belgium*
[3]*INRA, UMR1332 Biologie du fruit et Pathologie, INRA Bordeaux Aquitaine, Villenave d'Ornon Cedex, France*

Received 20 May 2019;
revised 10 July 2019;
accepted 25 July 2019.
**Correspondence* (Tel +32 9 331 38 00; fax +32 9 331 38 09; email dirk.inze@psb.vib-ugent.be)
†These authors contributed equally.

Keywords: *Arabidopsis thaliana*, *Zea mays*, orthology, data integration, networks, comparative transcriptomics, leaf development.

Summary

Leaf growth is a complex trait for which many similarities exist in different plant species, suggesting functional conservation of the underlying pathways. However, a global view of orthologous genes involved in leaf growth showing conserved expression in dicots and monocots is currently missing. Here, we present a genome-wide comparative transcriptome analysis between Arabidopsis and maize, identifying conserved biological processes and gene functions active during leaf growth. Despite the orthology complexity between these distantly related plants, 926 orthologous gene groups including 2829 Arabidopsis and 2974 maize genes with similar expression during leaf growth were found, indicating conservation of the underlying molecular networks. We found 65% of these genes to be involved in one-to-one orthology, whereas only 28.7% of the groups with divergent expression had one-to-one orthology. Within the pool of genes with conserved expression, 19 transcription factor families were identified, demonstrating expression conservation of regulators active during leaf growth. Additionally, 25 Arabidopsis and 25 maize putative targets of the TCP transcription factors with conserved expression were determined based on the presence of enriched transcription factor binding sites. Based on large-scale phenotypic data, we observed that genes with conserved expression have a higher probability to be involved in leaf growth and that leaf-related phenotypes are more frequently present for genes having orthologues between dicots and monocots than clade-specific genes. This study shows the power of integrating transcriptomic with orthology data to identify or select candidates for functional studies during leaf development in flowering plants.

Introduction

Zea mays and *Arabidopsis thaliana* are two important model organisms for monocots and dicots, respectively, which diverged 140–200 million years ago (Chaw *et al.*, 2004). Although they present numerous morphological and physiological differences, many developmental processes, such as leaf growth, show remarkable similarities (Li *et al.*, 2010; Nelissen *et al.*, 2016; Yu *et al.*, 2015). In both organisms, leaf growth is characterized by two major consecutive phases: a cell proliferation phase, during which cells divide, and a cell expansion phase, during which cells increase their volume. The interplay between cell division and cell expansion determines final leaf size (Hepworth and Lenhard, 2014). However, the direction of growth is different in both species. In Arabidopsis, the cellular growth pattern is dispersed, resulting in round mature leaves with reticulate veins, while in maize, growth is directed in a longitudinal fashion, generating narrow, elongated leaves with parallel veins (Nelissen *et al.*, 2016; Nelson and Dengler, 1997). As soon as the leaf emerges from the shoot apical meristem, all cells from the primordium start dividing. After a few days, cells at the tip of the leaf cease division and start to elongate, marking the beginning of cell expansion. The leaf is then composed of both dividing and expanding cells during a so-called transition period, characterized by a cell-cycle arrest front moving from tip to base until most cells are no longer dividing and only expanding (Andriankaja *et al.*, 2012; Donnelly *et al.*, 1999; Karidas *et al.*, 2015; Nelissen *et al.*, 2016; Pyke *et al.*, 1991).

Conservation of the cell division and cell expansion phases between both species suggests that the underlying molecular mechanisms are preserved. These conserved mechanisms involve orthologous genes that evolved from a gene in a common ancestor through a speciation event. Nevertheless, different events including gene duplication and functional divergence result in complex many-to-many orthology in plants, where potentially only a subset of orthologues have conserved functions whereas the other orthologues acquired different functions or are differently expressed. Sorting out the true functional orthologues which do share similar biological functions can be accomplished by combining transcriptomic data with orthology information (Movahedi *et al.*, 2012; Patel *et al.*, 2012). The resulting integrated network provides a powerful means to explore transcriptome-wide patterns of conservation and divergence between species (Movahedi *et al.*, 2011). Several studies have already described orthologous genes with a conserved expression pattern and/or function in two or more species (Hefer *et al.*, 2015; Libault *et al.*, 2010; Mustroph *et al.*,

2010; Narsai *et al*., 2011; Rensink *et al*., 2005). For example, members of the GROWTH-REGULATING FACTOR (GRF) transcription factor (TF) family play key roles in leaf growth regulation and development and have been shown to interact with the co-activator GRF-INTERACTING FACTOR (GIF) both in Arabidopsis and in maize (Kim and Kende, 2004; Lee *et al*., 2009; Nelissen *et al*., 2015; Vercruyssen *et al*., 2014; Zhang *et al*., 2008). EXPANSINs function in both Arabidopsis and maize as cell wall-loosening proteins facilitating cell expansion and stimulating cell enlargement when overexpressed or applied exogenously (Cho and Cosgrove, 2000; Geilfus *et al*., 2015). In these examples, the similar function of orthologous genes hints at a conservation of the molecular mechanisms that regulate leaf growth.

To identify the molecular processes active during leaf development, several genome-wide transcriptome studies addressing changes in gene expression during early leaf growth have been realized in Arabidopsis and maize (Avramova *et al*., 2015; Baerenfaller *et al*., 2012; Baute *et al*., 2016; Dubois *et al*., 2017; Li *et al*., 2010; Nelissen *et al*., 2018; Skirycz *et al*., 2011). These experiments provide a bird's-eye view on the expression of all genes possibly involved in leaf growth within a species. However, comparing gene expression data across multiple species requires accurate orthology information between the species. A network including orthology allows studying expression dynamics of different organisms at the gene level. Nonetheless, the current picture of all components involved in leaf growth shared between different plants is far from complete.

Here, we aimed at identifying common and divergent leaf growth-related regulatory processes in a monocot and a dicot plant. To do so, the expression conservation between Arabidopsis and maize was studied during early leaf development. We built an integrated network based on expression data linked with complex gene orthology information to study similarities between the two species, allowing to list potential leaf growth-regulating candidates. The level of conservation among Arabidopsis and maize TFs was also explored, indicating substantial conservation of transcriptional regulation.

Results

Similar expression patterns during leaf development in Arabidopsis and maize

To discover common or divergent expression patterns between Arabidopsis and maize during leaf development, we compared gene expression profiles in both species. Two publicly available gene expression data sets were used (Andriankaja *et al*., 2012; Nelissen *et al*., 2018). Both sets include time-course gene expression information from samples harvested during leaf development and correspond to the three main developmental phases: cell proliferation, cell expansion and the transition between these two phases (Figure S1a). For Arabidopsis, the expression information originated from six datapoints, in which two successive datapoints corresponded to one developmental phase. For maize, the expression information from samples harvested at nine datapoints was used, with four samples corresponding to the proliferating zone, three to the transitioning zone and two to the expanding zone. In total, expression levels were measured for 29 920 Arabidopsis and 39 323 maize genes of which, respectively, 4217 and 6495 differentially expressed (DE) genes (DEGs) were selected.

To identify major expression trends in both sets, genes were clustered based on their co-expression during leaf development. Hierarchical clustering was used to group the DEGs in a minimal set of clusters reporting coherent expression profiles within each cluster and distinct profiles between clusters (Methods, Figure S1b, S1c). To facilitate statistical data analysis, only clusters with at least 50 genes were retained. For the Arabidopsis data set, initial clustering of the DEGs yielded 11 clusters, of which six containing more than 50 genes with distinct expression trends were retained: clusters A1 to A6 that represent 97% of the DEGs. For the maize data set, the same approach resulted in the selection of seven out of eight clusters (clusters Z1 to Z7), representing 99% of the DEGs. In both sets of clusters, three trends were visible, corresponding to a peak of expression during proliferation, during transition or during expansion (Figure S1d, S1e). Additionally, in the maize data set, a fourth expression trend was visible corresponding to expression peaks during proliferation and expansion, forming an inverse profile of the transition-specific expression trend. To identify molecular processes active during leaf growth, we performed GO enrichment on all Arabidopsis and maize gene clusters (Figure S2). Clusters with decreasing expression (A1 and Z1) were enriched for terms such as cell division and DNA replication, processes that occur during early leaf development (Andriankaja *et al*., 2012). In A2, however, terms such as responses to stress (wounding, hypoxia), meristem development and leaf polarity determination were enriched, of which the latter process is known to occur in very early leaf development (Stahle *et al*., 2009). In clusters with a transition-specific profile (A3, A4, Z2, Z3 and Z4), different terms were found enriched in function of the cluster. Cluster A3 was enriched for terms related to cell division, whereas A4 was enriched for galactolipid and chlorophyll biosynthesis correlating with photosynthesis onset. Clusters Z2, Z3 and Z4 included genes leading to few enriched terms that suggested divergent biological processes such as root morphogenesis, fatty acid metabolic process and protein translation, respectively. Terms including photosynthesis and chlorophyll biosynthesis were found to be enriched in clusters with increasing expression during leaf development (A5, Z5 and Z6). This is to be expected since photosynthesis begins when leaves start to expand (Andriankaja *et al*., 2012; Van Dingenen *et al*., 2016). In contrast, A6 included terms related to plant defence [glucosinolate biosynthesis (Wittstock and Burow, 2010)], energy production (myo-inositol hexakisphosphate biosynthesis), auxin signalling and response to brassinosteroid. For Z7, no enriched terms were found.

To conclude, we mostly found similar expression trends for Arabidopsis and maize DEGs during leaf development with coinciding gene functions, suggesting that common pathways control leaf development in both species.

Integration of orthology and expression information identifies conserved orthologues expressed during leaf development

To identify conserved gene functions during leaf development in both species, gene orthology was determined and analysed. Orthologous genes were identified with the PLAZA integrative orthology method (Van Bel *et al*., 2012) that combines three orthology prediction methods: orthologous gene families inferred through sequence-based clustering with OrthoMCL (O); reconciled phylogenetic trees (T); and multispecies best hits and inparalogs (BHI) families (B) (Methods). Whereas the latter are related to a simple BLAST-based approach, OrthoMCL and phylogenetic trees are more advanced methods better capturing complex orthology relationships. For instance, using the B

evidence only, three orthologues of Arabidopsis *GRF3* could be identified, whereas including O and T evidence, twelve additional maize orthologues were identified (Figure 1a). To avoid missing any potential functionally conserved genes, we considered genes as orthologues when at least one type of evidence was found using PLAZA. At the genome-wide level, we found that 83.8% (22 987 genes) and 76.0% (29 874 genes), of the Arabidopsis and maize protein-encoding genes, respectively, have orthologues in the other species (Figure 1b, 1c). Different levels of complexity were observed in the constructed orthology network composed of a total of 7336 separated subnetworks (or orthologous groups): one-to-one (31%), one-to-many (28%) and many-to-many orthology (41%; Figure 1d).

To identify orthology groups showing similar expression patterns in Arabidopsis and maize, we used the integrative orthology information to compare the expression clusters from Arabidopsis and maize. To do so, the overlap between two clusters was estimated by determining the number of Arabidopsis genes of a given cluster having orthologues in a given maize cluster and by calculating the significance of this overlap (Figure 2, Methods). Of the 1811 genes of the Arabidopsis A1 cluster, there were 1709 genes with maize orthologues, and for the maize Z1 cluster that includes 2649 genes, there were 2407 Arabidopsis orthologues. The comparison of these two clusters showed that 862 Arabidopsis genes had 924 maize orthologues and that this 50.4% overlap (862/1709) was significant ($P < 0.001$). The same comparisons were done for all clusters, and we observed a significant overlap for eight cluster comparisons ($P < 0.001$).

Overall, when including significant and non-significant overlaps, four distinct scenarios could be distinguished (Figure 2). First, for five cluster pairs (A1-Z1, A5-Z5, A5-Z6, A6-Z5 and A6-Z6), a significant overlap of orthologous genes showing similar expression patterns was observed. This significant overlap suggests that in both species, gene function during leaf development is probably conserved among orthologues with similar expression. Among these five cluster pairs, 217 one-to-one orthology pairs showing conserved expression were identified, including 69 groups with 110 unknown genes, which suggest 69 new maize and 41 new Arabidopsis candidates that might regulate leaf development in both species in a similar way (Table S1). Furthermore, the 217 one-to-one pairs with conserved expression corresponded to 91.2% of all DE one-to-one orthology pairs.

Second, in three cases (A2-Z3, A3-Z6 and A4-Z6), there was a significant overlap between clusters with a divergent expression profile. For instance, in the A2-Z3 overlap (25 Arabidopsis and 22 maize genes), Arabidopsis genes showed high expression during proliferation, whereas for maize, the peak of expression occurred in the transition zone. The difference in expression was even more clear in the A3-Z6 and A4-Z6 significant overlaps. Clusters A3 and A4 showed high expression during transition, whereas in Z6 high expression occurred during cell expansion. This divergence could indicate that either the orthologous genes are involved in different processes or that they are part of similar processes, however regulated in a different manner during leaf development. One-to-one orthology was only found in the A3-Z6 and A4-Z6 overlaps, corresponding to, respectively, 0.8% (two groups) and 1.3% (three groups) of all DE one-to-one orthology pairs.

Third, seven of all cluster comparisons did not show any significant overlap although the respective clusters showed a conserved expression profile (A2-Z1, A3-Z2, A3-Z3, A3-Z4, A4-Z2, A4-Z3 and A4-Z4). Interestingly, six of these comparisons were between clusters that shared a transition-specific expression profile. The absence of significant overlap agrees with the poor GO enrichment overlap found for the five clusters (Figure S2). Finally, the other cluster comparisons between clusters with divergent profiles did not show any significant overlap.

Detection of conserved processes involved in leaf development

To identify the biological processes in which orthologues with conserved expression during leaf development take part, a GO enrichment analysis was done using the Arabidopsis shared

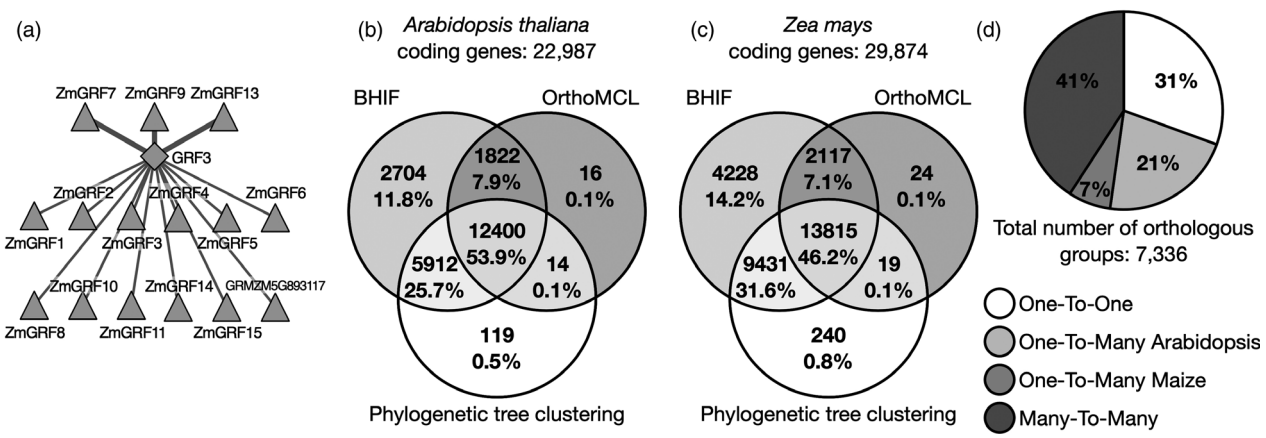


Figure 1 Overview of the properties of the orthologous network constructed between Arabidopsis and maize using the integrative orthology method in PLAZA. (a) Visual representation of Arabidopsis *GROWTH-REGULATING FACTOR 3* (*GRF3*) (diamond) and its orthologues in maize. Three maize orthologues (upper three triangles) are confirmed by multispecies Best-Hits-and-Inparalogs families (B) and reconciled phylogenetic trees (T), and twelve additional orthologues (lower 12 triangles) are predicted by T. Arabidopsis genes are represented by diamonds and maize genes by triangles. (b) and (c) Venn diagrams depicting the number and the percentage of Arabidopsis (b) and maize (c) genes sharing orthologues with the other species through the three evidence sources in PLAZA. (d) Pie chart showing the percentage of orthologous groups with one-to-one, one Arabidopsis gene with many maize orthologues (one-to-many Arabidopsis), one maize gene with many Arabidopsis orthologues (one-to-many maize) or many-to-many orthologous links.

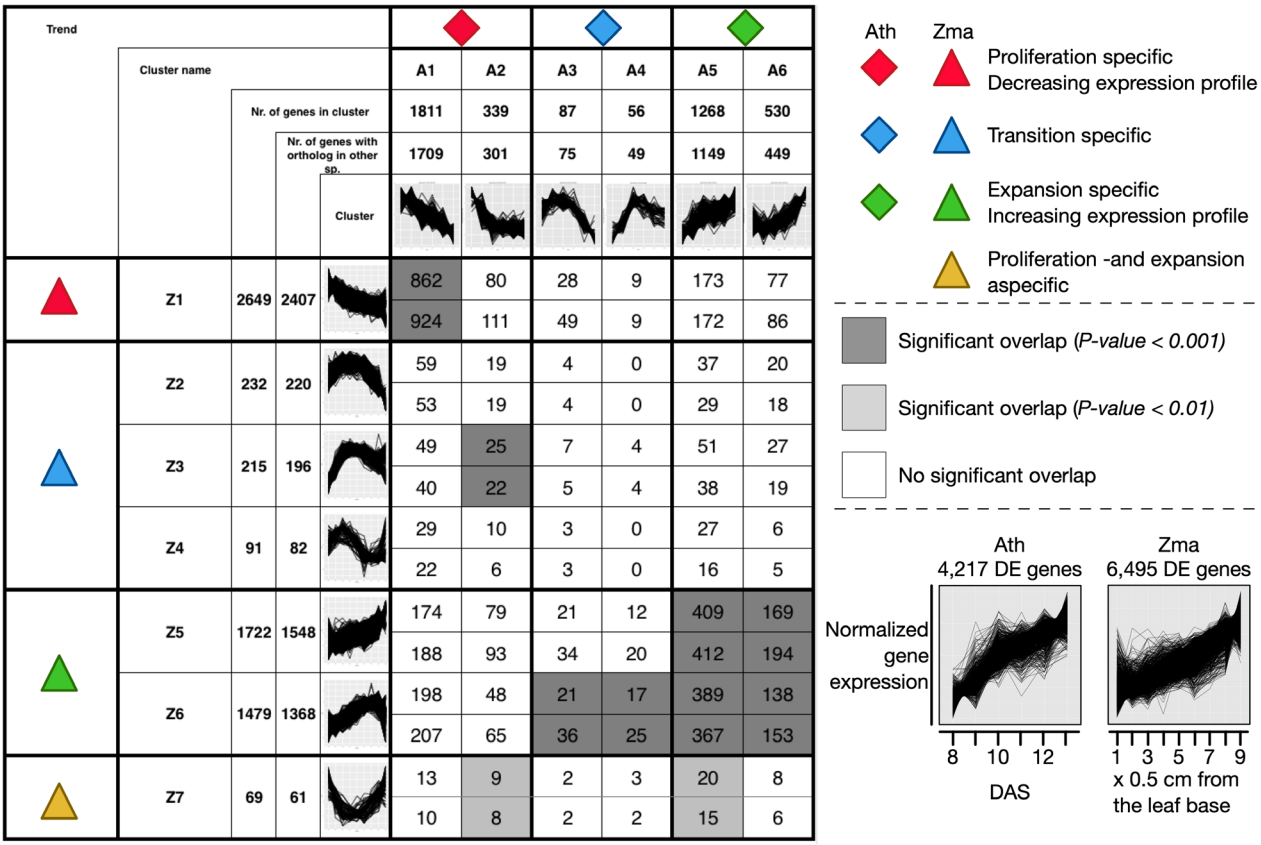


Figure 2 Clustered expression profiles of DEGs in Arabidopsis and maize during leaf development together with the orthology overlap between clusters. The six selected Arabidopsis clusters are depicted horizontally (A1–A6, diamonds) and the seven selected maize clusters are ordered vertically (Z1–Z7, triangles), all of them sorted according to their visual trend. Clusters A1, A2 and Z1 have high expression in the early stages, corresponding to the proliferation phase, and a decreasing expression as leaf development advances (red). Clusters A3, A4, Z2, Z3 and Z4 show high expression during transition and low expression at the beginning and later stages of leaf development (blue). Clusters A5, A6, Z5 and Z6 have an upward gene expression trend coinciding with the onset of the expansion phase (green). Cluster Z7 shows high expression during proliferation and expansion, and low expression during the transition phase (yellow). The total number of genes in each cluster and the number of DEGs with an DE orthologue are shown under or right of the cluster name for Arabidopsis or maize, respectively. For each comparison between two clusters, the overlap is shown for both the Arabidopsis (upper number) and maize genes (lower number). Eight significant overlaps ($P < 0.001$) are shown in dark grey, and two significant overlaps ($P < 0.01$) are shown in light grey. At the bottom right, two representative clusters of the Arabidopsis (left) and the maize (right) DEG set are shown. Normalized gene expression is depicted on the *x*-axis according to the different datapoints on the *y*-axis, which corresponds to 8–13 DAS for the Arabidopsis data set and to distance to the leaf base for the maize data set. Ath., *Arabidopsis thaliana*; Zma., Zea mays; Nr., number; sp., species; DE, differentially expressed; DAS, days after stratification.

orthologues from the five significant overlaps A1-Z1, A5-Z5, A5-Z6, A6-Z5 and A6-Z6 that exhibited conserved expression (Figure 3).

A significant overlap of orthologues was observed for the A1-Z1 comparison. The enriched terms were related to different steps of the cell cycle including the M phase (cell plate formation and mitotic cytokinetic process) and the S phase (regulation of DNA duplication). For example, we observed 16 members of the MINI-CHROMOSOME MAINTENANCE (MCM) complex and 19 members of the histone family. Interestingly, RNA methylation suggesting modification of the RNA metabolism and histone H3-K9 methylation resulting in gene silencing (Hu *et al*., 2019) were also found enriched in the orthologues showing a high expression during the proliferation phase.

The enriched terms in the A5-Z5 and A5-Z6 overlaps were mainly related to photosynthesis and energetic/carbohydrate metabolism. Within the genes part of the A5-Z5 overlap, 12 genes encoding glycoside hydrolases (including *AtBGLU8* and *AtBLGU9*), which are involved in cell expansion (Cosgrove, 2005; Minic and Jouanin, 2006), were present. Furthermore, 11 members of the glutathione S-transferase (GST) gene family (including *AtGSTU17* and *AtGSTU29*) were also found in the GO enrichment analysis. In leaves, GST transcripts have been reported to accumulate in older rather than in younger tissues (Gong *et al*., 2005), validating their allocation to the A5-Z5 overlap. In the shared Arabidopsis and maize orthologues from the A5-Z6 (389 Arabidopsis and 367 maize orthologues) overlap, aquaporins (11 genes all of which are either members of PLASMA MEMBRANE INTRINSIC PROTEIN or TONOPLAST MEMBRANE INTRINSIC PROTEIN families) were present. The aquaporin families have been shown to have an important role in expanding cells, causing turgor pressure by water uptake in Arabidopsis and maize (Chaumont *et al*., 1998; Ludevid *et al*., 1992).

In the A6-Z5 and A6-Z6 overlaps, instead of genes related to metabolism execution, we found genes related to regulatory processes such as hormone signalling, response to auxin and to

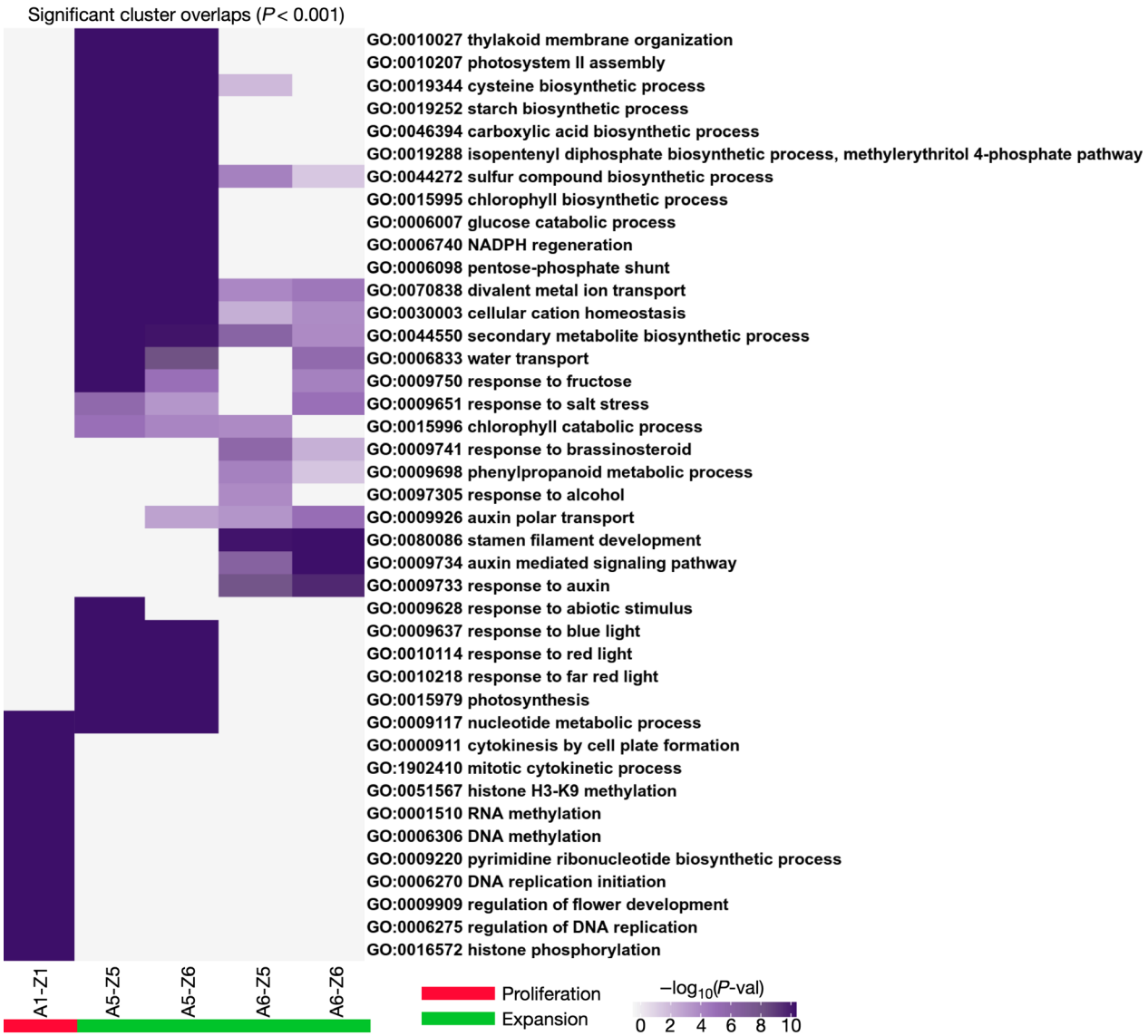


Figure 3 Heat map showing the top 10 enriched GO terms (biological process) per overlap with enrichment in at least one overlap for the four significant overlaps ($P < 0.001$) with conserved expression showing a peak in expression during cell expansion (green) and for the one significant overlap ($P < 0.001$) with conserved expression with an expression peak during cell proliferation (red).

brassinosteroids, and secondary metabolite biosynthesis (phenylpropanoid metabolic process). For example, the AUXIN-INDOLE-3-ACETIC ACID (AUX/IAA) protein family, among which *SHORT HYPOCOTYL 2* (*SHY2*) and *SOLITARY ROOT* (*SLR*) were found in these overlaps. Additionally, genes encoding eight members of the SMALL AUXIN UP RNA (SAUR) protein family, of which some are known to regulate plasma membrane H^+-ATPase activity resulting in cell elongation, were found in the A6-Z6 overlap (Spartz *et al*., 2014). Several genes related to cell wall biosynthesis were also found, such as *LACCASE17* (*LAC17*) and *CHITINASE PROTEIN-LIKE2* (*CTL2*), involved in lignin biosynthesis, and *XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE7* (*XTH7*) (Berthet *et al*., 2011; Sánchez-Rodríguez *et al*., 2012).

In Arabidopsis, although the overlaps including A5 or A6 genes showed a general increased expression during leaf growth, it was obvious that the enriched categories are different in both overlaps with A5 containing genes related to metabolism while A6 contains genes that are more related to regulatory processes (hormone signalling) and secondary metabolism. Interestingly, in maize, these two types of categories seem to be distributed in both clusters Z5 and Z6.

With this analysis, we were able to identify 1967 Arabidopsis genes with 2050 maize orthologues showing conserved expression and function, being part of a significant overlap. Conserved biological processes related, but not limited, to the cell cycle were found for 862 Arabidopsis and 924 maize orthologous genes with proliferation-specific expression in the A1-Z1 overlap. Interestingly, RNA and histone methylation were also enriched. A large part of the 1105 Arabidopsis and 1126 maize orthologous genes with expansion-specific expression (A5-Z5, A5-Z6, A6-Z5 and A6-Z6) were clearly related to photosynthesis and carbohydrate metabolism, but also to hormonal signalling. To conclude, we were able to show large functional conservation among orthologues in dicots and monocots with conserved expression.

Phenotypic analysis of Arabidopsis genes showing conserved and divergent expression profiles

To further explore the impact of conserved expression on gene function during leaf development, publicly available phenotypic data were mapped onto the orthology and expression information we combined. TAIR (Berardini *et al*., 2015) and RARGE II (Akiyama *et al*., 2013) were consulted to generate a list of 738 Arabidopsis genes that if mutated, lead to a leaf-related phenotype such as an increased leaf rosette size (Methods). The genes in A1, A5 and A6 that had a significant overlap were used in this analysis. For these clusters, three groups were defined as follows: (i) a conserved group containing Arabidopsis genes with at least one orthologue showing conserved expression, (ii) a divergent group in which Arabidopsis genes only had orthologues with divergent expression and (iii) a group of Arabidopsis genes lacking maize orthologues, thus representing species-specific genes. For example, cluster A5 included 119 genes with no maize orthologues and 756 which did have a DE maize orthologue. Of these 756 genes, 673 belonged to the conserved group and 83 to the divergent groups (Table S2). Phenotypic data were mapped onto these three groups for the three clusters. Globally, 4.0% (738/18 369) of the Arabidopsis genes lead to a leaf phenotype when mutated, while for A5 conserved and divergent groups, percentages of, respectively, 6.24% (42 genes; *P*-value 2.92E−02) and 7.23% (six genes; *P*-value 0.53) were found. For the A5 species-specific group, 3.36% of leaf-related phenotypes were found (four genes; *P*-value 0.87). These results suggest that DE A5 genes with DE maize orthologues during leaf development have a higher possibility to cause a leaf-related phenotype when mutated. For A1 and A6, the percentage of leaf-related phenotypes in the conserved group (3.58% and 2.86%, respectively) and the divergent group (4.84% and 5.88%, respectively) was greater as compared to the species-specific group (2.94% and 2.47%, respectively).

Over all groups, there were only nine genes that result in an increased leaf size when mutated, of which three could be confirmed (Brauner *et al*., 2014; Cheng *et al*., 2005; Gong *et al*., 2005; Storozhenko *et al*., 2002). In contrast, 49 genes were found that decreased leaf size when mutated, of which 13 could be confirmed (Chiang *et al*., 1995; Hobbie and Estelle, 1995; Kim *et al*., 2006; Lintala *et al*., 2007; Liu *et al*., 2015; Mattioli *et al*., 2008; Ohkama-Ohtsu *et al*., 2007; Peng *et al*., 2006; Šimková *et al*., 2012; Uematsu *et al*., 2012; Xia *et al*., 2010; Xiao *et al*., 2012; Yoshizawa *et al*., 2014). For both phenotypes, a larger number of genes were found in the conserved expression groups compared with the divergent expression groups, suggesting that expression conservation is an important property to identify genes involved in leaf development (Table S3).

In the A5 conserved group, five genes causing an increased leaf size through overexpression, corresponded to *AT1G04680*, *GLUTAMATE:GLYOXYLATE AMINOTRANSFERASE1* (*GGT1*), *1-AMINOCYCLOPROPANE-1-CARBOXYLIC OXIDASE* (*ATACO2*), *RESPONSIVE TO DESICCATION22* (*RD22*) and *STARCH-FREE1* (*STF1*). Apart from *GGT1* and *STF1*, which are known to decrease leaf size when knocked out (Brauner *et al*., 2014; Gong *et al*., 2005), these genes were only described in RARGE II and are thus interesting candidates for further characterization. Of these genes, we know that at least one maize orthologue showed conserved expression during leaf development and could thus also be involved in leaf size regulation. Additionally, in the A6 conserved group, TF DNA-BINDING WITH ONE ZINC FINGER2 (DOF2) was found to increase leaf size when overexpressed and is interestingly known to be expressed at the tip of the leaf in very early developmental stages (Gardiner *et al*., 2010). With this analysis, we conclude that for orthologues with conserved expression, a significant fraction is associated with known leaf phenotypes.

Apart from conservation analysis, also cases suggesting expression divergence of orthologues were investigated. For example, in the A5 divergent group, we identified GIBBERELLIN 3-OXIDASE 1 (GA3ox1), considered the last and rate-limiting enzyme in the gibberellin biosynthetic pathway leading to the production of bioactive gibberellins, that causes a semi-dwarfed phenotype in Arabidopsis when knocked out (Chiang *et al*., 1995; Fleet *et al*., 2003). Together with the expression of *GA3ox1,* which increases as leaf growth advances with a peak during cell expansion, the *ga3ox1-3* mutant phenotype demonstrates a function for this gene in the control of leaf expansion (Davies, 2004; Mitchum *et al*., 2006). In contrast, the maize *GA3ox1* orthologue exhibited the inverse expression pattern with a peak in expression during cell proliferation and a decrease in expression as leaf growth advances coinciding with the accumulation of gibberellin at the transition zone, controlling the shift towards cell expansion at this boundary (Nelissen *et al*., 2012). Although both orthologues may metabolically carry out the same biochemical function, the regulation of this function is different in Arabidopsis and maize due to the spatial organization of the leaf. In total, we could identify eight Arabidopsis genes of which four, three and one belonged to the A1, A5 and A6 divergent groups, respectively, that had divergent expression patterns compared with their maize orthologues and that exhibited an increase (one gene) or decrease (seven genes) in leaf size when mutated (Table S3). There were only four genes in the A5 (two genes) and A6 (two genes) groups with no orthologues that displayed an increase (one gene in A6NO) or decrease (two genes in A5NO and one gene in A6NO) in leaf size when mutated, suggesting an Arabidopsis-specific leaf-related function. These results reveal the potential to identify possible candidates that are either Arabidopsis-specific or gene candidates that have an altered expression profile, suggesting a change in regulation or function.

Visualization of the integrated orthologous expression network

To study the complexity of the genome-wide orthologous relations between clustered Arabidopsis and maize DEGs, a visual representation was generated using Cytoscape (Shannon *et al*., 2003). The complexity of the network was limited by only featuring the orthologous relations between DEGs of clusters A1-A6 and Z1-Z7 (Supplemental network file).

The filtered network included 2371 Arabidopsis and 2675 maize DEGs, visualized as nodes, that were connected through 4829 edges, representing orthology. Genes were coloured according to their expression profile trend. Additionally, edges were coloured when the orthologues were part of a significant overlap. This approach allowed studying transcriptional conservation between both species at both the gene level and the genome-wide level.

The network was divided into 1487 orthologous groups that can be seen as (parts of) gene families in which node colour visualizes conserved/divergent expression. There were 421 orthologous groups with conserved expansion-specific expression, of which 67.2% had a one-to-one orthology relation. Conserved proliferation-specific expression was observed for 501 orthologous groups, of which 64.7% had one-to-one orthology relation. Only four groups containing ten genes were found with conserved transition-

specific expression, of which three displayed one-to-one orthology. These data therefore revealed that 62.7% of the integrated orthology network showed expression conservation among orthologues. In contrast, 37.3% (561 groups) of the groups contained orthologues with divergent expression. Within these groups, only 161 corresponded to one-to-one orthology (28.7%), showing that orthologues with divergent expression have more complex orthology compared with completely conserved expression orthology groups. It seems thus more likely to observe orthologues with divergent expression in complex orthology groups with many-to-many orthology, suggesting neofunctionalization and subfunctionalization within such group. For example, among the 197 groups displaying one-to-two orthology (one Arabidopsis gene with two maize orthologues), for 59 groups one maize gene showed divergent expression (Table S4), suggesting a different role for the two orthologues. For instance, for Arabidopsis *ALCOHOL DEHYDROGENASE* (*ADH*), with high expression during proliferation, only *ZmADH1* had conserved expression, whereas *ZmADH2* had high expression during expansion. This example illustrates the duplication of a progenitor gene, after which the expression, and possibly the function, of *ZmADH2* could have deviated (Dennis *et al*., 1985).

Next, the conservation between DEGs acting as TFs or involved in chromatin-remodelling (CR), and their DE orthologues, which are expected to regulate the expression of numerous downstream genes, was analysed. There were 467 orthologues (encoding 211 Arabidopsis TFs and 256 maize TFs), which formed 106 orthologous groups corresponding to 47 TF and transcriptional regulator (TR) families (Table S5). These 47 families could be divided into four groups based on the number of orthologues with conserved expression: complete conservation; more than 50% conservation; less than 50% conservation; and complete divergence (Table S6). There were 19 TF and TR families, including 45 Arabidopsis genes and 56 maize orthologues, showing complete expression conservation. Among these families, two TF and TR groups were noticeable: a first group was related to chromatin-mediated transcription regulation [among which HIGH MOBILITY GROUP (HMG) genes, GNAT family members, SNF2 genes, SET genes and the ULTRAPETALA1 (ULT1) TF], while a second group consisted of more specific regulators of gene expression such as YABBY, GRF, E2F and NAC TF families (Figure 4). YABBY TFs are involved in the determination of the adaxial/abaxial leaf polarity and correspondingly, all DE TFs of this family showed high expression during proliferation (Stahle *et al*., 2009). The major cell-cycle regulators, the E2F/DP TFs, which regulate the G1-to-S phase transition (Berckmans and De Veylder, 2009; De Veylder *et al*., 2002), and the GRF TFs (Gonzalez *et al*., 2012) involved in the regulation of cell proliferation, had similar expression profiles corresponding to high expression during proliferation.

We observed 15 gene families, among which members of the TEOSINTE BRANCHED1/CYCLOIDEA/PCF (TCP) TF family, for which the expression conservation was higher than 50% (Figure S3). In Arabidopsis, class I and class II work antagonistically by activating and suppressing growth, respectively, balancing leaf developmental processes (Nicolas and Cubas, 2016). TCP orthologues with conserved expression included the class II *AtTCP3*, *AtTCP4* and *ZmTCP23*, and class I *AtTCP21* and *ZmTCP4* with high expression during expansion (Figure 5a). Divergent expression was observed in the group comprising *AtTCP19*, *ZmTCP19* and *ZmTCP21*. AtTCP19 and ZmTCP19 had conserved expression, whereas AtTCP19 and ZmTCP21 differed in expression, which identifies ZmTCP19 as the most likely functional orthologue of AtTCP19 (Figure 5a). Other TF families with more than 50% of the orthologues with conserved expression included, but were not limited to, the CCAAT, AUX/IAA and the C2H2 families.

Members of the MYB and bHLH family were categorized in the group with less than 50% conserved expression between orthologues. The rather low degree of conservation, 36.36% and 41.67%, respectively, can be explained by multiple duplication events in both species followed by functional divergence (Carretero-Paulet *et al*., 2010; Dubos *et al*., 2010; Feller *et al*., 2011). Only three TF families, i.e. LIM, TAZ and BRI1-EMS-SUPPRESSOR 1 (BES1), included orthologues lacking expression conservation, suggesting that these proteins might operate differently during leaf development in both species.

We conclude that TF families playing key roles in cellular processes required for leaf growth frequently have similar expression profiles in both conserved and divergent orthology groups, indicating that they might also regulate the expression of similar target genes in both organisms.

Integration of transcription factor binding site information to identify conserved regulators involved in leaf development

The presence of enriched TF binding sites (TFBSs) in the promoters of orthologues with conserved expression suggests the preservation of the mode of action of the TFs and their targets during leaf growth in dicots and monocots. Therefore, we investigated whether known Arabidopsis TFBSs linked with cell cycle (E2F) or leaf maturation (TCP) were overrepresented in the promoters of Arabidopsis DEGs and maize orthologues with conserved expression. The studied binding sites were previously determined using protein binding microarrays (Weirauch *et al*., 2014). First, the proof of concept was established using the well-characterized Arabidopsis E2F TFs known to bind the degenerative WTTSSCSS motif and of which many Arabidopsis targets are already known and described (Vandepoele *et al*., 2005). One specific instance of the degenerative E2F TFBSs, TTTCCCGC, was enriched in the promoters (500 bp upstream from the start codon) of A1 (47 genes, enrichment fold 2.04, *P*-value 2.2E−06) and Z1 (71 genes, enrichment fold 2.36 and *P*-value 1.26E−11; Figure S4), both containing genes with high expression during proliferation. Within this pool of genes, 12 Arabidopsis genes had 16 maize orthologues, indicating the presence of conserved E2F target genes (Table S7). Among these common targets were *MCM2* and *MCM5*, involved in chromatin organization, as well as *REPLICATION PROTEIN A 2B* (*RPA2*), involved in DNA replication, repair and recombination (Elmayan *et al*., 2005). These results were benchmarked using two previously published studies that identified E2F-regulated genes (Naouar *et al*., 2009) and E2F-bound genes (Verkest *et al*., 2014). The overlap between both studies resulted in 393 possible E2F targets. All 12 Arabidopsis genes were confirmed as previously described E2F targets, which validates our approach to identify putative, new TF targets in both species jointly.

To identify additional regulators and target genes involved in leaf development, the conservation of the known TCP TFBSs in both Arabidopsis and maize DEGs was studied. To do so, eight experimentally identified Arabidopsis TFBSs representing three class I TCPs (TCP16, TCP19 and TCP20) and five class II TCPs (TCP2, TCP3, TCP4, TCP5 and TCP24) were employed (Figure 5b; Weirauch *et al*., 2014). After motif enrichment analysis, two TFBSs representing TCP3 and TCP24, both classified as JAW-TCPs responsible for leaf curvature (Efroni *et al*., 2008), were enriched in clusters A5 and Z5 with high expression during expansion and

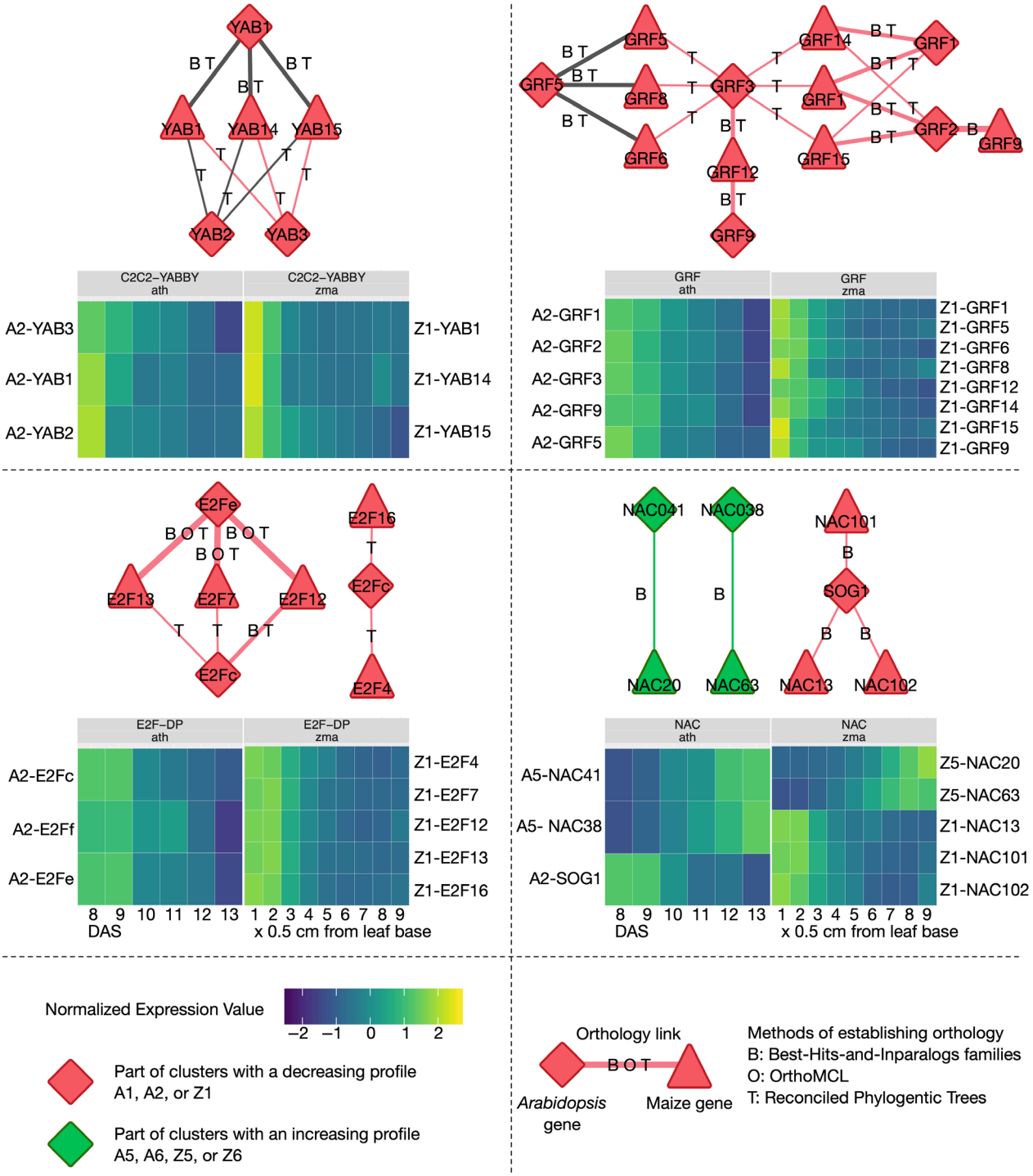


Figure 4 Overview of the DEGs from four TF families, YABBY, GRF, E2F and NAC, showing conserved expression patterns in both Arabidopsis and maize. In each panel, the top image represents the integrated orthology network in which Arabidopsis TFs (diamonds) and maize orthologues (triangles) are linked. The colour of the nodes corresponds to an increasing (green) or a decreasing (red) expression trend during leaf development. The thickness of the edge corresponds to the number of evidences found for that particular orthology link (B: best hits and inparalogs families, O: OrthoMCL and T: reconciled phylogenetic trees). The colour of the edges represents a significant orthology overlap between the respective Arabidopsis and maize genes. The heat map in each panel represents the normalized expression values of each DE TF throughout leaf development. Ath., *Arabidopsis thaliana*; zma., *Zea mays*;DAS, days after stratification.

A1 with proliferation-specific expression (Table S8). Among the set of 229 A1 and 272 A5 genes with an enriched TCP TFBS in their promoters, 34 genes shared 34 orthologues from a total of 279 Z5 genes, of which 25 Arabidopsis genes and 25 maize orthologues showed conserved expansion-specific expression (Figure 5c). We can thus consider these 68 genes (i) to be differentially expressed during leaf growth with a peak in expression during cell proliferation or expansion, (ii) to be

orthologues in Arabidopsis and maize, and (iii) to have a TCP binding site in their first 500 bp that is enriched in their respective clusters. From these three observations, 68 genes could thus be identified as putative TCP TF targets. Among the set of targets, *LOX2*, a known target of TCP4 and TCP20 (Danisman *et al*., 2012) and its maize ortholog, *GRMZM5G822593*, were identified as TCP3 and/or TCP24 targets (Figure 5c).

In conclusion, through the integration of conserved TFs active during leaf development and TF binding site information, detailed gene regulatory subnetworks could be identified offering new insights into target genes for specific TFs in both species. For the E2F TFs, we could show strong conservation of the TFBSs in both Arabidopsis and maize, while for the TCP family, the TFBS conservation was not that pronounced.

Discussion

Leaf growth has been studied extensively both in Arabidopsis and in maize and there are remarkable similarities, at cellular and molecular levels, between the seemingly very different outcomes (for a review, see Nelissen *et al*., 2016). Here, we aimed at answering the question to what extent there is a likely functional conservation between all orthologous genes expressed during leaf development in these distinct species.

With our analysis, we show that comparative analysis of orthologous gene expression profiles can help identifying potential new players involved in leaf growth. We found 217 one-to-one orthology groups with conserved expression, of which the gene function is likely to be similar in both species. Maize and/or Arabidopsis candidates could therefore be selected for functional characterization. Proof of concept is demonstrated with *LOW PSII ACCUMULATION 1* (*LPA1*), involved in photosystem II and chloroplast maintenance (Peng *et al*., 2006). It seems that the *LPA1* expression is conserved as it takes part in a one-to-one orthology with GRMZM2G383154, previously described as a chloroplast maintenance gene as well (Huang *et al*., 2013). Among these 217 orthology groups, 41 with unknown genes in both species were identified which could be potential leaf growth regulators (Table S1).

From the one-to-many orthology combined with the expression information, possible sub- or neofunctionalization of orthologous genes in both species could be shown by studying orthologous groups with divergent expression. This is best shown in one-to-two orthology, in which one of the two maize orthologues changed expression as a consequence of a duplication event, causing reduced selection to maintain the ancestral expression pattern and/or function, resulting in a tailored or new function for this orthologue. We were able to identify 59 one-to-two orthology groups, of which one of the two maize orthologues with divergent expression could have a new function (Table S4). Clearly, correctly sorting out these orthologues showing different levels of expression conservation is of key importance to select valid gene candidates when translating trait information from Arabidopsis to maize.

Furthermore, integration of transcriptomic data could help untangling many-to-many orthology and narrow down the genuine functional orthologues. For instance, *REGULATORY PARTICLE AAA-ATPASE 2A* (*RPT2a*), involved in the regulation of cell expansion duration (Ueda *et al*., 2004), is part of a complex orthologous group with 59 Arabidopsis genes and 77 maize orthologues. By including DEGs during leaf development into this orthology network, the most likely functional maize orthologue of *RPT2a*, *GRMZM2G056569*, could be identified since it exhibits the same expression profile. Another example involves *CYCLIN D3;1* (*CYCD3;1*) regulating the proliferation duration (Dewitte *et al*., 2007). *CYCD3;1* is part of an orthologous group including ten Arabidopsis genes and 23 maize orthologues. After integration of the expression data, two maize orthologues, *GRMZM2G140633* and *GRMZM2G133413*, with the same expression pattern, were found. These two genes could be more interesting candidates to study their role in cell proliferation in maize, compared with the other 21 maize orthologues (Supplemental network file). Narrowing down the orthologues to a set of potential functional orthologues by integrating gene expression data helps prioritizing and selecting potential functional orthologues simultaneously in Arabidopsis and maize for further functional studies.

To further aid the selection of potential new growth-regulating genes, we conducted a phenotypic analysis that revealed a stronger involvement of genes with orthologues in leaf-related phenotypes compared with specific-specific genes. Furthermore, this analysis yielded nine putative Arabidopsis candidates that are likely to increase leaf size when overexpressed (Table S3). Additionally, the phenotypic analysis allowed us to list four potential Arabidopsis-specific leaf-regulating genes that do not have a maize ortholog, but which do alter the leaf phenotype when mutated. Moreover, differences in leaf size and shape between both species could be partially shown as we identified eight Arabidopsis genes with nine maize orthologues exhibiting divergent expression involved in leaf growth.

Transcription factors and regulators that tightly control gene expression play an important role in plant development including leaf growth. Here, for 19 TF/TR families with DEGs, orthologues showed expression profiles that are similar during leaf growth in Arabidopsis and maize, suggesting that their transcriptional regulation is conserved. Some of the TFs have a well-known role in leaf development, such as the previously mentioned GRF, YABBY and E2F family members. Other genes having orthologues with conserved expression are involved in CR such as the Arabidopsis *ATXR6* gene and its maize ortholog, *PHD7*, members of the SET regulators (Baumbusch *et al*., 2001). ATXR6 interacts with the PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) and is up-regulated during the S phase of the cell cycle (Raynaud *et al*., 2006). The expression profile of *ATXR6* and *PHD7*, showing high expression during the cell proliferation phase, coincides therefore with that of the known *SET* genes.

In addition, we show evidence for expression conservation among putative orthologous targets of TFs, based on similar, enriched TFBSs in the promoters of these orthologues. This level of conservation is interesting, knowing the evolutionary distance between both species and suggests possible conservation of TFBSs in other plant species. For the E2F TFs, 12 Arabidopsis and 16 maize genes were identified as putative targets, of which all of the Arabidopsis targets were previously described as E2F targets. Even in more complex TF families such as the TCP family, conservation of the TFBS could be shown, albeit less evident. This low conservation might be because the TFBSs used in this study are Arabidopsis-centred and maize promoters might harbour adapted or even completely different TFBSs recruiting conserved TFs. Secondly, the size of the studied promoters (500 bp upstream from the start codon) was kept constant in both species, while maize promoters are considered larger and more diffuse as compared to Arabidopsis promoters since the majority of the maize genome is covered by transposable elements (Lisch, 2013).

In conclusion, through the cross-species analysis of gene expression profiles functional orthologous candidates with conserved expression could be identified for further functional characterization in both species during leaf development, bridging the knowledge between different model species for flowering plants.

Methods

Selection of differentially expressed genes and clustering

Normalized expression values from the transcriptome analysis of Arabidopsis (Andriankaja *et al*., 2012) and maize (Nelissen *et al*., 2018) were used to select DEGs. The data sets for both species were processed in a similar manner. First, fold changes and false discovery rates (FDRs) for each gene were calculated by comparing each datapoint to every other datapoint. For the Arabidopsis data set, with six datapoints, 15 comparisons ($6 \times (6 - 1)/2$) were done per gene, and for the maize data set including nine datapoints, 36 comparisons were done. Next, DEGs were selected based on $-2 \geq$ fold change ≥ 2 and FDR ≤ 0.05. The resulting DEGs in both data sets corresponded to the union of all comparisons. Both DE data sets were clustered according to hierarchical clustering in R version 3.4.0 (R Core Team, 2017). Distances between genes were calculated with the Euclidean distance measure, and complete linkage was used as agglomeration method. All cluster plots were generated with the ggplot2 R-package (Wickham, 2016).

GO enrichment

GO enrichment analysis was done using the PLAZA 3.0 Dicots platform. GO enrichment results were filtered based on *P* value $\leq 1E^{-3}$ and fold change ≥ 2. Only biological process (and molecular function in case of separate clusters) terms were considered. The heat maps that display the GO enrichment results were created in R version 3.4.4 with the ComplexHeatmap package (Gu *et al*., 2016). The rows with GO terms were clustered with hierarchical clustering (Euclidean distance measure, clustering method complete linkage). Finally, the top ten GO terms with enrichment in at least one cluster or cluster overlap were displayed.

Orthologue identification between Arabidopsis and maize

PLAZA was used to determine the orthologous genes between Arabidopsis and maize (Proost *et al*., 2015; Van Bel *et al*., 2012). This platform includes nine species: four dicots [*Arabidopsis thaliana* (TAIR10), *Arabidopsis lyrata* (JGI v1.0), *Populus trichocarpa* (JGI assembly release v3.0, annotation v3.0) and *Solanum lycopersicum* (ITAG 2.4 on genome build SL2.50)], four monocot species [*Zea mays* (AGPv3.0), *Sorghum bicolor* (v3.1), *Oryza sativa* ssp. *Japonica* (MSU RGAP 7) and *Brachypodium distachyon* (v3.1)] and, additionally, an out-group species, *Amborella trichopoda* (*Amborella* v1.0), to root the phylogenetic tree of those species. This PLAZA version allows to identify orthologues with four methods, corresponding to orthologous gene families inferred through sequence-based clustering with OrthoMCL (Li *et al*., 2003), phylogenetic trees, multispecies best hits and inparalogs families (Linard *et al*., 2011) and synteny information, although we did not use the latter method. In order to establish an orthologous relation between two genes, we used a threshold of one prediction method.

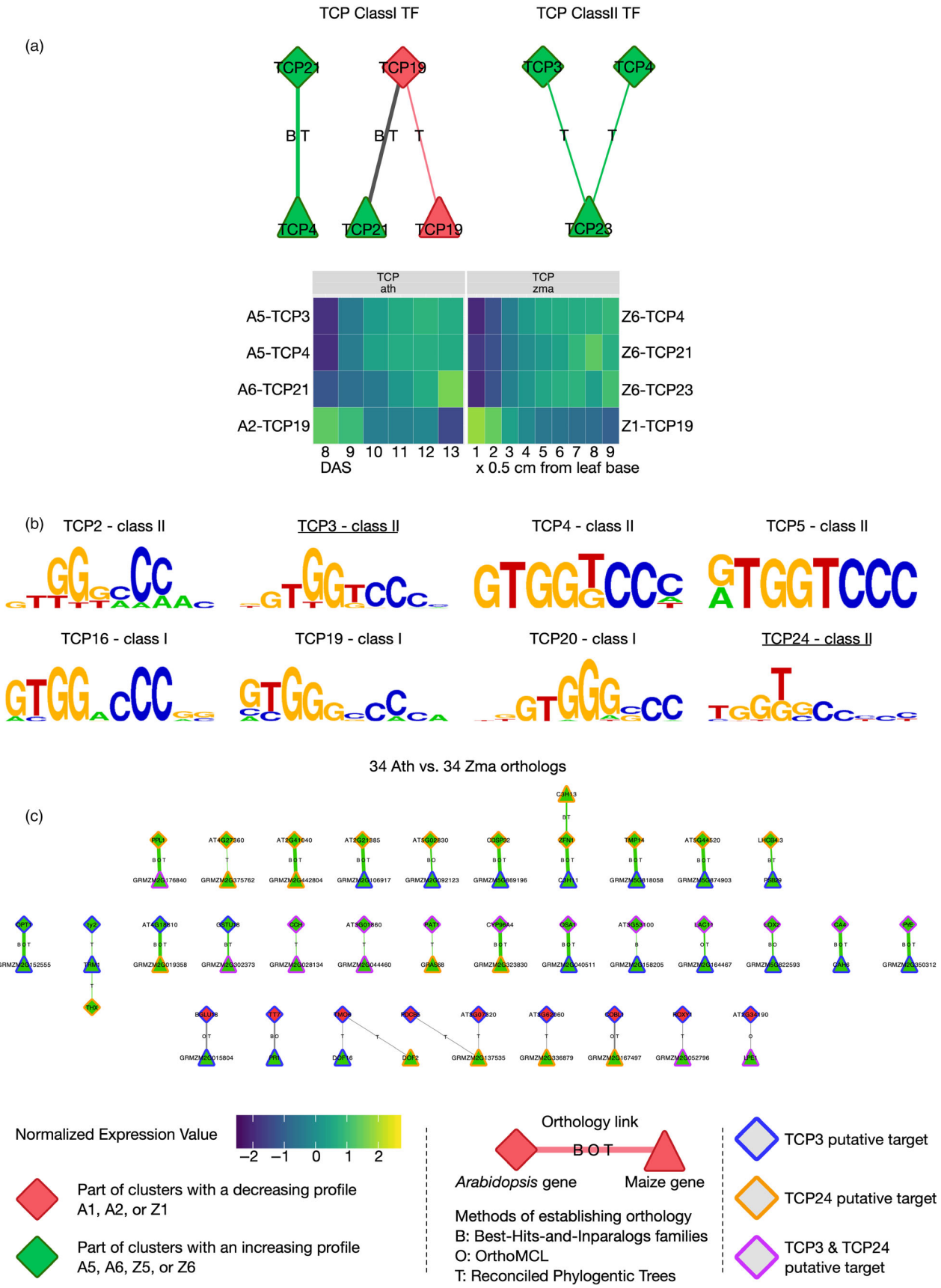
Significant overlap between Arabidopsis and maize clusters

We used a random sampling analysis to calculate the significant overlap. Each Arabidopsis cluster was compared to each maize cluster. The overlap between two clusters was determined as the group of orthologous Arabidopsis and maize genes present in those clusters. In order to determine significant overlaps between Arabidopsis and maize clusters, the *P*-value was calculated by comparing randomly sampled maize clusters to each existing Arabidopsis cluster and counting the overlap again. This was done 1000 times so the *P*-value could be determined at the 0.001 level.

In silico leaf-related phenotypic analysis

TAIR (Berardini *et al*., 2015) and RARGE II (Akiyama *et al*., 2013) were consulted in order to perform the phenotypic analysis. The TAIR germplasm list contained phenotypic information for 1841 Arabidopsis mutant lines. This gene list was filtered using the description of the phenotype based on the presence of the 'leaf' or 'leaves' keywords, and then manually curated for overexpressed or knocked out genes that lead to a leaf-related phenotype. This process resulted in a curated list of 291 genes. The RARGEII database included 17 809 Arabidopsis genes of which 2726 showed a phenotype. 453 genes were selected since they were annotated with rosette leaf terms according to the Plant Ontology structure: wrinkled, upturned, epinastic, coiled, increased or decreased in length, size and width, present in greater or fewer numbers in organism and abnormal. Genes involved in a pale, dark or variegated colouring of the rosette leaves were left out of the analysis. Combining curated data from both databases resulted in a list of 738 unique genes with leaf-related phenotypes when overexpressed or knocked out. For each gene group, the enrichment and significance of the leaf-related

Figure 5 Overview of the conserved expression among TCP TFs, their orthologues and their putative targets. (a) Overview of the Arabidopsis (diamonds) class I and class II TCP DEGs with their DE maize (triangles) orthologues during leaf growth. The heat map represents the normalized expression values of each DE TF throughout leaf development in both species. (b) Visual representation of eight experimentally verified TF binding sites (TFBSs) corresponding to eight TCP TFs, three class I TFs and five class II TFs. The TFBSs of the underlined TFs, TCP3 and TCP24, were found enriched in promoters of Arabidopsis and maize gene clusters (Table S8). Four different colours represent the four bases (yellow: G, red: T, blue: C and green: A), and the relative size of the bases corresponds to the frequency of occurrence. (c) Visual representation of the 34 Arabidopsis and 34 maize putative TCP3 and/or TCP24 targets that were identified based on the TFBS enrichment analysis. Genes (nodes) are coloured according to their expression trend: increasing profile (green) or decreasing profile (red) during leaf development, while the colour of the node edge shows the gene to be a target of TCP3 (blue), TCP24 (yellow) or both (purple). The thickness of the edge corresponds to the number of evidences found for that particular orthology link (B: best hits and inparalogs families, O: OrthoMCL and T: reconciled phylogenetic trees). The colour of the edges represents a significant orthology overlap between the respective Arabidopsis and maize genes. Ath., *Arabidopsis thaliana*; Zma., *Zea mays*; DE, differentially expressed.


(a)
TCP ClassI TF
TCP ClassII TF
TCP21
TCP19
TCP3
TCP4
B T
B T T
T T
TCP4
TCP21
TCP19
TCP23
TCP ath
TCP zma
A5-TCP3
A5-TCP4
A6-TCP21
A2-TCP19
Z6-TCP4
Z6-TCP21
Z6-TCP23
Z1-TCP19
8 9 10 11 12 13
DAS
1 2 3 4 5 6 7 8 9
x 0.5 cm from leaf base
(b)
TCP2 - class II
TCP3 - class II
TCP4 - class II
TCP5 - class II
TCP16 - class I
TCP19 - class I
TCP20 - class I
TCP24 - class II
34 Ath vs. 34 Zma orthologs
(c)
Normalized Expression Value
−2 −1 0 1 2
Part of clusters with a decreasing profile
A1, A2, or Z1
Part of clusters with an increasing profile
A5, A6, Z5, or Z6
Orthology link
B O T
Arabidopsis gene
Maize gene
Methods of establishing orthology
B: Best-Hits-and-Inparalogs families
O: OrthoMCL
T: Reconciled Phylogentic Trees
TCP3 putative target
TCP24 putative target
TCP3 & TCP24
putative target

phenotypes was calculated using the hypergeometric score and corrected for multiple testing with the Benjamini and Hochberg method (Benjamini and Hochberg, 1995).

Network visualization

Cytoscape 3.1 (Shannon *et al*., 2003) was used to visually represent the integrated comparative network. Orthologous information was used to create the network: genes were represented as nodes and their orthology relation as edges. Gene expression data together with species information and gene aliases were added as attribute format and visually represented by differently coloured nodes in the network. The .cys file is made available as supporting file.

Transcription factor binding site conservation

Cis-BP was consulted for all directly inferred TCP binding sites in position weight matrix (PWM) format of Arabidopsis (Weirauch *et al*., 2014). All TFBSs were mapped on the first 500 bp (including the 5′ UTR) of the promoter using the Cluster-Buster algorithm with settings: background model based on 3-kb input sequences, cluster score threshold 0 and pseudo-count 0.375; Frith *et al*., 2003). All mappings were ranked according to *P*-value, and the top 3000 hits for unique genes were used for calculating the enrichment. The enrichment of all TFBSs in each Arabidopsis and maize cluster was calculated employing the hypergeometric score and corrected for multiple testing with the Benjamini and Hochberg method (Benjamini and Hochberg, 1995).

Acknowledgements

We thank Annick Bleys for proofreading and preparing the manuscript. This work was supported by the European Research Council under the European Community's Seventh Framework Programme [FP7/2007-2013] under ERC grant agreement no [339341-AMAIZE]11 and by Ghent University ('Bijzonder Onderzoeksfonds Methusalem project' no. BOF08/01M00408 and Multidisciplinary Research Partnership 'Biotechnology for a Sustainable Economy' Grant 01MR0510W).

Author contributions

J.V., K.V., N.G. and D.I. designed the research methodology, performed data analysis and wrote the article. M.V.B. assisted in computing orthologues. S.R.K. helped with analysis of the PWMs and C.O.-C. generated the heat maps. V.S. preprocessed the transcriptomic data sets. H.N. assisted in the biological interpretation of the maize genes.

Conflict of interest

The authors declare no conflict of interest.

References

Akiyama, K., Kurotani, A., Iida, K., Kuromori, T., Shinozaki, K. and Sakurai, T. (2013) RARGE II: an integrated phenotype database of Arabidopsis mutant traits using a controlled vocabulary. *Plant Cell Physiol.* **55**, e4–e4.

Andriankaja, M., Dhondt, S., De Bodt, S., Vanhaeren, H., Coppens, F., De Milde, L., Mühlenbock, P. *et al*. (2012) Exit from proliferation during leaf development in *Arabidopsis thaliana*: a not-so-gradual process. *Dev. Cell*, **22**, 64–78.

Avramova, V., AbdElgawad, H., Zhang, Z., Fotschki, B., Casadevall, R., Vergauwen, L., Knapen, D. *et al*. (2015) Drought induces distinct growth response, protection and recovery mechanisms in the maize leaf growth zone. *Plant Physiol.* **169**, 1382–1396.

Baerenfaller, K., Massonnet, C., Walsh, S., Baginsky, S., Bühlmann, P., Hennig, L., Hirsch-Hoffmann, M. *et al*. (2012) Systems-based analysis of Arabidopsis leaf growth reveals adaptation to water deficit. *Mol. Syst. Biol.* **8**, 606.

Baumbusch, L.O., Thorstensen, T., Krauss, V., Fischer, A., Naumann, K., Assalkhou, R., Schulz, I. *et al*. (2001) The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res.* **29**, 4319–4333.

Baute, J., Herman, D., Coppens, F., De Block, J., Slabbinck, B., Dell'Acqua, M., Pè, M.E. *et al*. (2016) Combined large-scale phenotyping and transcriptomics in maize reveals a robust growth regulatory network. *Plant Physiol.* **170**, 1848–1867.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B*, **57**, 289–300.

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E. and Huala, E. (2015) The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis*, **53**, 474–485.

Berckmans, B. and De Veylder, L. (2009) Transcriptional control of the cell cycle. *Curr. Opin. Plant Biol.* **12**, 599–605.

Berthet, S., Demont-Caulet, N., Pollet, B., Bidzinski, P., Cézard, L., Le Bris, P., Borrega, N. *et al*. (2011) Disruption of LACCASE4 and 17 results in tissue-specific alterations to lignification of *Arabidopsis thaliana* stems. *Plant Cell*, **23**, 1124–1137.

Brauner, K., Hörmiller, I., Nägele, T. and Heyer, A.G. (2014) Exaggerated root respiration accounts for growth retardation in a starchless mutant of *Arabidopsis thaliana*. *Plant J.* **79**, 82–91.

Carretero-Paulet, L., Galstyan, A., Roig-Villanova, I., Martínez-García, J.F., Bilbao-Castro, J.R. and Robertson, D.L. (2010) Genome-wide classification and evolutionary analysis of the bHLH family of transcription factors in Arabidopsis, poplar, rice, moss, and algae. *Plant Physiol.* **153**, 1398–1412.

Chaumont, F., Barrieu, F., Herman, E.M. and Chrispeels, M.J. (1998) Characterization of a maize tonoplast aquaporin expressed in zones of cell division and elongation. *Plant Physiol.* **117**, 1143–1152.

Chaw, S.-M., Chang, C.-C., Chen, H.-L. and Li, W.-H. (2004) Dating the monocot–dicot divergence and the origin of core eudicots using whole chloroplast genomes. *J. Mol. Evol.* **58**, 424–441.

Cheng, N.-H., Pittman, J.K., Shigaki, T., Lachmansingh, J., LeClere, S., Lahner, B., Salt, D.E. *et al*. (2005) Functional association of Arabidopsis CAX1 and CAX3 is required for normal growth and ion homeostasis. *Plant Physiol.* **138**, 2048–2060.

Chiang, H.-H., Hwang, I. and Goodman, H.M. (1995) Isolation of the Arabidopsis GA4 locus. *Plant Cell*, **7**, 195–201.

Cho, H.-T. and Cosgrove, D.J. (2000) Altered expression of expansin modulates leaf growth and pedicel abscission in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA*, **97**, 9783–9788.

Cosgrove, D.J. (2005) Growth of the plant cell wall. *Nat. Rev. Mol. Cell Biol.* **6**, 850–861.

Danisman, S., van der Wal, F., Dhondt, S., Waites, R., de Folter, S., Bimbo, A., van Dijk, A.D.J. *et al*. (2012) Arabidopsis class I and class II TCP transcription factors regulate jasmonic acid metabolism and leaf development antagonistically. *Plant Physiol.* **159**, 1511–1523.

Davies, P.J. (2004) *Plant Hormones: Biosynthesis, Signal Transduction, Action!*. Berlin, Germany: Springer Science & Business Media.

De Veylder, L., Beeckman, T., Beemster, G.T.S., de Almeida Engler, J., Ormenese, S., Maes, S., Naudts, M. *et al*. (2002) Control of proliferation, endoreduplication and differentiation by the *Arabidopsis* E2Fa-DPa transcription factor. *EMBO J.* **21**, 1360–1368.

Dennis, E.S., Sachs, M.M., Gerlach, W.L., Finnegan, E.J. and Peacock, W.J. (1985) Molecular analysis of the alcohol dehydrogenase 2 (Adh2) gene of maize. *Nucleic Acids Res.* **13**, 727–743.

Dewitte, W., Scofield, S., Alcasabas, A.A., Maughan, S.C., Menges, M., Braun, N., Collins, C. *et al*. (2007) Arabidopsis CYCD3 D-type cyclins link cell proliferation and endocycles and are rate-limiting for cytokinin responses. *Proc. Natl Acad. Sci. USA*, **104**, 14537–14542.

Donnelly, P.M., Bonetta, D., Tsukaya, H., Dengler, R.E. and Dengler, N.G. (1999) Cell cycling and cell enlargement in developing leaves of Arabidopsis. *Dev. Biol.* **215**, 407–419.

Dubois, M., Claeys, H., Van den Broeck, L. and Inzé, D. (2017) Time of day determines Arabidopsis transcriptome and growth dynamics under mild drought. *Plant Cell Environ.* **40**, 180–189.

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C. and Lepiniec, L. (2010) MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* **15**, 573–581.

Efroni, I., Blum, E., Goldshmidt, A. and Eshed, Y. (2008) A protracted and dynamic maturation schedule underlies Arabidopsis leaf development. *Plant Cell*, **20**, 2293–2306.

Elmayan, T., Proux, F. and Vaucheret, H. (2005) Arabidopsis RPA2: a genetic link among transcriptional gene silencing, DNA repair, and DNA replication. *Curr. Biol.* **15**, 1919–1925.

Feller, A., Machemer, K., Braun, E.L. and Grotewold, E. (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J.* **66**, 94–116.

Fleet, C.M., Yamaguchi, S., Hanada, A., Kawaide, H., David, C.J., Kamiya, Y. and Sun, T.-P. (2003) Overexpression of AtCPS and AtKS in Arabidopsis confers increased ent-kaurene production but no increase in bioactive gibberellins. *Plant Physiol.* **132**, 830–839.

Frith, M.C., Li, M.C. and Weng, Z. (2003) Cluster-Buster: finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res.* **31**, 3666–3668.

Gardiner, J., Sherr, I. and Scarpella, E. (2010) Expression of DOF genes identifies early stages of vascular development in Arabidopsis leaves. *Int. J. Dev. Biol.* **54**, 1389–1396.

Geilfus, C.-M., Ober, D., Eichacker, L.A., Mühling, K.H. and Zörb, C. (2015) Down-regulation of ZmEXPB6 (*Zea mays* β-expansin 6) protein is correlated with salt-mediated growth reduction in the leaves of *Z. mays* L. *J. Biol. Chem.* **290**, 11235–11245.

Gong, H., Jiao, Y., Hu, W.-W. and Pua, E.-C. (2005) Expression of glutathione-*S*-transferase and its role in plant growth and development *in vivo* and shoot morphogenesis *in vitro*. *Plant Mol. Biol.* **57**, 53–66.

Gonzalez, N., Vanhaeren, H. and Inzé, D. (2012) Leaf size control: complex coordination of cell division and expansion. *Trends Plant Sci.* **17**, 332–340.

Gu, Z., Eils, R. and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, **32**, 2847–2849.

Hefer, C.A., Mizrachi, E., Myburg, A.A., Douglas, C.J. and Mansfield, S.D. (2015) Comparative interrogation of the developing xylem transcriptomes of two wood-forming species: *Populus trichocarpa* and *Eucalyptus grandis*. *New Phytol.* **206**, 1391–1405.

Hepworth, J. and Lenhard, M. (2014) Regulation of plant lateral-organ growth by modulating cell number and size. *Curr. Opin. Plant Biol.* **17**, 36–42.

Hobbie, L. and Estelle, M. (1995) The axr4 auxin-resistant mutants of *Arabidopsis thaliana* define a gene important for root gravitropism and lateral root initiation. *Plant J.* **7**, 211–220.

Hu, J., Manduzio, S. and Kang, H. (2019) Epitranscriptomic RNA methylation in plant development and abiotic stress responses. *Front Plant Sci.* **10**, 500.

Huang, M., Friso, G., Nishimura, K., Qu, X., Olinares, P.D.B., Majeran, W., Sun, Q. *et al*. (2013) Construction of plastid reference proteomes for maize and arabidopsis and evaluation of their orthologous relationships; the concept of orthoproteomics. *J. Proteome Res.* **12**, 491–504.

Karidas, P., Challa, K.R. and Nath, U. (2015) The tarani mutation alters surface curvature in Arabidopsis leaves by perturbing the patterns of surface expansion and cell division. *J. Exp. Bot.* **66**, 2107–2122.

Kim, J.H. and Kende, H. (2004) A transcriptional coactivator, AtGIF1, is involved in regulating leaf growth and morphology in *Arabidopsis*. *Proc. Natl Acad. Sci. USA*, **101**, 13374–13379.

Kim, K.-C., Fan, B. and Chen, Z. (2006) Pathogen-induced Arabidopsis WRKY7 is a transcriptional repressor and enhances plant susceptibility to Pseudomonas syringae. *Plant Physiol.* **142**, 1180–1192.

Lee, B.H., Ko, J.-H., Lee, S., Lee, Y., Pak, J.-H. and Kim, J.H. (2009) The Arabidopsis *GRF-INTERACTING FACTOR* gene family performs an overlapping function in determining organ size as well as multiple developmental properties. *Plant Physiol.* **151**, 655–668.

Li, L., Stoeckert, C.J. Jr. and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189.

Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., Kebrom, T.H. *et al*. (2010) The developmental dynamics of the maize leaf transcriptome. *Nat. Genet.* **42**, 1060–1067.

Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R.J., Franklin, L.D., He, J. *et al*. (2010) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J.* **63**, 86–99.

Linard, B., Thompson, J.D., Poch, O. and Lecompte, O. (2011) OrthoInspector: comprehensive orthology analysis and visual exploration. *BMC Bioinform.* **12**, 11.

Lintala, M., Allahverdiyeva, Y., Kidron, H., Piippo, M., Battchikova, N., Suorsa, M., Rintamäki, E. *et al*. (2007) Structural and functional characterization of ferredoxin-NADP + -oxidoreductase using knock-out mutants of Arabidopsis. *Plant J.* **49**, 1041–1052.

Lisch, D..(2013) How important are transposons for plant evolution? *Nat. Rev. Genet.* **14**, 49–61.

Liu, D., Wu, Z. and Hou, L. (2015) Loss-of-function mutation in SCY1 triggers chloroplast-to-nucleus retrograde signaling in *Arabidopsis thaliana*. *Biol. Plant.* **59**, 469–476.

Ludevid, D., Höfte, H., Himelblau, E. and Chrispeels, M.J. (1992) The expression pattern of the tonoplast intrinsic protein γ-TIP in *Arabidopsis thaliana* is correlated with cell enlargement. *Plant Physiol.* **100**, 1633–1639.

Mattioli, R., Marchese, D., D'Angeli, S., Altamura, M.M., Costantino, P. and Trovato, M. (2008) Modulation of intracellular proline levels affects flowering time and inflorescence architecture in Arabidopsis. *Plant Mol. Biol.* **66**, 277–288.

Minic, Z. and Jouanin, L. (2006) Plant glycoside hydrolases involved in cell wall polysaccharide degradation. *Plant Physiol. Biochem.* **44**, 435–449.

Mitchum, M.G., Yamaguchi, S., Hanada, A., Kuwahara, A., Yoshioka, Y., Kato, T., Tabata, S. *et al*. (2006) Distinct and overlapping roles of two gibberellin 3-oxidases in Arabidopsis development. *Plant J.* **45**, 804–818.

Movahedi, S., Van de Peer, Y. and Vandepoele, K. (2011) Comparative network analysis reveals that tissue specificity and gene function are important factors influencing the mode of expression evolution in Arabidopsis and rice. *Plant Physiol.* **156**, 1316–1330.

Movahedi, S., Van Bel, M., Heyndrickx, K.S. and Vandepoele, K. (2012) Comparative co-expression analysis in plant biology. *Plant Cell Environ.* **35**, 1787–1798.

Mustroph, A., Lee, S.C., Oosumi, T., Zanetti, M.E., Yang, H., Ma, K., Yaghoubi-Masihi, A. *et al*. (2010) Cross-kingdom comparison of transcriptomic adjustments to low-oxygen stress highlights conserved and plant-specific responses. *Plant Physiol.* **152**, 1484–1500.

Naouar, N., Vandepoele, K., Lammens, T., Casneuf, T., Zeller, G., Van Hummelen, P., Weigel, D. *et al*. (2009) Quantitative RNA expression analysis with Affymetrix Tiling 1.0R arrays identifies new E2F target genes. *Plant J.* **57**, 184–194.

Narsai, R., Rocha, M., Geigenberger, P., Whelan, J. and van Dongen, J.T. (2011) Comparative analysis between plant species of transcriptional and metabolic responses to hypoxia. *New Phytol.* **190**, 472–487.

Nelissen, H., Rymen, B., Jikumaru, Y., Demuynck, K., Van Lijsebettens, M., Kamiya, Y., Inzé, D. *et al*. (2012) A local maximum in gibberellin levels regulates maize leaf growth by spatial control of cell division. *Curr. Biol.* **22**, 1183–1187.

Nelissen, H., Eeckhout, D., Demuynck, K., Persiau, G., Walton, A., van Bel, M., Vervoort, M. *et al*. (2015) Dynamic changes in ANGUSTIFOLIA3 complex composition reveal a growth regulatory mechanism in the maize leaf. *Plant Cell.* **27**, 1605–1619.

Nelissen, H., Gonzalez, N. and Inzé, D. (2016) Leaf growth in dicots and monocots: so different yet so alike. *Curr. Opin. Plant Biol.* **33**, 72–76.

Nelissen, H., Sun, X.H., Rymen, B., Jikumaru, Y., Kojima, M., Takebayashi, Y., Abbeloos, R. *et al*. (2018) The reduction in maize leaf growth under mild drought affects the transition between cell division and cell expansion and cannot be restored by elevated gibberellic acid levels. *Plant Biotechnol. J.* **16**, 615–627.

Nelson, T. and Dengler, N. (1997) Leaf vascular pattern formation. *Plant Cell*, **9**, 1121–1135.

Nicolas, M. and Cubas, P. (2016) The role of TCP transcription factors in shaping flower structure, leaf morphology, and plant architecture. In *Plant Transcription Factors*, pp. 249–267. Cambridge, Massachusetts: Elsevier.

Ohkama-Ohtsu, N., Radwan, S., Peterson, A., Zhao, P., Badr, A.F., Xiang, C. and Oliver, D.J. (2007) Characterization of the extracellular γ-glutamyl transpeptidases, GGT1 and GGT2, in Arabidopsis. *Plant J.* **49**, 865–877.

Patel, R.V., Nahal, H.K., Breit, R. and Provart, N.J. (2012) BAR expressolog identification: expression profile similarity ranking of homologous genes in plant species. *Plant J.* **71**, 1038–1050.

Peng, L., Ma, J., Chi, W., Guo, J., Zhu, S., Lu, Q., Lu, C. *et al*. (2006) LOW PSII ACCUMULATION1 is involved in efficient assembly of photosystem II in *Arabidopsis thaliana*. *Plant Cell*, **18**, 955–969.

Proost, S., Van Bel, M., Vaneechoutte, D., Van de Peer, Y., Inzé, D., Mueller-Roeber, B. and Vandepoele, K. (2015) PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res.* **43**, D974–D981.

Pyke, K.A., Marrison, J.L. and Leech, A.M. (1991) Temporal and spatial development of the cells of the expanding first leaf of *Arabidopsis thaliana* (L.) Heynh. *J. Exp. Bot.* **42**, 1407–1416.

R Core Team. (2017) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: Foundation for Statistical Computing (http://www.R-project.org/).

Raynaud, C., Sozzani, R., Glab, N., Domenichini, S., Perennes, C., Cella, R., Kondorosi, E. *et al*. (2006) Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in Arabidopsis. *Plant J.* **47**, 395–407.

Rensink, W.A., Lee, Y., Liu, J., Iobst, S., Ouyang, S. and Buell, C.R. (2005) Comparative analyses of six solanaceous transcriptomes reveal a high degree of sequence conservation and species-specific transcripts. *BMC Genom.* **6**, 124.

Sánchez-Rodríguez, C., Bauer, S., Hématy, K., Saxe, F., Ibáñez, A.B., Vodermaier, V., Konlechner, C. *et al*. (2012) Chitinase-like1/pom-pom1 and its homolog CTL2 are glucan-interacting proteins important for cellulose biosynthesis in Arabidopsis. *Plant Cell*, **24**, 589–607.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N. *et al*. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.

Šimková, K., Moreau, F., Pawlak, P., Vriet, C., Baruah, A., Alexandre, C., Hennig, L. *et al*. (2012) Integration of stress-related and reactive oxygen species-mediated signals by Topoisomerase VI in *Arabidopsis thaliana*. *PNAS*, **109**, 16360–16365.

Skirycz, A., Claeys, H., De Bodt, S., Oikawa, A., Shinoda, S., Andriankaja, M., Maleux, K. *et al*. (2011) Pause-and-stop: the effects of osmotic stress on cell proliferation during early leaf development in *Arabidopsis* and a role for ethylene signaling in cell cycle arrest. *Plant Cell*, **23**, 1876–1888.

Spartz, A.K., Ren, H., Park, M.Y., Grandt, K.N., Lee, S.H., Murphy, A.S., Sussman, M.R. *et al*. (2014) SAUR inhibition of PP2C-D phosphatases activates plasma membrane H^+-ATPases to promote cell expansion in *Arabidopsis*. *Plant Cell*, **26**, 2129–2142.

Stahle, M.I., Kuehlich, J., Staron, L., Arnim, A.G.V. and Golz, J.F. (2009) YABBYs and the transcriptional corepressors LEUNIG and LEUNIG_HOMOLOG maintain leaf polarity and meristem activity in Arabidopsis. *Plant Cell*, **21**, 3105–3118.

Storozhenko, S., Belles-Boix, E., Babiychuk, E., Hérouart, D., Davey, M.W., Slooten, L., Van Montagu, M. *et al*. (2002) γ-Glutamyl transpeptidase in transgenic tobacco plants. Cellular localization, processing, and biochemical properties. *Plant Physiol.* **128**, 1109–1119.

Ueda, M., Matsui, K., Ishiguro, S., Sano, R., Wada, T., Paponov, I., Palme, K. *et al*. (2004) The HALTED ROOT gene encoding the 26S proteasome subunit RPT2a is essential for the maintenance of Arabidopsis meristems. *Development*, **131**, 2101–2111.

Uematsu, K., Suzuki, N., Iwamae, T., Inui, M. and Yukawa, H. (2012) Increased fructose 1, 6-bisphosphate aldolase in plastids enhances growth and photosynthesis of tobacco plants. *J. Exp. Bot.* **63**, 3001–3009.

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 Dingenen, J., Blomme, J., Gonzalez, N. and Inzé, D. (2016) Plants grow with a little help from their organelle friends. *J. Exp. Bot.* **67**, 6267–6281.

Vandepoele, K., Vlieghe, K., Florquin, K., Hennig, L., Beemster, G.T.S., Gruissem, W., Van de Peer, Y. *et al*. (2005) Genome-wide identification of potential plant E2F target genes. *Plant Physiol.* **139**, 316–328.

Vercruyssen, L., Verkest, A., Gonzalez, N., Heyndrickx, K.S., Eeckhout, D., Han, S.-K., Jégu, T. *et al*. (2014) ANGUSTIFOLIA3 binds to SWI/SNF chromatin remodeling complexes to regulate transcription during *Arabidopsis* leaf development. *Plant Cell*, **26**, 210–229.

Verkest, A., Abeel, T., Heyndrickx, K.S., Van Leene, J., Lanz, C., Van De Slijke, E., De Winne, N. *et al*. (2014) A generic tool for transcription factor target gene discovery in Arabidopsis cell suspension cultures based on tandem chromatin affinity purification. *Plant Physiol.* **164**, 1122–1133.

Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S. *et al*. (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*, **158**, 1431–1443.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed. Switzerland: Springer.

Wittstock, U. and Burow, M. (2010) Glucosinolate breakdown in Arabidopsis: mechanism, regulation and biological significance. *The Arabidopsis book/American Society of Plant Biologists*, **8**, e0134.

Xia, C., Wang, Y.J., Li, W.Q., Chen, Y.R., Deng, Y., Zhang, X.Q., Chen, L.Q. *et al*. (2010) The Arabidopsis eukaryotic translation initiation factor 3, subunit F (AteIF3f), is required for pollen germination and embryogenesis. *Plant J.* **63**, 189–202.

Xiao, J., Li, J., Ouyang, M., Yun, T., He, B., Ji, D., Ma, J. *et al*. (2012) DAC is involved in the accumulation of the cytochrome b6/f complex in Arabidopsis. *Plant Physiol.* **160**, 1911–1922.

Yoshizawa, E., Kaizuka, M., Yamagami, A., Higuchi-Takeuchi, M., Matsui, M., Kakei, Y., Shimada, Y. *et al*. (2014) BPG3 is a novel chloroplast protein that involves the greening of leaves and related to brassinosteroid signaling. *Biosci. Biotechnol. Biochem.* **78**, 420–429.

Yu, C.-P., Chen, S.C.-C., Chang, Y.-M., Liu, W.-Y., Lin, H.-H., Lin, J.-J., Chen, H.J. *et al*. (2015) Transcriptome dynamics of developing maize leaves and genomewide prediction of cis elements and their cognate transcription factors. *PNAS*, **112**, E2477–E2486.

Zhang, D.-F., Li, B., Jia, G.-Q., Zhang, T.-F., Dai, J.-R., Li, J.-S. and Wang, S.-C. (2008) Isolation and characterization of genes encoding GRF transcription factors and GIF transcriptional coactivators in Maize (*Zea mays* L.). *Plant Sci.* **175**, 809–817.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Visual representation of the two data sets and overview of the hierarchical clustering of the Arabidopsis and maize DEGs.

Figure S2 Heat map showing the top 10 enriched GO terms (BP and MF) per cluster with enrichment in at least one cluster.

Figure S3 Visual representation of the Arabidopsis class I (left) and class II (right) TCP gene family with the maize orthologues.

Figure S4 Schematic overview of the identified putative E2F targets benchmarking.

Table S1 Overview of the 217 one-to-one orthology groups with conserved expression and part of one of the significant overlaps (A1-Z2, A5-Z5, A5-Z6, A6-Z5, A6-Z6)

Table S2 Calculation of percentage, enrichment, and FDR of leaf-related phenotype data for the groups with conserved or divergent expression, and the Arabidopsis group with no orthologues of clusters A1, A5, and A6

Table S3 Group, ID, name, TF information, and phenotype description resulting from a knockout (KO) or an overexpression (OE) described in TAIR and or RARGE and confirmed by a publication for 49 genes, 10 genes, and 48 genes possibly

involved in a decrease or increase in leaf size or another leaf-related phenotype, respectively

Table S4 Overview of the 59 one-to-two orthology groups exemplifying possible neo -or subfunctionalization of the maize orthologue having divergent expression

Table S5 Overview of 211 DE Arabidopsis TFs and their 256 DE maize orthologues either with conserved or divergent expression throughout leaf development

Table S6 Overview of the DE TF families including Arabidopsis genes with their maize orthologues that are conserved, partially conserved, or divergent in expression throughout leaf development

Table S7 Overview of the identified E2F targets with conserved expression in Arabidopsis (12 genes) and maize (16 genes)

Table S8 Overview of the Arabidopsis and maize clusters in which TCP3 and TCP24 TF binding sites are enriched based on the hypergeometric score (HGS)

Supplemental network file with the full integrated network as well as the case-studies handled in this manuscript.

