

## ***SI Appendix***

*Title:* Transcriptome dynamics at *Arabidopsis* graft junctions reveal an inter-tissue recognition mechanism that activates vascular regeneration

*Authors:* Charles W Melnyk<sup>1,2</sup>, Alexander Gabel<sup>1,3</sup>, Thomas J Hardcastle<sup>4</sup>, Sarah Robinson<sup>5</sup>, Shunsuke Miyashima<sup>6</sup>, Ivo Grosse<sup>3,7,8</sup>, Elliot M Meyerowitz<sup>1,8,9</sup>

### *Author Affiliations:*

1- Sainsbury Laboratory, University of Cambridge, Bateman Street, Cambridge CB2 1LR, UK

2 – Department of Plant Biology, Swedish University of Agricultural Sciences, Almas allé 5, 756 51 Uppsala, Sweden

3 - Institute of Computer Science, Martin Luther University Halle–Wittenberg, 06120 Halle (Saale), Germany

4 – Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA, United Kingdom

5 - Institute of Plant Science, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland

6 – Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan

7 – German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany

8 – Division of Biology and Biological Engineering, California Institute of Technology, 1201 East California Boulevard, Pasadena, CA 91125, USA

9 – Howard Hughes Medical Institute, California Institute of Technology, 1201 East California Boulevard, Pasadena, CA 91125, USA

### *Correspondence:*

Charles W. Melnyk, Department of Plant Biology, Swedish University of Agricultural Sciences, Almas allé 5, 756 51 Uppsala, Sweden, [charles.melnyk@slu.se](mailto:charles.melnyk@slu.se)

## SI Materials and Methods

### Plant material and grafting

*Arabidopsis thaliana* accession Columbia was used throughout except where indicated. The *p35S::GFP-ER* (1), *pSUC2::GFP* (2), *pUBQ10::PM-tdTomato* (3), *pANT::H2B-YFP* (4), *pLOG4::n3GFP* (5), *pCASP1::NLS-GFP* (6), *p35S::DII-Venus* (7), *pDR5rev::GFP-ER* (8), *hca2* (9) and *p35S::HCA2-SRD5* (9) lines have been previously published. *pARR5::GFP* (10) is previously published and is in the *Ws* background. For the construction of *pHCA2::RFP*, a 2.9kb 5' upstream region of the *HCA2* gene (At5g62940) was cloned into pDONRp4-p1R donor vector, and recombined with tagRFPper into a destination vector by the Multisite Gateway system (11). The following primers were used for *pHCA2* cloning:

attB4\_HCA2(-)2958: GGGACAACCTTTGTATAGAAAAGTTGtcgatacggcgacagatatac

attB1\_HCA2ProEnd: ggggACTGCTTTTTTGTACAAACTTGttttgtgtctgtatgtttg.

*Arabidopsis thaliana* micrografting was performed according to a previously published protocol (12). Briefly, seven day old *Arabidopsis* seedlings were grown vertically on ½ strength Murashige and Skoog (MS) medium + 1% bacto agar (pH5.7; no sucrose) in short day conditions (8 hours of 80-100  $\mu\text{mol m}^{-2} \text{s}^{-1}$  light) at 20°C. Seedlings were placed on one layer of 2.5x4cm sterile Hybond N membrane (GE Healthcare) on top of two 8.5cm circles of sterile 3 Chr Whatman paper (Scientific Laboratory Supplies) moisten with sterile distilled water in a 9cm petri dish. In a laminar flow hood using a dissecting microscope, one cotyledon was removed (intact treatment) and, in some treatments, a transverse cut through the hypocotyl was made with a vascular dissecting knife (Ultra Fine Micro Knife; Fine Science Tools). Plants were left separated (separated treatment) or they were assembled by aligning the two cut halves and joining them together (grafted treatment), after which, the petri dishes was sealed with parafilm and placed vertically under short day conditions at 20°C. For grafting on a microscope coverslip to image the graft junction, a 10cm square Petri dish was modified by gluing a microscope coverslip in place of a section of plastic from the back. On top of the microscope cover slip was placed a 2.5x4 cm rectangle of Hybond N membrane. At the edges and base of the Petri dish three 3 x 8cm strips of Whatman paper were placed. Sterile water moistened both Whatman paper and Hybond N. After which, roots were placed on the Hybond N membrane and hypocotyls on the coverslip. Grafting then proceeded as above. Graft junctions were imaged through the coverslip with a Plan-Apochromat 20X/0.8 objective on a Zeiss LSM-700 or LSM-780 confocal microscope.

For auxin treatment assays, *pHCA2::RFP* and *pDR5::GFP* expressing plants were grown vertically on ½ MS medium + 1% bacto agar (pH5.7; no sucrose) in short day conditions (8 hours of 80-100  $\mu\text{mol m}^{-2} \text{s}^{-1}$  light) at 20°C. After 15 days, hypocotyls were cut using a vascular dissecting knife (Ultra Fine Micro Knife; Fine Science Tools) and placed on ½ MS + 1% bacto agar (pH5.7) media containing either DMSO, 1% sucrose or 1 $\mu\text{M}$  of the synthetic auxin 1-Naphthaleneacetic acid (NAA). After 1 or 2 days, cut hypocotyls were imaged on a LSM-780 confocal microscope. For root tip assays, *pHCA2::RFP* and *pDR5::GFP* expressing plants were grown vertically on ½ MS medium + 1% bacto agar (pH5.7; no sucrose) in short day conditions (8 hours of 80-100  $\mu\text{mol m}^{-2} \text{s}^{-1}$  light) at 20°C. After 10 days, plants were moved to ½ MS + 1% bacto agar (pH5.7; no sucrose) media containing either DMSO or 1 $\mu\text{M}$  NAA and imaged on a LSM-780 confocal microscope after one day.

### **Fluorescent assays and microscopy**

To test the effect of sugars on grafting, *Arabidopsis thaliana* Col-0 plants were grown on 1/2MS+ 1% bacto agar (pH5.7; no sucrose) for seven days in short day conditions. Wild type roots were grafted to scions expressing *pSUC2::GFP* using the protocol described above but either water or water containing 0.25%, 0.5% 1% or 2% sucrose was added to the grafted plates. Roots were observed for fluorescence 2-7 days after grafting with a Zeiss V12 dissecting microscope equipped with a GFP filter. Roots were scored daily and the same plants (n=24 for each sucrose concentration) were observed during the 7-day assay as previously described (13). To test the effect of *HCA2* upon grafting, wild type or mutant roots were grafted to scions expressing *pSUC2::GFP*. Roots were observed for fluorescence 3-7 days after grafting. Roots were scored daily and the same plants (n=71-72 for each genotype) were observed during the 7-day assay as previously described (13).

Fluorescent images were taken on a Zeiss LSM-700 or LSM-780 confocal microscope with a Zeiss Plan-Apochromat 20X/0.8 dry objective. A 488nm argon laser (Zeiss 780) or 488nm solid-state laser (Zeiss 700) was used for excitation of GFP and YFP. A 561nm solid-state laser was used for excitation of the tdTomato fluorescent protein. A T-PMT detector obtained bright-field transmitted light. Black and white fluorescent images of graft junctions were taken on a Zeiss V12 dissecting microscope fitted with a Hamamatsu EM-CCD camera and RFP and YFP filters. FIJI software (Fiji.sc) was used to process images. Image contrast and brightness were adjusted for controls and samples equally. For longitudinal images of the graft junction, z-stack projections are shown and made with the average intensity function in

FIJI from stacks containing the hypocotyl vascular tissues, mesophyll and epidermis. For images of *pHCA2::RFP* and *pDR5::GFP* hypocotyls, z-stack projections are shown and made with the average intensity function in FIJI from stacks containing the vascular tissues.

### **RNAseq sample and library preparation**

Wild type *Arabidopsis thaliana* accession Col-0 were grafted as above taking care to switch shoot and root between different plants. All grafting and cutting was performed in the morning to minimize circadian effects. For the 0 hour time points, plants were transferred from 1/2MS plates to the grafting plates and immediately harvested. Only intact plants were harvested at the 0 hour time since at this point, there would be insufficient time to reasonably expect the separated or grafted samples to be transcriptionally different (the time between cutting and freezing is less than two minutes). For all other time points, intact, grafted or separated plants were left on the grafting plates for the respective amount of time indicated. Tissues were harvested and care taken to separate grafts by gently pulling plants apart. Approximately 0.5mm of tissue was taken above or below each cut site and kept separate. Intact plants had 1mm of tissue taken in a similar location on the hypocotyl as separated or grafted plants. Grafted, separated or intact tissues were pooled into groups of approximately 24 tissues (1 plates with 24 plants) which were immediately placed in 96% ethanol on dry ice. After harvesting, microcentrifuge tubes were briefly centrifuged and the ethanol removed before storing at -80°C. Plants were grafted over two months to get sufficient material.

Tissues were ground in the microcentrifuge tube using a microcentrifuge pestle frozen in liquid nitrogen. RNA was extracted using an RNeasy Kit (Qiagen, UK) following the manufacturer's instructions including on column DNase digestion. RNA was eluted from the column with 50ul of sterile water. Quality and quantity of RNA was checked using an Agilent 2200 TapeStation and High Sensitivity (HS) RNA screentapes (Agilent, UK). After RNA extraction, two to four biological replicates were combined (50-100 plants) to get sufficient RNA. 90-100ng of RNA was used to prepare RNAseq libraries using the TruSeq® Stranded mRNA LT kit (Illumina, UK) according to the manufacturer's instructions. The final PCR was for 15 cycles and samples were resuspended in 23ul of distilled water. Quantity and quality of DNA libraries was checked on the Agilent 2200 TapeStation using D1000 screentapes (Agilent, UK). Each sample had two libraries prepared from grafted tissues or separated tissues at different times so that independent biological replicates were made and sequenced. Samples were diluted to 10nM and 11-12 barcoded samples randomly

mixed to make a total of 7 mixes for 7 flow lanes, one mix per lane. Samples were sequenced on the HiSeq 4000 platform (Illumina, UK) with Paired End 75bp transcriptome sequencing (BGI Tech Solutions, Shenzhen, China). RNAseq data are available on the Gene Expression Omnibus (GSE107203).

### **Iodine staining**

*Arabidopsis* seedlings were placed in a fixation solution (3.7% formaldehyde, 50% ethanol, 5% acetic acid) for 1 hour at room temperature, then transferred to 70% ethanol for 10 minutes. Afterwards, plants were transferred to 96% ethanol and stored at -20°C for up to a week. Samples were rehydrated in 50% ethanol for 1 hour at room temperature, then transferred to distilled water for 30 minutes. Samples were then transferred to a solution of Lugol solution (Sigma) and stained for 10 minutes at room temperature. Plants were rinsed with water, then mounted on microscope slides. Images were taken on a Zeiss Axioimager.M2 microscope with a PlanApochromat 20x objective and SPOT Flex camera (ImSol, UK).

### **Pairwise and baySeq analyses**

The reads acquired through high-throughput sequencing were quality trimmed with sickle (14) to increase the read quality before mapping. Reads were aligned to protein-coding gene sequences acquired from TAIR10 using Bowtie2. Read assignment was performed using the eXpress tool, which also provided effective gene lengths for use in normalisation. Library scaling factors were inferred from the sum of the number of reads assigned to the genes in the lowest seventy-five percentiles of expressed genes for each library (15). Figures were generated from library length normalised data (Dataset S4). Non-length normalised data are also available (Dataset S5).

Analyses of the data were carried out using the R package baySeq (16) and clustering based on the posterior probabilities acquired from this package and the clusterSeq package (17). For each timepoint, all possible patterns of differential expression between the graft types were considered, where a ‘pattern’ defines similarity and difference between different experimental conditions. For example,

*{Col\_cut\_bottomGenes=Col:Col\_bottomGenes=ungraftedGenes},{Col\_cut\_topGenes=Col:Col\_topGenes}*

defines a pattern in which gene expression is equivalent in the separated bottoms (*Col\_cut\_bottom*), the grafted bottoms (*Col:Col\_bottom*) and the intact plant (*ungrafted*), but

different to the equivalently expressed separated top (*Col\_cut\_top*) and grafted top (*Col:Col\_top*). The total number of possible patterns for five experimental conditions (as in this analysis) is fifty-two.

For a given timepoint, posterior likelihoods on the likelihood of each pattern of expression are calculated for every gene with greater than ten reads across all experimental conditions. The patterns were then modified to include orderings (denoted by < or >), for example, the pattern described would lead to the ordered pattern

*'{Col\_cut\_bottomGenes=Col:Col\_bottomGenes=ungraftedGenes}>{Col\_cut\_topGenes=Col:Col\_topGenes}'*

in which gene expression is equivalent in the separated bottoms, the grafted bottoms and the intact plant and greater than the equivalently expressed separated top and grafted top. In total, 541 ordered patterns exist in this data set. Posterior likelihoods for an ordered pattern were assigned to that of the unordered pattern for genes in which the (normalised) mean expressions within the equivalently expressed groups conformed to the ordering, and to zero otherwise.

Based on the posterior likelihoods for the ordered patterns, a similarity score  $s_{ij}$  was established between two genes  $i$  and  $j$  as the sum over the products of their likelihoods of each ordered pattern. A single-link agglomerative clustering of genes, in which a gene will join a cluster if it has a greater than 50% similarity to any gene within that cluster was then performed based on these similarity scores. We label each cluster according to the predominant ordered pattern with high likelihood amongst the genes that comprise it. The change in size of these clusters over time is shown for the major clusterings in Fig. 6.

We can also find likelihoods on comparisons between pairs of experimental conditions by summing the likelihoods over combinations of patterns. Fig. 3A shows the number of genes identified at each time point in a pairwise analysis between the grafted top and grafted bottom samples. The likelihood of symmetric expression (i.e., expression which is equivalent across the graft junction) is calculated as the sum of the likelihoods of all patterns in which the grafted top and grafted bottom samples are equivalent. Conversely, asymmetric expression is calculated as the sum of the likelihoods of all patterns in which the grafted top and grafted bottom samples are not equivalent. Additional sets can be formed by considering the ordering of the grafted top and grafted bottom samples. Sets of genes are identified at each time point with an FDR of less than 0.05 and a likelihood of symmetric/asymmetric expression greater

than 50%. Genes in this analysis were only included if they were differentially expressed relative to intact samples.

We considered the case where differential expression across the graft junction represents a normal gradient of expression by comparing the combined read counts from the top and bottom of both the grafted and separated cases to those observed intact plant, using baySeq to calculate all possible patterns of differential expression between these three cases (grafted, separated, and intact). When the library scaling factors are adjusted to reflect this summation, this is equivalent to considering average expression across the graft/separation. For each gene, we calculate the posterior likelihood that there is no differential expression in this averaged comparison but that there is differential expression when analysing the non-averaged data. We select the top genes meeting these criteria on a FDR of 5% and show the results in Fig. S2C.

### **Gene overlap analyses of up and down regulated genes**

To measure if the ratio of up- and down- regulated genes from a previously published dataset is significantly different to the ratio of up- and down-regulated genes in our grafting dataset we only took into account genes that are differentially expressed at a certain time point. A gene was called differentially expressed at a certain time point if the marginal likelihood, calculated by baySeq, was greater than 0.9 and if the absolute  $\log_2$ -foldchange was greater than 1. Hence, we only consider genes that are significantly two-fold up- or down- regulated based on the gene length normalised count data. This definition of differentially expressed genes was also used to filter the published datasets according to the expression values in our transcriptome dataset. Hence, some genes were filtered out from the original published datasets because they did not show a significant up- or down- regulation during a certain time point in our expression data based on our criteria. The histograms (Fig. 2, Fig. 3, Fig. 4, Fig. S4, Fig. S5, Fig. S6 and Fig. S7) show the relative number of up- and down- regulated genes from the published datasets during a certain time point and a certain condition (separated top, separated bottom, grafted top, grafted bottom) based on the number of genes in the published dataset after filtering. To calculate the significance of the difference of the ratios between the published DEGs and all up- and down- regulated genes, we used a two-sided Fisher's exact test. To correct for multiple testing we used the Benjamini-Yekutieli (BY) correction method. Hence, the asterisk above each bar highlights that the corrected p-value is below 0.05.

### **Dealing with probe ids from microarray datasets**

Due to the fact that some published datasets only used probe ids instead of gene ids to represent their differentially expressed genes we first had to match these probe ids to their corresponding gene ids. This step was done with the R package biomatr (18). If one probe id matched more than one gene id we used all the corresponding gene ids and tested afterwards if these genes were actually differentially expressed in our dataset. In some cases, one probe id was represented by more than one gene id. Hence, some gene sets contained slightly more gene ids than published probe ids. In contrast, some probe ids did not match to currently existing gene ids. Hence, some gene sets contained slightly fewer gene ids than published probe ids (Dataset S1).

### **Gene overlap analyses of gene sets involved in graft formation**

Grafting-specific genes (Fig. 7 and Fig. S9) were identified by taking clusters from the baySeq analysis that were specific to grafting (Dataset S2) and combining these clusters to generate a list of grafting-specific genes for which further analyses were performed. For calculating the significance of overlapping genes between the baySeq clusters and the published datasets a one-sided Fisher's exact test was applied, to prove if the overlap is greater than expected. The resulting p-values were corrected for multiple testing by using the Benjamini-Yekutieli method. This procedure was also applied to generate Table S1 to study the overlaps of symmetrically and asymmetrically expressed genes in the grafting dataset with previously published sugar-responsive genes.

### **GO enrichment analysis**

The gene ontology (GO) enrichment analysis on grafting-specific genes was done with a customized R script using the package GOstats (19). Gene ontology annotation was used from the Bioconductor package org.At.tair.db (20). The p-values calculated by a hypergeometric test were corrected for multiple testing with the Bonferroni correction. A GO category was called enriched if the corrected p-value was below 0.05.

A detailed description, the required data and R scripts to reproduce the clustering (hierarchical clustering and PCA), the statistical analyses regarding the overlap studies, and the GO enrichment analysis are available via the GitHub repository <https://github.com/AlexGa/GraftingScripts>.



### Breaking force measurements

Breaking force was calculated using a micro-extensometers as in (21) to apply force to either side of the graft junction until it broke. Briefly, grafted Col-0 plants were attached to a moving plate using tough tags (0.94X0.50 inches, white, 679 catalog no. TTSW-1000, DiversifiedBiotech) and a cyanoacrylate glue. The plates were moved apart and the force was measured using a force sensor (Futek LSB200 10g load cell, Futek Inc.). Upon breaking the force dropped. The maximum force measured before breaking was recorded. Images were captured using a Leica SP5 confocal microscope. During the experiment single z-plane images were captured and made into a movie in ImageJ (Movie S1, Movie S2).

To measure levels of tissue contamination between grafted top and bottom, grafted green fluorescent protein(*p35S::GFP*) and red fluorescent protein (*pUBQ10::PM-tdTomato*) expressing plants were pulled apart manually and imaged on a Zeiss LSM 700 confocal microscope. Z-stacks comprising the majority of the hypocotyl were made and regions approximately 0.5mm above and 0.5mm below the graft junction were made into projections using the average intensity function on FIJI software, and then the mean intensity quantified for red and green channels on FIJI. The mean intensity of one colour in the tissue tested for contamination was divided by the mean intensity of the same colour in the tissue expressing that transgenes to get a percentage of contamination. A region away from the cut site was also quantified to get a percentage of spectral overlap between red and green channels. The percent spectral overlap was then subtracted from the percentage contamination to get an overall percentage for how much contamination was present.

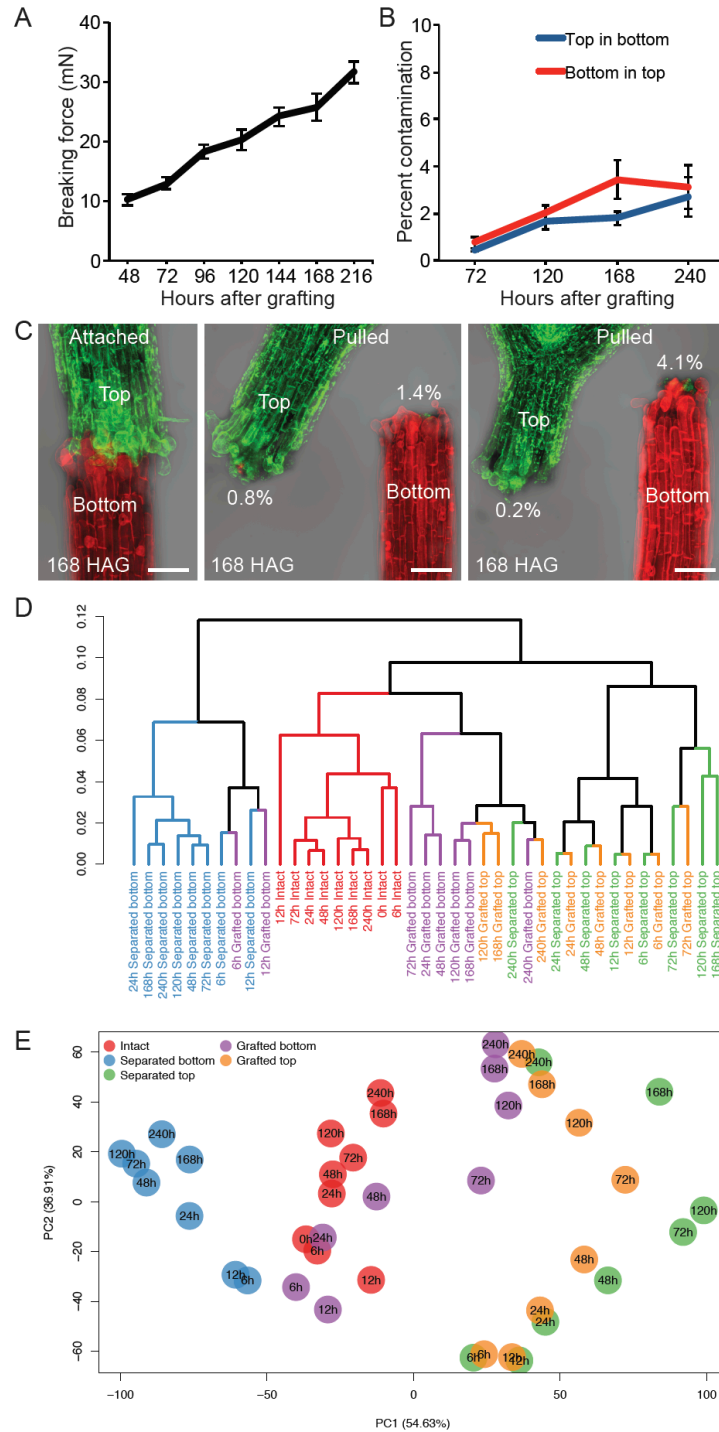
### Supplemental references

1. Nelson BK, Cai X, & Nebenfuhr A (2007) A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. *The Plant journal* 51(6):1126-1136.
2. Imlau A, Truernit E, & Sauer N (1999) Cell-to-cell and long-distance trafficking of the green fluorescent protein in the phloem and symplastic unloading of the protein into sink tissues. *Plant Cell* 11(3):309-322.
3. Segonzac C, *et al.* (2012) The shoot apical meristem regulatory peptide CLV3 does not activate innate immunity. *Plant Cell* 24(8):3186-3192.
4. Randall RS, *et al.* (2015) AINTEGUMENTA and the D-type cyclin CYCD3;1 regulate root secondary growth and respond to cytokinins. *Biol Open*.
5. De Rybel B, *et al.* (2014) Plant development. Integration of growth and patterning during vascular tissue formation in Arabidopsis. *Science* 345(6197):1255-1261.

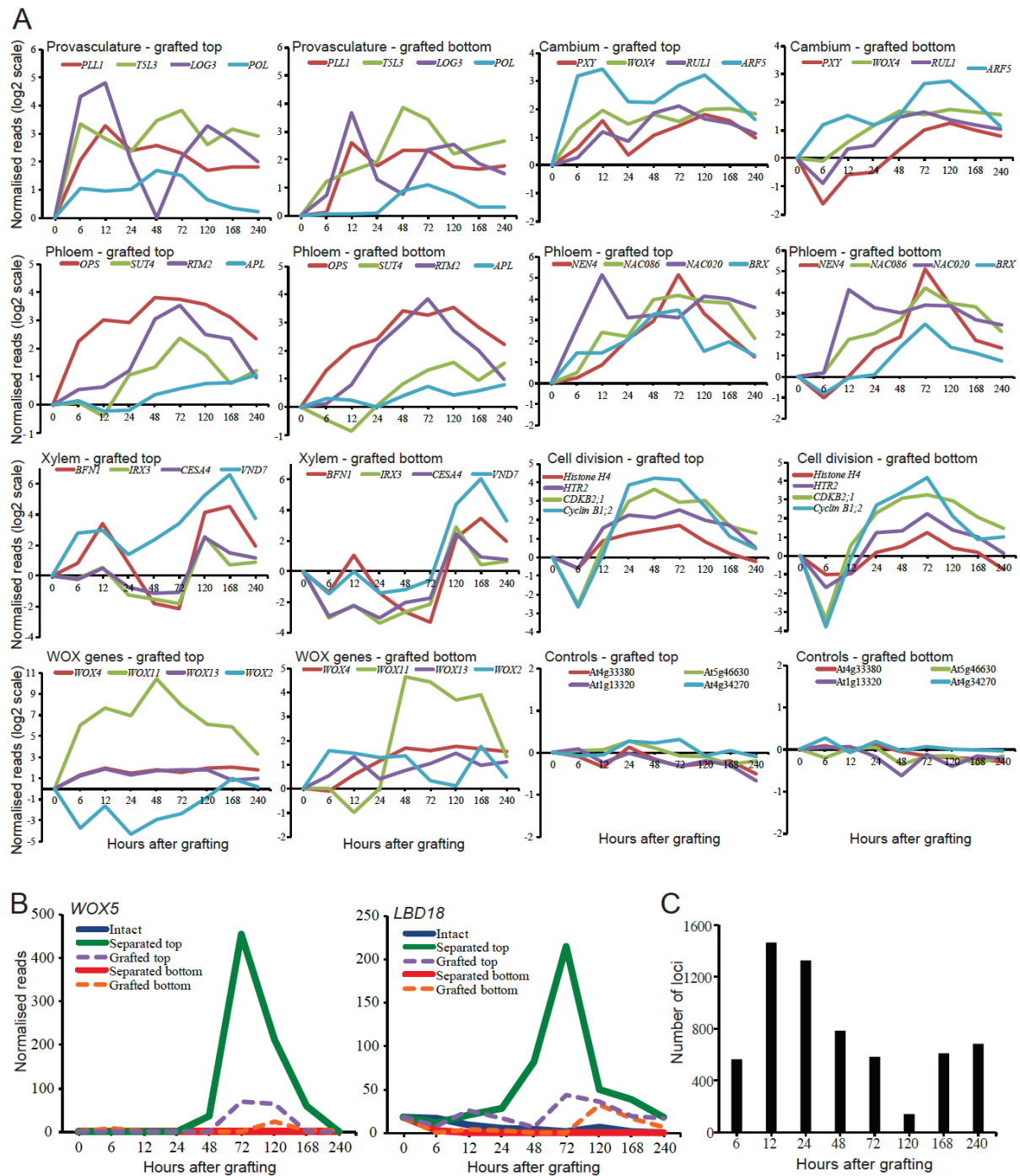
6. Roppolo D, *et al.* (2011) A novel protein family mediates Casparian strip formation in the endodermis. *Nature* 473(7347):380-383.
7. Brunoud G, *et al.* (2012) A novel sensor to map auxin response and distribution at high spatio-temporal resolution. *Nature* 482(7383):103-106.
8. Friml J, *et al.* (2003) Efflux-dependent auxin gradients establish the apical-basal axis of Arabidopsis. *Nature* 426(6963):147-153.
9. Guo Y, Qin G, Gu H, & Qu LJ (2009) Dof5.6/HCA2, a Dof transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in Arabidopsis. *Plant Cell* 21(11):3518-3534.
10. Yanai O, *et al.* (2005) Arabidopsis KNOXI proteins activate cytokinin biosynthesis. *Curr Biol* 15(17):1566-1571.
11. Siligato R, *et al.* (2016) MultiSite Gateway-Compatible Cell Type-Specific Gene-Inducible System for Plants. *Plant Physiol* 170(2):627-641.
12. Melnyk CW (2017) Grafting with Arabidopsis thaliana. *Methods Mol Biol* 1497:9-18.
13. Melnyk CW (2017) Monitoring Vascular Regeneration and Xylem Connectivity in Arabidopsis thaliana. *Methods Mol Biol* 1544:91-102.
14. Joshi NA & Fass JN (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). <https://github.com/najoshi/sickle>
15. Hardcastle TJ, Kelly KA, & Baulcombe DC (2012) Identifying small interfering RNA loci from high-throughput sequencing data. *Bioinformatics* 28(4):457-463.
16. Hardcastle TJ & Kelly KA (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics* 11:422.
17. Hardcastle TJ & Papatheodorou I (2017) ClusterSeq: methods for identifying co-expression in high-throughput sequencing data. *bioRxiv* 188581; doi: <https://doi.org/10.1101/188581>
18. Drost HG & Paszkowski J (2017) Biomart: genomic data retrieval with R. *Bioinformatics* 33(8):1216-1217.
19. Falcon S & Gentleman R (2007) Using GStats to test gene lists for GO term association. *Bioinformatics* 23(2):257-258.
20. Carlson M (2017) org.At.tair.db: Genome wide annotation for Arabidopsis), R package version 3.4.1.
21. Robinson S, *et al.* (2017) An automated confocal micro-extensometer enables in vivo quantification of mechanical properties with cellular resolution. *Plant Cell* 29: 2959-2973.
22. Osuna D, *et al.* (2007) Temporal responses of transcripts, enzyme activities and metabolites after adding sucrose to carbon-deprived Arabidopsis seedlings. *Plant J* 49(3):463-491.

**Table S1. Sugar response overlaps with asymmetry.** Symmetrically and asymmetrically differentially expressed genes were compared to previously published sugar-responsive genes (ref 22) and the percent overlap calculated (\* p<.05). HAG, hours after grafting.

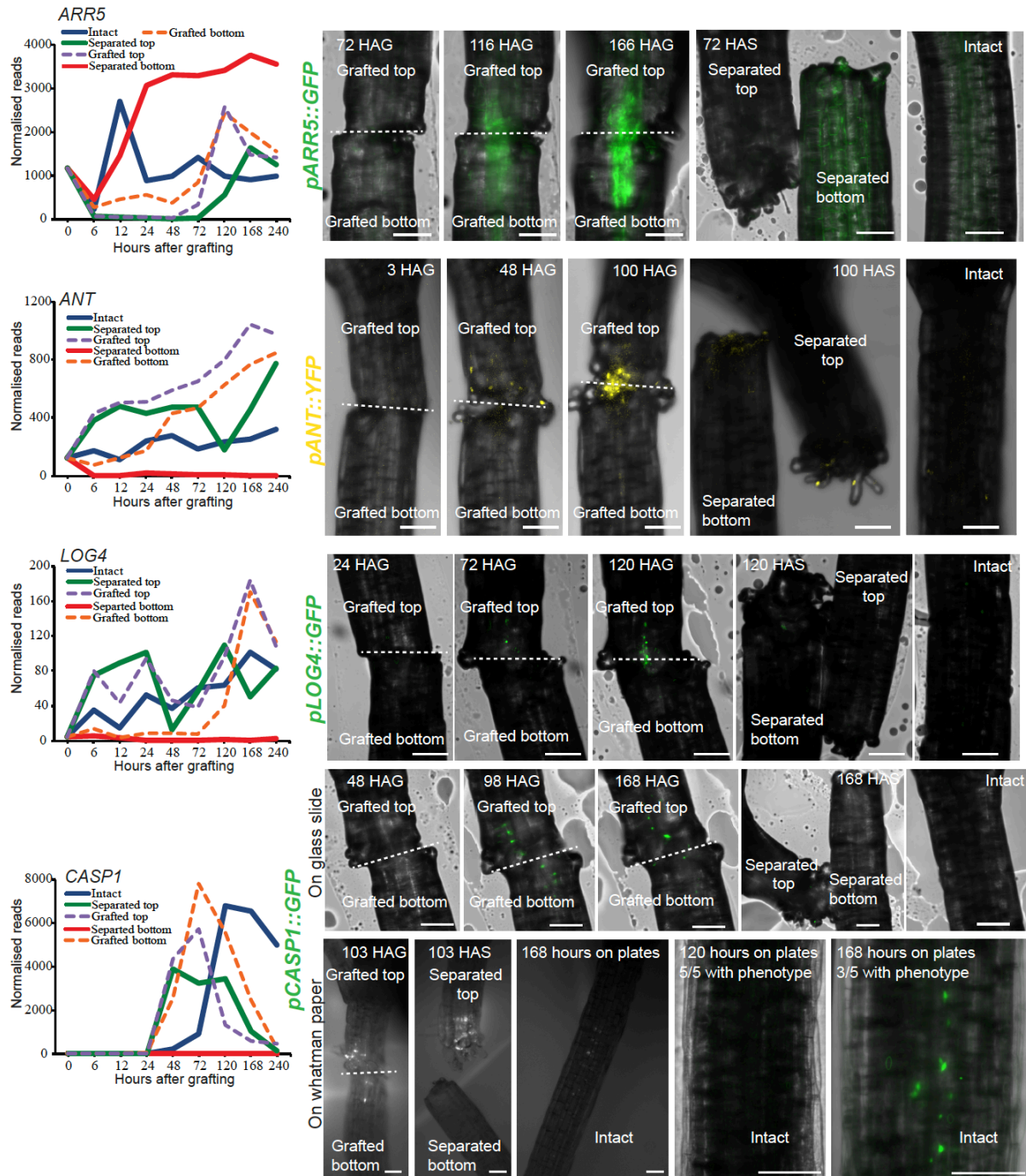
HAG	Sugar Induced <sup>22</sup>	Graft Top=Bottom	Number Overlap	% Overlap	Graft Top>Bottom	Number Overlap	% Overlap	Graft Top<Bottom	Number Overlap	% Overlap
6	2243	4988	263	5	6679	1403	21*	4971	99	2
12	2243	3473	165	5	7111	1444	20*	5657	231	4
24	2243	4135	171	4	6873	1414	21*	4942	195	4
48	2243	3689	197	5	6601	1241	19*	4915	218	4
72	2243	10421	1005	10*	2459	228	9*	2019	138	7
120	2243	15012	1063	7	1510	220	15*	941	115	12*
HAG	Sugar Repressed <sup>22</sup>	Graft Top=Bottom	Number Overlap	% Overlap	Graft Top>Bottom	Number Overlap	% Overlap	Graft Top<Bottom	Number Overlap	% Overlap
6	1998	4988	107	2	6679	53	1	4971	1563	31*
12	1998	3473	68	2	7111	112	2	5657	1526	27*
24	1998	4135	88	2	6873	72	1	4942	1525	31*
48	1998	3689	113	3	6601	93	1	4915	1427	29*
72	1998	10421	530	5	2459	111	5	2019	538	27*
120	1998	15012	979	7	1510	84	6	941	210	22*



**Fig. S1. Experimental approach.** (A) The breaking force in milliNewtons (mN) required to pull apart grafted plants measured by a micro-extensometer. (B-C) Contamination of top in bottom or bottom in top was less than 4% as measured by grafting green fluorescent protein-expressing plants (*p35S::GFP*) to tomato fluorescent protein-expressing plants (*pUBQ10::PM-tdTomato*) and measuring amounts of fluorescence at the different wavelengths of emission in the top segments relative to the bottom. Images show different plants prior to and after pulling with the percent contamination indicated. Scale bar is 100 $\mu$ m. (D) Hierarchical clustering of samples based on log10 transformed Transcripts Per Kilobase Million (TPM) values. Similarity between samples was measured by 1 – spearman correlation coefficient. (E) PCA of expression data shows clustering of similar samples. The grafted top and grafted bottom samples are very similar from 120 hours onwards.

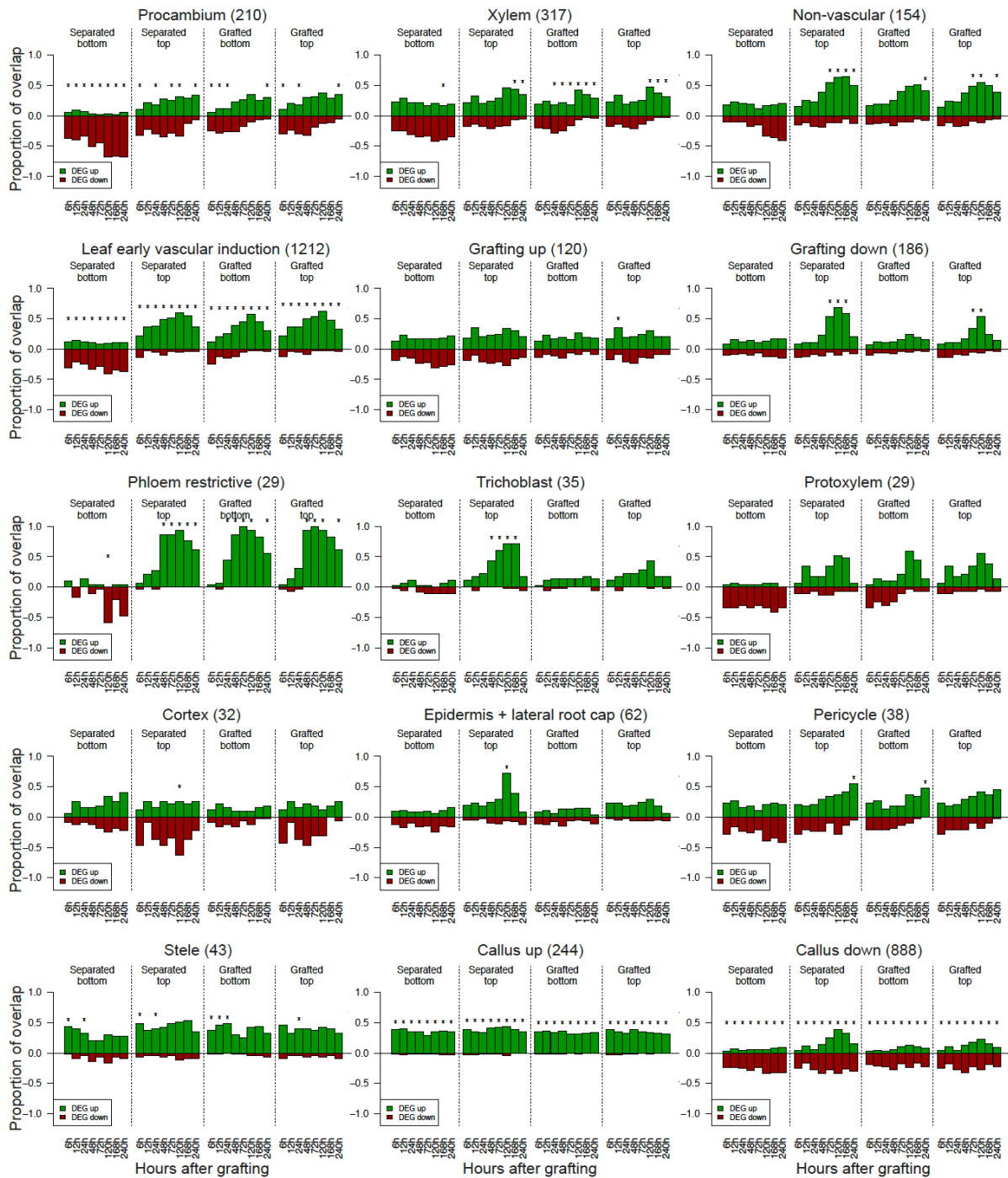


**Fig. S2. Transcriptional dynamics of genes associated with cambium, phloem, xylem and cell division.** (A) RNA accumulation profiles for various genes of interest were plotted over time as measured in grafted tops or grafted bottoms, normalised to intact samples and plotted on a  $\log_2$  scale. The grafting datasets could also be used to investigate the transcriptional dynamics of related genes, such as the sequential activation of WOX transcription factors at the graft junction. (B) Expression levels of a primary root specific transcript (of the *WOX5* gene) or a lateral root specific transcript (from the *LBD18* gene) were plotted over time for intact, separated and grafted samples. (C) Number of loci in the RNAseq datasets where the average behaviour across top/bottom equals the intact expression (topSeparated + bottomSeparated = topGraft + bottomGraft = intact).

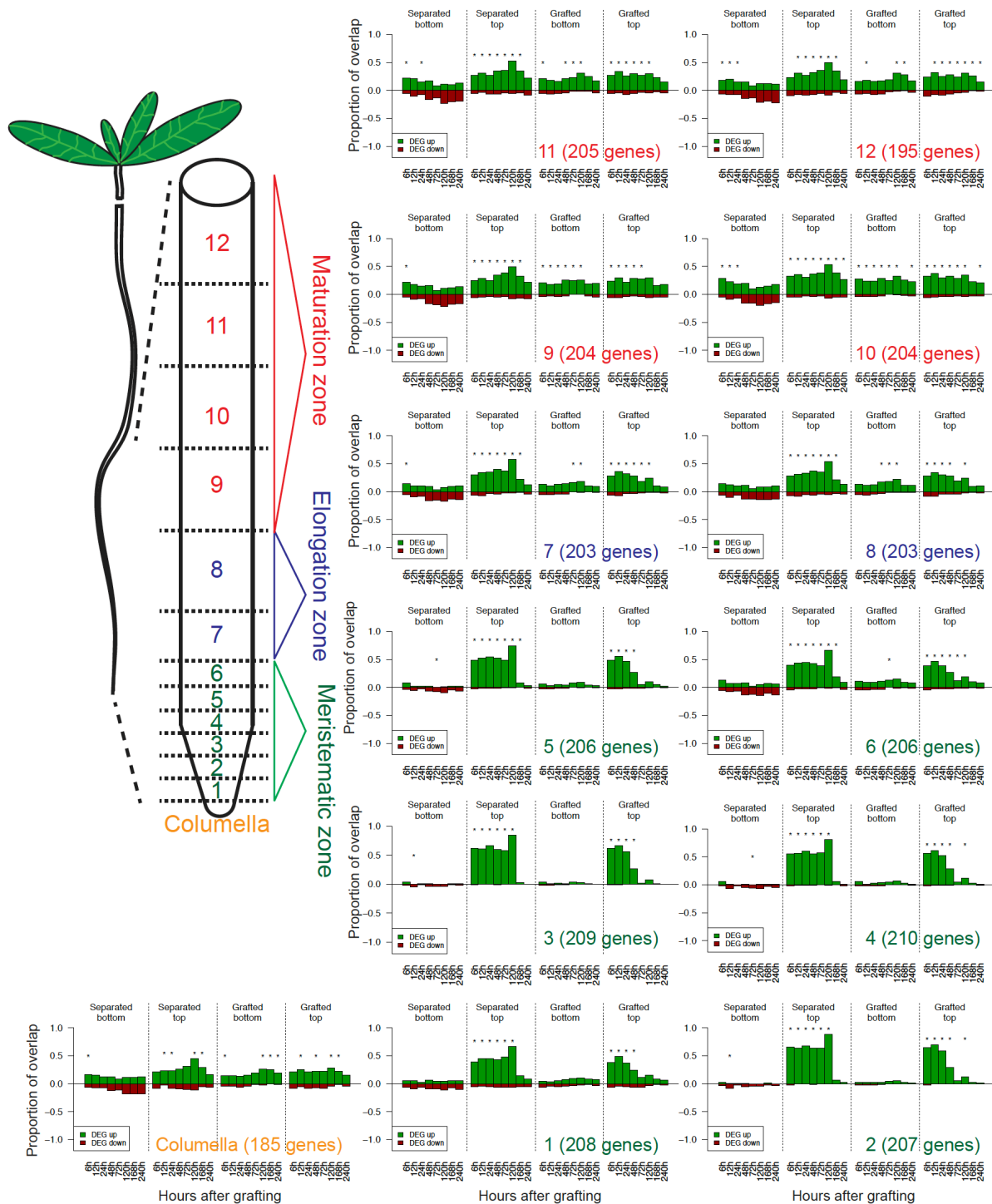


**Fig. S3. Comparison between RNAseq transcriptome expression profiles and transcriptional fluorescent reporters.** RNAseq expression profiles for various genes upregulated during graft formation were plotted for intact, separated and grafted samples (left panels). Transcriptional reporter-expressing plants were cut and separated, cut and grafted, or left intact. After cutting, plants were imaged and z-projections made at various time points (right panels). For *pCASP1::GFP*, we did not observe a signal in intact plants grafted on glass slides (see Materials and Methods), but observed a signal with 3/5 plants 7 days after grafting on Whatman-nylon membrane, the same condition used for transcriptome library preparation. HAG, hours after grafting. HAS, hours after separation. Dashed lines denote the graft junction. Scale bar is 100µm.



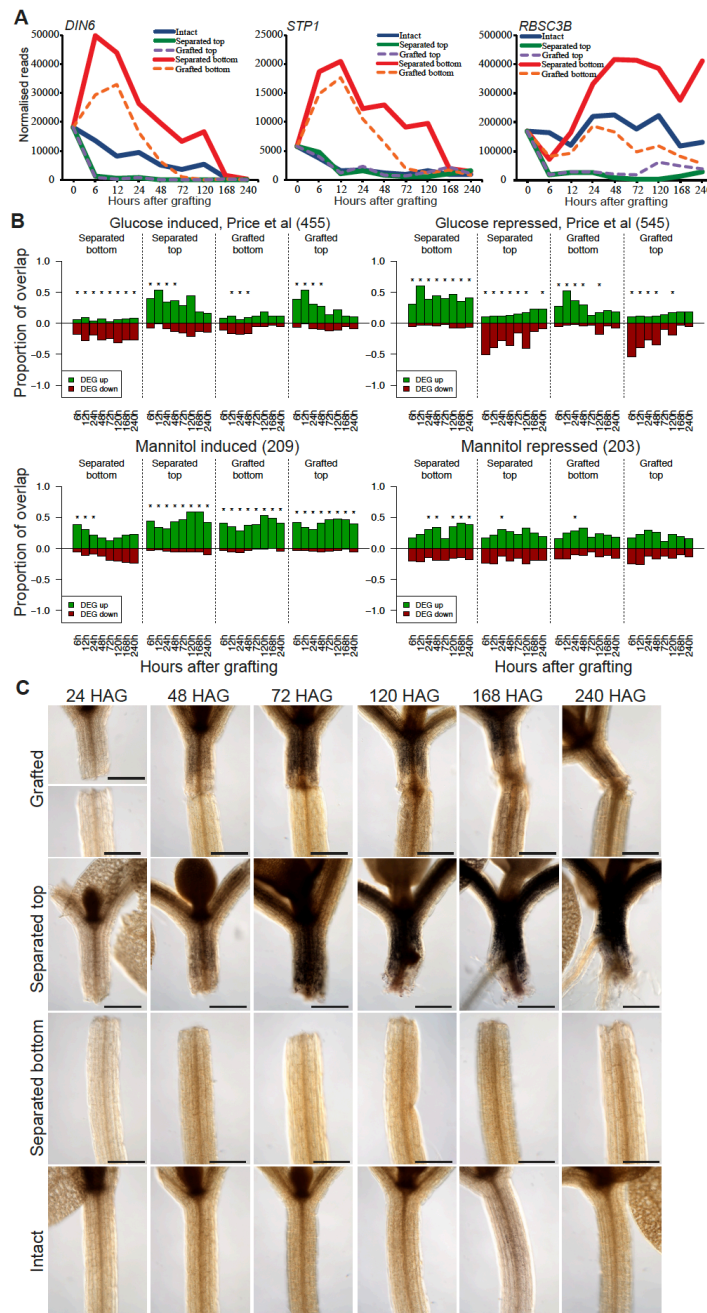


**Fig. S4. Transcriptional overlap between previously published datasets and the grafting datasets.** Genes whose transcripts are associated with various cell types or biological processes were taken from previously published datasets (see Dataset S1) and compared to the datasets generated here. The number in brackets represents the number of cell type-specific or process-specific genes identified in the previous dataset, and overlap is presented as a ratio out of 1.0 for differentially expressed genes (DEG) up- or down-regulated in our dataset relative to intact samples. An asterisk represents a significant difference ( $p < 0.05$ ) between the ratio of up- and down-regulated genes in a previously published transcriptome dataset compared to the ratio of all up- and down-regulated genes in our grafting dataset at a certain time point.

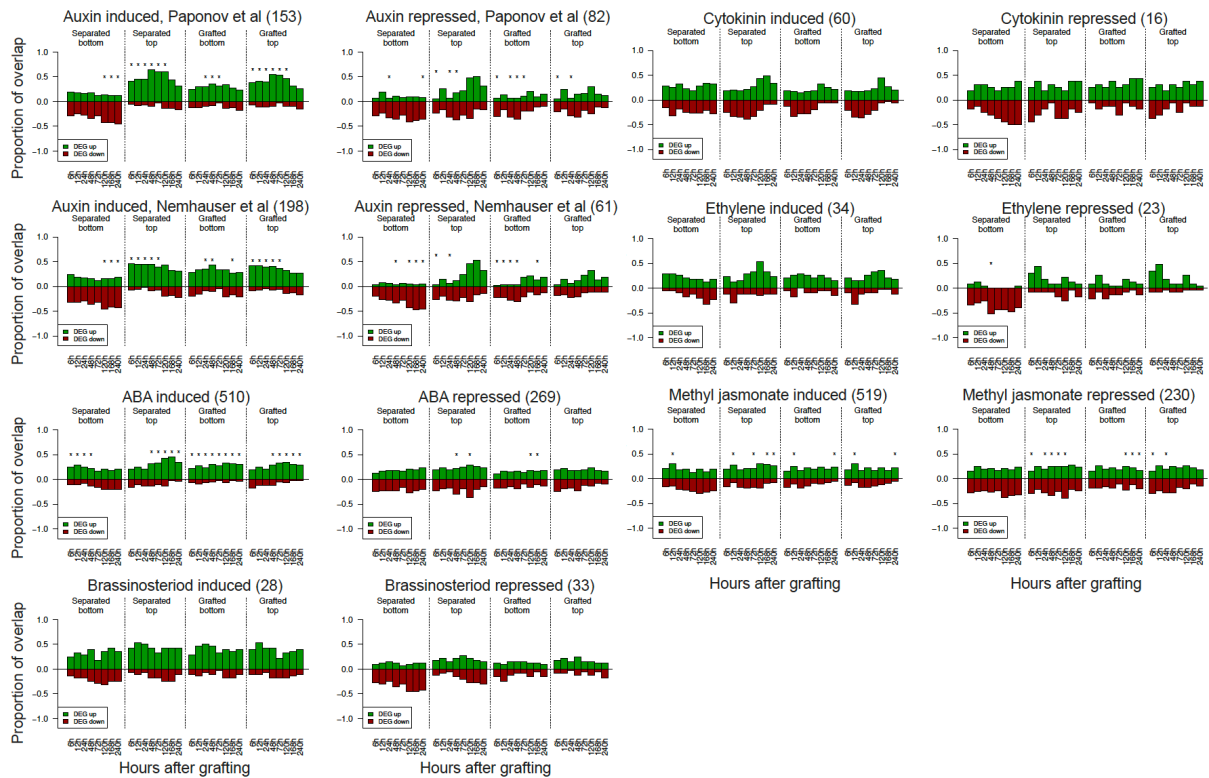


**Fig. S5. Transcriptional overlap between previously published root transcriptomic datasets and the grafting RNA-seq dataset.** Genes whose transcripts are associated with various root regions were taken from previously published datasets (see Dataset S1) and compared to the datasets generated here. Root layer is indicated in the bottom right of each histogram with reference to the cartoon. The number in brackets represents the number of root layer-specific genes identified in the previous dataset, and overlap is presented as a ratio out of 1.0 for differentially expressed genes (DEG) up- or down-regulated in our dataset relative to intact samples. An asterisk represents a significant difference ( $p < 0.05$ ) between the ratio of up- and down-regulated genes in a previously published transcriptome dataset compared to the ratio of all up- and down-regulated genes in our grafting dataset at a certain time point.

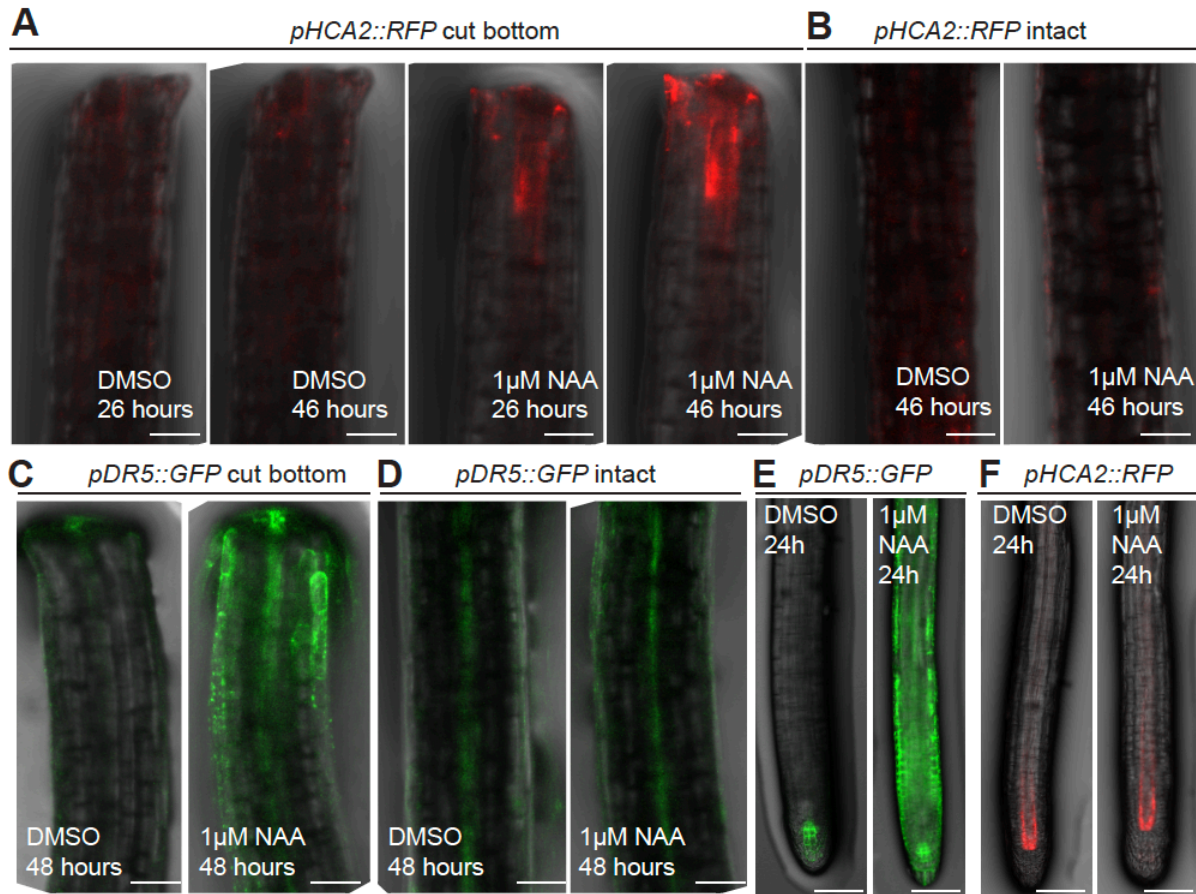




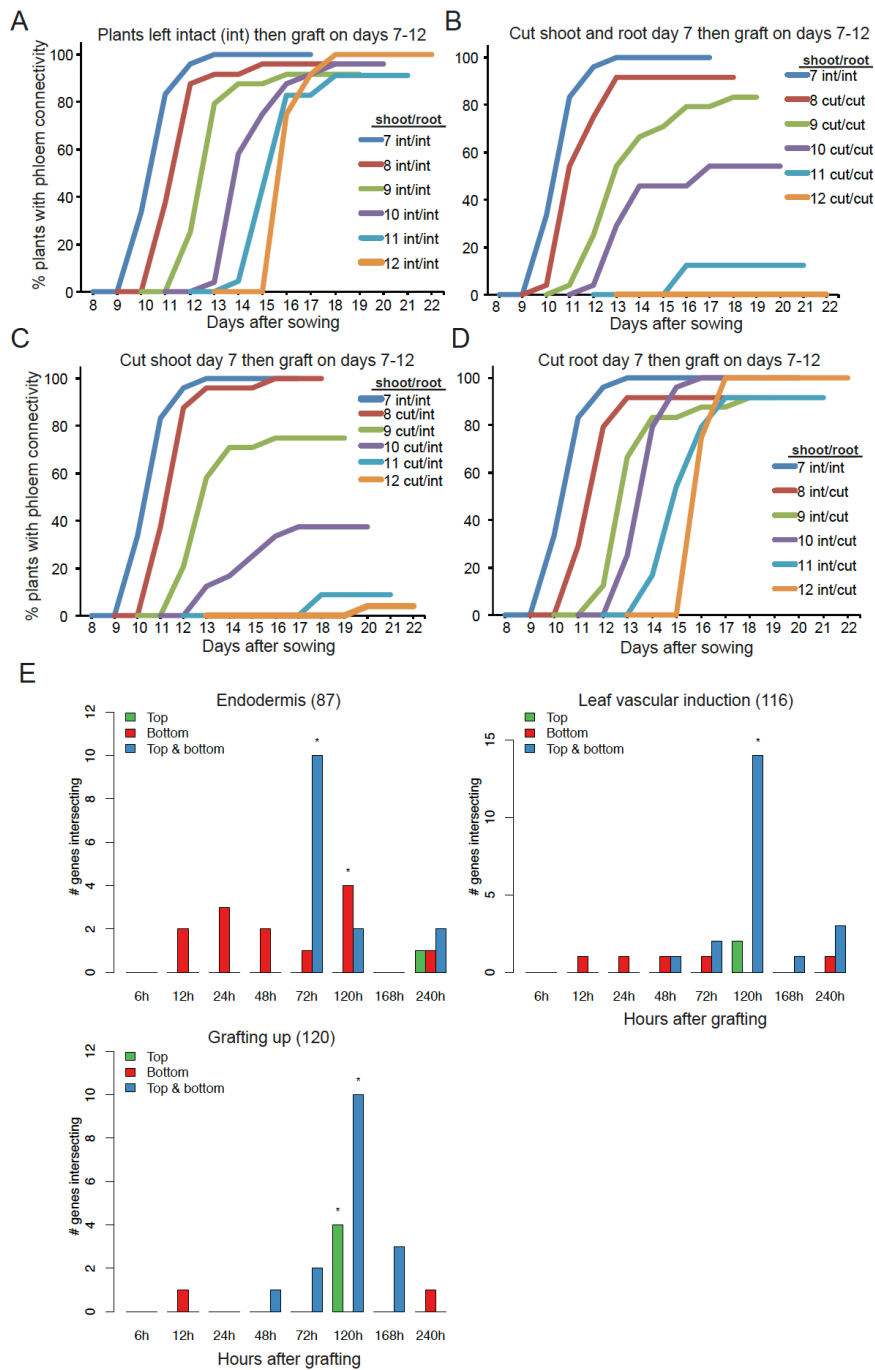
**Fig. S6. Asymmetry is a feature observed with sugar response and starch accumulation.** (A) Expression profiles for RNAs coded by sugar-repressed genes (*DIN6*, *STP1*) or a photosynthetic gene (*RBSC3B*) were plotted for intact, separated and grafted samples. (B) Transcriptional overlap between previously published mannitol-induced, mannitol-repressed, sugar-induced or sugar-repressed genes and our datasets. The numbers in brackets represents the number of glucose or mannitol-responsive genes identified in the previous dataset, and overlap is presented as a ratio out of 1.0 for differentially expressed genes (DEG) up or down regulated in our dataset relative to intact samples. An asterisk represents a significant difference ( $p < 0.05$ ) between the ratio of up- and down- regulated genes in a previously published transcriptome dataset compared to the ratio of all up- and down- regulated genes in our grafting dataset at a certain time point. (C) Lugol staining of grafted plants at various time points reveals dark brown staining associated with starch accumulation. Upper grafted panels are the same as those presented in Fig. 3 and are shown here to compare with controls. HAG, hours after grafting. Scale bar is 250 $\mu$ m.



**Fig. S7. Transcriptional overlap between previously published hormone responsive RNA datasets and the grafting RNA datasets.** Genes whose differential expression is associated with various hormone responses were taken from previously published datasets (see Dataset S1) and compared to the genes represented in the RNA-seq datasets generated. The number in brackets represents the number of cell type-specific or process-specific genes identified in the previous dataset, and overlap is presented as a ratio out of 1.0 for differentially expressed genes (DEG) up- or down-regulated in our dataset relative to intact samples. An asterisk represents a significant difference ( $p < 0.05$ ) between the ratio of up- and down-regulated genes in a previously published transcriptome dataset compared to the ratio of all up- and down-regulated genes in our grafting dataset at a certain time point.



**Fig. S8. Cutting combined with exogenous auxin treatment activates *HCA2* expression.** (A) Separated hypocotyl bottoms containing *pHCA2::RFP* activated expression upon treatment of the synthetic auxin, NAA, within 26 hours and increased expression with time. *pHCA2::RFP* did not activate with DMSO control treatment after 46 hours. (B) Intact hypocotyls did not increase *HCA2* expression upon NAA treatment. (C) Separated hypocotyl bottoms containing *pDR5::GFP* increased expression upon treatment of the synthetic auxin, NAA, after 48 hours. (D) Intact hypocotyls slightly increased *DR5* expression upon NAA treatment after 48 hours. (E-F) 24 hours of NAA treatment strongly increased *pDR5::GFP* expression in the primary root tip, but did not increase *pHCA2::RFP* expression. (A-F) Scale bar is 100 $\mu$ m.



**Fig. S9. A subset of genes is differentially expressed only during graft formation.** (A) 7 days old *pSUC2::GFP* and wild-type Col-0 *Arabidopsis* were transferring to grafting plates and kept intact (intact treatment) until grafted together 0-5 days after transferring. GFP movement from *pSUC2::GFP* shoots to Col-0 roots was monitored over 7 days. (B) Shoots and roots were cut at 7 days after germination and grafted 0-5 days after cutting. (C) Shoots were cut 7 days after germination and grafted 0-5 days after cutting. Roots were kept intact until immediately before grafting. (D) Roots were cut 7 days after germination and grafted 0-5 days after cutting. Shoots were kept intact until immediately before grafting. (E) Genes differentially expressed in grafted tops and grafted bottoms show overlap with previously published genes whose transcripts are associated with the endodermis, vascular induction and grafting (see Dataset S1 for treatment information). An asterisk represents significant high overlap ( $p < 0.05$ ) of previously published gene sets that are also differentially expressed in the grafted samples at a certain time point.

**Dataset S1.** Details of previously published datasets used to compare to the grafting datasets.

**Dataset S2.** Numbers and categories associated with the Bayseq analysis. Categories are defined as grafted (Col:Col), separated (Col\_cut) or intact (ungrafted). The second tab shows clusters used to identify genes specifically up or down regulated by grafting.

**Dataset S3.** GO analysis for biological process (BP). Shown are the top 20 BP GO terms for the grafting-specific genes. Time point selected are those that contain the most genes in each cluster: grafted bottom samples (48hrs) or grafted top + bottom samples (120 hrs). The second tab shows the top 20 BP GO terms for genes upregulated in both grafted tops, grafted bottoms, separated tops and separated bottoms. Time points selected are those immediately after grafting.

**Dataset S4.** Length-normalised reads for all the protein-coding genes in the datasets. By entering the ATG number of interest, a plot is made which shows a differential gene expression profile for the gene of interest.

**Dataset S5.** Normalised reads for all the protein-coding genes in the datasets. By entering the ATG number of interest, a plot is made which shows a differential gene expression profile for the gene of interest.