Supporting Information for

The Rate of Meristem Maturation Determines Inflorescence Architecture in Tomato

Soon Ju Park*, Ke Jiang*, Michael C. Schatz, and Zachary B. Lippman†

Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.

* These authors contributed equally to this work.

† Corresponding author: Zachary B. Lippman (lippman@cshl.edu)

This file includes:

SI Text

Supporting references

Fig. S1 to S16

Table S1 to S7

SI Text : SI Materials and Methods

Plant material and growth conditions

S. lycopersicum cv. M82 (as a reference wild type unbranched domesticated tomato), the inflorescence branching mutant *compound inflorescence* (*s: s-n5568*) in the M82 background (1), the mutant *single flower truss* (*sft-7187*) (2) in the M82 background, and the green-fruited wild species *S. peruvianum* (LA0103) were used in this study. *S. peruvianum* accession LA0103 was selected for analysis from 28 accessions representing four wild tomato species (*S. chilense, S. habrochaites, S. peruvianum* and *S. pennellii*), as this species and accession showed consistent inflorescence branching (3) and flowering time. All genotypes for meristem profiling were grown in 72 cell flats under natural light in a greenhouse supplemented with artificial light from high-pressure sodium bulbs (50μmol/m^2/sec; 16h/8h) at Cold Spring Harbor Laboratory. Daytime temperature was 78°F and nighttime temperature was 65°F, with a relative humidity of 40-60%.

Meristem imaging

To image live meristems using a stereomicroscope, shoot apices were dissected from seedlings of all stages and genotypes. Older leaf primordia (larger than 150μm) were trimmed off under a Nikon SMZ1500 microscope. The meristem images were taken immediately after dissection with an integrated Nikon digital camera attached to the stereomicroscope. The images were recaptured by Z-series manually and merged to create focused images. For scanning electron microscopy (SEM), shoot apices were dissected and collected in 70% ethanol and dehydrated in an ethanol series to 100%

ethanol over 5 days. The samples were critically point dried, and leaf primordia were removed. Meristems including up to plastochron 2 (P2) were sputter-coated with argon gas, and SEM images were captured on a Hitachi S-3500N scanning electron microscope.

Tissue collection and RNA extraction

Vegetative meristem stages were defined by plastochron index (4) and a time scale based on "days after germination" (DAG) as EVM (5^{th} leaf initiated; 7DAG), MVM (6^{th} leaf initiated; 10DAG), LVM (7^{th} leaf initiated; 13DAG) and SYM (2^{nd} leaf initiated in the first produced sympodial shoot; 19DAG) in *S. lycopersicum* cv. M82 and *compound inflorescence* (*s*). Both genotypes reach the reproductive transition and flowering after developing between 7-10 leaves, but the majority (>60%) transition to flowering upon developing 8 leaves in our growth conditions (Fig S1, Table S5). In *S. peruvianum*, the flowering transition is longer and ranges from 11-17 leaves; however, more than 60% of plants transition to flowering after producing 13-14 leaves (Table S5). Thus, for *S. peruvianum*, the MVM (9^{th} leaf initiated; 15DAG) and LVM (11^{th} leaf initiated; 19DAG) were collected according to the plastochron ratio matching the corresponding stages in M82. Although the variation in flowering time can affect precise harvesting of the vegetative meristems (EVM, MVM, and LVM), each stage in each genotype is essentially normalized, because more than 70 meristems comprise each RNA sample for two biological replicates (see below). As well, variation in flowering time does not affect the precise harvesting of the TM, FM and SIM, as each reproductive stage is defined and harvested according to morphological markers, including meristem size and position relative to the last leaf formed: TM (broader and taller meristem than vegetative stages

with a smaller last formed leaf; 15DAG), FM and SIM (shaped as apical and lateral domes, respectively; 16-17DAG) in *S. lycopersicum* cv. M82 and *s* mutants. Importantly, meristem morphologies were indistinguishable in *S. peruvianum*, even though reproductive stages developed later in this species: TM (20-23DAG), SIM and FM (22-26DAG) (Fig. S9).

According to the defined time scale, shoot apices from each genotype were dissected from seedlings within 1 hour (always between 10-11am), and immediately fixed in 100% acetone, followed by vacuum infiltration (5, 6). Tomato shoot apical meristems (SAMs) including up to P2 were microdissected by surgical blades under a stereomicroscope (7). (Fig. 2*A, E, F*). More than 70 meristems were collected for each biological replicate. Meristems were dried for 3 minutes at room temperature to remove remaining acetone, and then thoroughly ground in a mixer mill MM300 (Retsch) with a tungsten bead (3mm diameter). A total of 1.5 - 5ug total RNA was extracted from each collection using the PicoPure™ RNA Extraction kit (ARCTURUS), including DNase treatment with RNase-free DNase (Qiagen) according to the manufacturer's instructions. The quality of total RNA was assessed using a Bioanalyzer 2100 (Agilent), and only high quality RNA was used in subsequent experiments.

Quantitative RT-PCR

First strand cDNA synthesis was performed using 120ng of total RNA with the SuperScript® III First-Strand Synthesis System with oligo dT_{20} (Invitrogen). Quantitative PCRs were performed with one micro-liter of cDNA using Phusion® High-fidelity DNA

ploymerase (NEB), iQ™ SYBR® Green Supermix (Bio-Rad), and gene-specific primers. Primer sequences are available upon request. Semi-quantitative RT-PCR was used for validating stage and meristem specificity of samples. Stage-specific marker genes were examined in all biological replicates to confirm temporal precision of meristem harvesting: *SP* was used as a marker for the TM and SYM (8, 9); *S* and *AN* were sequentially enriched in the SIM and FM stages, respectively (1); *FA* and *LeT6* were used as general meristem markers (10, 11); *LA* marked expression of leaf primordia (12); *SP5G* was used to test for contamination from young expanding leaves (13) and *UBIQUITIN* (*Solyc01g056940*) was used as a reference. Real-time quantitative RT-PCR was used to verify the expression dynamics detected by mRNA-Seq experiments. Two biological replicates for each stage and genotype were analyzed using CFX96™ Real-time PCR System (Bio-Rad). Quantitative data were calculated from the number of PCR cycles (Ct) and normalized against *UBIQUITIN* using the qbase PLUS Data-Analysis Software. Representative dynamically expressed genes, including (i) four stage-specific genes (*S, AN* (1)*, SlAP1/MC* (14)*,* and *SP* (8), (ii) six wave pattern genes (*SlFIL, LeT6* (11)*, SPGB* (15)*, FA* (10)*, SlWUS,* and *SlFUL*), and (iii) four lowly expressed genes (*SlTCP5, SlSCL32, LA* (12)*,* and *SlAO1L*), showed identical temporal dynamics compared to mRNA-seq data (Fig. S7). Additional qRT-PCRs were performed for two newly discovered, lowly expressed, marker genes (TM: *Solyc03g063550*, FM: *Solyc10g050990*) to validate marker gene identification (Fig. 2*D*). Reciprocal best BLAST hit (RBBH) and phylogenetic analyses were used to identify the most likely orthologs from the tomato gene annotation (iTAG2.3 version) (16). Tomato genes were

named after known Arabidopsis orthologous genes such as "*SlFIL* (*Solanum lycopersicum FILAMENTOUS FLOWER*)".

Library preparation

Poly-A containing mRNAs (20~80ng) purified from 1~3ug total RNA (Invitrogen) were used for mRNA-seq library construction according to the mRNA sequencing sample preparation guide (Illumina). The cDNAs of approximately 250bp - 350bp size range were purified by gel elution and used as templates to enrich mRNA-seq libraries. A minimal PCR amplification (16-17 cycles) was performed using PE PCR primers 1.0 and 2.0 (Illumina) and Phusion polymerase PCR mix (NEB). The libraries were quantified using a Bioanalyzer 2100 (Agilent) and sequenced on an Illumina GAIIx instrument. Solexa paired-end (PE) 50 base sequencing were performed as described on the Illumina GAII platform (Table S1). Two biological replicates were sequenced for each stage of each genotype. Samples were randomized across Illumina sequencing flowcells and lanes within flowcells. All read data (FASTQ files) have been released to the public without restrictions through the Sol Genomics Network.
(ftp://ftp.solgenomics.net/user_requests/LippmanZ/public_releases/)

Read mapping and analysis

All sequencing results were qualified by base intensities, nucleotide distributions, GC contents, and quality scores for every cycle produced by the GAIIx. Reads were aligned against predicted protein coding sequences (CDSs) from tomato iTAG version 2.3 (http://solgenomics.net/organism/solanum_lycopersicum/genome) (16) with the short

read mapping software Bowtie (17). The following parameters were used: maximum of two mismatches within the first 25 bases and unlimited mismatches in the remaining 25 bases; however, the sum of base qualities of mismatches in the later 25 bases was not allowed to exceed 70. The acceptable insert size between read-pairs was between 100 bp and 400 bp. Only unique alignments were retained for further analyses (only ~4% of reads showed mapping to multiple locations in each library). Raw read counts were summarized based on unique alignments on a gene-by-gene basis, then normalized against library sizes and CDS lengths to calculate Reads Per Kilobase of exon Model per million mapped reads (RPKM) values for use in DDI analyses (see section below on Digital Differentiation Index analyses) (18). Proportions of mapped reads for all samples are summarized in Table S1. Genes showing less than average 3 RPKM/sample (summed over all stages) were removed, as suggested by previous studies (19). For example, only genes showing higher than 42 (3 RPKM X 7 stages X 2 replicates) total RPKM across 2 replicates of 7 stages of *S. lycopersicum* cv. M82 were retained. The retained genes consisted of the detected genes in our samples and the raw read counts of the genes served as the foundation for all further analyses. The entire process was completed using the R statistical package with custom R scripts (20).

Identification and analyses of dynamically expressed genes

Statistical tests of differential gene expression based on raw read counts between each pair of samples involving five stages of *S. lycopersicum* primary meristems (EVM, MVM, LVM, TM, and FM) were conducted using R. Replicates and tag-wise dispersion factors were used in a modified exact test implemented by edgeR to test differential

expression in two sample comparisons based on raw read counts, according to the edgeR handbook (21). The false discovery rate (FDR) was corrected by the Benjamini and Hochberg's procedure (22). Out of all detected genes, 3919 genes showed greater than two-fold change at a *P*-value ≤ 0.05. These two-fold differentially expressed genes were selected as a core set for further analyses (Table S2).

Mean RPKM values over two biological replicates of the 3919 genes were clustered using the K-means clustering algorithm implemented in the R package Mfuzz (23). After plotting the sum of squares within clusters against cluster numbers (from 1 to 50), we determined that 20 clusters are appropriate to capture the majority of dynamic trends represented by these genes, because no new major dynamic patterns of gene expression were found with increasing cluster numbers. Major types of temporal expression patterns were classified manually, including stage-specific peaks, waves, and gradients (Fig. S4). For each cluster, genes were annotated using Mercator and MapMan annotation pipelines (24). Enrichment of selected functional categories in each cluster was tested using Fisher's exact tests against the genome-wide gene annotation. *P* values of enrichment tests were converted into $-\log_{10}P$, scaled for one functional category across all clusters, and plotted as a color gradient heat-map (Fig. S5). The expression dynamics of all transcription factors (TFs) among the 3919 genes were also clustered according to their temporal expression patterns, resulting in three major clusters. The numbers of TFs in each TF family belonging to the three clusters were summarized in Fig. S6.

***compound inflorescence* (*s*) data analyses**

Because *compound inflorescence* (*s*) is an isogenic mutant in the M82 background, we expected very low sequence divergence between the mutant and *S. lycopersicum*. Therefore, the same read mapping and quantification pipeline was used for *compound inflorescence* (*s*). Tests of differential expression based on raw read counts between *S. lycopersicum* and *s* mutants were conducted with the edgeR package in a stage-by-stage comparison in the order of MVM, LVM, TM, SIM and FM. Only genes showing larger than two-fold change at a *P*-value ≤ 0.05 were selected and summarized in Table S6. Enrichments of selected functional categories in differentially expressed genes were tested using Fisher's exact tests against the genome-wide gene annotation. The *P* values of enrichment tests were converted into $-\log_{10}P$ and scaled within one functional category across stages to a zero to one range, then plotted in the form of a heat-map (Fig. S10).

To investigate changes in temporal expression dynamics in *compound inflorescence* (*s*), genes showing dynamic expression in five stages of *S. lycopersicum* (MVM, LVM, TM, SIM, and FM) were identified from the differential expression tests described above. Their expression dynamics were clustered using K-means clustering algorithm with K=20. Genes from clusters that show expression peaks in TM, SIM and FM stages were identified in *S. lycopersicum*, and the same genes were also identified in *compound inflorescence* (*s*). Scaled expression levels were plotted in both lines to illustrate the change of dynamics of these clusters in *compound inflorescence* (*s*) (Fig. 3*I-K* and Fig. S12). Differential temporal dynamics between *S. lycopersicum* and *compound inflorescence* (*s*) was tested using a permutation test. For each cluster, an 'observed'

distance between two genotypes was calculated based on the Pearson correlation of average expression over all genes within this cluster across five stages. The genotypic identities of genes were then randomly shuffled, creating a permuted sample with two 'genotypic' groups. The distance between these two groups were also calculated based on Pearson correlations for average expression over all genes. This process was carried out 1000 times to generate an empirical distribution of distances between the two 'genotypic' groups. Finally, the 'observed' distance, which reflects how different the temporal dynamics are between the two true genotypes, was compared to the empirical distribution to see how likely the 'observed' distance between genotypes is due to chance. An 'observed' distance found only for a few instances (<=10, P<=0.01) in 1000 permutations indicates a very significant difference of temporal dynamics between *S. lycopersicum* cv. M82 and *compound inflorescence* (*s*), which was deemed a 'disrupted' cluster in the mutant. TM, SIM and FM specific clusters were identified as disrupted clusters in *compound inflorescence* (*s*) (Fig. 3*I-K*). All clustering, permutation tests and plots were completed in R with custom R scripts.

Digital Differentiation Index analyses

The Digital Differential Index (DDI) algorithm was used to index and predict meristem maturation states (25). DDI takes a collection of samples with known or pre-determined maturation states as calibration points and identifies marker genes that exhibit an expression peak at each calibration point. These genes characterize the calibration points molecularly. DDI then queries the marker gene expressions in uncharacterized or unknown samples to quantify their maturation states relative to the calibration points.

Each marker gene gives a 'maturation score' as an estimation of maturation state for the unknown sample when its expression level is compared between each calibration sample and unknown sample. Collectively, a set of marker genes generates a distribution of maturation scores for samples of unknown identity. At the same time, a Student's t-test of average maturation score difference between calibration and unknown samples was conducted for each unknown meristem sample, yielding a *P* value for the significance of the maturation state difference. For each prediction, this *P* value was obtained for comparisons between the unknown sample and temporarily successive calibration points, in order to generate a 'gradient' of meristem similarity (plotted in heat-maps in the form of scaled $1/(-\log_{10}P)$). For example, to predict the maturation state of *S. lycopersicum* SIM using the first replicate of *S. lycopersicum* EVM, MVM, LVM, TM and FM as calibration points, *P*-values were calculated for maturation state comparisons SIM *vs.* EVM, SIM *vs.* MVM, SIM *vs.* LVM, SIM *vs.* TM and SIM *vs.* FM. The *P* values were then transformed into $1/(-\log_{10}P)$ and scaled across five values into a zero to one range (scaling was done for each prediction independently). Because smaller *P*-values indicate larger differences in maturation scores, the scaled $1/(-\log_{10}P)$ values quantify the relative similarity of a SIM to each of the five calibration points. In addition, the second replicate of *S. lycopersicum* LVM, TM and FM were treated as 'unknown' samples and queried in the same way as the SIM. The resulting *P* values were transformed, scaled and plotted in the heat-maps as references along with SIM results.

DDI selects marker genes based on the fold changes between their maximum and minimum expression levels among calibration points. By default, a two-fold change

between maximum and minimum is required by DDI for being a marker gene. We experimented with this parameter using 2, 3 and 4 fold changes, and found no differences among the results. We used the first biological replicate of *S. lycopersicum* as calibration points to index maturation states for five stages of primary meristems. Based on the indices, the predicted maturation states of the second replicate verified the accuracy and preciseness of DDI analysis (Fig. S8*A*).

We applied the DDI algorithm with three-fold change to select marker genes and predict *compound inflorescence* (*s*) meristem states using five stages of *S. lycopersicum* meristems (MVM, LVM, TM, SIM and FM) as calibrations (Fig. 4*A*-*C*). To verify that DDI captures the molecular signature of each stage, including the SIM, we used the first biological replicate of *S. lycopersicum* as calibration points to index meristems from five stages (MVM, LVM, TM, SIM and FM) and predicted the maturation states of meristems from the second replicate (Fig. S8*B*). Scaled mean RPKM values from *compound inflorescence* (*s*) meristems were queried based on *S. lycopersicum* calibrations. Corresponding *S. lycopersicum* meristems were also queried in the same way as references. All *P* values from t-tests of maturation score differences are listed in Table S4. All marker genes selected by DDI are listed in Table S3.

***S. peruvianum* data analyses**

S. peruvianum is an out-crossing, self-incompatible, green-fruited wild tomato species. The divergence between *S. peruvianum* and *S. lycopersicum* at genomic and transcriptomic levels are largely unknown. Therefore, to determine whether *S.*

peruvianum mRNA reads could be mapped and quantified using the *S. lycopersicum* transcriptome as a reference, we estimated the level of sequence divergence in the protein coding regions. Paired-end 50bp Illumina sequencing reads from *S. peruvianum* meristem mRNA samples were first screened for base-calling qualities to filter out bases with quality under 36. Reads were then mapped to the *S. lycopersicum* cv. Heinz reference transcriptome (http://solgenomics.net/organism/solanum_lycopersicum/genome) without the paired-end relationships maintained using the short read alignment programs BWA, allowing four mismatches for every 50bp (26). BWA was used in this step to obtain more accurate mapping and to accommodate for possible small insertion-deletions (indels) between the two tomato species. Reads from different samples were mapped separately and the alignments were merged into a single BAM file. *S. peruvianum* SNPs and other variants were identified from BWA/SAMtools/bcftools pipeline with default parameters, following coordinate sorting, duplicates marking and removal, indel realignment and mapping/base quality recalibration implemented by the GATK pipeline (Picard: http://picard.sourceforge.net) (27). Allele frequencies were estimated from read variant frequencies using the EM-estimation method implemented by SAMtools/bcftools (30). Homozygous and heterozygous SNPs/variants were also distinguished by GATK application UnifiedGenotyper (28). Overall, about ~70% SNPs/variants are homozygous in *S. peruvianum*, suggesting that fixed differences between *S. peruvianum* and *S. lycopersicum* are predominant over heterozygous variants within *S. peruvianum* (Fig. S14*A*). The GATK application FastaAlternateReferenceMaker was used to reconstruct *S. peruvianum* protein coding regions by replacing the reference sites with the identified alternative nucleotides (27). For subsequent mapping analysis, heterozygous sites in *S.*

peruvianum were replaced with the most abundant alternative allele because BWA and Bowtie do not support ambiguous bases. The entire process was completed with SAMtools and custom Perl scripts (30).

Because gene reconstruction is based on read mapping, the divergence estimated from the reconstructed genes is likely underestimated due to the exclusion of those reads that are too divergent to map. To avoid this bias, *de novo* assembly of *S. peruvianum* meristem mRNA reads were performed with the Trinity assembly pipeline (31). The *de novo* assembly was done with three different K-mers (20, 25, 30) with paired-end relationship incorporated (default for all other parameters). These three assemblies resulted in similar assembly qualities, so the assembly from K-mer 25 was selected for further analyses. The assembly has an N50 value of 663 bp and captures 51 Mbp of the transcriptome. Out of 738916 assembled contigs with an average depth of coverage greater than five and length greater than 50 bp, 13840 contigs can be aligned uniquely with annotated reference genes. The *de novo* assembly of the *S. peruvianum* transcriptome and SNP variations were released to the public without restrictions through the Sol Genomics Network. (ftp://ftp.solgenomics.net/unigene_builds/single_species_assemblies/Solanum_peruvianum).

Both reconstructed and *de novo* assembled *S. peruvianum* genes were aligned with corresponding *S. lycopersicum* cv. Heinz reference genes by BLASTn (32). The resulting alignments were refined by ClustalW (33) and manual inspection. Pair-wise sequence divergence was estimated based on nucleotide differences and corrected for multiple

substitutions using the Jukes-Cantor nucleotide substitution model (34), whereas heterozygous sites in *S. peruvianum* were ignored in the calculation. The average divergence in protein coding regions between *S. peruvianum* and the reference is 0.007049 based on reconstructed *S. peruvianum* genes, and 0.009993 based on *de novo* assembled contigs. The median divergences estimated from the two approaches were 0.005522 and 0.008824, respectively. Some genes with high divergence in *S. peruvianum* were not captured by read mapping but were represented by *de novo* assembly, leading to a higher average divergence estimated by comparing de novo assembled transcripts to reference transcriptome. Average divergence between *S. lycopersicum* cv. M82, *compound inflorescence* (*s*) and the reference are 0.002909 and 0.002691, respectively, whereas both medians are zero based on estimation from reconstructed genes, suggesting extremely low divergence within *S. lycopersicum*. Overall, the protein coding region divergence between our samples and the Heinz reference reflected both intraspecific and interspecific genetic distances. More importantly, divergence between *S. peruvianum* and *S. lycopersicum* is below 1% in protein coding regions, which allowed us to use the same read mapping process as *S. lycopersicum* to quantify gene expression. Still, one additional mismatch was permitted for the first 25 bases of *S. peruvianum* reads in the mapping process by Bowtie to accommodate more divergent genes. To confirm that mapping *S. peruvianum* reads to *S. lycopersicum* reference transcriptome does not introduce biases in expression level estimation in the wild species, a new round of read mapping using Bowtie and the same parameters were carried out against the new reconstructed *S. peruvianum* reference. The expression levels were estimated based on the new mapping and compared with expression level estimated from mapping using *S.*

lycopersicum as reference. The correlation coefficient is 0.9973658 between expression levels (overall expression summed over all meristem maturation stages) estimated from two mapping processes (Fig. S14*B*). Differential expression tests based on raw read counts yielded 7744 differential expressed genes between *S. peruvianum* and *S. lycopersicum* using mapping results from the reconstructed *S. peruvianum* transcriptome (7728 genes detected in mapping with *S. lycopersicum* as reference). Over 97% of the differential expression genes estimated from mapping processes of two different references overlapped, suggesting that using the *S. lycopersicum* cv. Heinz reference annotation does not affect overall expression level estimation of *S. peruvianum*.

Using the read counts from mapping *S. peruvianum* reads to *S. lycopersicum* cv. Heinz reference and applying our criterion of 3 RPKM/sample across five stages (sum of RPKM >= 15), 18188 genes were detected in *S. peruvianum* meristem samples, 94.1% of which were also detected in *S. lycopersicum*, suggesting that we captured highly comparable transcriptomes across tomato species.

Estimation of species divergence based on transcriptomes

The RPKM values from *S. peruvianum* read mapping were combined with *S. lycopersicum* and *compound inflorescence* (*s*) mean RPKM values to evaluate the overall transcriptome divergence among them. A Principle Component Analysis (PCA) was conducted using the expression profiles of the three genotypes (summed expression over all meristem stages) (Fig. S15). The first principle component (PC1) clearly reflected the species divergence, where *S. peruvianum* was distant from the tightly grouped *S.*

lycopersicum and *compound inflorescence* (*s*). In addition, the sequence divergences among the three genotypes (median divergence: *S. lycopersicum vs. compound inflorescence*: 0; *S. lycopersicum vs. S. peruvianum*: 0.007; *compound inflorescence vs. S. peruvianum*: 0.00682) are proportional to the overall meristem transcriptome divergence revealed by PC1.

The raw read counts of *S. peruvianum* were also subjected to tests of differential expression between *S. lycopersicum* and *S. peruvianum* using the edgeR package in stage-by-stage comparisons (Table S7). The results revealed that genome-wide mean and median log-fold-change of expression between the two species are -0.00619 and -0.00710, respectively. Compared to mean (0.1863) and median (-0.003) log-fold-change between *S. lycopersicum* cv. M82 and the isogenic mutant *compound inflorescence* (*s*). *S. peruvianum* does not show significant and systematic underestimation of expression levels. In terms of the relationship between expression levels and sequence divergence, regression analysis showed that the expression is slightly lower in *S. peruvianum* when the sequence divergence is large with a slope of -6.04, indicating that higher divergence may contribute to the minor underestimation of expression in *S. peruvianum* due to less mapped reads in genes with higher divergence (Fig. S14*C*). However, this does not affect our results because the majority of the dynamically expressed genes do not show high species divergence (Fig. S14*C*). In fact, the 684 DDI-selected marker genes used for predicting *S. peruvianum* meristem maturation states showed a mean divergence of 0.0108 and median divergence of 0.009. In addition, new DDI analyses with only genes showing divergence less than 2% did not change our results and conclusions.

As revealed by the above analyses, both sequence divergence and tests of differential expression suggested that species divergence of *S. peruvianum* from domesticated tomatoes does not affect mapping and quantification in our expression profiling. Therefore, scaled RPKM values from *S. peruvianum* meristems were also subjected to the clustering analyses and permutation tests described in *compound inflorescence* (*s*) data analyses, which revealed that *S. peruvianum* showed both similar and different dynamic disruptions compared to *compound inflorescence* (*s*) (Fig. 3*I-K*); however the trajectory of change from *S. lycopersicum* was similar to *s* (Fig. S12). To investigate the maturation states of *S. peruvianum* meristems relative to *S. lycopersicum*, scaled RPKM values from *S. peruvianum* meristems were also subjected to the DDI analysis. First, five stages of *S. peruvianum* meristems were queried as unknown samples using five stages of *S. lycopersicum* meristems (MVM, LVM, TM, SIM and FM) as calibrations (Fig. 4*D-F*). The second replicates of corresponding meristems of *S. lycopersicum* were queried in the same way as references. The *P* values from maturation score difference tests were transformed, scaled and plotted as described above in the *compound inflorescence* (*s*) analysis section (Fig. 4*D-F*).

Supporting References

1. Lippman ZB*, et al.* (2008) The making of a compound inflorescence in tomato and related nightshades. *PLoS Biol* 6(11):e288.
2. Lifschitz E*, et al.* (2006) The tomato FT ortholog triggers systemic signals that regulate growth and flowering and substitute for diverse environmental stimuli. *Proc Natl Acad Sci U S A* 103(16):6398-6403.
3. Peralta IE & Spooner DM (2005) Morphological characterization and relationships of wild tomatoes (Solanum L. Section Lycopersicon). *Monogr. Syst. Bot., Missouri Bot Gard* 104:227-257.
4. Erickson RO, and Michelini, F.J. (1957) The plastochron index. *Am. J. Bot.* 44:297-305.
5. Ruzin SE (1999) *Plant Microtechnique and Microscopy.* (Oxford University Press, New York).
6. Goldsworthy SM, Stockton PS, Trempus CS, Foley JF, & Maronpot RR (1999) Effects of fixation on RNA extraction and amplification from laser capture microdissected tissue. *Mol Carcinog* 25(2):86-91.
7. Fleming AJ, Mandel T, Roth I, & Kuhlemeier C (1993) The patterns of gene expression in the tomato shoot apical meristem. *Plant Cell* 5(3):297-309.
8. Pnueli L*, et al.* (1998) The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. *Development* 125(11):1979-1989.
9. Thouet J, Quinet M, Ormenese S, Kinet JM, & Perilleux C (2008) Revisiting the involvement of SELF-PRUNING in the sympodial growth of tomato. *Plant Physiol* 148(1):61-64.
10. Molinero-Rosales N*, et al.* (1999) FALSIFLORA, the tomato orthologue of FLORICAULA and LEAFY, controls flowering time and floral meristem identity. *Plant J* 20(6):685-693.
11. Chen JJ, Janssen BJ, Williams A, & Sinha N (1997) A gene fusion at a homeobox locus: alterations in leaf shape and implications for morphological evolution. *Plant Cell* 9(8):1289-1304.
12. Ori N*, et al.* (2007) Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. *Nat Genet* 39(6):787-791.
13. Shalit A*, et al.* (2009) The flowering hormone florigen functions as a general systemic regulator of growth and termination. *Proc Natl Acad Sci U S A* 106(20):8392-8397.
14. Vrebalov J*, et al.* (2002) A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. *Science* 296(5566):343-346.
15. Pnueli L*, et al.* (2001) Tomato SP-interacting proteins define a conserved signaling system that regulates shoot architecture and flowering. *Plant Cell* 13(12):2687-2702.
16. The International Tomato Genome Sequencing Consortium (2011) The official annotation for the tomato genome is provided by the International Tomato Annotation Group (ITAG), a multinational consortium, funded in part by the EU-SOL project.
17. Langmead B, Trapnell C, Pop M, & Salzberg SL (2009) Ultrafast and memory-

efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
18. Mortazavi A, Williams BA, McCue K, Schaeffer L, & Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621-628.
19. Blekhman R, Marioni JC, Zumbo P, Stephens M, & Gilad Y (2010) Sex-specific and lineage-specific alternative splicing in primates. *Genome Res* 20(2):180-189.
20. Team RDC (2011) R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria).
21. Robinson MD, McCarthy DJ, & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.
22. Benjamini Y & Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1):289-300
23. Kumar L & M EF (2007) Mfuzz: a software package for soft clustering of microarray data. *Bioinformation* 2(1):5-7.
24. Thimm O*, et al.* (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* 37(6):914-939.
25. Efroni I, Blum E, Goldshmidt A, & Eshed Y (2008) A protracted and dynamic maturation schedule underlies Arabidopsis leaf development. *Plant Cell* 20(9):2293-2306.
26. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754-1760.
27. McKenna A*, et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297-1303.
28. DePristo MA*, et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491-498.
29. Koboldt DC*, et al.* (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics* 25(17):2283-2285.
30. Li H*, et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
31. Grabherr MG*, et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*.
32. Altschul SF*, et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-3402.
33. Thompson JD, Higgins DG, & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673-4680.
34. Jukes TH & Cantor CR (1969) *Evolution of protein molecules.* (Academic Press, New York).
35. Kaufmann K, Pajoro A, & Angenent GC (2010) Regulation of transcription in plants: mechanisms controlling developmental switches. *Nat Rev Genet*

11(12):830-842.
36. Fornara F & Coupland G (2009) Plant phase transitions make a SPLash. *Cell* 138(4):625-627.
37. Kaufmann K, *et al.* (2010) Orchestration of floral initiation by APETALA1. *Science* 328(5974):85-89.
38. Wigge PA, *et al.* (2005) Integration of spatial and temporal information during floral induction in Arabidopsis. *Science* 309(5737):1056-1059.

Supporting Figure Legends

Fig. S1. Identification and classification of seven stages of shoot apical meristem (SAM) development in tomato. (A) Scanning electron micrographs (SEM) from domesticated tomato (*Solanum lycopersicum* cv. M82) representing sequential stages of maturation of the primary shoot meristem (PSM): the Early, Middle, and Late Vegetative Meristems (EVM: 5th leaf initiated; MVM: 6th leaf initiated; LVM: 7th leaf initiated), the Transition Meristem (TM: 8th leaf initiated), and the Floral Meristem (FM). The two sympodial meristems of tomato are also presented: the Sympodial Inflorescence Meristem (SIM) and the Sympodial shoot meristem (SYM: 2nd leaf initiated). The termination of the PSM, and the initiation of the SIM and SYM occur within 72 hours and nearly simultaneously. The SIM is manifested as a large dome much more rapidly than the SYM, and so the first produced SYM on each plant becomes prominent and harvestable only after the flower of the PSM begins to form and the first SIM has already initiated another SIM. **(B)** Flowering time as measured number of leaves formed by the PSM from *Solanum lycopersicum* cv. M82. Leaves generated by the PSM were counted from 60 plants after the PSM terminated in the first terminal flower of the first inflorescences (19DAG) in our growth condition (see "Plant material and growth conditions**"** in *SI text*). Note than more than 60% of plants flower at 8 leaves. The stages of the PSM were defined by leaf primordia number, which also corresponded to a specified number of days following germination (DAG: days after germination). L=leaf number counted from 1st leaf. Scale bar, 50μm.

Fig. S2. Flow chart of computational pipeline for mRNA-seq data analysis.

Fig. S3. Temporal expression dynamics during meristem maturation for tomato homologs comprising three sub-classes of the MADS-box domain gene family and the *SQUAMOSA PROMOTER BINDING PROTEIN LIKE* (*SPL*) gene family. (A) Expressional dynamics of tomato homologs of three subclasses of MADS-box genes related to: i) the vegetative to reproductive phase transition: *SHORT VEGETATIVE PHASE-like* (*SVP-like*), ii) inflorescence/floral meristem identity: *APETALA1-like* (*AP1-like*), iii) and floral organ identity: *SEPALLATA-like* (*SEP-like*). Sub-classes were identified through a combined phylogenetic analysis involving all annotated MADS-box genes in tomato and Arabidopsis (*SI Text*). (**B**) Expression dynamics of tomato *SPL-like* genes grouped according to whether they are dynamically expressed (left) or not (right), as determined by differential expression analyses comparing all stages of wild type (Table S2). Tomato gene identifiers are listed, including genes not detected by mRNA-seq according to our read count threshold (see *SI text*). Numbers in parentheses indicate the number of detected genes in each category out of the total number of annotated genes in the tomato genome.

Fig. S4. K-means (K=20) clustering of temporal expression patterns of 3,919 dynamically expressed genes among five stages of PSM maturation in *S. lycopersicum*. Black lines indicate scaled mean expression values. Gray areas indicate standard deviation. Clusters were reordered and named according to the manually classified patterns: Peaks (c1 to c5), Waves (c6 to c15) and Gradients (c16 to c20). The

number of genes in each cluster follows each cluster name. (P=peak, W=wave, G=gradient)

Fig. S5. Enrichment of selected functional groups in each cluster as determined by MapMan classifications. Clusters are ordered according to the three major dynamic patterns represented by colored bars above the heat-map: peaks (c1 to c5, orange), waves (c6 to c15, green) and gradients (c16 to c20, blue). Scaled $-\log_{10}P$ values are shown in the heat-map. *P* values are derived from Fisher's exact tests of the proportions of selected terms in clusters against the proportions in the genome annotation (*SI Text*).

Fig. S6. Temporal expression dynamics of transcription factors in *S. lycopersicum* meristems. (**A**) Clustering of dynamically expressed transcription factors (TFs) into three major groups (*SI Text*). Dendrogram and heat-map based on scaled mean expression levels for each group ("G") are shown for the 410 TFs that are significantly differentially expressed among five stages over PSM maturation (Table S2). (**B**) Distributions of TF genes among 43 families. Families with more than seven members dynamically expressed are shown to the left, and families with less than seven members dynamically expressed are shown to the right. The total number of TF members annotated are shown in parentheses to the right of each TF family name.

Fig. S7. Quantitative RT-PCR (qRT-PCR) validation of temporal expression dynamics as detected by RNA-seq. The RNA-seq results from four stage-enriched "peak" genes (top), six "wave" pattern genes (middle), and four additional dynamically

and lowly expressed genes (bottom) were selected for validation by qRT-PCR. Red lines indicate mean normalized RNA-seq values for each stage, and blue lines indicate corresponding qRT-PCR dynamics obtained from a third biological replicate. Tomato gene identifiers are listed along with the closest Arabidopsis homolog, where applicable.

Fig. S8. Digital Differentiation Index (DDI) quantification of meristem maturation in *S. lycopersicum.* (**A and B**) One biological replicate of mRNA-seq data from each of five stages of primary shoot meristem maturation (A), and another five stages in which the SIM is added as an additional stage (B) was used for calibration, and a second biological replicate was predicted. Scaled $1/(-\log_{10}P)$ values are shown in the heat-map where larger *P* values indicate high similarity in meristem maturation as reflected by darker color along a gray scale. *P* values are derived from 841 (A) and 685 (B) DDI-selected marker genes used to estimate maturation scores of calibration samples and query samples. The reciprocal analysis in which replicate one was queried using replicate two as calibration showed the same results.

Fig. S9. Stereoscope images comparing corresponding stages of meristem ontogeny between *S. lycopersicum* (domesticated tomato), the highly branched *s* mutant, and the branched wild species *S. peruvianum*. Vertically arranged images represent equivalent developmental stages among the three genotypes. Note the nearly indistinguishable meristem morphologies among all three genotypes prior to inflorescence branching, and that *s* and *S. peruvianum* are identical from wild type until the TM stage, which grows only slightly taller in *s* and *S. peruvianum* compared to *S.*

lycopersicum, and the FM becomes slightly wider. Note that there is no ontogenetic change between all three genotypes reflecting continued meristem activity from the PSM or SIMs. Later stage inflorescence images are shown to highlight the first flower (red arrowhead) representing termination of the primary shoot meristem in *S. lycopersicum*, and highlighting the late-stage manifestation of indeterminacy and branching in *s* and *S. peruvianum*, suggesting a slower progression to floral termination. Red arrowheads in *s* and *S. peruvianum* point to the putative first developing FM in each genotype. DAG= days after germination, L=leaf number counted from 1st leaf, Size bars, 100 μm.

Fig. S10. Enrichment of selected functional groups from the differentially expressed genes between *S. lycopersicum* and *s* as determined by Mapman classifications. Scaled -$\log_{10}P$ values are shown in the heat-map, which is ordered according to meristem stage. *P* values are derived from Fisher's exact tests of the proportions of selected terms in gene sets against the proportions in the genome annotation.

Fig. S11. Selected electronic Fluorescent Pictograph (eFP) browser images depicting gene expression dynamics and differences across five meristem stages from *S. lycopersicum* and the inflorescence branching mutant *compound inflorescence* (*s*). Top colored windows list Arabidopsis genes with known roles in vegetative meristem (green), transition meristem (yellow), inflorescence and floral meristem (orange) and floral organ specification (pink) (35-38). The most closely related tomato homologs of several of these genes are shown below with associated gene identifiers. Relative gene expression calculated from RPKM values are shown as a heat-map, and red asterisks

indicate those genes showing differential expression, and therefore altered temporal dynamics, between wild type *S. lycopersicum* and *s* (Table S6). Note that the tomato homologs of Arabidopsis *PISTILLATA* (*PI*) and *APETALA3* (*AP3*) are not detected, consistent with the FM being targeted for harvesting in its earliest stages of development before floral organ primordia are visible. Numbers in parentheses indicate total read counts summed across all five stages to provide a reference for absolute expression level for each gene in each genotype.

Fig. S12. Disruption of K means clustering patterns in *s* and *S. peruvianum* compared to *S. lycopersicum*. Temporal expression patterns of 3,922 dynamically expressed genes in *S. lycopersicum* of five meristem stages (MVM, LVM, TM, SIM and FM) were grouped into 20 clusters by K-means (K=20) clustering. Scaled expression values for each of the 20 *S. lycopersicum* clusters (blue lines) were plotted with expression values for corresponding genes from *s* (green lines) and *S. peruvianum* (red lines) across five meristem stages for comparison of dynamics. Expression values of genes in *S. lycopersicum*, *s* and *S. peruvianum* were scaled independently (*SI Text*). Lines indicate scaled mean RPKM expression values. Error bars indicate standard deviations.

Fig. S13. Inflorescence branching in four wild species of tomato. The green-fruited wild species *S. chilense* (accession LA1967), *S. habrochaites* (accession LA1223), and *S. peruvianum* (accession LA0103) consistently generate between two to three branches per inflorescence (referred to as "forked" inflorescences), whereas *S. pennellii* (accession LA1272) produces a mix of branched and unbranched inflorescences. Red arrowheads

point to inflorescence branches, and white arrowhead in *S. pennellii* indicates an unbranched inflorescence.

Fig. S14. Analyses of *S. peruvianum* sequence divergence and expression levels.

(A) Frequency distribution of *S. peruvianum* SNP/variant allele frequencies. Allele frequencies were estimated from read variant frequencies using the EM algorithm implemented by SAMtools/bcftools (see *SI text*) (B) Correlation between logRPKMs (summed over all 5 stages) estimated from mapping against *S. lycopersicum* cv. Heinz annotated CDSs and *S. peruvianum* reconstructed CDSs. The correlation coefficient between expression levels is 0.997 between the two mapping processes. X-axis, log values of RPKM sum using *S. lycopersicum* cv. Heinz as the reference; y-axis, *S. peruvianum* as the reference; each dot is one gene; green dots, dynamically expressed genes; red dots, remaining genes. (C) Relationship between sequence divergence and expression level changes (*S. lycopersicum vs. S. peruvianum*): x-axis, sequence divergence; y-axis, log values of fold changes; each dot is one gene; green dots, dynamically expressed genes; red dots, remaining genes; dark blue line, linear regression fitted line.

Fig. S15. Principal Component Analysis showing overall transcriptome distances between *S. lycopersicum*, *compound inflorescence* (*s*) and *S. peruvianum*. The first two principle components from a Principal Component Analysis (PCA) of the expression profiles are shown. The variance explained by the two principal components are shown in parentheses. Only genes with sum of normalized read counts larger than 90 (average 3

counts/sample, MVM, LVM, TM, SIM, FM, 5 stages for all three) are included. The *s* mutant originated as an isogenic mutation from the *S. lycopersicum* cv. M82 background, and group together on the first principal component compared to *S. peruvianum,* which is distant.

Fig. S16. Branching of the primary inflorescence in the *sft* mutant. (**A**) Representative primary inflorescence of *sft* mutants (allele *sft-7187*) showing branching prior to reverting to a vegetative shoot that goes on to produce leaves and infrequent flowers. The red arrowheads point to branches and the yellow arrow indicates reversion of one of the inflorescence branches to a vegetative shoot. (**B and C**) Percentage of inflorescence branching in *sft* mutant and wild type (WT) plants in both greenhouse and field conditions. The percentage of primary *sft* inflorescences that branch in greenhouse conditions is higher than field conditions, likely because the flowering transition of tomato, although day neutral, is highly sensitive to reduced light intensity and quality from artificial lighting. WT, white bars; *sft-7187* mutant (C-terminal deletion allele), grey bars. Plants were grown in the greenhouse in early spring (B) and in the field during the summer at Cold Spring Harbor, New York (C). **(D)** Differential expression dynamics of the *SINGLE FLOWER TRUSS* (*SFT*) and *SELF PRUNING* (*SP*) genes during meristem maturation in *S. lycopersicum*, *s*, and *S. peruvianum*. Mean RPKM normalized RNA-seq values from two biological replicates (lines graphs) are shown. Red and green asterisks indicate significantly differentially expressed stages in *s* compared to wild type, and *S. peruvianum* compared to wild type, respectively (two fold change, $P \leq 0.05$, Table S6 and S7). Semi-quantitative RT-PCR validations for all genotypes and stages from a third

biological replicate are shown below the line graphs.

Table S1. Meristem samples collected, mRNA-seq library sizes, and read mapping statistics. The second sheet shows correlations between replicates. [This table is provided as a separate file in MS Excel (.xls) Format]

Table S2. Differentially expressed genes within *S. lycopersicum* during meristem maturation from 10 pair-wise comparisons involving five stages in the primary shoot meristem [This table is provided as a separate file in MS Excel (.xls) Format]. Only those genes showing greater than two fold change and P value ≤ 0.05 are listed. Presence of a gene satisfying our criteria for differential expression are indicated by an "Y" in the column corresponding to the specific pair-wise comparison listed in column header. The total mean normalized RPKM expression values are also shown (column L). Additional columns show cluster association (column M), whether a particular gene belongs to a Transcription Factor (TF) family (column N), and the closest *Arabidopsis thaliana* homolog (column O), Bin number (column P) and description (column Q) annotated by MapMan, and functional annotation of each gene (column R).

Table S3. Marker genes selected by DDI algorithm [This table is provided as a separate file in MS Excel (.xls) Format]. The first sheet contains 841 (3FC) marker genes for *S. lycopersicum* DDI analyses (Fig. 2). The second sheet contains 685 (3FC) marker genes for *s* and *S. peruvianum* DDI analyses using *S. lycopersicum* samples for calibration (Fig. 4).

Table S4. Mean maturation scores derived from the DDI algorithm and associated *P* values resulting from Student's t-test of the maturation scores [This table is provided as a separate file in MS Excel (.xls) Format]. Maturation scores are average scores over all marker genes used in each prediction. *P* values are presented in the form of scaled $1/(-\log_{10}P)$, which were used to generate the heat-map in Fig. 2 and Fig. 4. Zero indicates a small *P* value (small $1/(-\log_{10}P)$ value), resulting from a very small *P* calculation that is below the detection level in the R statistical package (e.g. 1-e38). First sheet: *S. lycopersicum* Primary Shoot Meristem (PSM) stages used a calibration to predict SIM and SYM. The second sheet: *S. lycopersicum* MVM to FM (five stages including SIM) used as calibration to predict three stages (TM, SIM, and FM) in compound inflorescence (*s*). Third sheet: *S. lycopersicum* MVM to FM (five stages including SIM) used as calibration to predict three stages (TM, SIM, and FM) in *S. peruvianum*.

Table S5. Flowering time for *S. lycopersicum, s*, and *S peruvianum* as measured by the number of leaves produced by the PSM before floral termination.

Leaf number was scored when plants were completely past the reproductive transition stages (M82 and *s* mutant, 19 DAG; *S. peruvianum*, 30 DAG). Percentage values were calculated from 60 plants of M82, 91 plants of *s*, and 157 plants of *S. peruvianum*. SD = Standard Deviation.

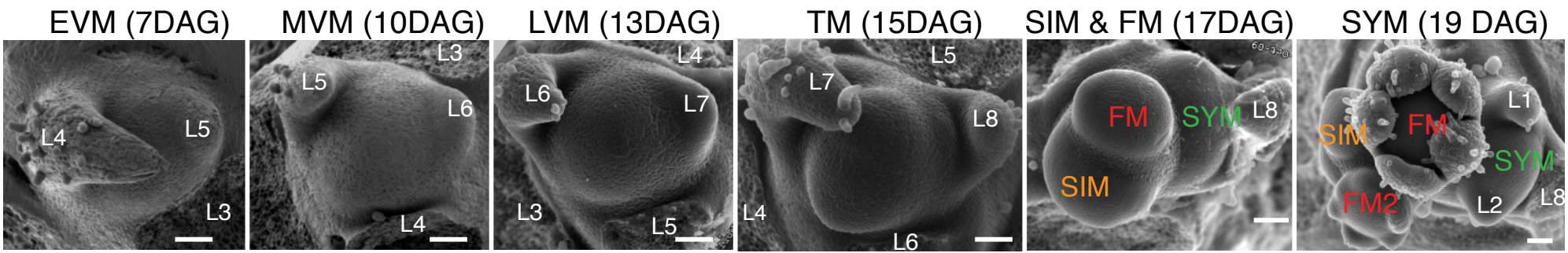
Table S6. List of differentially expressed genes between *S. lycopersicum* and *s* resulting from a stage-by-stage pair-wise comparison from MVM to FM [This table

is provided as a separate file in MS Excel (.xls) Format]. Only genes showing larger than two fold change and *P* value ≤ 0.05 are listed. Presence of a gene satisfying our criteria for differential expression are indicated by an "Y" in the column corresponding to the specific comparison. Additional columns show the sum of mean normalized RPKM expression values across five stages for each genotype (column G, H), whether a particular gene belongs to a Transcription Factor (TF) family (column I), the closest *Arabidopsis thaliana* homolog (column J), Bin number (column K) and description (column L) annotated by MapMan, and functional annotation (column M). The second sheet shows only TFs ordered according to TF family names. The remaining sheets show genes satisfying our criteria in individual stages, with fold changes and *P* values.

Table S7. List of differentially expressed genes between *S. lycopersicum* and *S. peruvianum* resulting from a stage-by-stage pair-wise comparison from MVM to FM [This table is provided as a separate file in MS Excel (.xls) Format]. Only genes showing larger than two fold change and *P* value ≤ 0.05 are listed. Presence of a gene satisfying our criteria for differential expression are indicated by a "Y" in the column corresponding to the specific comparison. Additional columns show the sum of mean normalized RPKM expression values across five stages for each genotype (column G, H), whether a gene belongs to a Transcription Factor (TF) family (column I), Bin number (column J) and description (column K) annotated by MapMan, the closest *Arabidopsis thaliana* homolog (column L), and functional annotation (column M). The second sheet shows only TFs ordered according to TF family names. The remaining sheets show genes satisfying our criteria in individual stages, with fold changes and *P* values.

Fig. S1

A


EVM (7DAG)
MVM (10DAG)
LVM (13DAG)
TM (15DAG)
SIM & FM (17DAG)
SYM (19 DAG)
L4
L5
L3
L3
L5
L6
L4
L4
L6
L7
L3
L5
L5
L7
L8
L6
FM
SYM
L8
SIM
SIM
FM
L1
SYM
FM2
L2
L8

B

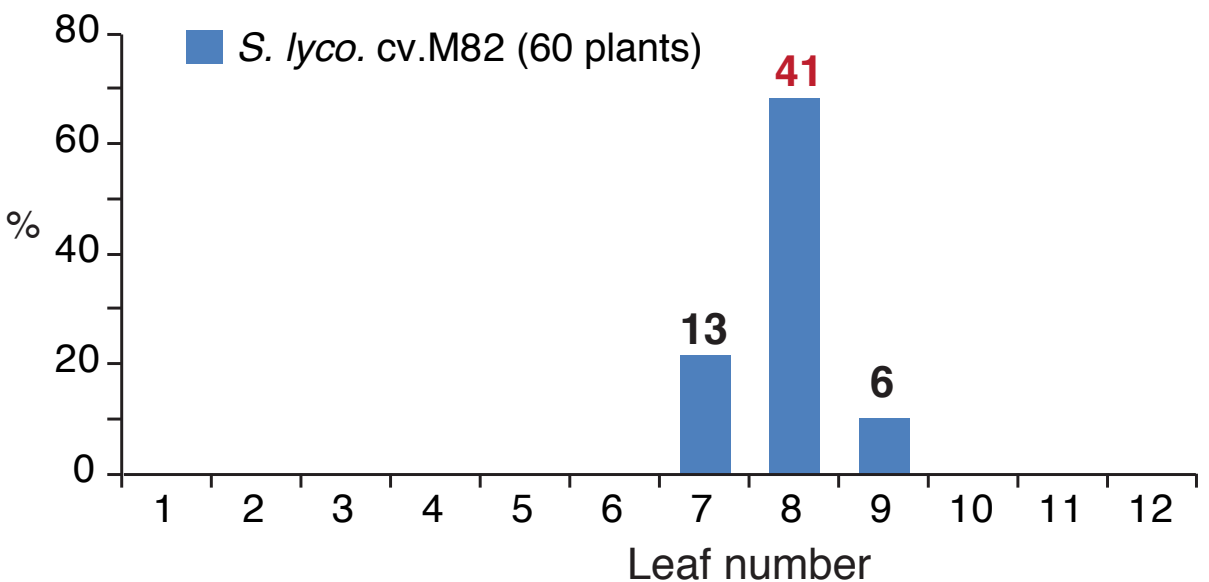

S. lyco. cv.M82 (60 plants)
%
80
60
40
20
0
41
13
6
1
2
3
4
5
6
7
8
9
10
11
12
Leaf number

Fig. S2

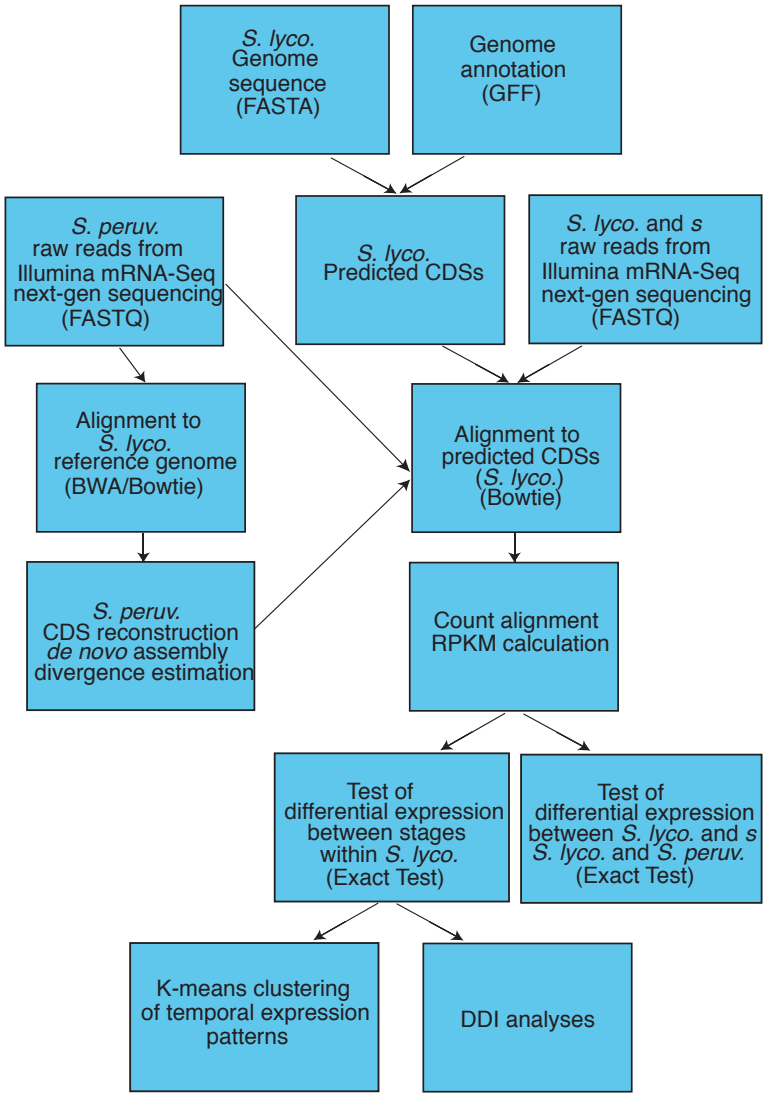
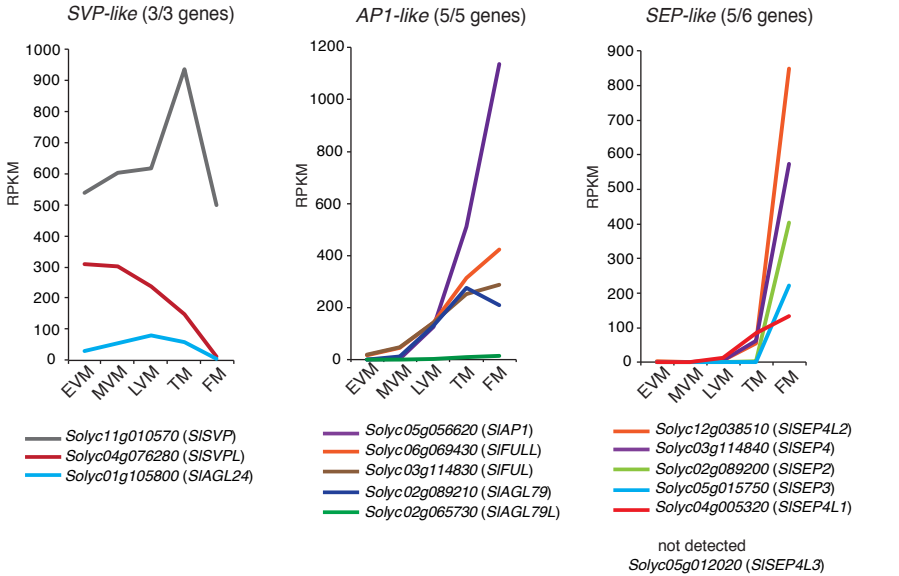

S. lyco. Genome sequence (FASTA)
Genome annotation (GFF)
S. peruv. raw reads from Illumina mRNA-Seq next-gen sequencing (FASTQ)
S. lyco. Predicted CDSs
S. lyco. and *s* raw reads from Illumina mRNA-Seq next-gen sequencing (FASTQ)
Alignment to *S. lyco.* reference genome (BWA/Bowtie)
Alignment to predicted CDSs (*S. lyco.*) (Bowtie)
S. peruv. CDS reconstruction *de novo* assembly divergence estimation
Count alignment RPKM calculation
Test of differential expression between stages within *S. lyco.* (Exact Test)
Test of differential expression between *S. lyco.* and *s* *S. lyco.* and *S. peruv.* (Exact Test)
K-means clustering of temporal expression patterns
DDI analyses

Fig. S3

A


SVP-like (3/3 genes)
AP1-like (5/5 genes)
SEP-like (5/6 genes)
RPKM
EVM
MVM
LVM
TM
FM
Solyc11g010570 (SlSVP)
Solyc04g076280 (SlSVPL)
Solyc01g105800 (SlAGL24)
Solyc05g056620 (SlAP1)
Solyc06g069430 (SlFULL)
Solyc03g114830 (SlFUL)
Solyc02g089210 (SlAGL79)
Solyc02g065730 (SlAGL79L)
Solyc12g038510 (SlSEP4L2)
Solyc03g114840 (SlSEP4)
Solyc02g089200 (SlSEP2)
Solyc05g015750 (SlSEP3)
Solyc04g005320 (SlSEP4L1)
not detected
Solyc05g012020 (SlSEP4L3)

B

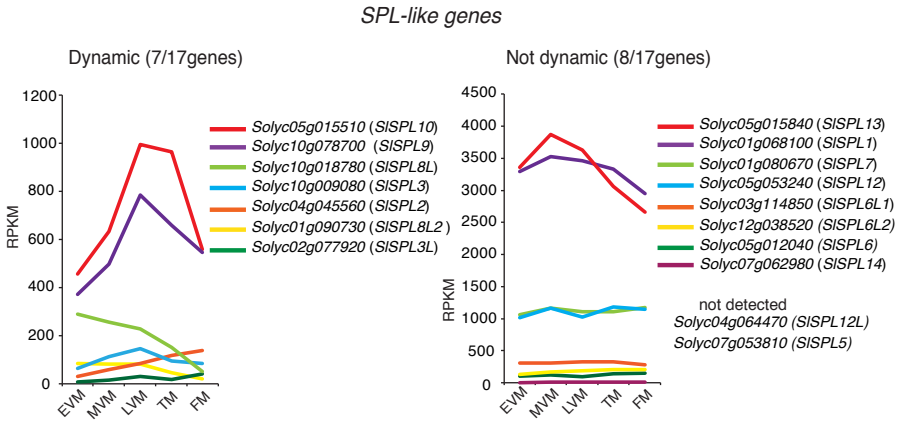

SPL-like genes
Dynamic (7/17genes)
Not dynamic (8/17genes)
RPKM
EVM
MVM
LVM
TM
FM
Solyc05g015510 (SlSPL10)
Solyc10g078700 (SlSPL9)
Solyc10g018780 (SlSPL8L)
Solyc10g009080 (SlSPL3)
Solyc04g045560 (SlSPL2)
Solyc01g090730 (SlSPL8L2)
Solyc02g077920 (SlSPL3L)
Solyc05g015840 (SlSPL13)
Solyc01g068100 (SlSPL1)
Solyc01g080670 (SlSPL7)
Solyc05g053240 (SlSPL12)
Solyc03g114850 (SlSPL6L1)
Solyc12g038520 (SlSPL6L2)
Solyc05g012040 (SlSPL6)
Solyc07g062980 (SlSPL14)
not detected
Solyc04g064470 (SlSPL12L)
Solyc07g053810 (SlSPL5)

Fig. S4

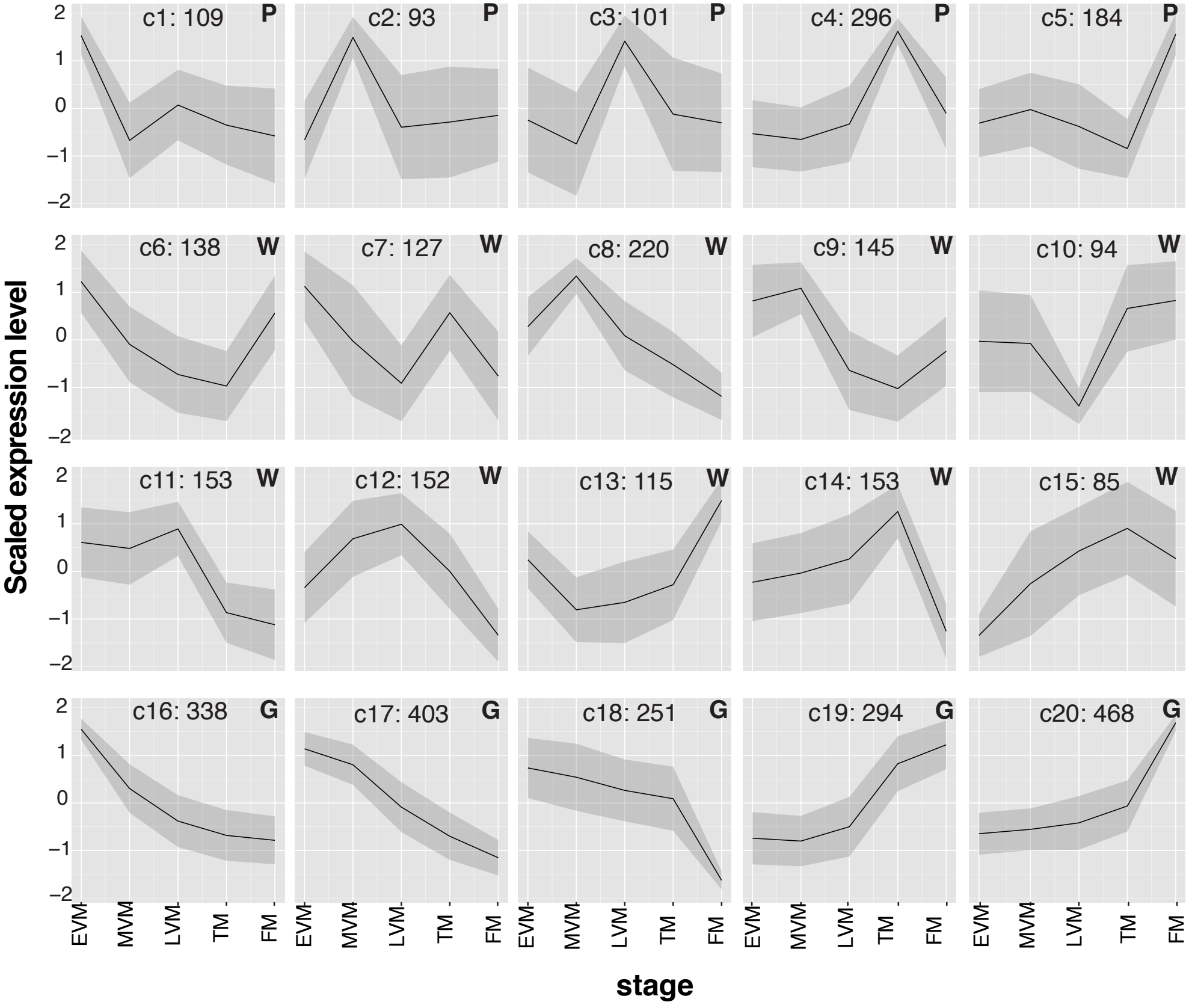

c1: 109 P
c2: 93 P
c3: 101 P
c4: 296 P
c5: 184 P
c6: 138 W
c7: 127 W
c8: 220 W
c9: 145 W
c10: 94 W
c11: 153 W
c12: 152 W
c13: 115 W
c14: 153 W
c15: 85 W
c16: 338 G
c17: 403 G
c18: 251 G
c19: 294 G
c20: 468 G
Scaled expression level
2
1
0
−1
−2
EVM
MVM
LVM
TM
FM
stage

Fig. S5

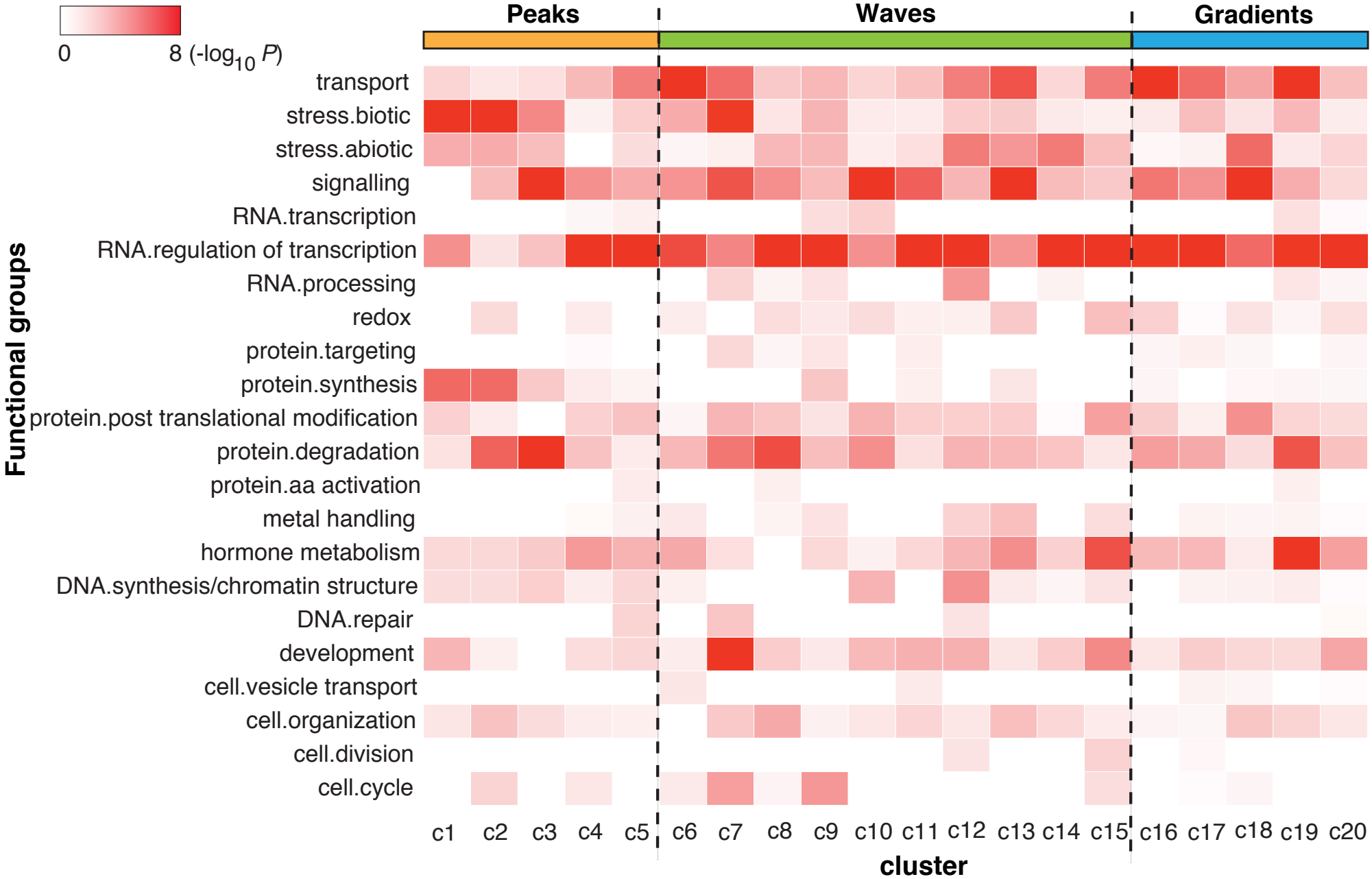

0
8 (-log10 P)
Peaks
Waves
Gradients
Functional groups
transport
stress.biotic
stress.abiotic
signalling
RNA.transcription
RNA.regulation of transcription
RNA.processing
redox
protein.targeting
protein.synthesis
protein.post translational modification
protein.degradation
protein.aa activation
metal handling
hormone metabolism
DNA.synthesis/chromatin structure
DNA.repair
development
cell.vesicle transport
cell.organization
cell.division
cell.cycle
c1 c2 c3 c4 c5 c6 c7 c8 c9 c10 c11 c12 c13 c14 c15 c16 c17 c18 c19 c20
cluster

Fig. S6

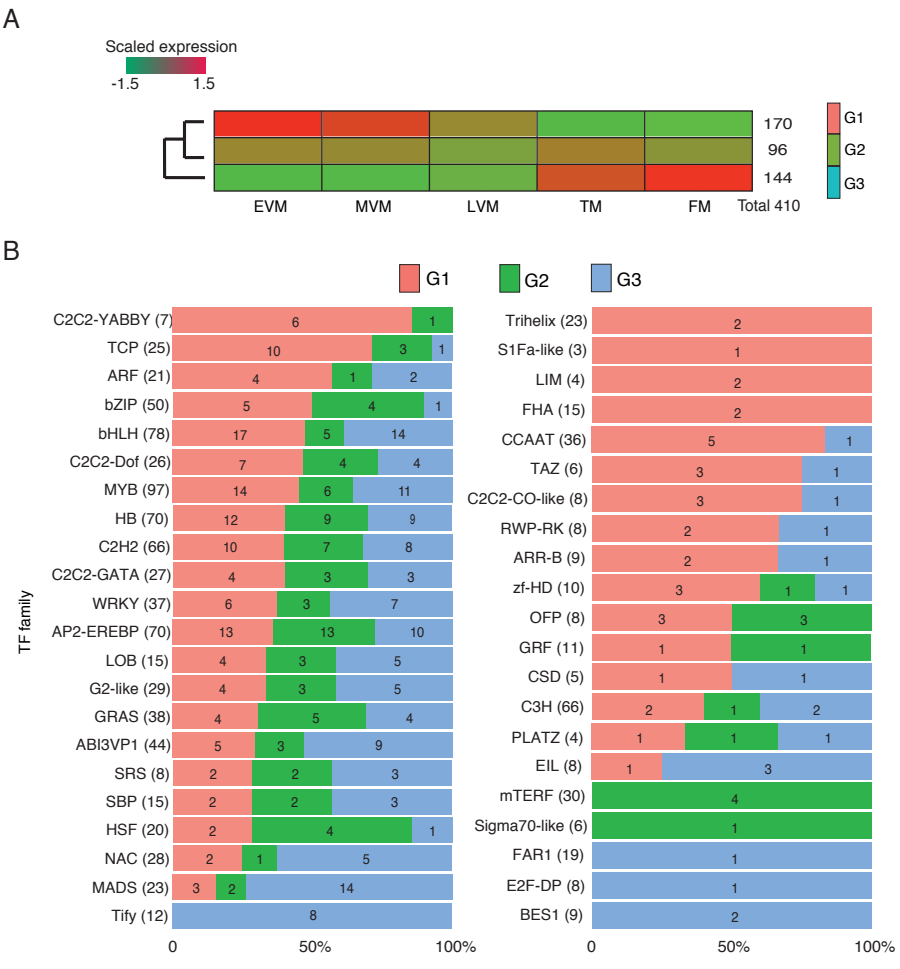

A
Scaled expression
-1.5
1.5
170
96
144
G1
G2
G3
EVM
MVM
LVM
TM
FM
Total 410
B
G1
G2
G3
TF family
C2C2-YABBY (7)
TCP (25)
ARF (21)
bZIP (50)
bHLH (78)
C2C2-Dof (26)
MYB (97)
HB (70)
C2H2 (66)
C2C2-GATA (27)
WRKY (37)
AP2-EREBP (70)
LOB (15)
G2-like (29)
GRAS (38)
ABI3VP1 (44)
SRS (8)
SBP (15)
HSF (20)
NAC (28)
MADS (23)
Tify (12)
Trihelix (23)
S1Fa-like (3)
LIM (4)
FHA (15)
CCAAT (36)
TAZ (6)
C2C2-CO-like (8)
RWP-RK (8)
ARR-B (9)
zf-HD (10)
OFP (8)
GRF (11)
CSD (5)
C3H (66)
PLATZ (4)
EIL (8)
mTERF (30)
Sigma70-like (6)
FAR1 (19)
E2F-DP (8)
BES1 (9)
0
50%
100%
0
50%
100%

Fig. S7

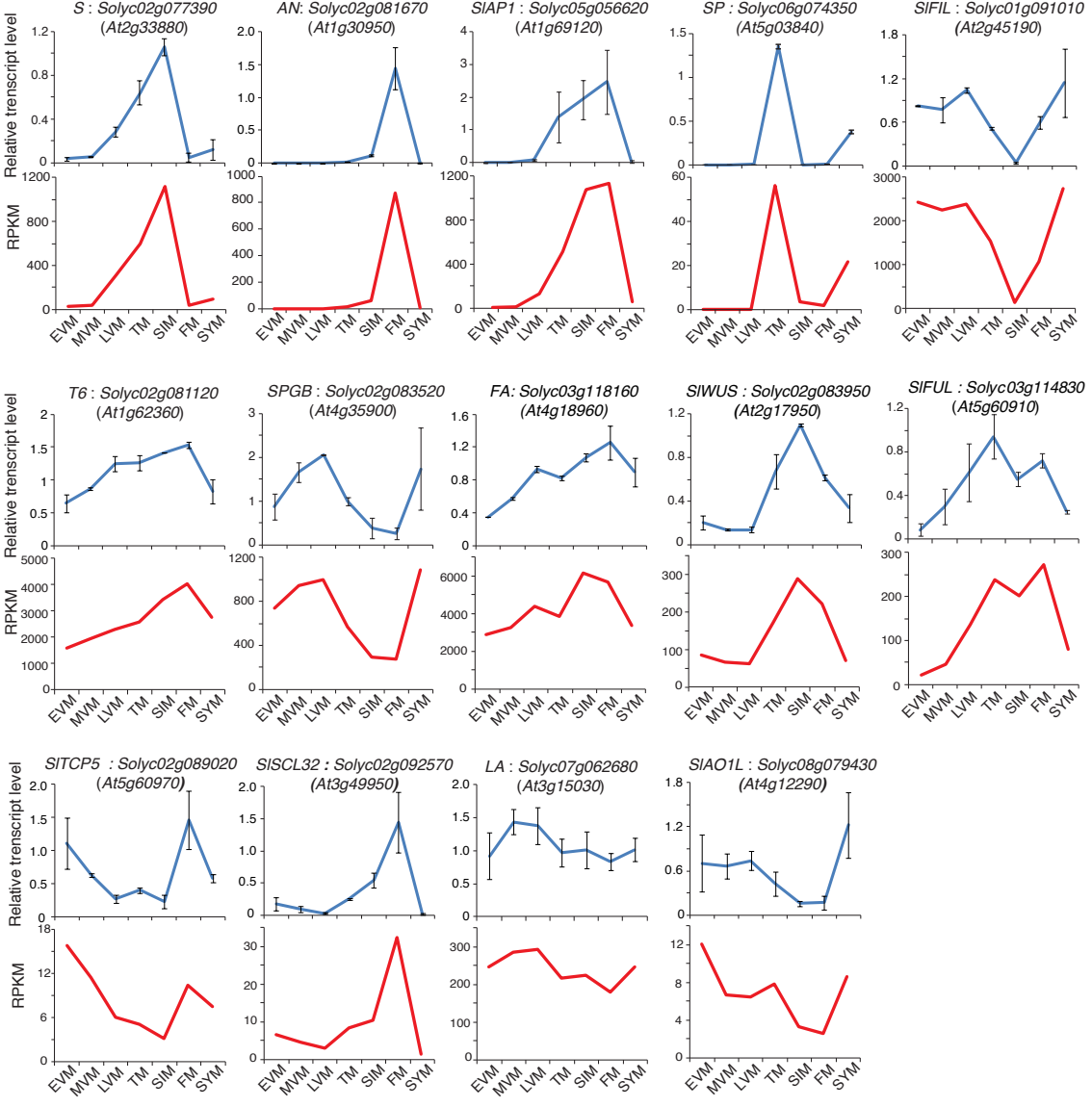

S : Solyc02g077390 (At2g33880)
AN: Solyc02g081670 (At1g30950)
SlAP1 : Solyc05g056620 (At1g69120)
SP : Solyc06g074350 (At5g03840)
SlFIL : Solyc01g091010 (At2g45190)
T6 : Solyc02g081120 (At1g62360)
SPGB : Solyc02g083520 (At4g35900)
FA: Solyc03g118160 (At4g18960)
SlWUS : Solyc02g083950 (At2g17950)
SlFUL : Solyc03g114830 (At5g60910)
SlTCP5 : Solyc02g089020 (At5g60970)
SlSCL32 : Solyc02g092570 (At3g49950)
LA : Solyc07g062680 (At3g15030)
SlAO1L : Solyc08g079430 (At4g12290)
Relative trenscript level
RPKM
EVM MVM LVM TM SlM FM SYM

Fig. S8

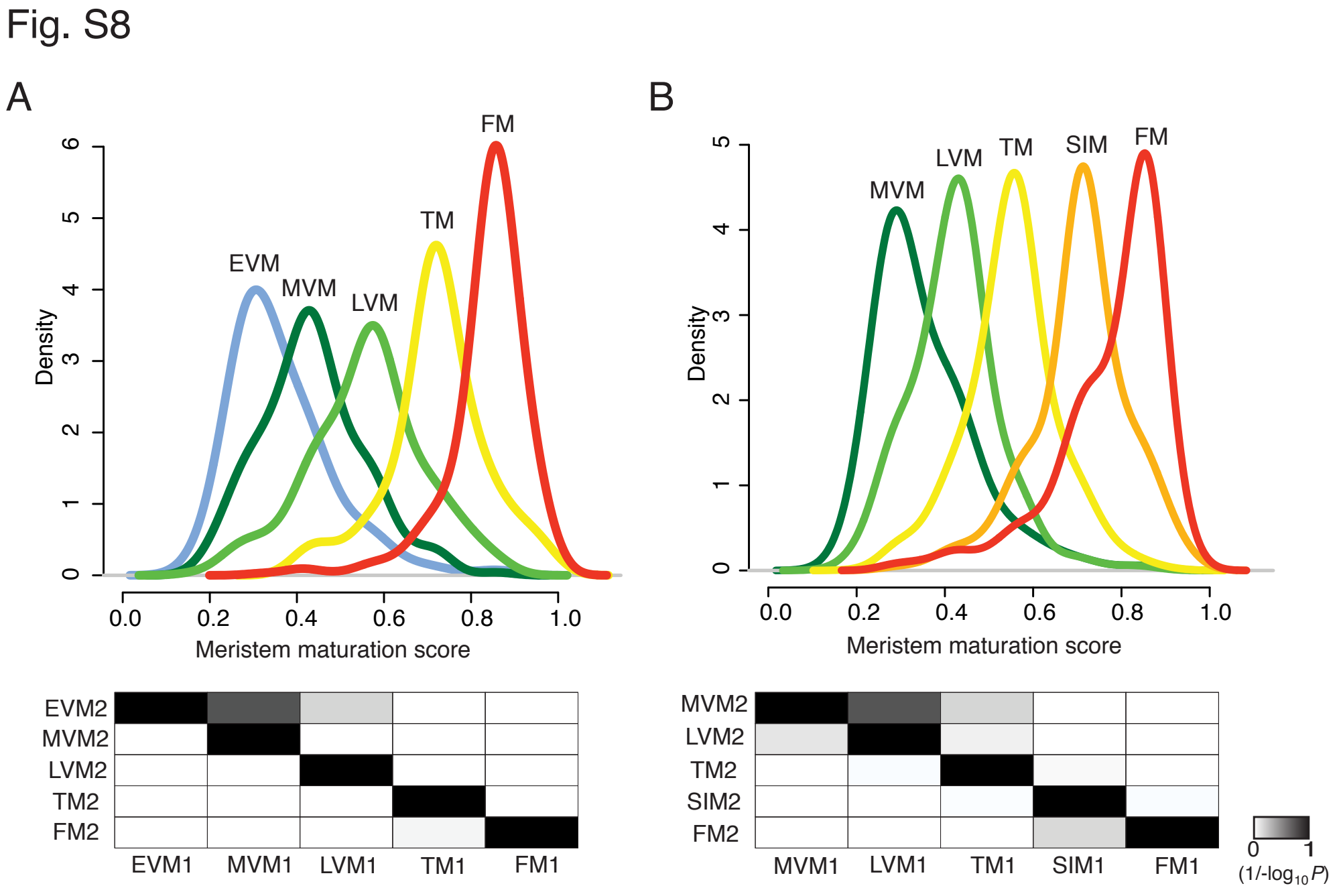

A
B
EVM
MVM
LVM
TM
FM
SIM
Density
Meristem maturation score
0.0
0.2
0.4
0.6
0.8
1.0
EVM2
MVM2
LVM2
TM2
FM2
SIM2
EVM1
MVM1
LVM1
TM1
FM1
SIM1
0
1
(1/-log10 P)

Fig. S9

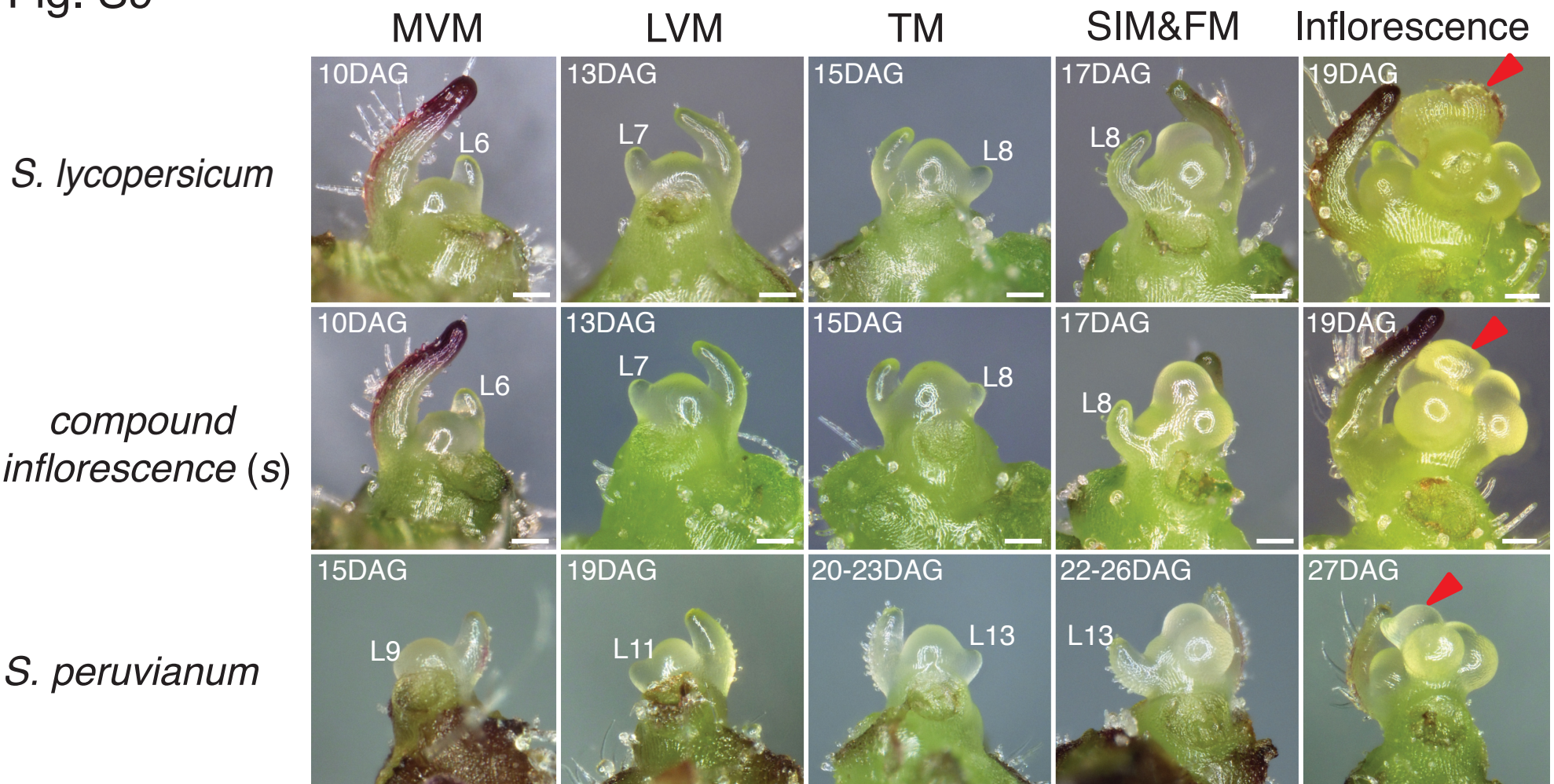

MVM
LVM
TM
SIM&FM
Inflorescence
S. lycopersicum
compound inflorescence (s)
S. peruvianum
10DAG
L6
13DAG
L7
15DAG
L8
17DAG
L8
19DAG
10DAG
L6
13DAG
L7
15DAG
L8
17DAG
L8
19DAG
15DAG
L9
19DAG
L11
20-23DAG
L13
22-26DAG
L13
27DAG

Fig. S10

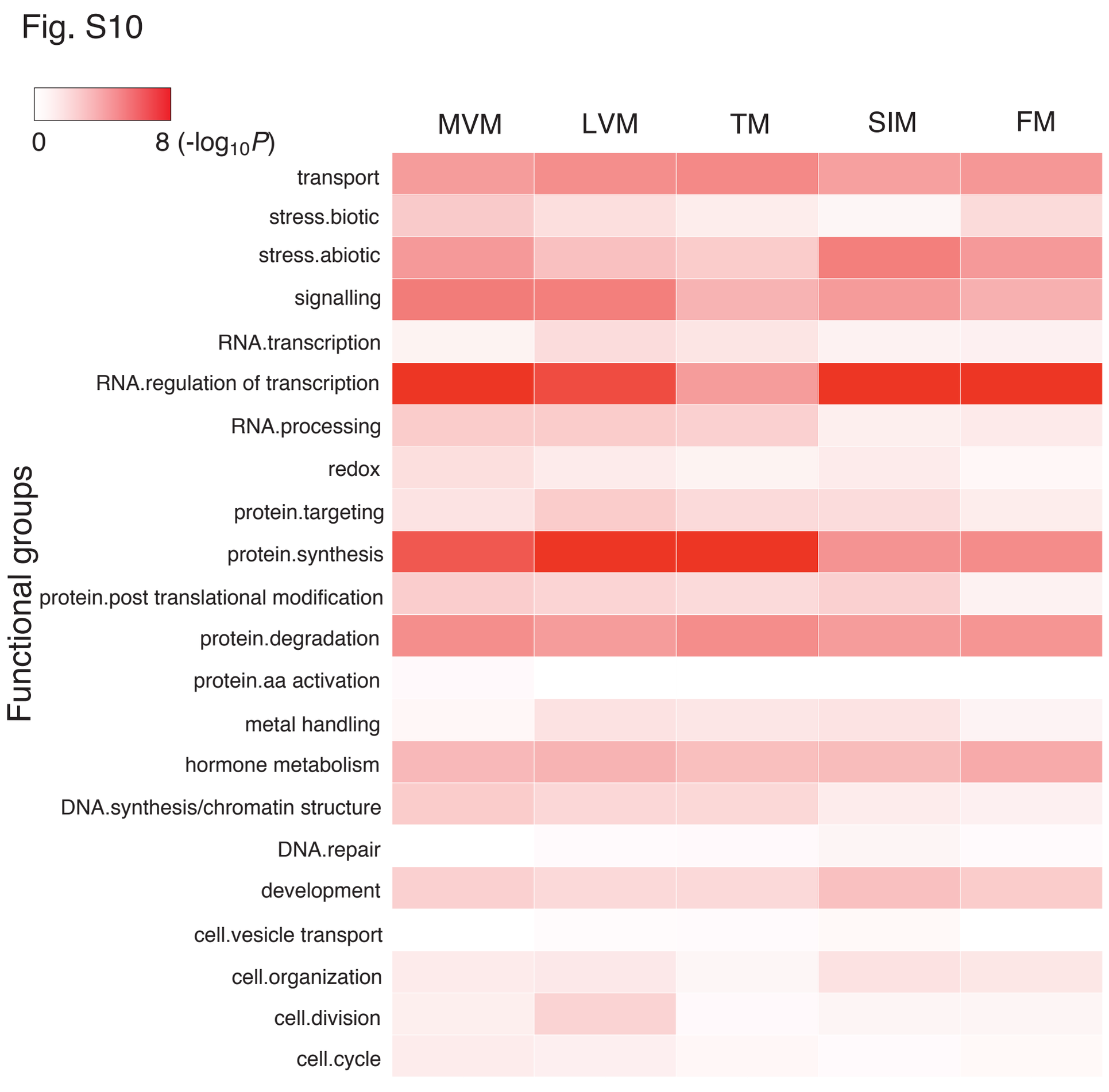

0
8 (-log10P)
MVM
LVM
TM
SIM
FM
Functional groups
transport
stress.biotic
stress.abiotic
signalling
RNA.transcription
RNA.regulation of transcription
RNA.processing
redox
protein.targeting
protein.synthesis
protein.post translational modification
protein.degradation
protein.aa activation
metal handling
hormone metabolism
DNA.synthesis/chromatin structure
DNA.repair
development
cell.vesicle transport
cell.organization
cell.division
cell.cycle

Fig. S11

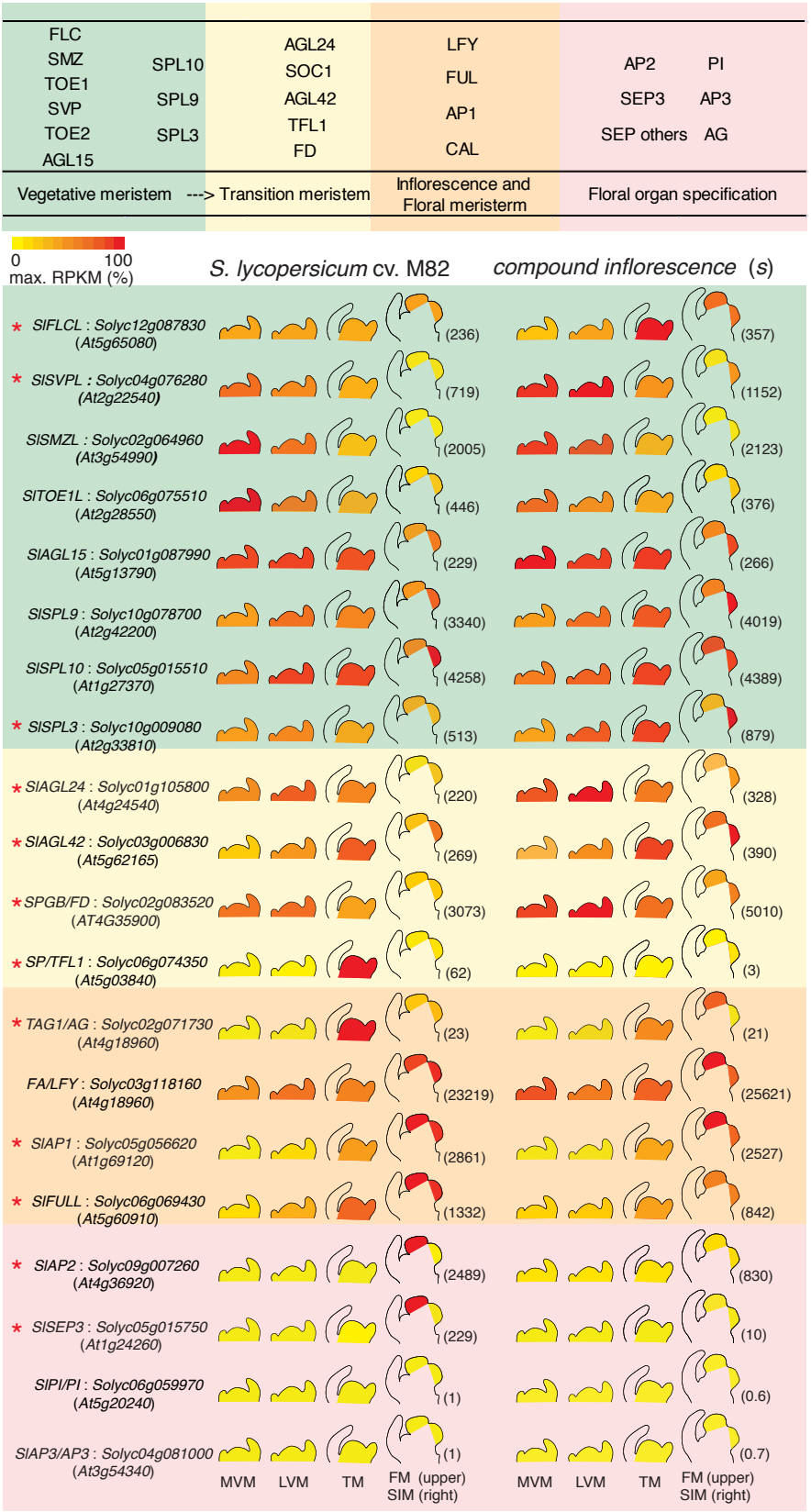

FLC
SMZ
TOE1
SVP
TOE2
AGL15
SPL10
SPL9
SPL3
AGL24
SOC1
AGL42
TFL1
FD
LFY
FUL
AP1
CAL
AP2
PI
SEP3
AP3
SEP others
AG
Vegetative meristem
---> Transition meristem
Inflorescence and Floral meristerm
Floral organ specification
0
100
max. RPKM (%)
S. lycopersicum cv. M82
compound inflorescence (s)
* SlFLCL : Solyc12g087830 (At5g65080)
(236)
(357)
* SlSVPL : Solyc04g076280 (At2g22540)
(719)
(1152)
SlSMZL : Solyc02g064960 (At3g54990)
(2005)
(2123)
SlTOE1L : Solyc06g075510 (At2g28550)
(446)
(376)
SlAGL15 : Solyc01g087990 (At5g13790)
(229)
(266)
SlSPL9 : Solyc10g078700 (At2g42200)
(3340)
(4019)
SlSPL10 : Solyc05g015510 (At1g27370)
(4258)
(4389)
* SlSPL3 : Solyc10g009080 (At2g33810)
(513)
(879)
* SlAGL24 : Solyc01g105800 (At4g24540)
(220)
(328)
* SlAGL42 : Solyc03g006830 (At5g62165)
(269)
(390)
* SPGB/FD : Solyc02g083520 (AT4G35900)
(3073)
(5010)
* SP/TFL1 : Solyc06g074350 (At5g03840)
(62)
(3)
* TAG1/AG : Solyc02g071730 (At4g18960)
(23)
(21)
FA/LFY : Solyc03g118160 (At4g18960)
(23219)
(25621)
* SlAP1 : Solyc05g056620 (At1g69120)
(2861)
(2527)
* SlFULL : Solyc06g069430 (At5g60910)
(1332)
(842)
* SlAP2 : Solyc09g007260 (At4g36920)
(2489)
(830)
* SlSEP3 : Solyc05g015750 (At1g24260)
(229)
(10)
SlPI/PI : Solyc06g059970 (At5g20240)
(1)
(0.6)
SlAP3/AP3 : Solyc04g081000 (At3g54340)
(1)
(0.7)
MVM
LVM
TM
FM (upper)
SIM (right)
MVM
LVM
TM
FM (upper)
SIM (right)

Fig. S12

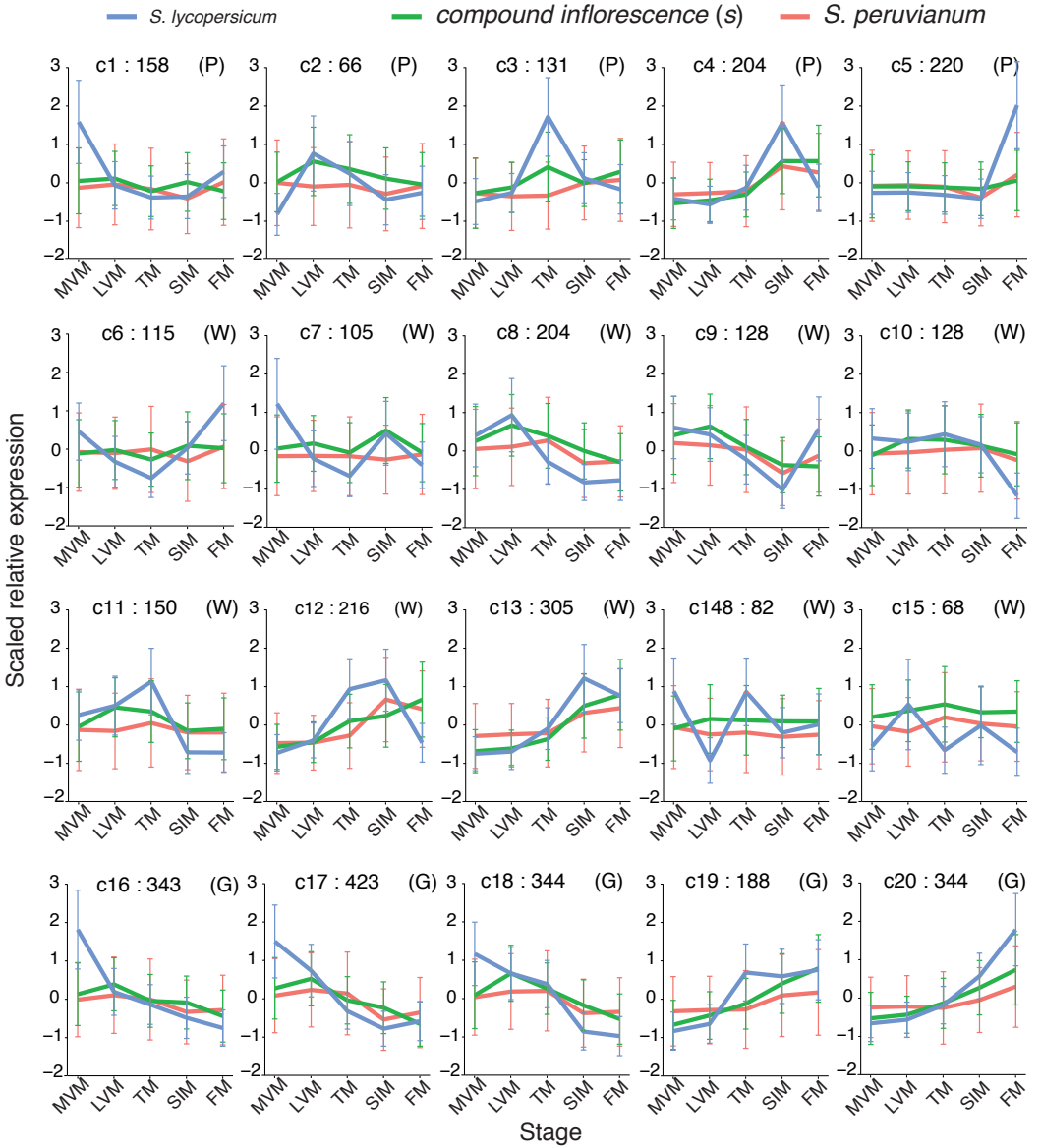

S. lycopersicum
compound inflorescence (s)
S. peruvianum
c1 : 158 (P)
c2 : 66 (P)
c3 : 131 (P)
c4 : 204 (P)
c5 : 220 (P)
c6 : 115 (W)
c7 : 105 (W)
c8 : 204 (W)
c9 : 128 (W)
c10 : 128 (W)
c11 : 150 (W)
c12 : 216 (W)
c13 : 305 (W)
c148 : 82 (W)
c15 : 68 (W)
c16 : 343 (G)
c17 : 423 (G)
c18 : 344 (G)
c19 : 188 (G)
c20 : 344 (G)
MVM
LVM
TM
SIM
FM
Scaled relative expression
Stage

Fig. S13

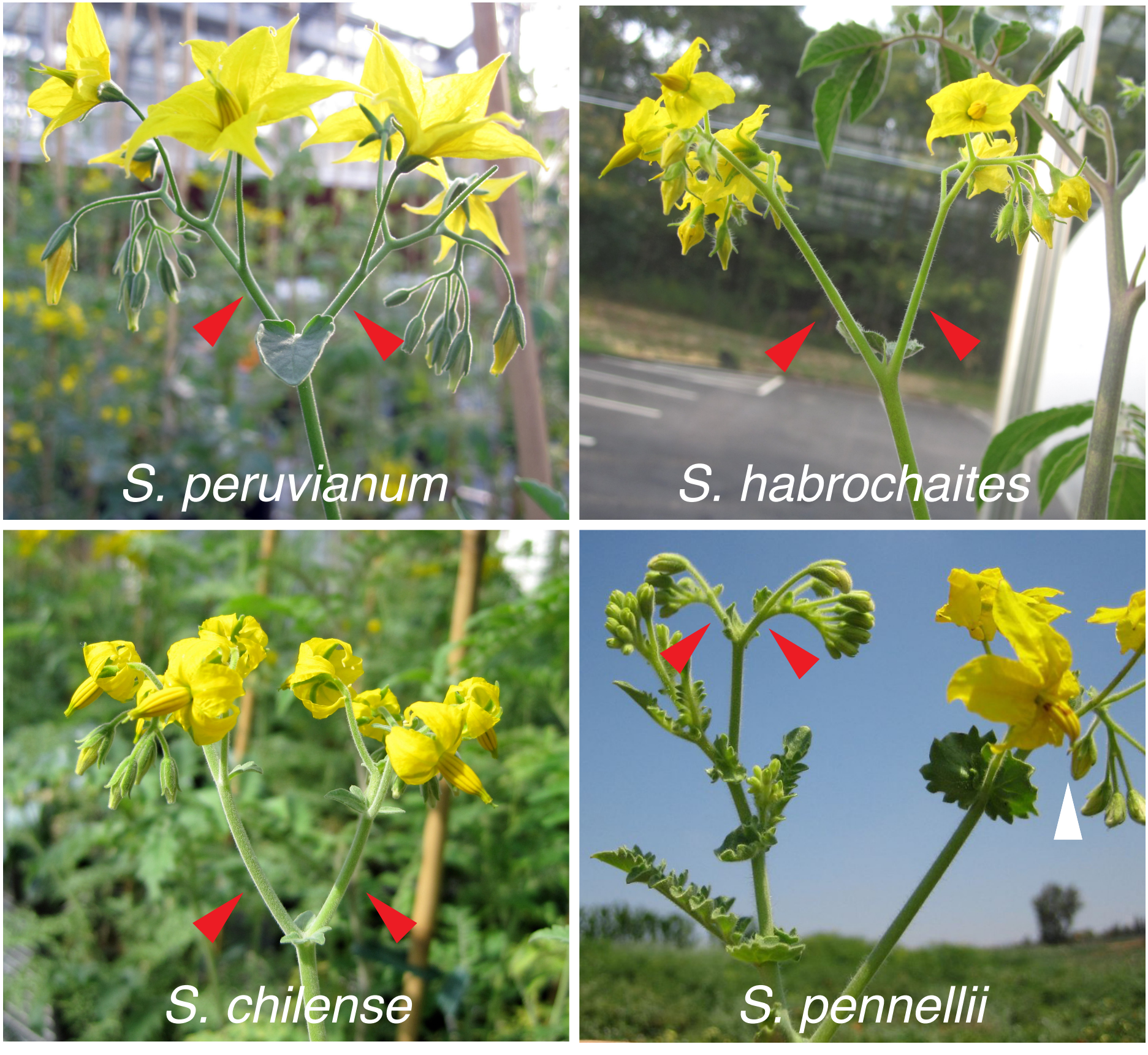


Fig. S14

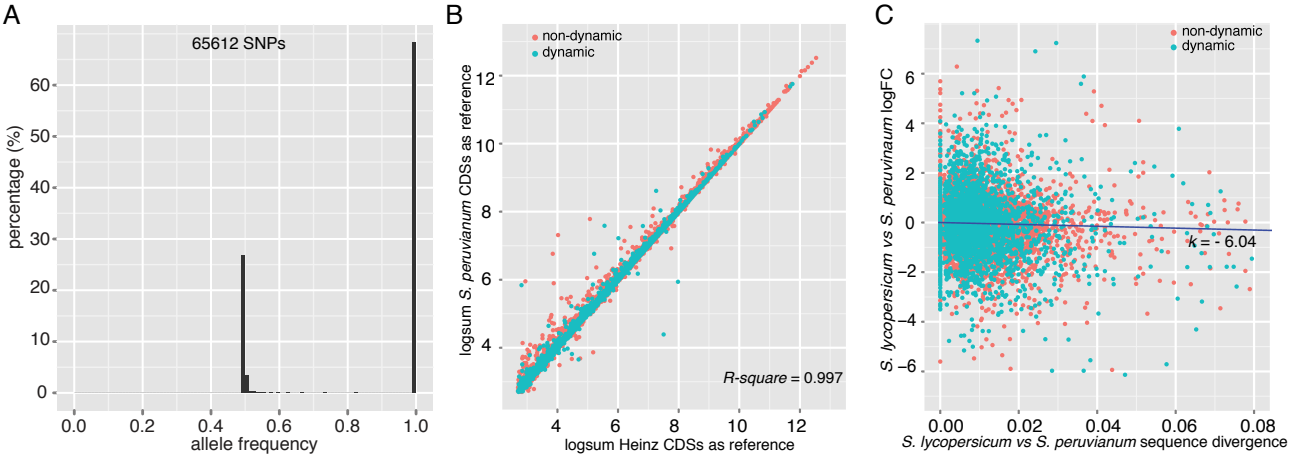

A
65612 SNPs
percentage (%)
allele frequency
B
non-dynamic
dynamic
logsum S. peruvianum CDSs as reference
logsum Heinz CDSs as reference
R-square = 0.997
C
non-dynamic
dynamic
S. lycopersicum vs S. peruvinaum logFC
S. lycopersicum vs S. peruvianum sequence divergence
k = - 6.04

Fig. S15

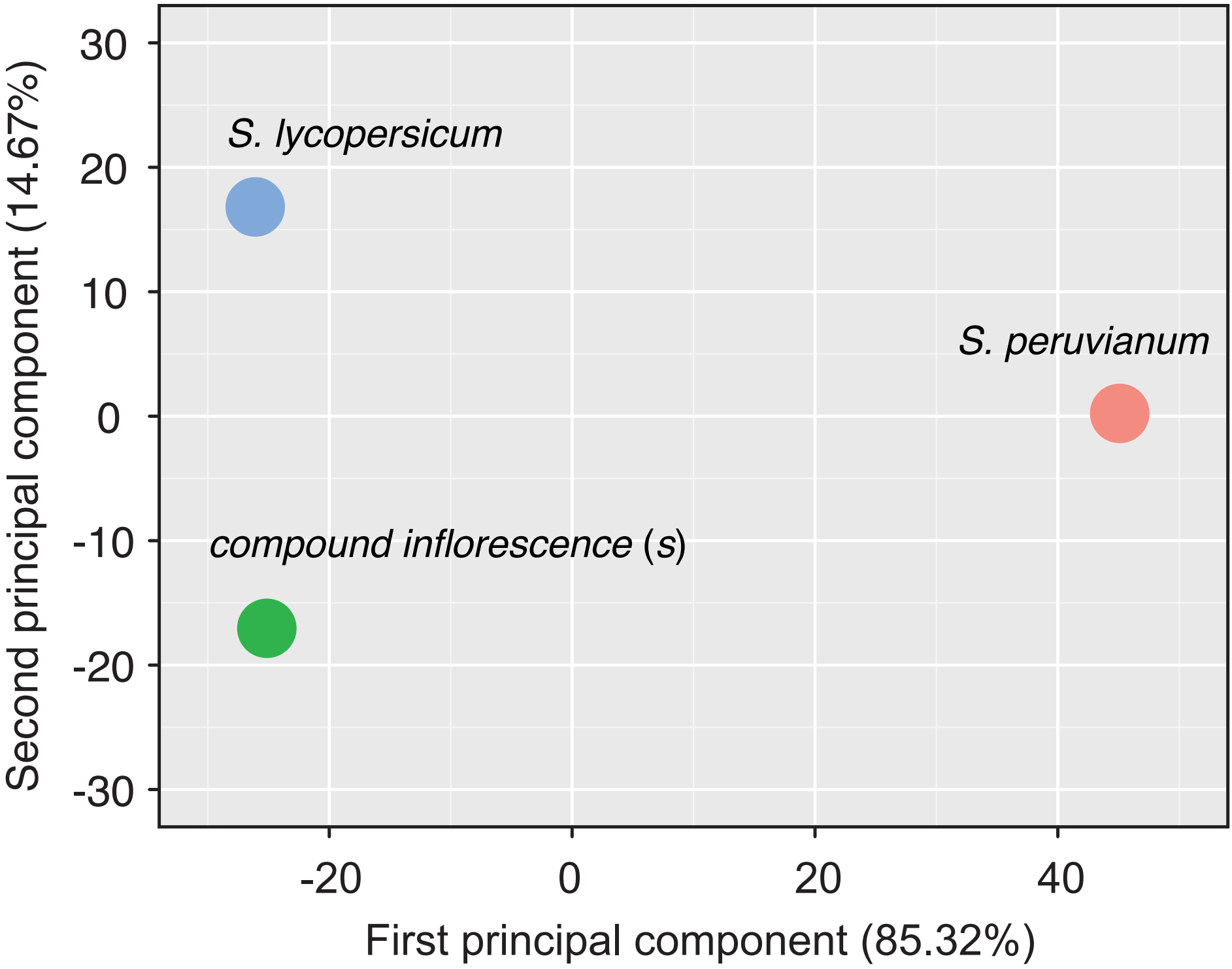

30
20
10
0
-10
-20
-30
Second principal component (14.67%)
S. lycopersicum
S. peruvianum
compound inflorescence (s)
-20
0
20
40
First principal component (85.32%)

Fig. S16

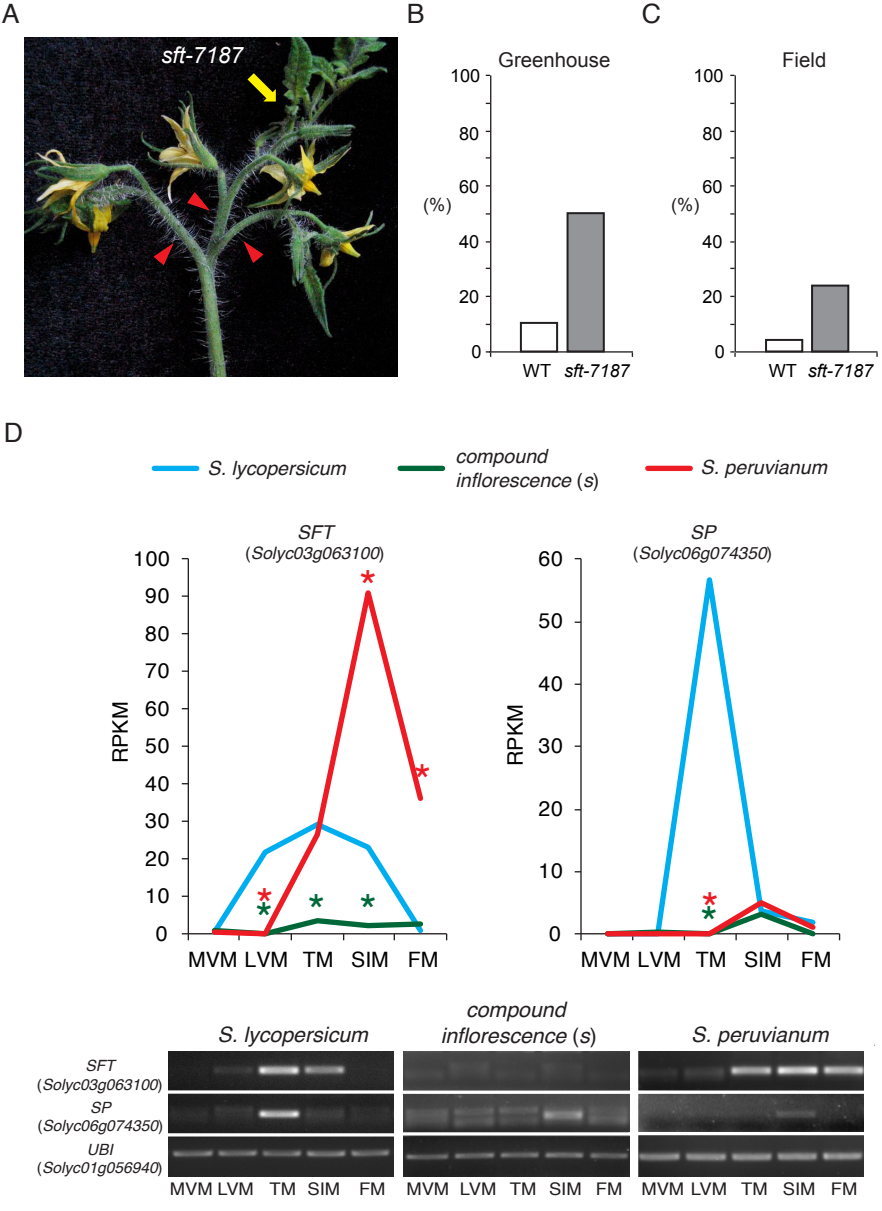

A
sft-7187
B
Greenhouse
100
80
60
40
20
0
(%)
WT
sft-7187
C
Field
100
80
60
40
20
0
(%)
WT
sft-7187
D
S. lycopersicum
compound inflorescence (s)
S. peruvianum
SFT
(Solyc03g063100)
RPKM
100
90
80
70
60
50
40
30
20
10
0
MVM LVM TM SIM FM
SP
(Solyc06g074350)
RPKM
60
50
40
30
20
10
0
MVM LVM TM SIM FM
S. lycopersicum
compound inflorescence (s)
S. peruvianum
SFT
(Solyc03g063100)
SP
(Solyc06g074350)
UBI
(Solyc01g056940)
MVM LVM TM SIM FM
MVM LVM TM SIM FM
MVM LVM TM SIM FM

Table S5

Genotype	7 leaves	8 leaves	9 leaves	10 leaves	11 leaves	12 leaves	13 leaves	14 leaves	15 leaves	16 leaves	17 leaves	Averages leaf No. (SD)
***S. lycopersicum* (M82)**	21.7 %	68.3%	10%	0	0	0	0	0	0	0	0	7.75 (±0.56)
***s* (M82)**	4.4 %	63.7 %	30.77 %	1.1 %	0	0	0	0	0	0	0	8.29 (±0.54)
***S. peruvianum* (LA0103)**	0	0	0	0.64 %	6.37 %	13.38 %	28.03 %	33.76 %	13.38 %	3.18 %	1.27 %	13.47 (±1.26)