Robustness of organ morphology is associated with modules of co-expressed genes related to plant cell wall

Diego A. Hartasánchez[1,2,#], Annamaria Kiss[1], Virginie Battu[1], Mathilde Dumond[1], Charline Soraru[1], Abigail Delgado-Vaquera[1], Florian Massinon[1], Marina Brasó-Vives[3,4], Corentin Mollier[1], Nelly Dubrulle[1], Fabien Sénéchal[5], Marie-Laure Martin-Magniette[6,7], Arezki Boudaoud[1,8,*,#], Françoise Monéger[1,*,#]

1 Laboratoire Reproduction et Développement des Plantes, Université de Lyon, ENS de Lyon, CNRS, INRAE, UCBL, Lyon, France
2 Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
3 Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
4 Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland
5 UMRT INRAE 1158 BioEcoAgro - BIOPI Biologie des Plantes et Innovation, Université de Picardie, 80039 Amiens, France
6 Université Paris-Saclay, CNRS, INRAE, Institute of Plant Sciences Paris-Saclay (IPS2), 91405 Orsay, France
7 Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA-Paris, 75005 Paris, France
8 LadHyX, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau Cedex 91128 France

* co-last
# corresponding

Abstract

Reproducibility in organ size and shape is a fundamental trait of living organisms. The mechanisms underlying such robustness remain, however, to be elucidated. Taking the sepal of Arabidopsis as a model, we investigated whether variability of gene expression plays a role in variation of organ morphology. To address this question, we produced a dataset composed of both transcriptomic and morphological information obtained from 27 individual sepals from wild-type plants. Although nearly identical in their genetic background, environment, and developmental stage, these sepals exhibited appreciable variability in both morphology and transcriptome. We identified modules of co-expressed genes in sepals, three of which correlated significantly with morphology, revealing biologically relevant gene regulatory networks. Interestingly, cell-wall related genes were overrepresented in two of these three modules. Additionally, we found that highly variable genes were unexpectedly enriched in cell-wall related processes. We then analyzed sepal morphology from 16 cell-wall mutants and found that the more a gene is expressed in wild-type, the more variable the morphology of the corresponding mutant. Altogether, our work unravels the contribution of cell-wall related genes to the robustness of sepal morphology. More generally, we propose that canalizing traits during development could rely on the modulation of highly expressed genes.

Introduction

Many organisms have a remarkably constant size and shape, yet, at the cellular level, cell growth and shape can be highly variable. How does reproducible (robust) organ morphology emerge from stochastic cell behavior and what is the impact of molecular variability in this process? Waddington and Schmalhausen brought this important question to prominence in the 1940s referring to "canalization" as the process through which stabilizing selection would tend to select the norm over deviations from it and hence ensure the stability of morphogenetic systems against environmental factors or mutations (Schmalhausen, 1949; Waddington, 1953, 1959). Perhaps in part due to the elusiveness of the concept of canalization (Gibson & Wagner, 2000), this term has been substituted to a large extent by "robustness", which is the invariant expression of phenotype in the face of environmental and/or genetic perturbations (Félix & Barkoulas, 2015). In plants, molecular mechanisms have been found to modulate morphogenetic robustness to environmental perturbations, as in the case of heat-shock proteins (Queitsch et al., 2002), and also to genetic changes such as whole genome duplications (Lachowiec et al., 2016). Robustness, however, also refers to developmental stability despite systemic internal noise (Alvarez-Buylla et al., 2008). Indeed, gene expression is a stochastic (random) process attributed to a combination of external and internal noise (Elowitz et al. 2002). In a multicellular context, such as that of Arabidopsis, gene expression appears to be extremely variable in time and space (Joseph et al., 2015; Araújo et al., 2017; Meyer et al., 2017). In fact, even when measuring variability in gene expression at the whole-organism level, there are some genes that exhibit very high variability between individuals. This variability itself has been observed to differ between day and night, for example, in Arabidopsis seedlings (Cortijo et al. 2019) and between developmental stages in *C. elegans* (Zalts & Yanai, 2017) and Drosophila (Liu et al., 2020). Development is, then, not only robust to gene expression variability, but also, quite likely dependent on it.

Here, we focus on how variability in gene expression at the organ level is associated with organ morphology. A classical system for the study of robustness of organ morphology has been the Drosophila wing (Debat et al., 2011), but other systems have also been used, for example, cichlid fish to study bilateral symmetry (Klingenberg et al., 2007). Recently, the Arabidopsis sepal has emerged as a good system to study organ morphology for several reasons (Roeder, 2021). Arabidopsis has a large number of flowers on each inflorescence and sepals are very accessible for dissection, imaging and experimentation. The sepals of Arabidopsis exhibit very reproducible morphology. Hong et al. (2016) found that spatiotemporal averaging of cell growth rates across development is fundamental for robust sepal morphology in Arabidopsis, which depends on the expression of *FTSH4*. Similarly, the expression of *DRMY1* is essential for both

timing of organ initiation and adequate sepal growth (Zhu et al., 2020). Despite this progress, our comprehension of mechanisms underlying robustness is far from complete.

The association between genotype and phenotypic robustness can be approached in many different ways. A routinely used approach is based on the characterization of the effects of a mutation on the phenotype of the organism. However, its main drawback is that because genes are part of complex regulatory networks, it may be hard to draw reliable information from the phenotype of one- or two-gene mutants, especially when studying complex traits (Chen et al., 2018). Genome-wide association studies are powerful (Wu et al., 2016) but require large sample sizes in order to achieve adequate statistical power (Lee & Lee, 2021) and are susceptible to false associations (Sasaki et al., 2021). Searching for quantitative-trait loci associated with robustness has identified the gene *ERECTA*, which is believed to transduce microenvironmental variation into phenotypic differentiation (Hall et al., 2007). Regarding the choice of scale, single-cell gene expression data has enabled the characterization of complex organs with multiple different cell-types such as the root (Dorrity et al., 2021) and ovules (Hou et al., 2021) in Arabidopsis. Complementarily, transcriptomic data at the organismal level, such as that obtained from whole seedlings, has provided very valuable insight regarding variability in gene expression (Cortijo et al., 2019) highlighting the role of modules of co-expressed genes (Cortijo et al., 2020). Both of these approaches, however, have their shortcomings. On the one hand, single-cell RNA sequencing has limited power to detect lowly expressed genes. On the other hand, transcriptomes obtained from whole individuals combine different structures, tissues and cell-types. Here, we have chosen an intermediate scale with a whole organ, the sepal, which has relatively few cell-types, and have performed high coverage RNA sequencing to obtain high quality transcriptomic data at the organ level, aiming at identifying modules of co-expressed genes that could be linked to phenotypic robustness. To identify these modules in wild-type plants, we have evaluated wild-type variation in sepals by obtaining gene expression data and morphology from 27 sepals in environmentally controlled conditions and with identical genetic background. Taking advantage of molecular and phenotypic variation among a high number of wild-type sepals, we extracted relevant biological information supporting the role of cell-wall related genes in morphological robustness.

Results

A unique dataset to analyze the association between gene expression and morphology

Our goal was to analyze the variability of both the morphology and the transcriptome of individual sepals from wild type plants grown in standard conditions. On a single inflorescence (flowering stem) of Arabidopsis, flower buds initiate at regular time intervals in a continuous manner. Each flower bud contains four sepals that are formed in a defined order, the first one to emerge being called the abaxial sepal (it forms on the flower-bud side furthest away from the meristem). Previous studies (Hervieux et al., 2016; Hong et al., 2016) have shown that arrest of sepal growth begins at stage 11 (according to staging of Arabidopsis flower development described in Smyth et al., 1990). We therefore chose early stage 11 to collect sepals and to minimize stage heterogeneity. Indeed, early stage 11 is rather transient, as we rarely find two flowers at this stage in a given inflorescence. We then collected 30 individual abaxial sepals, each from 30 secondary inflorescences, taken from three different Col-0 wild-type plants, labeled D, E and F (plants labeled A, B and C had been used initially for a pilot experiment to set up the protocol), grown simultaneously in experimentally controlled standard conditions, and used them to recover their 3D shape as well as their RNA to perform RNA-Seq analysis (Figure 1A). Each sepal was imaged under a confocal microscope using autofluorescence. Immediately following imaging, the sepal was frozen in liquid nitrogen for RNA extraction. Sepal images were analyzed with in-house python scripts and using MorphographX (Barbier de Reuille et al, 2015, see Materials and Methods for details).

Sepal morphology was characterized by measuring curvilinear length (Length), curvilinear width (Width), surface area (Area), longitudinal curvature (LongCurv) and transversal curvature (TransCurv) (Figure 1B). From these measurements, we additionally calculated the aspect ratio (AspRatio) by dividing Length over Width, and the Gaussian curvature (GaussCurv) by multiplying LongCurv by TransCurv. The parameters that we have used here to describe sepal morphology were selected after considering several parameters and carefully evaluating their reproducibility, in particular by measuring the same sepal several times (see Materials and Methods). Other parameters that had originally been measured, such as curvature at the borders, were excluded because they were not reliable. For this reason, curvature was measured from the central zone of the sepal (Figure 1B). From our transcriptome pre-analysis (described in the Materials and Methods' section Transcriptome Normalization and Filtering) we detected three outliers (all from one plant, plant F) which were not further analyzed. All results shown hereafter are for 27 sepal samples.

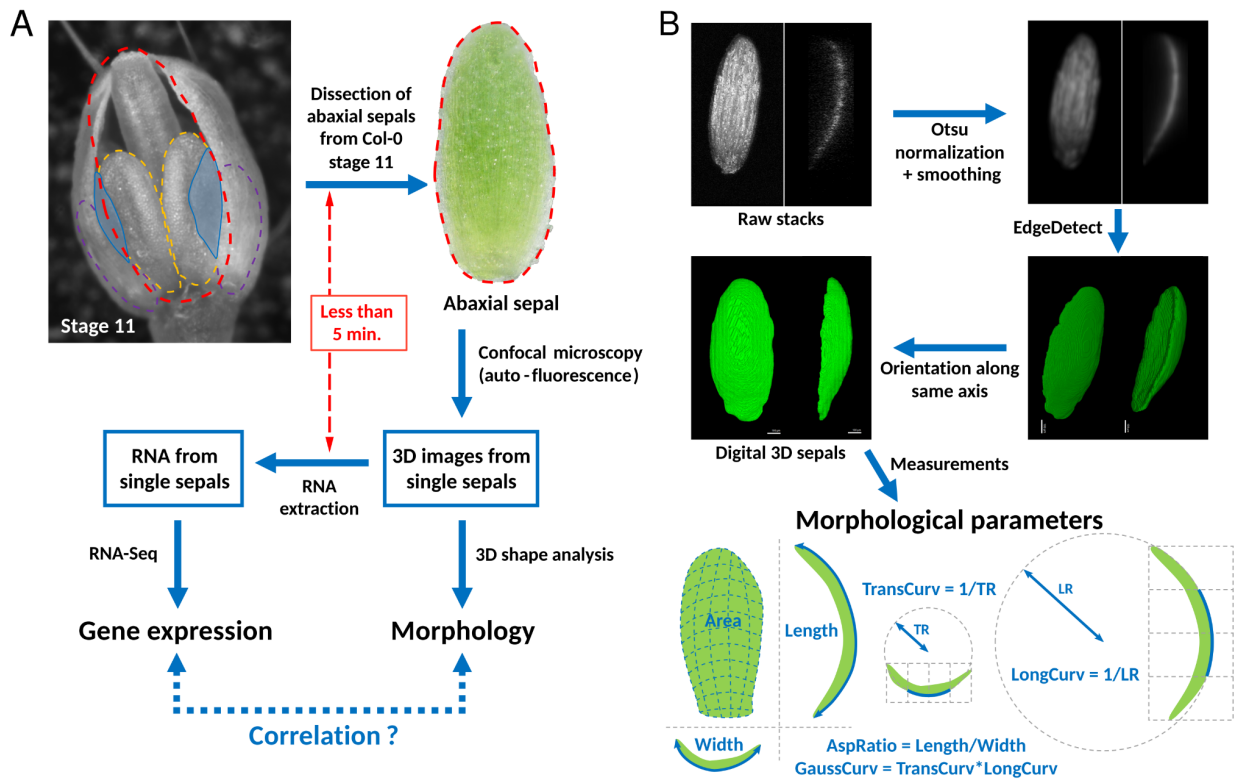

Figure 1: Protocol used for sample collection, data generation and pipeline of image analysis.

(A) Flower buds at early stage 11 were identified as such when petals (in blue) were longer than the lateral stamen (outlined by purple dashed lines, not visible on the image) but shorter than the median stamen (yellow dashed lines). The abaxial sepal was dissected and put on agar to avoid dehydration. Autofluorescence enabled imaging of the entire sepal using a confocal microscope with the lowest possible laser power and resolution to minimize laser exposure. Within less than 5 minutes, the sepal was frozen in liquid nitrogen. After grinding, RNA was extracted from individual sepals and used for RNA-Seq.

(B) In order to retrieve precise morphological parameters for each sepal sample, raw images obtained from the confocal microscope were processed. Otsu normalization and smoothing were performed on the images, followed by applying the EdgeDetect function in MorphographX (Barbier de Reuille et al, 2015) and fixing sepal orientation. From the digital 3D images, morphological parameters were extracted, namely, length, width, aspect ratio, area, and transversal, longitudinal and Gaussian curvatures (see Materials and methods for details). Transversal (longitudinal) curvature was measured from the central zone of the sepal by dividing the total width (length) of the sepal in four equal sections as shown in the figure, then, by fitting the circumference that best matched the sepal curvature in the middle two sections and taking the inverse of the radius of said circumference.

Variability in wild-type sepal morphology is mostly explained by variation in area and aspect ratio

We analyzed the seven morphological parameters abovementioned (Length, Width, AspRatio, Area, LongCurv, TransCurv and GaussCurv) within plants and across our 27 sepals (Figure 2A). Analysis of correlations between each pair of parameters shows that Area is the most informative parameter since it is strongly positively correlated with Length and Width (as expected, even though it is an independent measurement) and negatively correlated with all three curvature measurements, particularly with GaussCurv. These correlations are statistically significant when considering all 27 samples, and also when evaluating plants independently, despite the expected smaller statistical significance found for plant F due to its smaller sample size (from exclusion of the three outlier sepals according to the RNA-Seq data). In general, anticorrelations observed between size and curvature measurements indicate that larger sepals tend to be less curved in their central region.

Results from a principal component analysis (PCA) on morphological measurements (Figures 2B & 2C) confirm that despite identical or nearly identical genetic backgrounds, environment and developmental stages between sepal samples, we do observe some level of variability among our samples. We do not observe any clustering by plant and variability within individual plants is comparable across them (again, slightly lower for plant F). The first principal component accounts for 56.5% of the variance and separates Length, Width and Area from LongCurv, TransCurv and GaussCurv, in agreement with correlations observed in Figure 2A. The second principal component, explaining 22.3% of the variance, is mostly dominated by AspRatio, which has no correlation with Area or curvature parameters. From these observations we can conclude that area and aspect ratio are complementary measurements that describe sepal shape and size to a large extent. The third principal component explains 14.1% of the variance and groups all parameters, including Area and curvature parameters, in the same direction with the strongest contribution coming from Length (Figure 2 - Figure supplement 1).

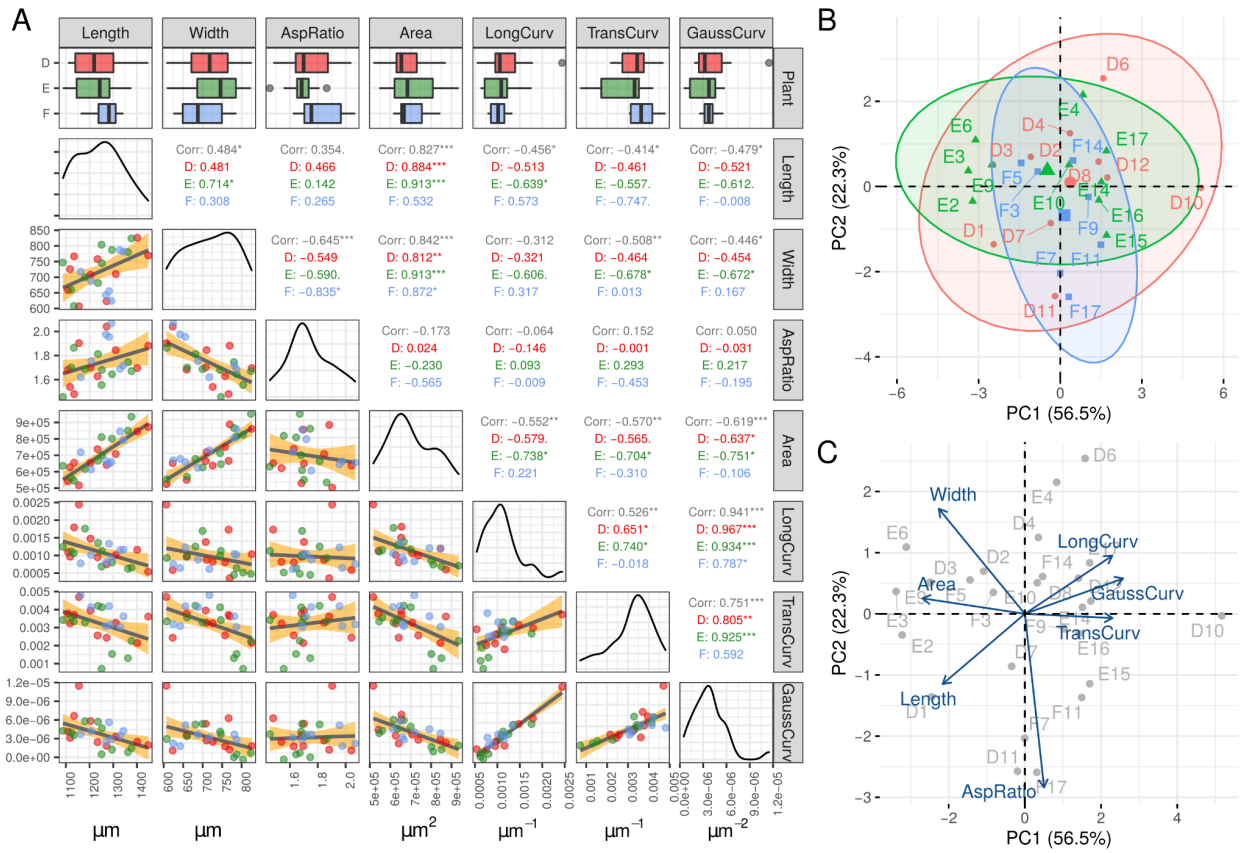

Figure 2. Pairwise comparison and principal component analysis of morphological parameters.

(A) Comparison between morphological parameters. First row shows boxplots for each parameter and each plant (D in red, E in green, F in blue), with medians represented with a black line. Colored boxes extend from the first to the third quartile, while whiskers extend a further 1.5 times the interquartile range. Scatter plots for each pairwise comparison show each of the samples with its corresponding color. The gray line with yellow shading corresponds to a linear model adjustment. Pearson correlation coefficients for each pair are shown for all samples pulled together (in gray) and for each plant independently (in green, red and blue) with corresponding p-values shown with asterisks (*** implies p-value < 0.001, ** implies p-value < 0.01, * implies p-value < 0.05, and . implies p-value < 0.1). The diagonal shows the probability density plot for each parameter.

(B) Principal component analysis (PCA) for all samples according to their morphological parameters, showing the percentage of variance explained by the first (PC1) and second (PC2) principal components. Ellipses show 95% confidence ellipses for each set of 2D normally distributed samples.

(C) Loadings for the PCA shown in B.

Analysis of gene co-expression reveals functional modules associated with morphology

Having analyzed the morphological variability present in wild-type sepals under the same controlled growth conditions, we investigated if this variability across samples could be associated with differences in gene expression that we could detect in our RNA-Seq data. We first evaluated if the variability in morphology across samples is comparable to variability in gene expression. We thus calculated the squared coefficient of variation (CV^2) for each morphological parameter and for each gene from the RNA-Seq data. CV^2 (the variance divided by the square of the mean) is a dimensionless quantity that evaluates variability and that allows for comparison between measurements of different nature. Regarding morphological parameters, CV^2 values for Length, Width and AspRatio are lower than CV^2 values for curvature parameters, with CV^2 for area being between those two groups (Figure 3A). Although there can potentially be intrinsic differences between size and curvature parameters, a possible interpretation of this difference is that the size of the sepal is much more constrained than its curvature. When comparing CV^2 between morphology and gene expression, as depicted in Figure 3A, we find that Length, Width and AspRatio exhibit less variation (as measured by CV^2) than gene expression. This observation could be a consequence of buffering of noise in gene expression when the latter is translated to phenotype (morphology, in this case).

Since robustness in morphology is thought to result from gene co-regulation (Lachowiec et al. 2016), we searched for modules of co-expressed genes whose expression correlated with morphological parameters. We used the Weighted Correlation Network Analysis (WGCNA) package (Langfelder & Horvath, 2008) to identify modules of co-expressed genes among our 27 samples. We then evaluated whether some of these modules were significantly associated with morphology (see Methods, Modules of co-expressed genes). Clustering samples by their gene expression profiles reveals that sepals from the same plant do not have a more similar gene expression profile among them compared to sepals from other plants (Figure 3B and Figure 3 - Figure supplement 1 (PCA)), indicating that the 27 sepals can be considered as independent samples. Tree reconstruction based on gene expression does not yield a clear pattern when normalized morphological parameters are plotted against it (Figure 3B), suggesting a complex link between gene expression and sepal morphology.

Clustering genes in modules according to their expression patterns across samples and searching for correlations between module eigengene expression (theoretical gene expression profile that is representative of the module) and morphological parameters can allow us to detect gene regulatory networks linked to morphology. We identified 16 modules of co-expressed genes of varying size, ranging from 63 to 2919 genes (Figure 3C, right panel). The

list of genes expressed in sepals along with the module each gene belongs to is provided in Supplementary Table 1. We found 3 modules of co-expressed genes whose eigengene correlated with a significant p-value ($p < 0.05$) and a correlation absolute value above 0.45 with at least one of our seven morphological parameters (Figure 3C, left panel). The magenta module has the strongest correlation value (with Width), followed by the cyan module (with Width) and the orange module (with AspRatio). A visual representation of correlations between modules and morphological parameters is shown in Figure 4 - Figure supplement 1. Normalized gene expression values for genes with high module membership and high trait significance (in absolute values) in the magenta, orange and cyan modules, correlate well with 1/Width, Width and AspRatio, respectively. Genes with low module membership and trait significance do not correlate well among each other or with morphological parameters. Additional modules with relatively high correlation values, such as the green and khaki modules correlate most strongly with Length and GaussCurv, respectively.

We performed gene ontology (GO) enrichment analysis (see Materials and Methods) on all modules (Supplementary Table 2) using the 14,085 gene set as a background, and focusing on the top 3 correlated modules. The magenta module showed an enrichment in cell wall organization or biogenesis and other GO subcategories, such as xylan, cell wall polysaccharide, cell wall macromolecule and hemicellulose metabolic processes, and also secondary cell wall biogenesis, among other terms. The orange module also showed GO enrichment in cell wall organization or biogenesis, but no further GO terms surpassed the FDR < 0.05 threshold with the 14,085 gene set background. To have slightly more power to detect relevant GO terms in the orange module, we performed GO enrichment analysis with the whole genome of Arabidopsis as background (results for the top 3 modules are shown in Supplementary Table 3). Doing so, we found that supramolecular fiber organization, cell morphogenesis, and polysaccharide and pectin biosynthetic processes are enriched in the orange module. The cyan module showed an enrichment in thylakoid membrane organization, chloroplast rRNA processing and peptide biosynthesis, as well as translation. In addition, the green module was enriched in photosynthesis-related GO terms and the khaki module, in response to oxygen levels.

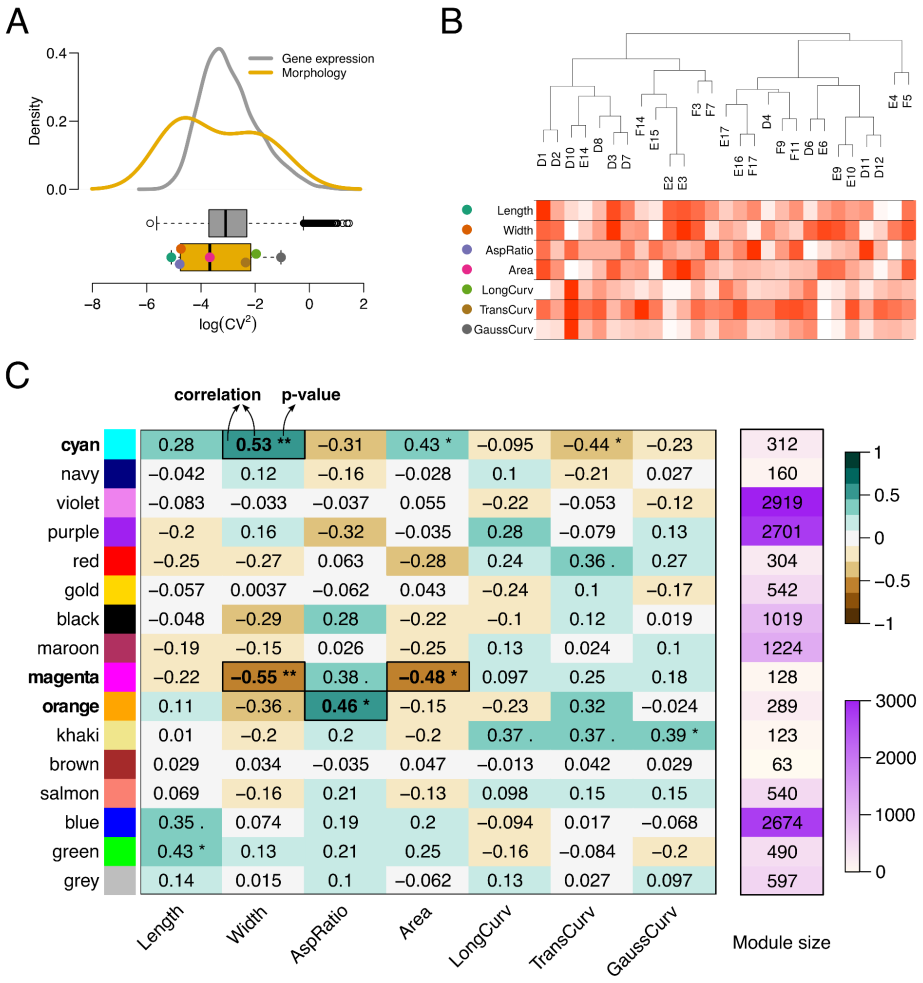

Figure 3: Modules of co-expressed genes associated with morphological parameters

(A) Density profile and boxplots of the squared coefficient of variation (log scale) of gene expression for each of the 14085 genes (gray) and for seven morphological parameters (yellow). Points over the upper box plot correspond to outliers. Points over the lower boxplot show the location of CV^2 of each parameter (color legend in B).

(B) Sample dendrogram from gene expression data reconstructed with WGCNA (Langfelder & Horvath, 2008) and relative magnitude of each morphological parameter measured for each sample (normalized across samples, with intense red depicting the maximum value and white, the minimum value).

(C) Module-trait associations obtained with WGCNA (Langfelder & Horvath, 2008). Modules are randomly assigned a color and are of different sizes, shown on the right panel on a floralwhite-to-violet color scale. Pearson correlation values between module eigengene expression and morphological parameters (left panel) are shown in black numbers and in a brown-to-green color scale with corresponding p-values shown with asterisks (** implies p-value < 0.01, * implies p-value < 0.05, and . implies p-value < 0.1). Module-trait pairs with correlation absolute values above 0.45 and p-values below 0.05 are highlighted with black rectangles with the corresponding module names in bold.

In order to ascertain the biological relevance of the reconstructed modules, we evaluated if genes associated with a particular biological function were co-expressed across our samples. As can be observed in Figures 4A & 4B, gene expression values of cell-wall related genes (associated with GO term “cell wall organization or biogenesis”) are particularly well correlated between each other within the magenta module. Highlighting those transcription factors within the module which are also cell-wall related, we find *KNAT7* and *NAC007,* which are known to be involved in secondary cell-wall deposition (Wang 2020). *KNAT7* and *NAC007* have indeed a high module membership value and a strong negative correlation with Width (Figure 4B and Figure 4 - Figure supplement 1C). We then reconstructed the gene regulatory network within this module using GENIE3 (Huynh-Thu et al., 2010). Figure 4C shows the gene regulatory network corresponding to cell-wall related genes in the magenta module with the gene encoding transcription factor *KNAT7* as a central node.

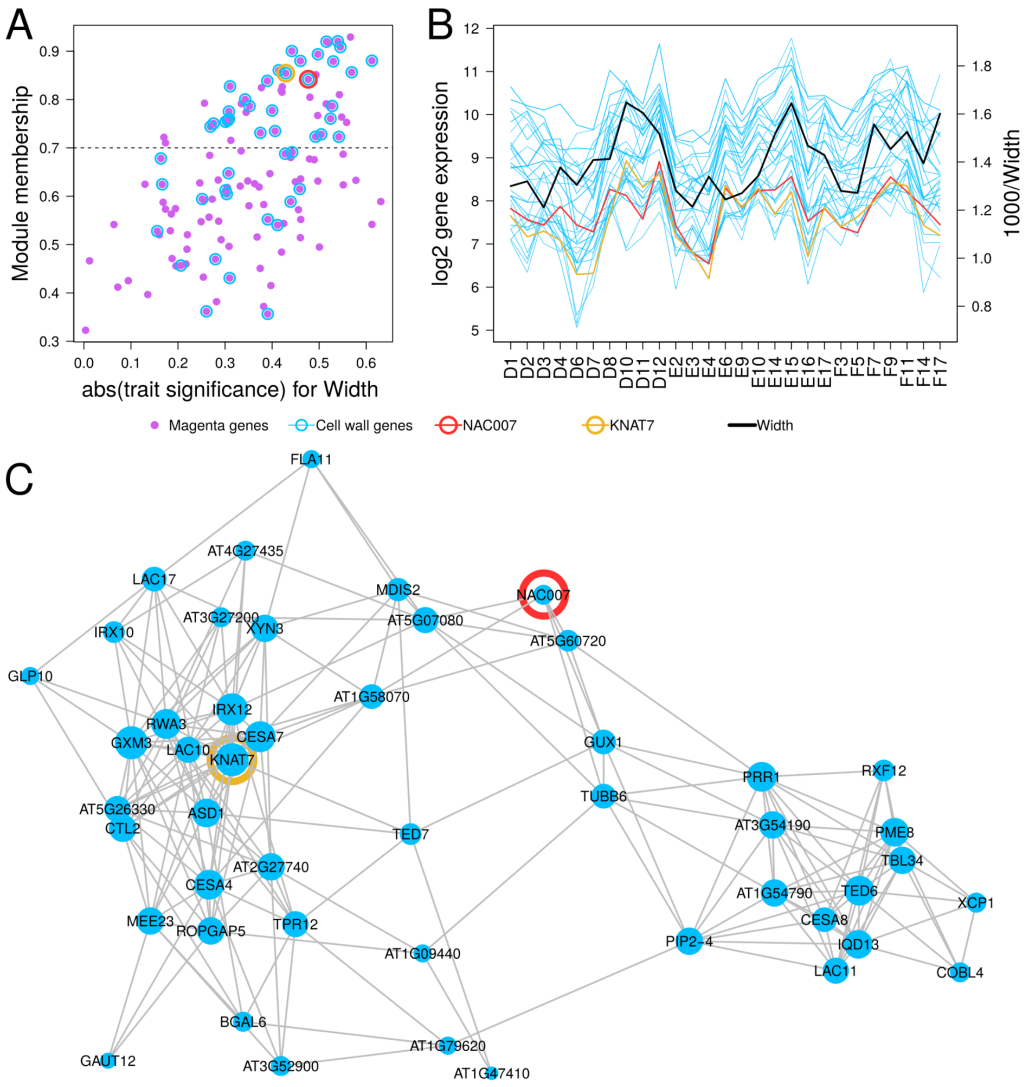

Figure 4. Properties of the magenta module and underlying gene regulatory network.

(A) Module membership against trait significance in absolute value with respect to width for all genes from the magenta module. Higher module membership value indicates a gene that is more central in the module, while higher trait significance indicates that the expression of the gene has a higher correlation with width. Cell-wall genes are highlighted with blue rings, and genes encoding transcription factors *NAC007* and *KNAT7* with red and yellow rings, respectively.

(B) Gene expression values across 27 samples for cell-wall related genes within the magenta module with module membership above 0.7 (black dotted line in A). Genes encoding transcription factors *NAC007* and *KNAT7* are shown in red and yellow, respectively. The black line corresponds to 1000/Width values for each sample (right axis), highlighting the negative correlation between gene expression of genes with high module membership in the magenta module and Width. The right axis has been chosen for Width measurements to overlap with gene expression values in order to visually show the negative correlation among them.

(C) Gene regulatory network of all cell-wall related genes in the magenta module (see Methods/Gene regulatory network reconstruction). *NAC007* and *KNAT7* are highlighted in red and yellow, respectively. Node size is proportional to the degree (number of connections) of each node.

Highly variable genes are enriched in response to stimulus and cell-wall functions

We have classified our set of 14,085 genes according to their variability by measuring their CV^2 across 27 samples. Given the higher variability (both technical and biological) found in lowly expressed genes, one can correct CV^2 to account for this association (Cortijo et al., 2019). However, the gene expression threshold that we have set has eliminated most lowly expressed genes from our dataset and essentially, there is no correlation between CV^2 values and average gene expression. Hence, we have ordered genes by their raw CV^2 values and have selected highly variable genes (HVGs: top 5%, top 5-to-10%, and top 10-to-15%), and lowly variable genes (LVGs: bottom 5%, bottom 5-to-10%, and bottom 10-to-15%) in order to see if these groups of genes were enriched in particular biological functions (Figure 5A).

To this aim, we searched for enrichment of GO categories in each of these subsets of genes focusing on high hierarchy GO terms using the PANTHER17.0 online tool (Mi et at., 2021; see Materials and methods). The top 5% HVGs show a very different GO enrichment profile compared to the bottom 5% LVGs (Figure 5B). Top HVGs are enriched in response to stimulus, to chemical, to stress, to lipid, to wounding and defense response, whereas bottom LVGs are enriched in cellular and metabolic processes, vesicle-mediated transport and cellular localization. These findings are consistent with other studies that have found HVGs to be enriched in response to environment genes and LVGs enriched in housekeeping functions (Cortijo et al., 2019; Zheng et al., 2004). An unexpected finding, however, was that within the top 5-to-10% and top 10-to-15% HVGs there is an enrichment in carbohydrate metabolic process and cell-wall organization or biogenesis genes (Figure 5B).

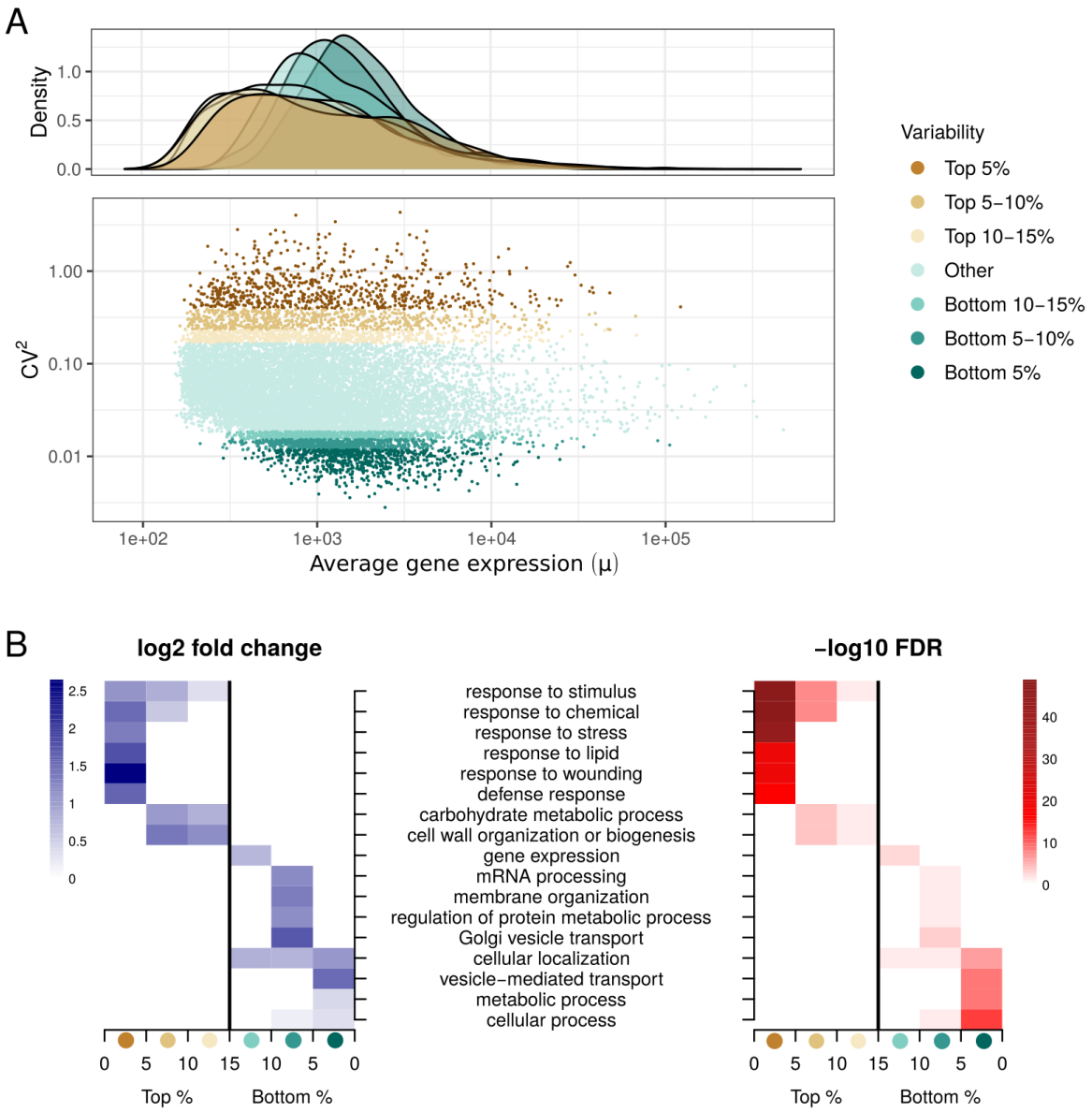

Figure 5. GO enrichment analysis of genes according to their variability.

A) Squared coefficient of variation (CV^2) against average (μ) of gene expression across 27 samples for 14,085 genes. Highly variable genes (top 5%, top 5-10%, and top 10-15%) and lowly variable genes (bottom 5%, bottom 5-10%, and bottom 10-15%) are identified (bottom). Probability density function for each CV^2 classification (top).

B) GO enrichment results for each highly variable and lowly variable gene categories. Only top hierarchy GO terms that appear enriched in any of the shown categories are included. Log2 fold change enrichment (left) and -log10 false discovery rate (right) are shown for each category.

A curated cell-wall related gene list to assess gene ontology

Highly variable genes are enriched in cell-wall functions. Additionally, cell-wall related genes are over-represented in modules of coexpressed genes which have the strongest correlation with morphological parameters. These two observations might be related. Importantly, they are both driven by GO annotations which are known to be incomplete and biased (Timmons et al., 2015). In order to use an alternative approach to assess cell-wall related gene expression in sepals, we have compiled a non-exhaustive but rather comprehensive list of cell-wall related genes centered on genes encoding proteins involved in structure, biosynthesis and cell-wall remodeling. To produce this list, first, we referred to a review describing all the enzymes acting on the cell wall (Frankova and Fry, 2013) and identified the corresponding genes in the TAIR Arabidopsis database (www.tair.com). Second, we used the CAZy database (www.cazy.org) describing families of structurally-related catalytic and carbohydrate-binding modules of enzymes that degrade, modify or create glycosidic bonds to find corresponding genes in the Arabidopsis genome. Finally, we manually enriched the list with proteins known to directly interact with those enzymes in the cell wall, such as pectin methylesterase inhibitors or interactors of cellulose synthase, also adding the extensins, which are non-enzymatic proteins highly abundant in the cell wall and important for its biosynthesis. After removing duplicates, we recovered a list of 1585 genes (Supplementary Table 4) that we will refer to as “cell-wall related gene list”.

Our cell-wall related gene list was compiled from the whole Arabidopsis genome. Among these genes, 718 are expressed in sepal since they are within our 14,085 sepal gene set. We tested for enrichment of cell-wall related genes within this list across our modules and confirmed that the magenta and orange modules are enriched in cell-wall related genes (Figure 6A). Additionally, CV^2 values of genes within our cell-wall related gene list are higher compared to the entire gene dataset (Figure 6B) confirming our results on variability of expression based on GO term enrichment.

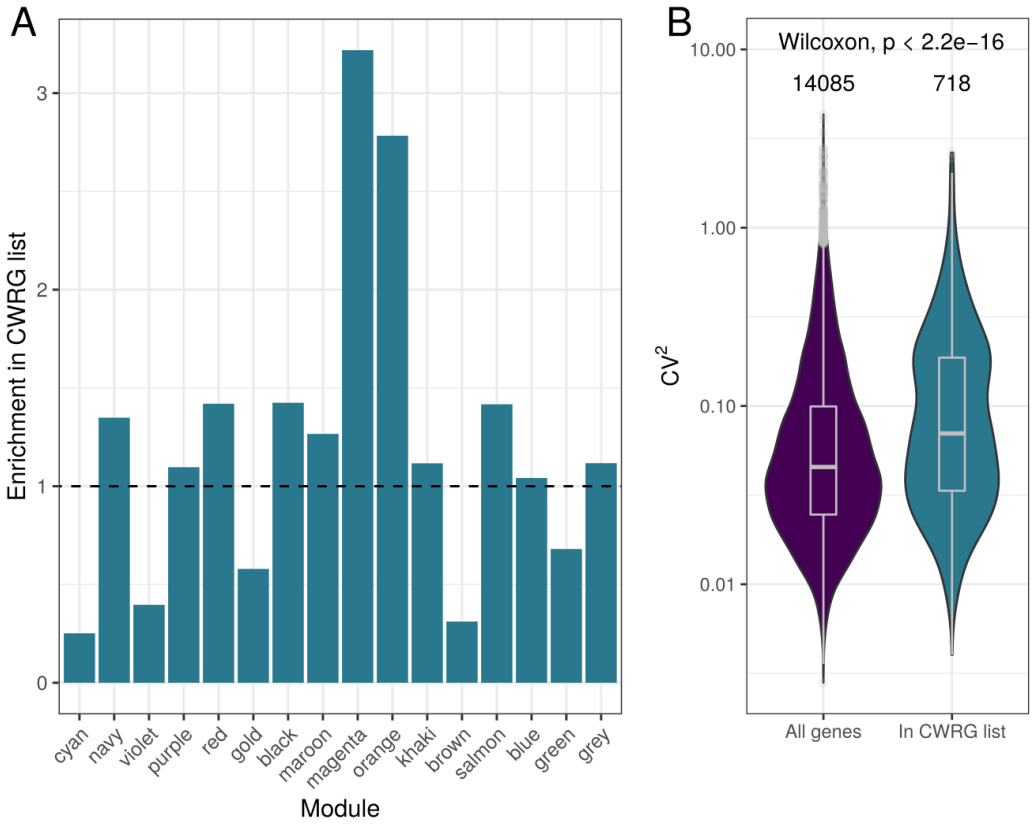

Figure 6: Cell wall-related gene (CWRG) distribution in modules and CV^2 distribution

A) Enrichment in genes within our cell-wall related gene list across all modules. Enrichment values are calculated by dividing the observed number of CWRGs in each module by the expected number according to module size.

B) Squared coefficient of variation (CV2) distribution for our entire 14,085-gene dataset (purple) compared to that of the 718 CWRGs found in the dataset (blue). p-value of Wilcoxon test between distribution is shown on top.

Variability of cell-wall mutant morphology correlates with average expression of corresponding genes in wild-type plants

To further explore the link between cell-wall related gene expression and sepal morphology, we decided to exploit an unpublished mutant dataset previously produced in our team (Dumond, 2017). This mutant dataset had been generated in the context of sepal phenotype exploration and contained data for 16 of the 718 cell-wall related genes expressed in sepal. The mutants had been selected based on the level of expression of the corresponding genes in sepals at stage 12 (from publicly available datasets, not from our RNA-Seq data) and on availability of mutants with T-DNA insertions in exons. Because it is difficult to compare

developmental stages between mutants and wild-type, we focused on sepals that had ceased growing (stage 12). For each of these mutants, we obtained length, width, aspect ratio, and area from 50 to 60 sepals. Given the large number of sepals to be imaged, the protocol for these measurements was different from the one used to obtain measurements described above: the sepals were flattened prior to imaging and pictures were taken using a stereomicroscope. So, although we do not consider measurements from these sepals comparable to the ones obtained for our wild-type plants, we can compare the variation present in mutant phenotypes with our gene expression data for the corresponding knockout genes.

We first hypothesized that expression of a given gene in wild-type plants would predict the morphology of the corresponding mutant. We thus examined the relation between gene expression in wild-type and mutant morphological parameters (Figure 7 for Area and Figure 7 - Figure supplement 1 for Length, Width and AspRatio). Unexpectedly, the morphology of mutant lines was not correlated with the level of expression of the corresponding gene in wild-type plants (Figure 7, top right). We then hypothesized that genes with higher variability in wild-type would tend to be more important for robust morphogenesis, and hence, when knocked out, would have a stronger effect on sepal average morphology or on variability in morphology. Again, this was not supported by the data (Figure 7, left column). Finally, we hypothesized that knocking down genes highly expressed in wild-type would prove more difficult to cope with than for genes with lower expression. Accordingly, mutant phenotypic variability would correlate with gene expression in wild-type. Indeed, there is a strong correlation between the CV^2 values of the mutant phenotypes (for area, length and width, but not aspect ratio) and the log mean expression in the wild type of the corresponding knocked out genes (Figure 7 bottom right and Figure 7 - Figure supplement 1), supporting the latter possibility. The correlations shown are quite robust to removal of one or several mutants from the dataset. Accordingly, our results indicate that expression of a cell-wall related gene in wild-type predicts variability (and not average) of morphology in the corresponding mutant.

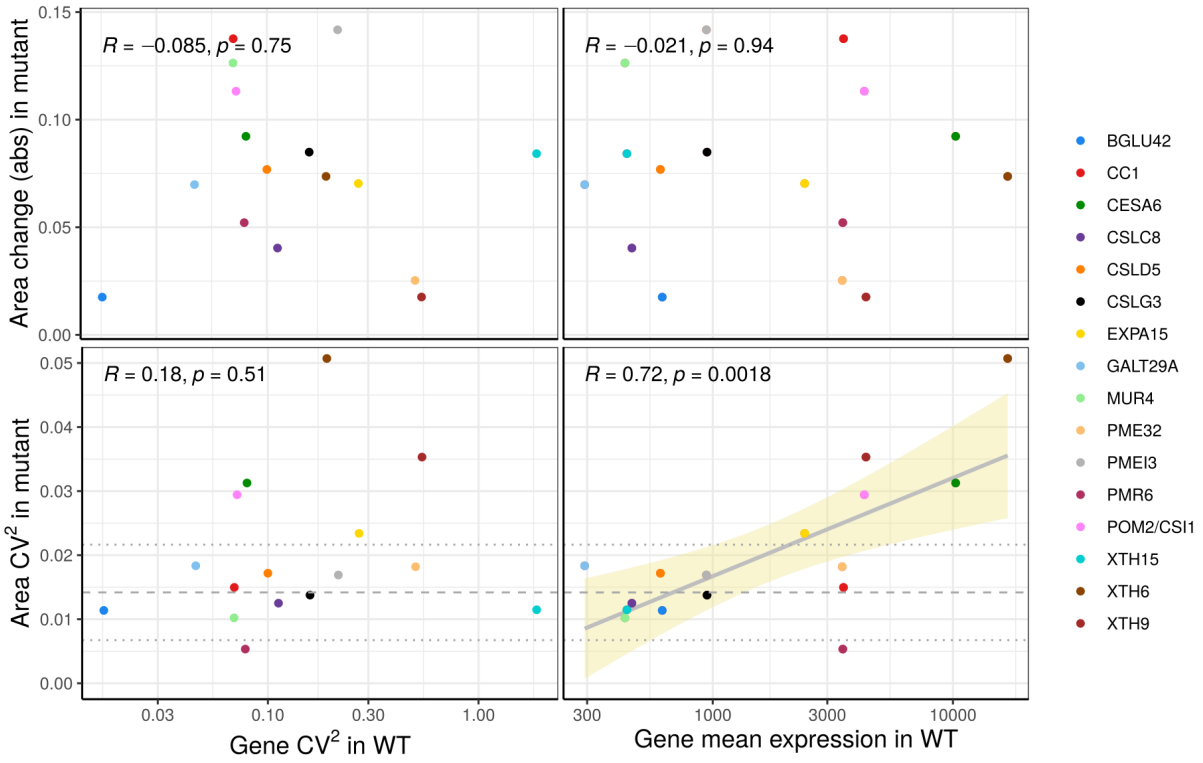

Figure 7: Mutant sepal area (effect compared to wild-type and variability) against gene expression (mean and variation) of the corresponding gene in wild-type.

Top row y-axis: effect of the mutation on sepal area, corrected by sepal area in wild-type plants for each batch. Bottom row y-axis: coefficient of variation of sepal area in knockout mutants. Left column x-axis: coefficient of variation of each of the corresponding mutant genes in wild-type Col-0 plant from the RNA-Seq data across 27 sepal samples. Right column x-axis: mean gene expression of the corresponding mutant genes in wild-type Col-0 plant from the RNA-Seq data across 27 sepal samples. Each point corresponds to one gene. Pearson correlation coefficient R and p-values are shown on the top left corner in each plot. Gray solid line shows linear model adjustment with standard error in tan shade. Gray dashed line corresponds to average CV^2 in area measurements in wild-type Col-0 control batches with dotted lines showing average plus/minus one standard deviation.

Discussion

Summary of results

Robustness of organ size and shape is thought to be the result of the complex interaction of genes within gene regulatory networks and environmental cues. Gene expression variability can be considered, *a priori*, as a factor affecting robustness. Interestingly, Hong et al. (2016) reported that cell-to-cell variability of growth is required for robustness of sepal shape and size, opening the question of the role of variability in gene expression in this process. Our work is an effort to characterize organ morphology and gene expression with enough sample size to be able to evaluate the role of gene expression variability in ensuring organ robustness. Cell-wall related genes appear to be essential to this process. In this paper we have successfully evaluated variability in wild-type Arabidopsis sepal morphology and its links with inter-sepal variability in gene expression.

Our image analysis pipeline allows for automated, precise and quantitative analysis of sepal shape compatible with large-scale analysis, which can be used to characterize new mutants as well as to re-examine known mutants. One advantage of the protocol is that it exploits the natural autofluorescence of plant tissue and does not require introducing any fluorescent protein by genetic transformation, like in previous work on leaves (Biot et al., 2016). Furthermore, our 3D-shape based analysis has not only confirmed some expected correlations between parameters, such as length and area, but also revealed the previously undescribed negative correlation between area and curvature, which was not achievable in 2D-shape based analyses (Bensmihen et al. 2008).

We have identified modules of co-expressed genes correlated with shape parameters and recovered a biologically relevant gene regulatory network. We have found that those modules which correlate most strongly with shape parameters appear to be enriched in cell-wall related genes and peptide synthesis. Additionally, cell-wall related genes are also found to be highly variable across our samples. To test for a causal link between gene expression variability and robustness, we have evaluated sepal morphology in cell-wall related mutants. We have found that genes that are more highly expressed in wild-type sepals are those that, when knocked out, yield more variability in morphology. More generally, we propose that gene regulatory networks that ensure organ robustness in wild type plants are more affected when highly expressed genes are knocked out.

Control of organ morphology: our approach against classical approaches

Sepals are an important organ of angiosperm flowers. As with many other plant structures and in particular flower organs, their morphology has been studied from a broad variety of approaches. Importantly, mutant-based approaches or condition-based approaches focus on morphological differences in sepals between different mutants or different growth conditions, respectively. Although some signaling pathways regulating organ size and shape have been identified (reviewed in Powell & Lenhard, 2012), the mechanisms behind morphological robustness have remained elusive. More recently, a screen for mutants that disrupted the robustness of sepal size and shape led to the identification of genes involved in reproducibility of sepal morphogenesis (Hong et al., 2016; Zhu et al., 2020). This work revealed that spatio-temporal averaging of cellular variability as well as precise timing of organ initiation, are required for precision in organ size. The mechanisms underlying organ robustness, however, remain largely enigmatic.

Mutant-based approaches have certain limitations, one of which being the existence of compensatory mechanisms. The mutation of a given gene can affect the expression of distant loci and in some cases, upregulate genes that compensate for the loss of the mutant gene function, thereby masking or attenuating the mutant phenotype. This probably accounts for the fact that many genes lack observable phenotypes when mutated (reviewed by Kontarakis & Stainier, 2020). For example, inhibition of cell proliferation is often compensated by increased cell expansion, resulting in a strongly reduced effect on whole organ size (De Veylder et al., 2001; Tsukaya, 2002; Tsukaya, 2003; Beemster et al., 2003). Horiguchi et al. (2006) reported that mutation leading to reduced cell number in leaf leads to increased cell size and results in conserved leaf area. Compensation for a mutation can also occur through changes in gene expression, which is known as transcriptional adaptation. An example of this has been reported by Sénéchal et al. (2015) who showed that mutation of the gene *PMEI4* encoding an inhibitor of pectin methylesterase leads to the concomitant decreased expression of the pectin methylesterase gene *PME17* and a higher expression of the subtilase SBT3.5 that could potentially increase the amount of processed active PME17 protein. In the same line, Hocq et al. (2020) found an increased polygalacturonase activity in two knockout mutants for polygalacturonase encoding genes, *PGLR1* and *PGLR2*, and showed that this unexpected result was due to transcriptional compensation as four genes of the same family were overexpressed compared to the wild type, leading to increased polygalacturonase activity and subsequent increased root length. In these examples, protein activity could be measured because of its enzymatic nature, allowing the detection of induced compensation. However, it is likely that compensation is often under-appreciated because of the lack of simple assays for protein activity, for example in the case of transcription factors. As a consequence, the mutant

phenotype may not directly reveal the underlying molecular mechanisms and the use of alternative approaches is essential.

Instead of searching for particular mutants in which the robustness of sepal size and shape is affected, our approach has been, rather, to use variability in gene expression and morphology in wild-type plants of the same accession to unravel gene regulatory networks underlying robustness of organ morphology. We then referred to mutants, not to search for candidate genes involved in morphology, but to test the way in which knockout mutants of genes putatively involved in robustness, such as cell-wall related genes, affected morphology. Our results provide evidence that knockouts of highly expressed genes tend to increase phenotypic variability. In particular, highly-expressed cell-wall related genes in wild-type plants induce larger morphological variation when knocked out than less expressed genes.

Modules and enriched gene ontology categories

To recover modules of co-expressed genes, we used the WGCNA package (Langfelder & Horvath, 2008) which allows this with relatively few parameters to adjust. Other tools, such as sPLS (Lê Cao et al., 2008), also allow for the association of two-block data, such as gene expression and morphology. However, sPLS recovers individual genes instead of modules. Given the risk of false positives when focusing on small subsets of genes, we considered that associating morphology to modules of co-expressed genes instead of individual genes was a more conservative approach.

Based on our morphology analysis, we predicted that area and aspect ratio were the most informative parameters to describe sepal size and shape. However, our WGCNA results show width as the most highly correlated parameter. Nevertheless, when viewing all statistically significant correlations, we observe that modules that correlate with width also correlate with area (magenta and cyan modules) and that the third most significantly correlated module (orange) correlates with aspect ratio, which fits in well with the aforementioned prediction. Looking at enriched GO terms for each module (Supplementary Table 3) we find that two modules are strongly enriched in cell-wall related genes (magenta and orange), and another one (cyan), in chloroplast-related functions and peptide biosynthesis. The cell wall is known to strengthen the plant body and to play a key role in plant growth. During evolution, plant cells have acquired the capacity to synthesize walls made of polysaccharides, to assemble them into a strong fibrous network and to regulate cell-wall expansion during growth (for a review, see e.g. Cosgrove, 2005). The polysaccharides that contribute to the biomass and to the cell wall are made up of sugars produced in chloroplasts, as end products of photosynthesis. Therefore, these results are quite coherent.

The contribution of different sets of co-regulated genes to sepal morphology can be thought of in isolation but connections between modules can also be established, albeit with a certain degree of speculation. For example, the magenta module correlates negatively with width, whereas the cyan module correlates positively with width. One could speculate that genes in the magenta module could have repressive effects on growth, i.e., secondary cell wall might be involved in the growth repression; and that the genes in the cyan module could have a contrary effect, i.e., increased peptide biosynthesis associated with increased growth. Additionally, the fact that the orange module, also enriched in cell-wall related genes, correlates best with aspect ratio and not with width like the magenta module, could indicate that different mechanisms are at work to control size and shape. In fact, the alignment of cortical microtubules is important for the anisotropy of sepal growth and in determining the aspect ratio of mature sepals (Hervieux et al., 2016). Interestingly, supramolecular fiber and cytoskeleton organization genes appear to be overrepresented in the orange module.

Whereas other studies have aimed at reconstructing complete and consistent GRNs based on reviews of the literature (e.g. Espinosa-Soto et al., 2004; La Rota et al., 2011), we have shown that wild-type variability in gene expression allows the recovery of gene sets that likely function together. An illustration that this is likely the case, is that we find two transcription factors which are known to dimerize, BZIP34 and BZIP61, in the orange module. If we focus on the magenta module, which contains genes related to the cell wall, we find that the top two transcription factors of this module (ordered by module membership), NAC007 and KNAT7, are known to be involved in secondary cell-wall deposition (Wang et al., 2020). As expected, the identified direct targets of NAC007 (O'Malley, et al., 2016) are enriched in the magenta module. Although DAP-Seq data is not currently available for KNAT7, some targets of KNAT7, such as *IRX8* and *IRX10* (Li et al., 2012), which belong to the magenta module, have been identified. In addition, three *CESA* genes that have been shown to be upregulated in the *knat7* loss-of-function mutant (Li et al., 2012) also belong to the magenta module. Interestingly, the *knat7* mutant has reduced seed size (Renard et al., 2020) and a repressor version of KNAT7 (supposed to mimic the mutant) induces a dwarf phenotype (Qin et al., 2020), supporting a link between KNAT7 and growth control. A speculative mechanism could be that KNAT7 positively regulates secondary cell-wall deposition, as shown by Wang et al. (2020), which is known to rigidify the tissues (Zhong & Ye, 2015) and could potentially lead to growth arrest. This is consistent with the magenta module being negatively correlated with width.

Cell-wall related genes and robustness of organ morphology

Results from our evaluation of phenotypic variation in cell-wall related mutants point towards the effect of knockouts of highly expressed cell-wall genes being difficult to buffer, with

a higher variation in length, width and area in higher-expressed-gene knockout mutants. There is knowledge about capacitors of phenotypic variation, such as heat-shock protein HSP90 (Queitsch et al., 2002), which limits the manifestation of cryptic genetic variation allowing for developmental stability (Sangster et al., 2008). Although, as discussed above, there might be compensation mechanisms affecting the phenotypic outcome of cell-wall knockout mutants (e.g. Sénéchal et al., 2015; Hocq et al., 2020), focusing not on magnitude of mutation effect but variability of effect, might help unravel the function of poorly characterized genes. Our observation that cell-wall related genes appear to be involved in buffering mechanisms that allow for developmental robustness, adds to evidence of cell-wall related genes being important for organ size and proportions (Weiss et al., 2005).

Despite having grown our three plants in experimentally controlled conditions and all of them having grown healthy, we observe that genes involved in response to stimulus are highly variable among our sepal samples. Although we cannot completely exclude the possibility that expression of response to stimulus genes were induced by sepal dissection and imaging, it is unlikely since we managed to keep the duration of sepal manipulation below five minutes. Our interpretation, in line with other studies (Araújo et al., 2017, Cortijo et al., 2019), is that this variability is not in itself due to differences in gene expression as a response to strong external stimuli experienced by different sepals, but rather, the evidence of underlying basal gene expression variability allowing the plant to cope with small microenvironmental differences. This basal variability is thought to have been selected to allow for the possibility of adaptation to environmental change (Queitsch et al., 2002). We extend this interpretation to cell-wall related genes, which also show significant variability of expression. We consider that reproducibility in morphology is achieved in part thanks to underlying regulatory and compensatory mechanisms. Our results indicate that cell-wall related genes could be fundamental for these mechanisms, ensuring robust development. More generally, our work sheds light on the links between expression variability, gene regulatory networks, and developmental robustness.

Materials and Methods

Plant Material

Sibling Col-0 plants were grown on soil at 20°C in short day conditions (8 h light/16 h dark) for 20 days before being transferred to long day conditions at 22°C (16 h light/8 h darkness). Sepals were dissected from secondary inflorescences after at least 10 siliques were formed. Buds were dissected under the binocular to identify those at the beginning of stage 11 as described by Smyth et al. 1990 (stigmatic papillae formed, petals being longer than the short stamen but shorter than the long stamen). Each abaxial sepal was transferred to 0.8% agarose to avoid dehydration, imaged with a confocal microscope and immediately transferred to liquid nitrogen. The manipulation and the confocal imaging lasted for less than 5 min to minimize the impact of the manipulation on the transcriptome.

Confocal Imaging

Sepals were examined in a Leica SP5 confocal microscope equipped with a 10X objectif. Samples were imaged using laser 488 nm set up at 20% of the maximum power and emission from 498 to 735 nm was recovered. The acquisition time was no more than 1min 40s.

Segmentation and extraction of geometrical parameters

Images are pretreated in the xy plane using a Gaussian filter of sigma of the order of the z direction voxel size. Then a normalizing procedure is performed on the whole filtered sample. The linear normalization is constructed in such a way that it superimposes the background value and the Otsu threshold value of the output images. The sepal's contour detection procedure is based on the Edge Detect Morphology process of the MorphoGraphX software. First, a rough contour is assessed as the intersection of the top-down and the bottom-up contours given by Edge Detect. Then a dilation in the xy direction is performed in order to recover the weakly marked margin cells of the sepals.

The principal directions of the sepal contour are computed, and sepals are placed in the frame where their center of mass is in the origin (their first principal axis is along the Oy axis and the second principal axis is along the Ox axis). The length measurements are done on the longitudinal and transversal principal sections using a python script. The sepal's abaxial and adaxial surfaces are separated by watershed segmentation of the maximal curvature map constructed on the sepal's whole surface. Then the area of the abaxial surface is computed as areas of cells on a tissue surface mesh in MorphoGraphX. The morphological parameters measured are length, width, area, longitudinal curvature and transversal curvature. Aspect ratio is length divided by width and Gaussian curvature is the product of longitudinal and transversal curvatures.

In order to validate geometrical measurements, we compared a sample of 10 images from 10 different sepals with a sample of 10 images of the same sepal that was manipulated between consecutive shootings. For each of our parameters, same-sample variability was verified to be less than 5% of the variability measured across different samples. Other parameters that had

originally been measured, such as curvature at the sepal periphery, were excluded because they did not pass this validation criterion.

RNA extraction and sequencing

Each sepal was ground individually using a plastic conical pestle fitting the bottom of the Eppendorf tube, which was dipped in liquid nitrogen. RNA extraction was subsequently performed using the Arcturus PicoPure RNA Isolation Kit from Thermo Fisher Scientific following the manufacturer's instructions. The libraries were sequenced by the HELIXIO company on Illumina NextSeq 500 using single read sequencing of 75bp in length.

Quantification of gene expression

Raw reads were pseudo-aligned to cDNA and non-coding RNA sequences from the TAIR10 *Arabidopsis thaliana* reference genome (Ensembl release 47) using kallisto (Bray et al., 2016). Transcript abundances obtained were converted to gene abundances using tximport (Soneson et al., 2015). Gene counts per million were obtained using the edgeR package (Robinson et al., 2010) in R (R Core Team, 2021). Only genes with at least 5 counts per million in at least 14 out of 27 sepal samples were retained and counts per million were normalized using the a trimmed mean of M-values (TMM) method in edgeR, resulting in 14,085 gene expression values for 14,085 genes. We validated our bulk RNA-Seq results by comparing them to RNA-Seq samples obtained from Arabidopsis seedlings from Cortijo et al. (2019) (Figure 1 - Figure supplement 1).

WGCNA

Parameters used for WGCNA were: soft-thresholding power for adjacency matrix reconstruction, 11; minimum module size, 50; module merging threshold, 0.35.

GO enrichment analysis

Gene ontology enrichment analysis was performed with PANTHER17.0 (Mi et al., 2021) at http://pantherdb.org (last accessed on 2022-03-30) with the following parameters: Analysis Type: PANTHER Overrepresentation Test (Released 2022-02-02); Annotation Version and Release Date: GO Ontology database DOI:10.5281/zenodo.5725227 Released 2020-11-01; Test Type: FISHER; Correction: FDR. Background used was our 14,085 gene dataset (Supplementary Table 2), and Arabidopsis whole genome (Supplementary Table 3). Top hierarchy GO terms were obtained by filtering the PANTHER output with the following criteria: only positive enrichment terms were kept, only top hierarchy terms in each block (as output by PANTHER) were kept according to total number of genes associated with each GO term, terms were ordered according to their false discovery rate (FDR) starting from the smallest rate, only the top five resulting terms were kept for each highly-variable group (HVG) or lowly-variable group (LVG) group unless a term in the top five in one group also appeared in another group, in which case it was kept in both. Log2 fold change values (Supplementary Table 5) and log10 FDR (Supplementary Table 6) were calculated for each of these terms in each HVG or LVG group.

Gene regulatory networks

Gene regulatory network reconstruction was performed with GENIE3 (Huynh-Thu et al., 2010) with the Random Forest method and default parameters. The weighted adjacency matrix

obtained was modified by setting all adjacent values above 0.045 to 1 and below 0.045 to 0. The 0.045 threshold was chosen to improve visibility and to include all cell-wall related genes in the magenta module. Graph visualization was performed by using the graph_from_adjacenty_matrix function from the igraph (version 1.2.11) package and the ggnet2 function in R. Genes under the GO term "cell wall organization or biogenesis" (GO:0071554) were extracted from Agrigo2.0 (Tian et al., 2017) at http://systemsbiology.cau.edu.cn/agriGOv2/ (last accessed on 2022-02-20) and are listed in Supplementary Table 7.

Variability of gene expression

Squared coefficient of variation values were calculated by dividing the variance over the square of the mean of measurements across 27 sepal samples.

Mutants

We selected 16 genes among the cell-wall related gene list that are expressed in sepals and based on availability of T-DNA insertions in exons and considered one mutant allele for each of these genes (Supplementary Table 8). Seeds were obtained from well-characterized stocks (Sénéchal, 2013; Bringmann et al., 2012; Endler et al., 2015) or from SAIL and SALK collections maintained at NASC (Tzafrir et al., 2003; Alonso et al., 2003). Mutant plants were genotyped following O'Malley et al. (2015), except for *pom2-5* and *prc1* which have distinctive phenotypes. Plants were grown as described above. We assessed the final shape of sepals (stage 14, according to Smyth et al. 1990), by taking pictures of flattened sepals under a dissection scope (Hong et al., 2016).

Extraction of morphological parameters of mutant lines

We analyzed sepal contours and quantified area, length, width, and aspect ratio following Hong et al. (2016). Briefly, we flattened the sepals between two slides and took photographs with a black background under a dissecting microscope. We used Python scripts (Van Rossum & Drake, 2009) to segment and align sepals, and to extract morphological parameters.

Data availability

Raw RNA-Seq data for 30 sepal samples are available at the European Nucleotide Archive (ENA) under the accession number [Under progress]. Code used for the analysis and figure generation is available at https://github.com/diegoharta/SepalMorphology/. Source data for all figures are provided as Supplementary Material. Data analysis and figures were performed in R (R Core Team, 2021).

Acknowledgements

We wish to acknowledge A. Roeder for constructive discussion and for suggestions on a first version of the manuscript, A. Ahmed for work on a pilot project, PLATIM for provision of microscope facilities, Helixio for their excellent service and assistance in the RNA sequencing, J. Pelloux for sharing mutants and discussion on cell wall, Y. Long, P. Das, A. Fruleux and S. Bovio

for comments and suggestions, and A. Lacroix, J. Berger, P. Bolland, H. Leyral and I. Desbouchages for assistance with plant growth and logistics. This work was funded by the French National Research Agency (ANR) through a European ERA-NET Coordinating Action in Plant Sciences (ERA-CAPS) grant (Grant No. ANR-17-CAPS-0002-01 V-Morph) and through a direct grant (Grant No. ANR-17-CE20-0023-02 WALLMIME), by Fond de Recherche ENS Lyon (Projet émergent FLORIVAR), and by BAP INRAE (Projet FLORIVAR).

Competing interests

The authors have no conflicts of interest to declare.

References

Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H., Shinn, P., Stevenson, D. K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C. C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., … Ecker, J. R. (2003). Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science (New York, N.Y.), 301(5633), 653–657. https://doi.org/10.1126/science.1086391

Alvarez-Buylla, E. R., Chaos, A., Aldana, M., Benítez, M., Cortes-Poza, Y., Espinosa-Soto, C., Hartasánchez, D. A., Lotto, R. B., Malkin, D., Escalera Santos, G. J., & Padilla-Longoria, P. (2008). Floral morphogenesis: stochastic explorations of a gene network epigenetic landscape. PloS one, 3(11), e3626. https://doi.org/10.1371/journal.pone.0003626

Araújo, I. S., Pietsch, J. M., Keizer, E. M., Greese, B., Balkunde, R., Fleck, C., & Hülskamp, M. (2017). Stochastic gene expression in Arabidopsis thaliana. Nature communications, 8(1), 2132. https://doi.org/10.1038/s41467-017-02285-7

Barbier de Reuille, P., Routier-Kierzkowska, A. L., Kierzkowski, D., Bassel, G. W., Schüpbach, T., Tauriello, G., Bajpai, N., Strauss, S., Weber, A., Kiss, A., Burian, A., Hofhuis, H., Sapala, A., Lipowczan, M., Heimlicher, M. B., Robinson, S., Bayer, E. M., Basler, K., Koumoutsakos, P., Roeder, A. H., … Smith, R. S. (2015). MorphoGraphX: A platform for quantifying morphogenesis in 4D. eLife, 4, 05864. https://doi.org/10.7554/eLife.05864

Beemster, G. T., Fiorani, F., & Inzé, D. (2003). Cell cycle: the key to plant growth control?. Trends in plant science, 8(4), 154–158. https://doi.org/10.1016/S1360-1385(03)00046-3

Bensmihen, S., Hanna, A. I., Langlade, N. B., Micol, J. L., Bangham, A., & Coen, E. S. (2008). Mutational spaces for leaf shape and size. HFSP journal, 2(2), 110–120. https://doi.org/10.2976/1.2836738

Biot, E., Cortizo, M., Burguet, J., Kiss, A., Oughou, M., Maugarny-Calès, A., Gonçalves, B., Adroher, B., Andrey, P., Boudaoud, A., & Laufs, P. (2016). Multiscale quantification of morphodynamics: MorphoLeaf software for 2D shape analysis. Development (Cambridge, England), 143(18), 3417–3428. https://doi.org/10.1242/dev.134619

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nature biotechnology, 34(5), 525–527. https://doi.org/10.1038/nbt.3519

Bringmann, M., Li, E., Sampathkumar, A., Kocabek, T., Hauser, M. T., & Persson, S. (2012). POM-POM2/cellulose synthase interacting1 is essential for the functional association of cellulose synthase and microtubules in Arabidopsis. The Plant cell, 24(1), 163–177. https://doi.org/10.1105/tpc.111.093575

Chen, D., Yan, W., Fu, L. Y., & Kaufmann, K. (2018). Architecture of gene regulatory networks controlling flower development in Arabidopsis thaliana. Nature communications, 9(1), 4534. https://doi.org/10.1038/s41467-018-06772-3

Cortijo, S., Aydin, Z., Ahnert, S., & Locke, J. C. (2019). Widespread inter-individual gene expression variability in Arabidopsis thaliana. Molecular systems biology, 15(1), e8591. https://doi.org/10.15252/msb.20188591

Cortijo, S., Bhattarai, M., Locke, J., & Ahnert, S. E. (2020). Co-expression Networks From Gene Expression Variability Between Genetically Identical Seedlings Can Reveal Novel Regulatory Relationships. Frontiers in plant science, 11, 599464. https://doi.org/10.3389/fpls.2020.599464

Cosgrove D. J. (2005). Growth of the plant cell wall. Nature reviews. Molecular cell biology, 6(11), 850–861. https://doi.org/10.1038/nrm1746

Debat, V., Bloyer, S., Faradji, F., Gidaszewski, N., Navarro, N., Orozco-Terwengel, P., Ribeiro, V., Schlötterer, C., Deutsch, J. S., & Peronnet, F. (2011). Developmental stability: a major role for cyclin G in drosophila melanogaster. PLoS genetics, 7(10), e1002314. https://doi.org/10.1371/journal.pgen.1002314

de Oliveira Silva, F. M., de Ávila Silva, L., Araújo, W. L., Zsögön, A., & Nunes-Nesi, A. (2017). Exploiting Natural Variation to Discover Candidate Genes Involved in Photosynthesis-Related Traits. Methods in molecular biology (Clifton, N.J.), 1653, 125–135. https://doi.org/10.1007/978-1-4939-7225-8_9

De Veylder, L., Beeckman, T., Beemster, G. T., Krols, L., Terras, F., Landrieu, I., van der Schueren, E., Maes, S., Naudts, M., & Inzé, D. (2001). Functional analysis of cyclin-dependent kinase inhibitors of Arabidopsis. The Plant cell, 13(7), 1653–1668. https://doi.org/10.1105/tpc.010087

Dorrity, M. W., Alexandre, C. M., Hamm, M. O., Vigil, A. L., Fields, S., Queitsch, C., & Cuperus, J. T. (2021). The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. Nature communications, 12(1), 3334. https://doi.org/10.1038/s41467-021-23675-y

Dumond, M. (2017). From cellular variability to shape reproducibility : mechanics and morphogenesis of Arabidopsis thaliana sepal Morphogenesis. Université de Lyon.

Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic gene expression in a single cell. Science (New York, N.Y.), 297(5584), 1183–1186. https://doi.org/10.1126/science.1070919

Endler, A., Kesten, C., Schneider, R., Zhang, Y., Ivakov, A., Froehlich, A., Funke, N., & Persson, S. (2015). A Mechanism for Sustained Cellulose Synthesis during Salt Stress. Cell, 162(6), 1353–1364. https://doi.org/10.1016/j.cell.2015.08.028

Espinosa-Soto, C., Padilla-Longoria, P., & Alvarez-Buylla, E. R. (2004). A gene regulatory network model for cell-fate determination during Arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. The Plant cell, 16(11), 2923–2939. https://doi.org/10.1105/tpc.104.021725

Félix, M. A., & Barkoulas, M. (2015). Pervasive robustness in biological systems. Nature reviews. Genetics, 16(8), 483–496. https://doi.org/10.1038/nrg3949

Gibson, G., & Wagner, G. (2000). Canalization in evolutionary genetics: a stabilizing theory? BioEssays : news and reviews in molecular, cellular and developmental biology, 22(4), 372–380. https://doi.org/10.1002/(SICI)1521-1878(200004)22:4<372::AID-BIES7>3.0.CO;2-J

Hall, M. C., Dworkin, I., Ungerer, M. C., & Purugganan, M. (2007). Genetics of microenvironmental canalization in Arabidopsis thaliana. Proceedings of the National Academy

of Sciences of the United States of America, 104(34), 13717–13722. https://doi.org/10.1073/pnas.0701936104

Hervieux, N., Dumond, M., Sapala, A., Routier-Kierzkowska, A. L., Kierzkowski, D., Roeder, A. H., Smith, R. S., Boudaoud, A., & Hamant, O. (2016). A Mechanical Feedback Restricts Sepal Growth and Shape in Arabidopsis. Current biology : CB, S0960-9822(16)30180-4. Advance online publication. https://doi.org/10.1016/j.cub.2016.03.004

Hocq, L., Guinand, S., Habrylo, O., Voxeur, A., Tabi, W., Safran, J., Fournet, F., Domon, J. M., Mollet, J. C., Pilard, S., Pau-Roblot, C., Lehner, A., Pelloux, J., & Lefebvre, V. (2020). The exogenous application of AtPGLR, an endo-polygalacturonase, triggers pollen tube burst and repair. The Plant journal : for cell and molecular biology, 103(2), 617–633. https://doi.org/10.1111/tpj.14753

Hong, L., Dumond, M., Tsugawa, S., Sapala, A., Routier-Kierzkowska, A. L., Zhou, Y., Chen, C., Kiss, A., Zhu, M., Hamant, O., Smith, R. S., Komatsuzaki, T., Li, C. B., Boudaoud, A., & Roeder, A. H. (2016). Variable Cell Growth Yields Reproducible OrganDevelopment through Spatiotemporal Averaging. Developmental cell, 38(1), 15–32. https://doi.org/10.1016/j.devcel.2016.06.016

Horiguchi, G., Fujikura, U., Ferjani, A., Ishikawa, N., & Tsukaya, H. (2006). Large-scale histological analysis of leaf mutants using two simple leaf observation methods: identification of novel genetic pathways governing the size and shape of leaves. The Plant journal : for cell and molecular biology, 48(4), 638–644. https://doi.org/10.1111/j.1365-313X.2006.02896.x

Hou, Z., Liu, Y., Zhang, M., Zhao, L., Jin, X., Liu, L., Su, Z., Cai, H., & Qin, Y. (2021). High-throughput single-cell transcriptomics reveals the female germline differentiation trajectory in Arabidopsis thaliana. Communications biology, 4(1), 1149. https://doi.org/10.1038/s42003-021-02676-z

Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., & Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. PloS one, 5(9), e12776. https://doi.org/10.1371/journal.pone.0012776

Joseph, B., Corwin, J. A., & Kliebenstein, D. J. (2015). Genetic variation in the nuclear and organellar genomes modulates stochastic variation in the metabolome, growth, and defense. PLoS genetics, 11(1), e1004779. https://doi.org/10.1371/journal.pgen.1004779

Klingenberg, C. P., Barluenga, M., & Meyer, A. (2002). Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. Evolution; international journal of organic evolution, 56(10), 1909–1920. https://doi.org/10.1111/j.0014-3820.2002.tb00117.x

Kontarakis, Z., & Stainier, D. (2020). Genetics in Light of Transcriptional Adaptation. Trends in genetics : TIG, 36(12), 926–935. https://doi.org/10.1016/j.tig.2020.08.008

Lachowiec, J., Queitsch, C., & Kliebenstein, D. J. (2016). Molecular mechanisms governing differential robustness of development and environmental responses in plants. Annals of botany, 117(5), 795–809. https://doi.org/10.1093/aob/mcv151

Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics, 9, 559. https://doi.org/10.1186/1471-2105-9-559

La Rota, C., Chopard, J., Das, P., Paindavoine, S., Rozier, F., Farcot, E., Godin, C., Traas, J., & Monéger, F. (2011). A data-driven integrative model of sepal primordium polarity in Arabidopsis. The Plant cell, 23(12), 4318–4333. https://doi.org/10.1105/tpc.111.092619

Lê Cao, K. A., Rossouw, D., Robert-Granié, C., & Besse, P. (2008). A sparse PLS for variable selection when integrating omics data. Statistical applications in genetics and molecular biology, 7(1), . https://doi.org/10.2202/1544-6115.1390

Lee, T., & Lee, I. (2021). Genome-Wide Association Studies in Arabidopsis thaliana: Statistical Analysis and Network-Based Augmentation of Signals. Methods in molecular biology (Clifton, N.J.), 2200, 187–210. https://doi.org/10.1007/978-1-0716-0880-7_9

Li, E., Bhargava, A., Qiang, W., Friedmann, M. C., Forneris, N., Savidge, R. A., Johnson, L. A., Mansfield, S. D., Ellis, B. E., & Douglas, C. J. (2012). The Class II KNOX gene KNAT7 negatively regulates secondary wall formation in Arabidopsis and is functionally conserved in Populus. The New phytologist, 194(1), 102–115. https://doi.org/10.1111/j.1469-8137.2011.04016.x

Liu, J., Frochaux, M., Gardeux, V., Deplancke, B., & Robinson-Rechavi, M. (2020). Inter-embryo gene expression variability recapitulates the hourglass pattern of evo-devo. BMC biology, 18(1), 129. https://doi.org/10.1186/s12915-020-00842-z

Meyer, H. M., Teles, J., Formosa-Jordan, P., Refahi, Y., San-Bento, R., Ingram, G., Jönsson, H., Locke, J. C., & Roeder, A. H. (2017). Fluctuations of the transcription factor ATML1 generate the

pattern of giant cells in the Arabidopsis sepal. eLife, 6, e19131. https://doi.org/10.7554/eLife.19131

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L. P., Mushayamaha, T., & Thomas, P. D. (2021). PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic acids research, 49(D1), D394–D403. https://doi.org/10.1093/nar/gkaa1106

O'Malley, R. C., Huang, S. C., Song, L., Lewsey, M. G., Bartlett, A., Nery, J. R., Galli, M., Gallavotti, A., & Ecker, J. R. (2016). Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell, 165(5), 1280–1292. https://doi.org/10.1016/j.cell.2016.04.038

Powell, A. E., & Lenhard, M. (2012). Control of organ size in plants. Current biology : CB, 22(9), R360–R367. https://doi.org/10.1016/j.cub.2012.02.010

Qin, W., Yin, Q., Chen, J., Zhao, X., Yue, F., He, J., Yang, L., Liu, L., Zeng, Q., Lu, F., Mitsuda, N., Ohme-Takagi, M., & Wu, A. M. (2020). The class II KNOX transcription factors KNAT3 and KNAT7 synergistically regulate monolignol biosynthesis in Arabidopsis. Journal of experimental botany, 71(18), 5469–5483. https://doi.org/10.1093/jxb/eraa266

Queitsch, C., Sangster, T. A., & Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. Nature, 417(6889), 618–624. https://doi.org/10.1038/nature749

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Renard, J., Niñoles, R., Martínez-Almonacid, I., Gayubas, B., Mateos-Fernández, R., Bissoli, G., Bueso, E., Serrano, R., & Gadea, J. (2020). Identification of novel seed longevity genes related to oxidative stress and seed coat by genome-wide association studies and reverse genetics. Plant, cell & environment, 43(10), 2523–2539. https://doi.org/10.1111/pce.13822

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England), 26(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Roeder, A. (2021). Arabidopsis sepals: A model system for the emergent process of morphogenesis. Quantitative Plant Biology, 2, E14. doi:10.1017/qpb.2021.12

Sangster, T. A., Salathia, N., Lee, H. N., Watanabe, E., Schellenberg, K., Morneau, K., Wang, H., Undurraga, S., Queitsch, C., & Lindquist, S. (2008). HSP90-buffered genetic variation is common in Arabidopsis thaliana. Proceedings of the National Academy of Sciences of the United States of America, 105(8), 2969–2974. https://doi.org/10.1073/pnas.0712210105

Sasaki, E., Köcher, T., Filiault, D. L., & Nordborg, M. (2021). Revisiting a GWAS peak in Arabidopsis thaliana reveals possible confounding by genetic heterogeneity. Heredity, 127(3), 245–252. https://doi.org/10.1038/s41437-021-00456-3

Schmalhausen, I. I. (1949). Factors of evolution: the theory of stabilizing selection. Univ. of Chicago Press, Chicago, IL.

Schurr, U., Walter, A., & Rascher, U. (2006). Functional dynamics of plant growth and photosynthesis--from steady-state to dynamics--from homogeneity to heterogeneity. Plant, cell & environment, 29(3), 340–352. https://doi.org/10.1111/j.1365-3040.2005.01490.x

Sénéchal, F., Mareck, A., Marcelo, P., Lerouge, P., & Pelloux, J. (2015). Arabidopsis PME17 Activity can be Controlled by Pectin Methylesterase Inhibitor4. Plant signaling & behavior, 10(2), e983351. https://doi.org/10.4161/15592324.2014.983351

Sénéchal, F. (2013). Rôles des pectines méthylestérases (PMEs) dans le développement chez Arabidopsis thaliana. Étude de leur régulation par les inhibiteurs (PMEIs) et protéases de type subtilisines (SBTs). Université de Picardie Jules Verne, Amiens.

Smyth, D. R., Bowman, J. L., & Meyerowitz, E. M. (1990). Early flower development in Arabidopsis. The Plant cell, 2(8), 755–767. https://doi.org/10.1105/tpc.2.8.755

Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research, 4, 1521. https://doi.org/10.12688/f1000research.7563.2

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., & Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic acids research, 45(W1), W122–W129. https://doi.org/10.1093/nar/gkx382

Timmons, J. A., Szkop, K. J., & Gallagher, I. J. (2015). Multiple sources of bias confound functional enrichment analysis of global -omics data. Genome biology, 16(1), 186. https://doi.org/10.1186/s13059-015-0761-7

Tsukaya H. (2002). Interpretation of mutants in leaf morphology: genetic evidence for a compensatory system in leaf morphogenesis that provides a new link between cell and organismal theories. International review of cytology, 217, 1–39. https://doi.org/10.1016/s0074-7696(02)17011-2

Tsukaya H. (2003). Organ shape and size: a lesson from studies of leaf morphogenesis. Current opinion in plant biology, 6(1), 57–62. https://doi.org/10.1016/s1369526602000055

Tzafrir, I., Dickerman, A., Brazhnik, O., Nguyen, Q., McElver, J., Frye, C., Patton, D., & Meinke, D. (2003). The Arabidopsis SeedGenes Project. Nucleic acids research, 31(1), 90–93. https://doi.org/10.1093/nar/gkg028

Van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. Scotts Valley, CA: CreateSpace.

Waddington, C. H. (1953). Assimilation of an acquired character. Evolution 7(2): 118–126.

Waddington, C. H. (1959). Canalization of development and genetic assimilation of acquired characters. Nature, 183(4676), 1654–1655. https://doi.org/10.1038/1831654a0

Wang, S., Yamaguchi, M., Grienenberger, E., Martone, P. T., Samuels, A. L., & Mansfield, S. D. (2020). The Class II KNOX genes KNAT3 and KNAT7 work cooperatively to influence deposition of secondary cell walls that provide mechanical support to Arabidopsis stems. The Plant journal : for cell and molecular biology, 101(2), 293–309. https://doi.org/10.1111/tpj.14541

Weiss, J., Delgado-Benarroch, L., & Egea-Cortines, M. (2005). Genetic control of floral size and proportions. The International journal of developmental biology, 49(5-6), 513–525. https://doi.org/10.1387/ijdb.051998jw

Wu, S., Alseekh, S., Cuadros-Inostroza, Á., Fusari, C. M., Mutwil, M., Kooke, R., Keurentjes, J. B., Fernie, A. R., Willmitzer, L., & Brotman, Y. (2016). Combined Use of Genome-Wide Association Data and Correlation Networks Unravels Key Regulators of Primary Metabolism in Arabidopsis thaliana. PLoS genetics, 12(10), e1006363. https://doi.org/10.1371/journal.pgen.1006363

Zalts, H., & Yanai, I. (2017). Developmental constraints shape the evolution of the nematode mid-developmental transition. Nature ecology & evolution, 1(5), 113. https://doi.org/10.1038/s41559-017-0113

Zheng, Y., Roberts, R. J., & Kasif, S. (2004). Segmentally variable genes: a new perspective on adaptation. PLoS biology, 2(4), E81. https://doi.org/10.1371/journal.pbio.0020081

Zhong, R., & Ye, Z. H. (2015). Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. Plant & cell physiology, 56(2), 195–214. https://doi.org/10.1093/pcp/pcu140

Zhu, M., Chen, W., Mirabet, V., Hong, L., Bovio, S., Strauss, S., Schwarz, E. M., Tsugawa, S., Wang, Z., Smith, R. S., Li, C. B., Hamant, O., Boudaoud, A., & Roeder, A. (2020). Robust organ size requires robust timing of initiation orchestrated by focused auxin and cytokinin signalling. Nature plants, 6(6), 686–698. https://doi.org/10.1038/s41477-020-0666-7

Supplementary Figures

Figure 1 - Figure supplement 1: Validation of our RNA-Seq data by comparing squared coefficient of variation (top) and average gene expression (bottom) for our 14085 genes (in red) with equivalent data (15646 genes) for Arabidopsis seedlings from Cortijo et al., 2019 (in black). Results are shown for each "original dataset" (solid lines) and for genes in "common" (dashed lines) between both datasets (13013 genes). CV^2 values are comparable and have a strong correlation (top right). Mean gene expression values extracted from read counts show very similar distributions and a strong correlation (bottom right) but higher coverage in our case.

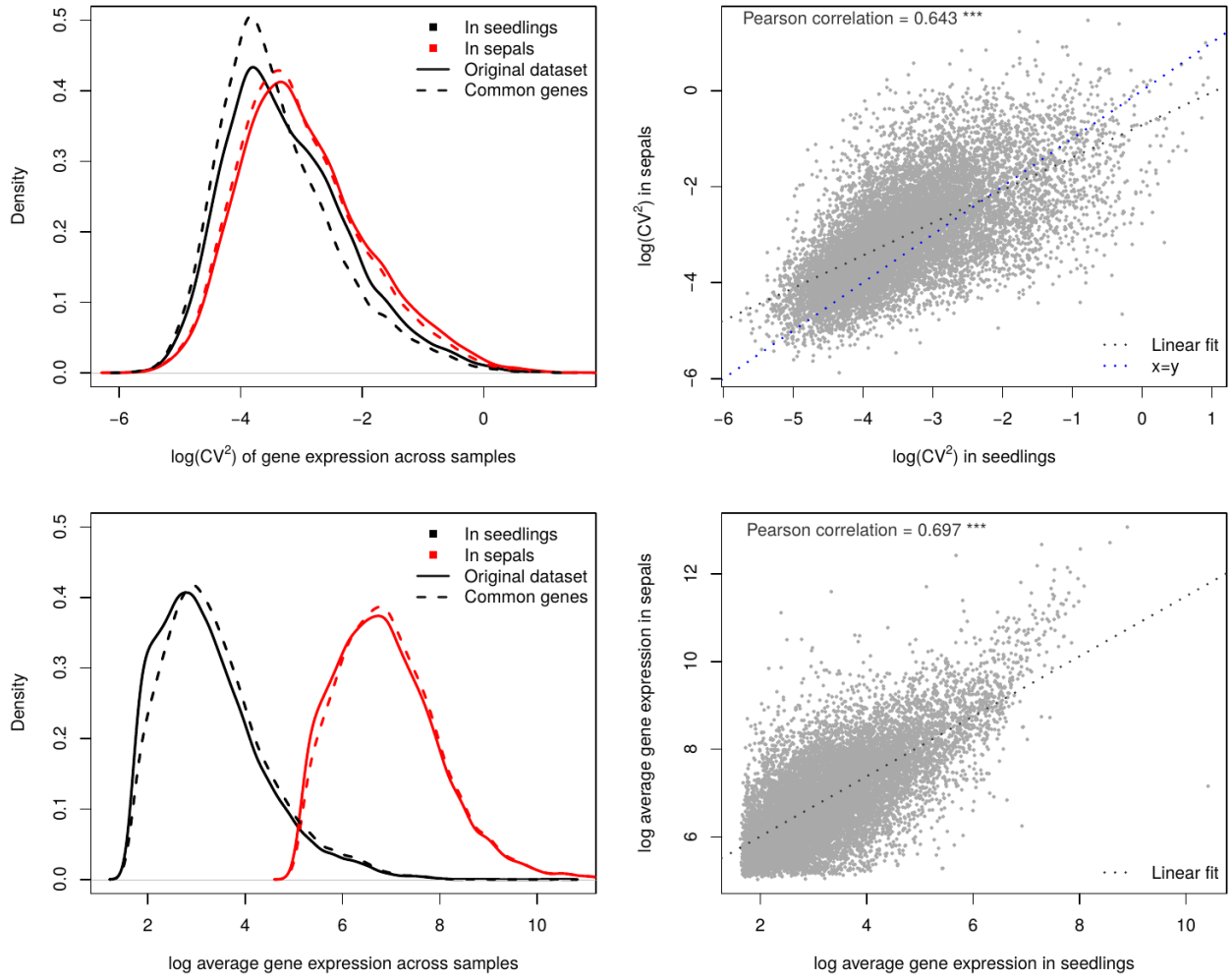

Figure 2 - Figure supplement 1: PCA and loadings on morphological parameters. Top: PC1 vs PC3. Bottom: PC2 vs. PC3.

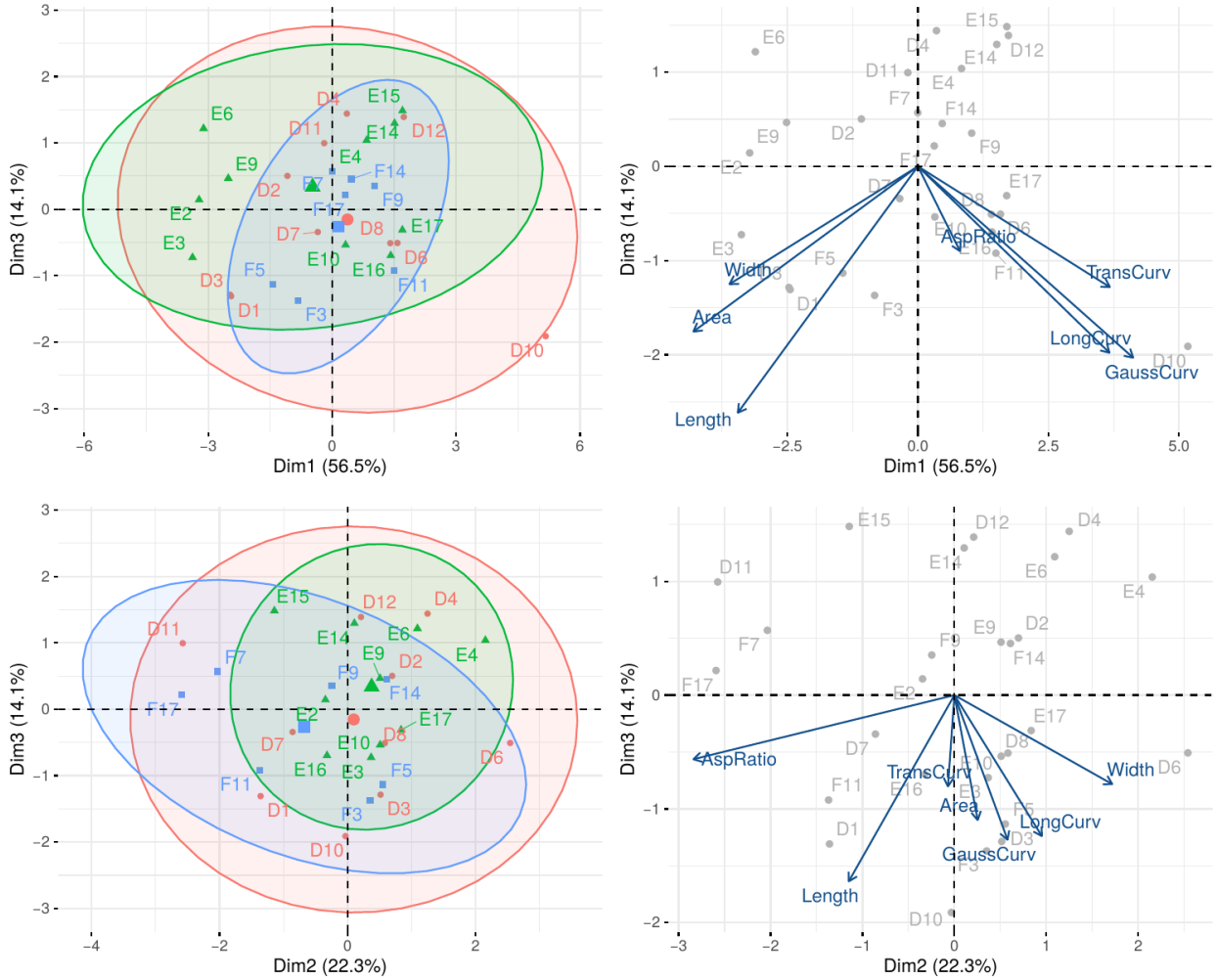

Figure 3 - Figure supplement 1: PCA on RNA-Seq for 14085 genes across 27 samples.

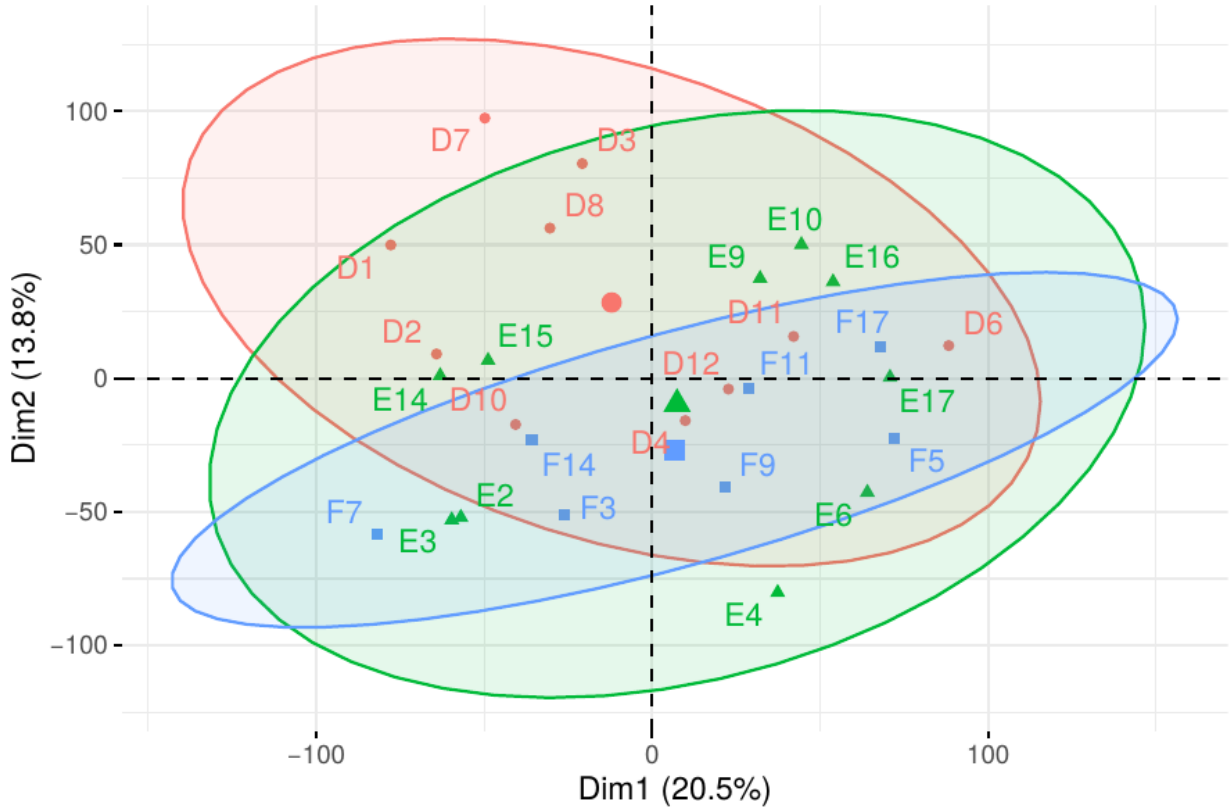

Figure 4 - Figure supplement 1: Module membership against trait significance for top correlated morphological parameter and normalized log2 gene expression values for 27 samples, for magenta, orange, cyan and green modules. Gene expression values are compared with the corresponding trait parameter with a scale chosen so as to depict high correlation with gene expression for the highlighted genes. A,D,E,F) Only genes with module membership above 0.7 and trait significance above 0.35 for the corresponding parameter (filled circles in the left plot) are plotted in the gene expression plot (right plot). B) Only genes with module membership below 0.7 and trait significance below 0.35 for the corresponding parameter (filled circles in the left plot) are plotted in the gene expression plot (right plot). C) Only genes with module membership above 0.7, trait significance above 0.35 for the corresponding parameter (filled circles in the left plot), and under GO term "cell wall organization and biogenesis" are plotted in the gene expression plot (right plot). Transcription factors NAC007 and KNAT7 are highlighted in red and yellow, respectively.

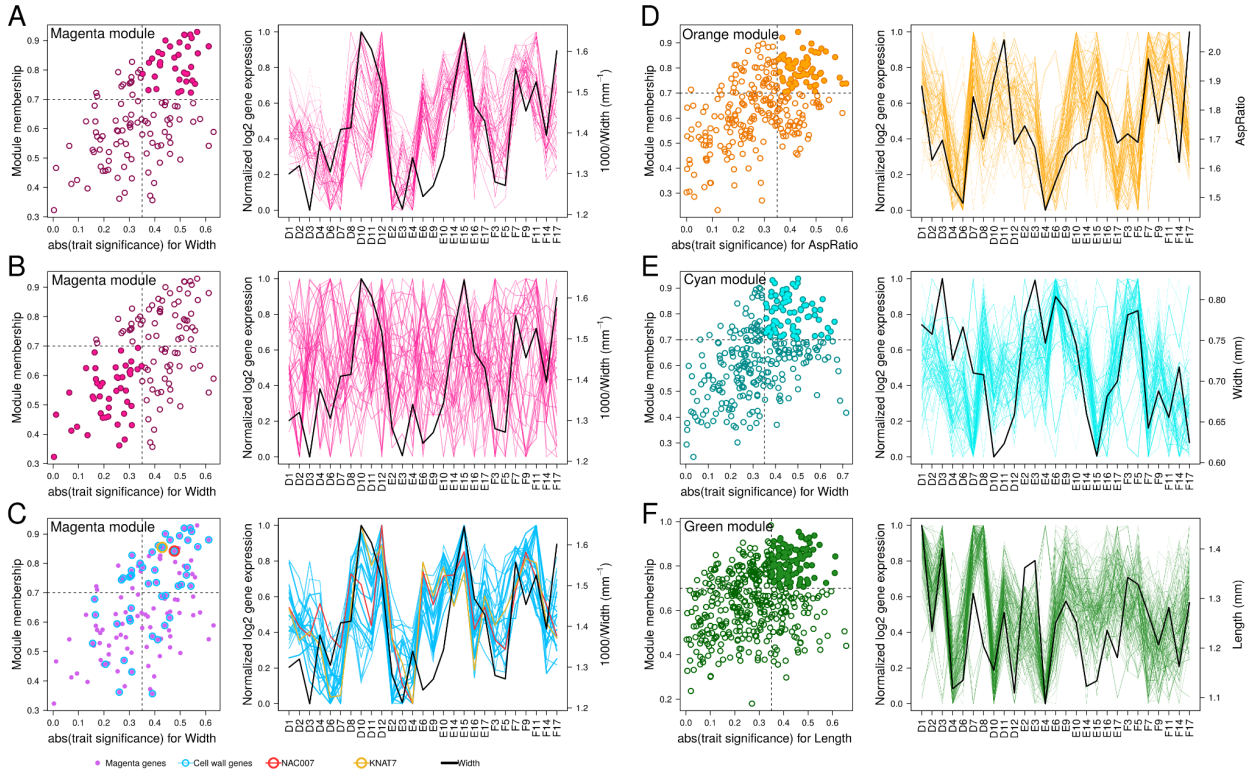
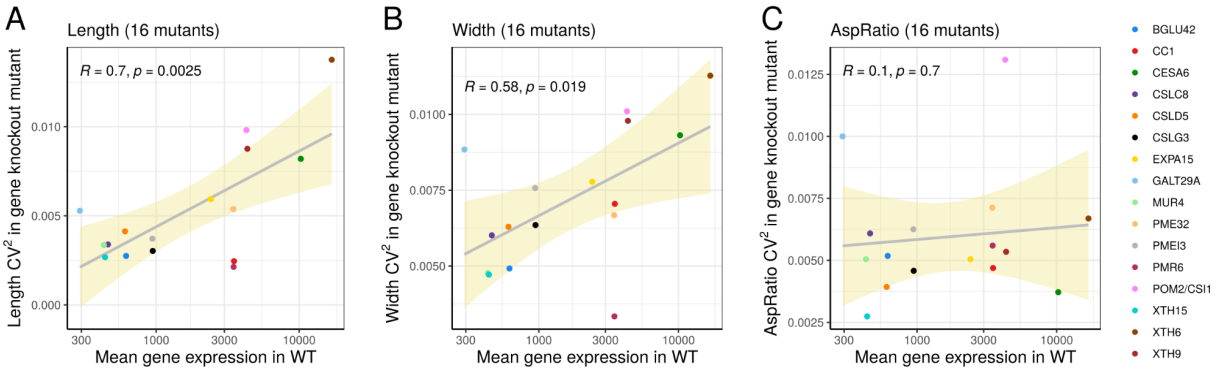
Figure 7 - Figure supplement 1: Mutant sepal variation in phenotype for 16 cell-wall related mutants against mean gene expression of the corresponding knocked-out gene in wild-type for length (A), width (B) and aspect ratio (C).