PNAS
www.pnas.org

Supplementary Information for:

Evolution of vascular plants through redeployment of ancient developmental regulators

Kuan-Ju Lu[1]†, Nicole van 't Wout Hofland[1]†, Eliana Mor[2,3], Sumanth Mutte[1], Paul Abrahams[1], Hirotaka Kato[1], Klaas Vandepoele[2,3], Dolf Weijers[1]* and Bert De Rybel[1,2,3]*

Correspondence to: beryb@psb.vib-ugent.be, dolf.weijers@wur.nl

This PDF file includes:

Supplementary text: Materials and Methods
Figures S1 to S15
Legends for Tables S1 to S6
Legends for Datasets S1 to S2
SI References

Other supplementary materials for this manuscript include the following:

Tables S1 to S6
Datasets S1 to S2

Supplementary Information Text: Materials and Methods

Sequence data

Genomic sequences of green algae and land plant species *Cyanidioschyzon merolae*, *Ostreococcus lucimarinus*, *Chlamydomonas reinhardtii, Physcomitrella patens*, *Selaginella moellendorffii, Amborella trichopoda, Oryza sativa*, *Zea mays, Arabidopsis thaliana, Solanum lycopersicum and Populus tremulus* were retrieved from the comparative genomics database PLAZA 3.0, PLAZA 2.5 and PicoPLAZA (33-36) (*http://bioinformatics.psb.ugent.be/plaza*). Genomic data of *Marchantia polymorpha* was accessed via Phytozome v12 (37) (*https://phytozome.jgi.doe.gov/pz/portal.html*). *Klebsormidium nitens* genomic sequences were collected from the *Klebsormidium nitens* NIES-2285 genome project (38) (*http://www.plantmorphogenesis.bio.titech.ac.jp/~algae_genome_project/klebsormidium/*).
Transcriptomic data of *Culcita macrocarpa, Equisetum diffusum, Picea engelmanni, Marchantia polymorpha, Phaeoceros carolinianus* (*Phc*), *Leiosporoceros dussii* (*Ldu*) and a wide range of lycophytes, bryophytes and algae were derived from the transcriptome database OneKP (39) (**Table S3**). For extensive LHW and TMO5 analyses full length protein sequences of TMO5, TMO5-LIKE1 (T5L1), TMO5-LIKE2 (T5L2), TMO5-LIKE3 (T5L3), TMO5-LIKE4 (T5L4), LHW, LHW-LIKE1 (LL1) and LHW-LIKE2 (LL2) were used. PPR regions of LHW-LIKE3 (LL3) were excluded from the search query. Transcriptome assembly outputs resulting from tBLASTn searches were translated into protein sequences prior to alignments. Overlap between differentially expressed gene sets from different species were performed using the Workbench in PLAZA Dicots 4.0 (36).

Sequence alignments and phylogenetic tree construction

Full-length plant proteins were aligned using MAFFT version 7 (40) (*http://mafft.cbrc.jp/alignment/server/*), L-INS-i algorithm using default parameters. Alignments were cleaned using TrimAl (41) where the positions that have more than 20% gaps were removed prior to the phylogeny construction. MSA in the Figures were visualized by Jalview v2.10.2 using clustal based color coding. Both 'Modeltest-ng' and 'Partition finder v2' indicated the 'JTT' with GAMMA rate heterogeneity as the best model of evolution for both TMO5 and LHW (42). IQtree was used for the phylogenetic tree construction based on the Maximum Likelihood (ML) (43) with 1000 rapid bootstrap replicates and tree branches tested by SH-aLRT method. bHLH domain containing sequences that are neither TMO5 or LHW, from four different land plant species, were used as outgroup sequences. The obtained trees were visualized using iTOL v4 (itol.embl.de; see **Fig. S1** and **Fig. S2**). The files containing all sequences used in FASTA format are present as **Dataset S1** for TMO5 and **Dataset S2** for LHW.

Plant material and growth conditions

All seeds were surface sterilized and grown on ½ MS plates. After a two-day stratification at 4°C, seedlings were grown under long day conditions (16 hours light, 8 hours dark) at 22°C. Arabidopsis ecotype Columbia-0 (Col-0) was used as wild-type. *lhw, lhw ll1, tmo5 and tmo5 t5l1* mutants were generated by and obtained from (44, 45). *Marchantia polymorpha* strain Takaragaike-1 (Tak-1) was used as wild type and in all transformations and physiological analyses in this study. The growth condition for *Marchantia* is as follow: 50-60 μmol photos $m^{-2}s^{-1}$ white LED light at 22°C with continuous light.

***Marchantia polymorpha* transformation**

For gain-of-function and loss-of-function mutant generation, the agrobacterium dependent thallus transformation was performed as previously described (46): each 14-day-old Tak-1 thallus was dissected into 8 pieces after removing the apical meristem niches and transferred onto half-strength Gamborg B5 (1/2 B5) medium with 1% sucrose and 1% Daishin agar plates for 3 days. The thallus pieces were further transferred and incubated with agrobacterium containing indicated plasmids in 50 ml 0M51C medium with 200μM Acetosyringone (4'-Hydroxy-3',5'-dimethoxyacetophenone) in 200-ml flasks with agitation at 130 rpm for another 3 days. The thalli were further filtered and wished

with water and transferred onto 1/2 B5 plates containing respective antibiotics for selection. Independent T1 lines were isolated and single G1 lines from independent T1 line were generated by subcultivating single gemmalings which emerged asexually from single initial cells (47). The next generation of G1, called G2 generation, was used for analyses.

Cloning

Complementation vectors expressing p*TMO5* or p*LHW* were constructed through conventional cloning of pPLV28 (48). 3 kb fragments upstream of the *TMO5* or *LHW* ATG were amplified from genomic DNA using Phusion Flash PCR Master Mix (Thermo Scientific). All further cloning procedures were performed using Seamless Ligation Cloning Extract (SLiCE), with 15 homologous bases. Using primers with flanking LIC sites, a YFP was inserted in the LIC site of p*LHW*::LIC, creating p*LHW*::YFP. cDNA of *Populus trichocarpa, Oryza sativa, Picea abies and Selaginella moellendorffii* was used to amplify orthologs, while *MpTMO5*, *MpLHW*, *KnLHW*, *EgLHW*, *EdTMO5*, *SwLHW*, *CmaLHW* and *Cma2008862* were gene synthesized by GeneArt or GenScript. *KnTMO5* CDS was amplified from genomic DNA (see below) by stitching PCR. The cDNAs of TMO5 orthologs were amplified and introduced in pGIIB-p*TMO5::*LIC-NOSt. The cDNAs of LHW orthologs, excluding the stop codon, were amplified and cloned into pGIIB-p*LHW::*LIC-YFP-NOSt. BiFC plasmids were constructed by cloning the orthologs into a modified pPLV22 or pPLV27 vector containing a p*35S*::LIC-nYFP or p*35S*::LIC-cYFP respectively. All constructs were verified by sequencing and transformed into Arabidopsis Col-0 plants by simplified floral dipping (49). All *MpTMO5* related constructs and the *MpLHW* promoter construct used for generating Marchantia transformants were cloned by PCR amplification from the genomic DNA with primers listed in **Table S6**. For *Mp*LHW overexpression and GR lines, the *MpLHW* CDS was synthesized (**Table S5**) and sub-cloned by PCR amplification with primers listed in **Table S6**. The amplified fragments were cloned into pENTR-D-TOPO vector (Invitrogen) according to the user manual and further subcloned into pMpGWB vectors as previously described (50). For FRET-FLIM analysis, *MpTMO5* genomic fragment and *MpLHW*, *KnTMO5* and *KnLHW* coding sequence were amplified by PCR with primers listed in **Table S6** and cloned into pMON999 vectors with sCFP3A and sYFP2 to create *Mp*TMO5-sCFP3A and *Mp*LHW-sYFP2 constructs by SLiCE cloning method (51). For hornwort LHW (PhLHW, PhcLHW, and LduLHW) complementation analysis, the CDS of each gene was synthesized by IDT gBlock and directly fused with pGIIB-pLHW::LIC-YFP-NOSt vector by SLiCE method.The Arabidopsis *TMO5* and *LHW* constructs were adapted as previously described (44). For genomic DNA analysis, small pieces of thalli were collected by forceps and homogenized with liquid nitrogen in the Retsch MM400 machine (Retsch GmbH) and DNA was extracted with 2x CTAB buffer (2% CTAB, 1.4M NaCl, 100 mM Tris-HCl pH 8.0, and 20 mM EDTA). The desired fragments were amplified by primers listed in **Table S6** and were further analyzed by sequencing (sequencing primers are indicated in **Table S6**). Sequence alignment was performed on Benchling (benchling.com) by using genomic sequence obtained from phytozome (37) (*https://phytozome.jgi.doe.gov/*) (*Marchantia polymorpha* v3.1).

Complementation assays

For complementation assays, plasmids carrying the LHW orthologs were transformed into *lhw* plants by simplified floral dipping (49). The rescue was analyzed in T1 as well as T2 plants. Approximately 10 individual seedlings were screened per T2 line. Plasmids carrying the TMO5 orthologs were transformed to *tmo5 t5l1* plants, homozygous for *tmo5*, heterozygous for *t5l1* (obtained by crossing *tmo5* mutant plants with *tmo5 t5l1* double mutant plants). In T1 lines, a rescue was considered when 80% or more of the independent lines showed complementation. Homozygous *tmo5 t5l1* double mutant backgrounds were selected in *Mp*TMO5 and *Kn*TMO5 T2 lines by screening vascular pattern in seedlings that lost the transgene by segregation.

DNA extraction Klebsormidium

Klebsormidium, freshly grown on BCD agar medium, was inoculated in 200ul undiluted Edwards solution (200 mM Tris-HCl pH 7.5, 250 mM NaCl, 25 mM EDTA, and 0.5% SDS) for 1 hour at 90°C. After centrifugation, the supernatant was diluted 20 times to obtain a working stock Klebsormidium DNA.

Bimolecular Fluorescence Complementation (BiFC)
Agrobacterium tumefaciens containing BiFC plasmids were cultured overnight at 28°C/250 rpm in 5ml LB medium containing 50 µg/ml kanamycin, 25 µg/ml rifampicin, 2 µg/ml tetracycline and 200 µM acetosyringone. The bacteria were collected by centrifugation (4000rpm, 10 minutes) and resuspended in MMA infiltration medium (20 g/l sucrose, 5 g/l MS-salts, 2 g/l MES, pH 5.6) containing 200 µM acetosyringone to an optical density (OD_{600}) of 0.3. BiFC samples were mixed in a 1:1 ratio to a total OD_{600} of 0.6. Samples were incubated for 1-2 hours at room temperature (RT) under continuous shaking. The abaxial side of the two youngest, fully expanded leaves of 5 – 6 weeks old *Nicotiana benthamiana* plants were infiltrated with a 1 ml needless syringe. Infiltrated leaves were harvested after 2 – 3 days and imaged with a confocal microscope.

FRET-FLIM
The FRET-FLIM analysis was performed as previously described (44, 52). FRET-FLIM measurements were performed on a Leica SP8 system with an HyD detector. A diode-pulse laser is used to generate the single-photon excitation with 40 Mhz. FRET between sCFP3A and sYFP2 was detected by monitoring donor emission. Images with a frame size of 128 × 128 pixels were acquired, and the average count rate was around 10_4 photons per second for an acquisition time of ±60s. Donor FLIMs (sCFP3A) were analyzed with SPCImage 3.10 software (Becker & Hickl) using a two-component decay model. Several cells (n > 15) were analyzed, and the average FLIMs of different combinations were exported for generating a box plot. Statistical significance of differences between samples was determined using a two-tailed Student's t test.

Plant imaging and Quantification of vascular cell numbers
For confocal imaging of monarch or diarch root architecture, 5 – 7 day-old-seedlings were stained with 10 µg/ml Propidium Iodide. Confocal imaging was done on a Leica SP5 confocal microscope, with a HyD detector. Modified Pseudo Schiff – Propidium Iodine (mPS-PI) staining was performed as described previously (53). For Marchantia gemmae imaging, 4-week-old thalli were fixed and stained with mPS-PI method. The gemma cups were hand-dissected before imaging on a Leica SP8 confocal microscope. Radial cross-sections and quantification of vascular cell numbers were obtained using a SP5 confocal microscope on 5-day-old roots that were stained using mPS-PI. Sections were taken at the middle of the root meristem: half way between the QC and the first elongating cortex cell. The data was visualized using BoxPlotR.

Histological analysis and sectioning
All histological analysis, sectioning and counterstaining using toluidine blue and ruthenium red was performed as described previously (54).

RNA-sequencing
For *Arabidopsis thaliana*, seeds of Col-0, p*RPS5A*::TMO5-GR, p*RPS5A*::LHW-GR and p*RPS5A*::TMO5-GR x p*RPS5A*::LHW-GR (dGR) (44) were surface sterilized and stratified for 2 days in the cold room then were plated on MS plates with nylon meshes (pore size 100 µm). All plates were incubated in a climate chamber at 22 °C with 16h/8h day/night cycle for 7 days and then transferred to MS plates containing 10 µM dexamethasone for 1hr. All seedling roots (about 1.5 cm) of each line were harvested and subjected to RNA isolation. For *Marchantia polymorpha*, gemmalings of Tak-1, p*EF*::*Mp*TMO5-GR, p*EF*::*Mp*LHW-GR and p*EF*::dGR were transferred to 1/2 B5 medium plates with nylon mesh (pore size 100 µm) and incubated in an incubator at 22°C with 16h/8h day/night cycle for 10 days then transferred to 1/2 B5 liquid medium containing 10 µM dexamethasone for 1hr. All gemmalings of each line were harvested and subjected to RNA isolation. Total RNA (at least 500 ng) of each sample with was delivered to BGI (HongKong, China) for f RNA-sequencing. Raw reads were mapped to the respective genomes using HISAT2 (55) with additional parameters "--trim5 10 --dta". Read counts were calculated using feature Counts implemented in the subread package (v1.6.0). The obtained raw read counts were normalized and differentially expressed genes (Padj <0.01) were identified using DEseq2 (56) implemented in R Bioconductor package. To determine the overlap of induced genes by combined overexpression of TMO5 and LHW between Arabidopsis and Marchantia (**Table S4**), we used the PLAZA Dicots 4.0 Workbench (36). First, starting from the *Mp* double-GR genes, Arabidopsis orthologs were

identified using the integrative orthology method requiring at least one evidence, considering Tree-based orthologs, Best Hit Family and Orthologous gene family information, yielding a set of 1322 Arabidopsis genes. Next, this set was compared with the *At* double-GR genes, yielding an overlap of 145 genes (9%, 145/1647).

Edu staining

S-phase cells were visualized using a Click-iT EdU Imaging Kit (Life Technologies), according to the manufacturer's instruction and as previously described with some modifications (57). Gemmae from p*EF*::double-GR or Tak-1 were transferred to liquid half strength B5 medium and cultured for 16 hours under 50-60 μmol photos $m^{-2}s^{-1}$ white LED light at 22°C with continuous light. 10 μM EdU and DEX or DMSO were then added into the liquid medium and incubated the gemmalings for another 2 hours. After DEX treatment, explants were transferred to a tube containing half strength B5 medium with 10 μM 5-ethynyl-2-deoxyuridine (EdU) and incubated for 4 h at the same light conditions. Incorporation of EdU was terminated by fixing the explants with 3.7 % formaldehyde solution in phosphate-buffered saline (PBS) for 20 min. The explants were washed twice with PBS, permeabilized with 0.5 % Triton X-100 in PBS for 20 min, and washed twice with PBS. EdU signals were captured using a Leica SP5-II confocal laser scanning microscope system, with excitation at 561 nm and detection at 565–600 nm.

Upset Plots

Upset plots allow intuitive quantitative analysis of sets, their intersections, and aggregates of intersections. Compared to Venn diagrams, which become challenging to read with more than three sets of data are used, upset plots show the size and properties of aggregates and intersections, and a duality between the visualization of the elements in a dataset and their set membership. For a detailed explanation on the use of upset plots, please see https://caleydo.org/tools/upset/.

RT-PCR

For confirming BiFC experiment, we extracted total RNA from *Nicotiana Benthamiana* leaves 48 hours after infiltration by RNase Mini Kit (QIAGEN) based on the user manual. 1 μg total RNA was used to perform reverse transcription with iScript cDNA synthesis kit (Bio-Red) according to the user manual. PCR was performed with specific primer pairs listed in Table 6.

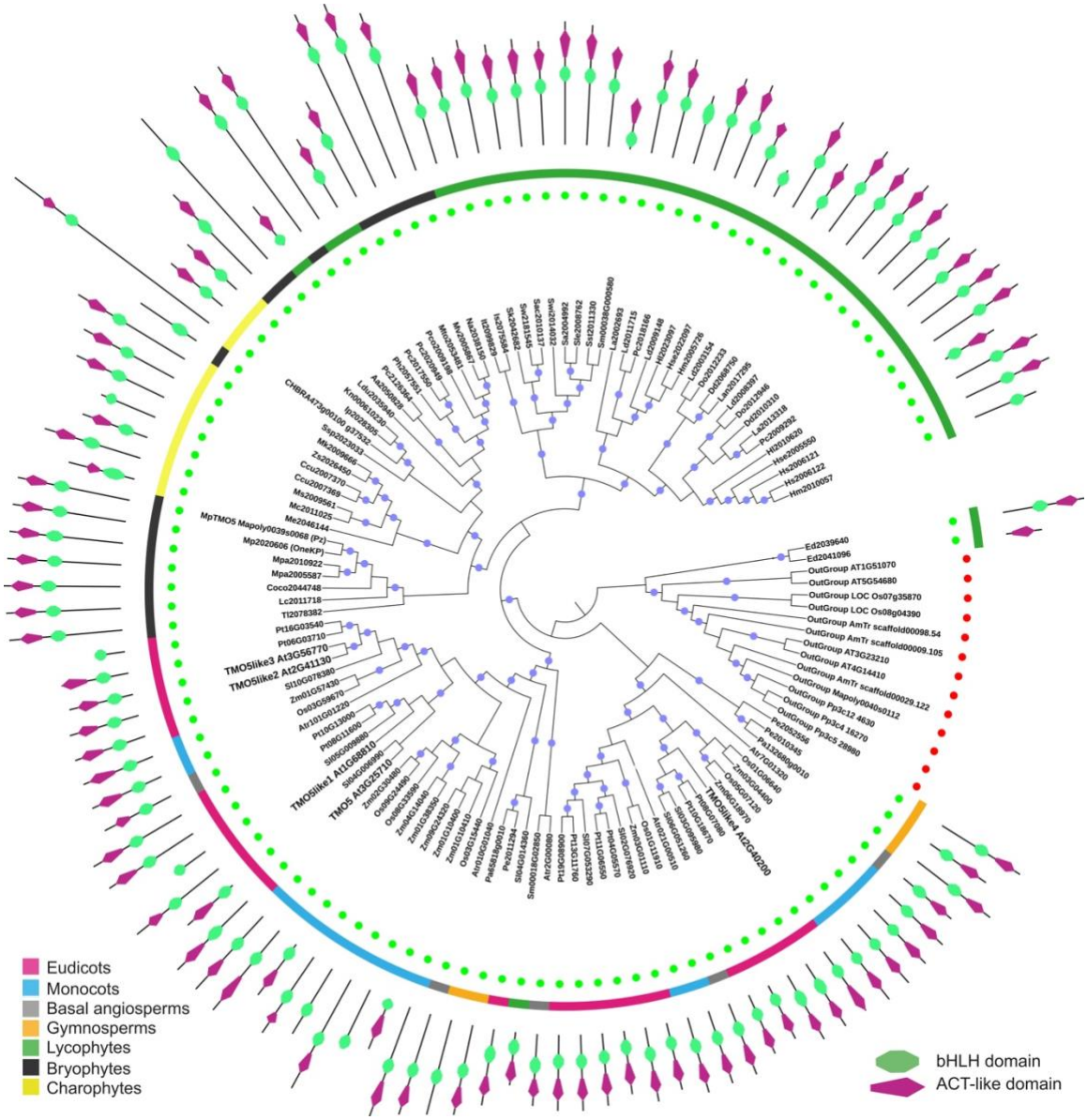

Fig. S1. The phylogenetic tree of TMO5 orthologs in the plant kingdom. The phylogenetic tree was rooted with a bHLH outgroup. The presence of the blue dot indicates bootstrap values over 75. The green and red dot ring shows the reciprocal BLAST analysis, a green dot indicates a successful reciprocal BLAST; a red dot indicates the reciprocal BLAST did not trace back to the original template. The colored ring indicates different clades of species which is referred to the legend. Additionally, the domain architecture is shown for each protein in which the green shape represents the bHLH domain and the magenta shape the ACT-like domain. Note that the *Mp*TMO5 Mapoly0039s0068 from the Phytozome database and Mp2020606 from the OneKP database refer to the same *Mp*TMO5 protein, but for correctness, both entries are retained in this tree. Also note that presence of the ACT-like domain based on predicted protein folding is stable across all clades.

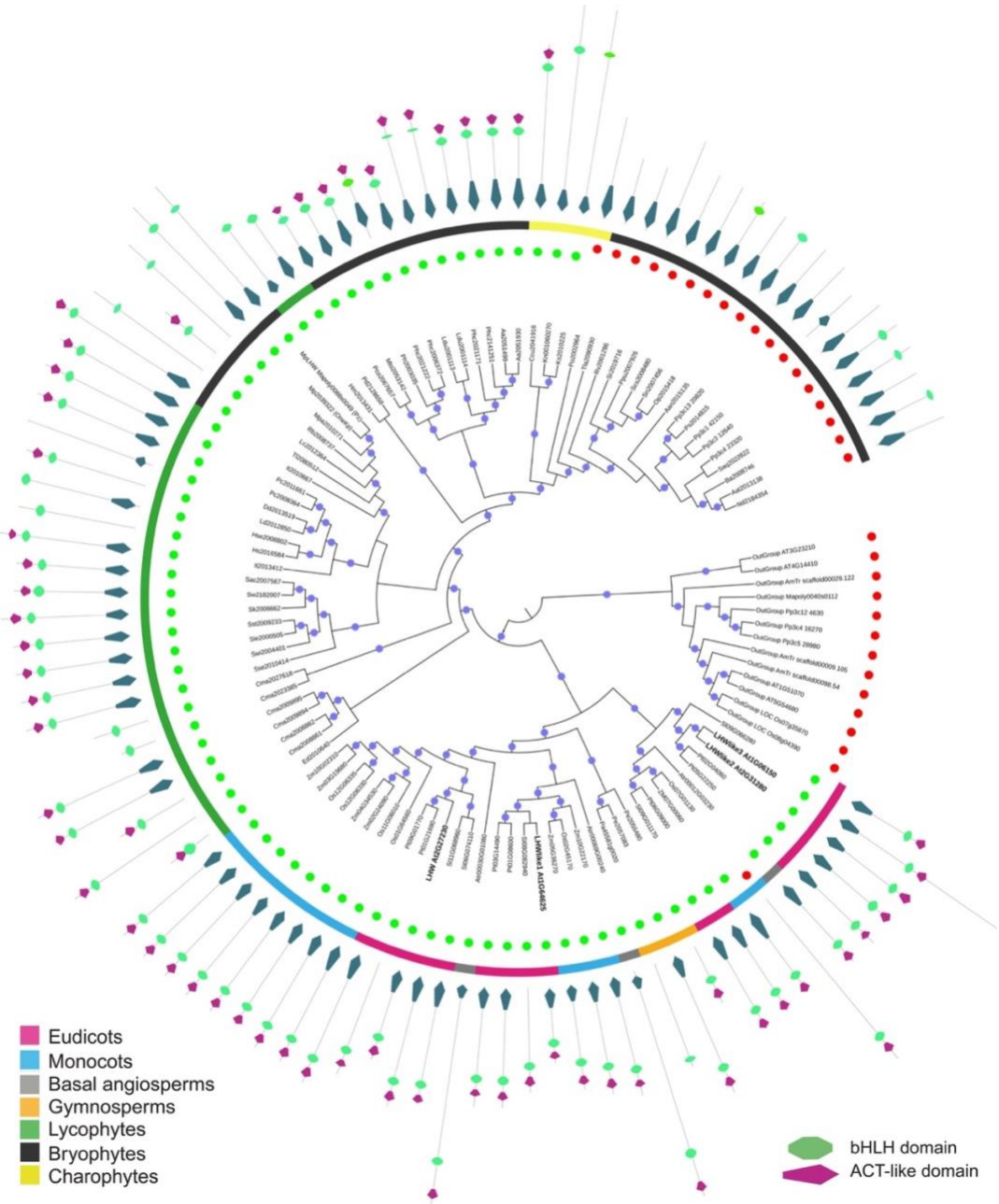

Fig. S2. The phylogenetic tree of LHW orthologs in the plant kingdom. The phylogenetic tree was rooted with a bHLH out group. The presence of the blue dot indicates bootstrap values over 75. The green and red dot ring shows the reciprocal BLAST analysis, a green dot indicates a successful reciprocal BLAST; a red dot indicates the reciprocal BLAST did not trace back to the original template. The colored ring indicates different clades of species which is referred to the legend. Additionally, the domain architecture is shown for each protein in which the green shape represents the bHLH domain; the magenta shape the ACT-like domain and the dark green shape the GAF-like domain. Note that the *Mp*LHW Mapoly0088s0049 from the Phytozome database and Mp2039322 from the OneKP database refer to the same *Mp*LHW protein, but for correctness, both entries are retained in this tree. Also note that presence of the ACT-like domain based on predicted protein folding is variable in the non-tracheophytes and stabilizes in the tracheophytes.

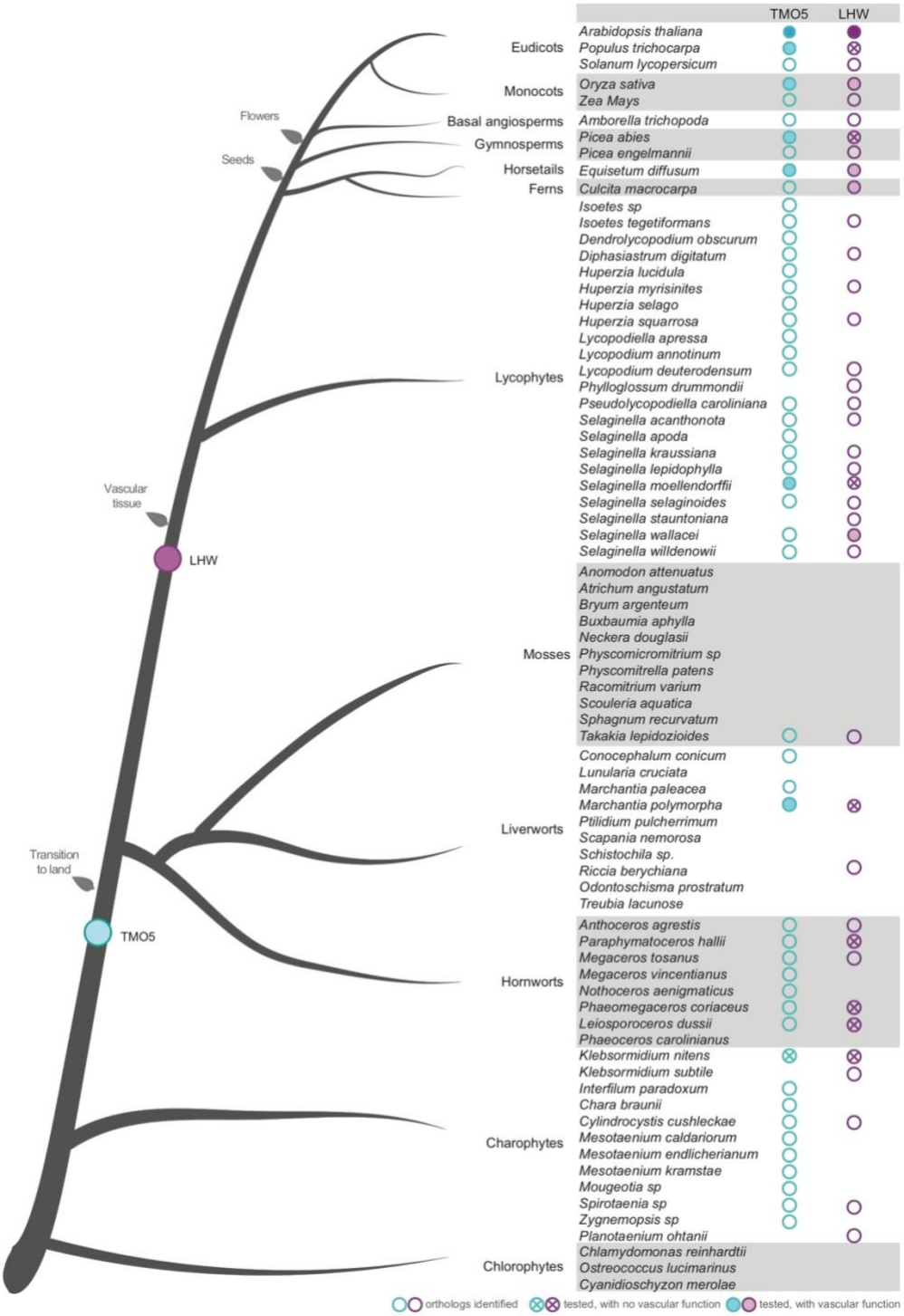

Fig. S3. Summary of the presence and PRD function of TMO5 and LHW orthologs in the plant kingdom. All species we analyzed in this study were listed in the right block. The colored circles indicate orthologous genes were identified. Filled circles indicate one of the identified orthologs was tested and rescued the mutant phenotype in the inter-species complementation assay (see main Figure 2); Crossed circles indicate the tested gene did not rescue the respective Arabidopsis mutant.

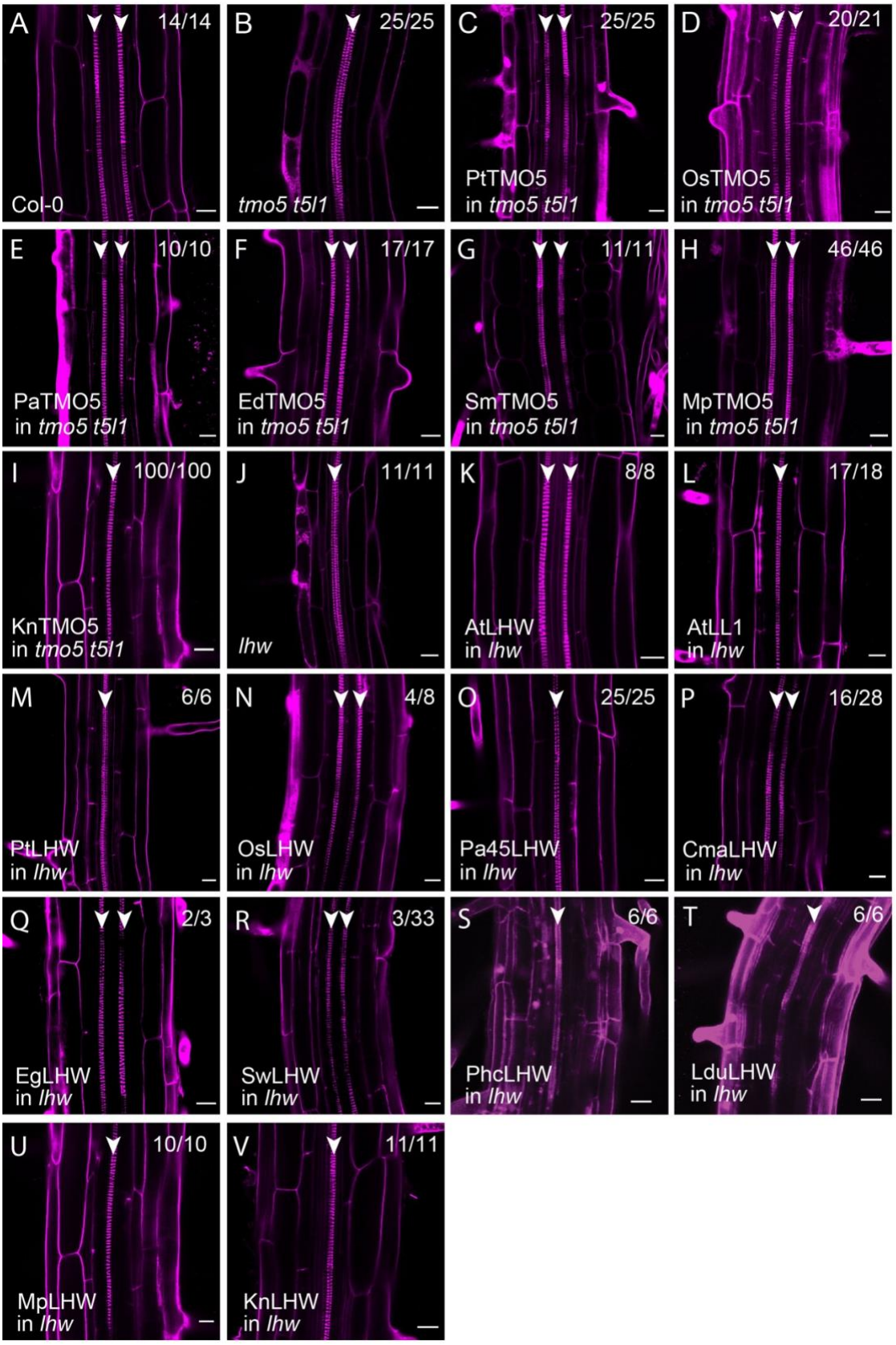
A
14/14
Col-0
B
25/25
tmo5 t5l1
C
25/25
PtTMO5
in tmo5 t5l1
D
20/21
OsTMO5
in tmo5 t5l1
E
10/10
PaTMO5
in tmo5 t5l1
F
17/17
EdTMO5
in tmo5 t5l1
G
11/11
SmTMO5
in tmo5 t5l1
H
46/46
MpTMO5
in tmo5 t5l1
I
100/100
KnTMO5
in tmo5 t5l1
J
11/11
lhw
K
8/8
AtLHW
in lhw
L
17/18
AtLL1
in lhw
M
6/6
PtLHW
in lhw
N
4/8
OsLHW
in lhw
O
25/25
Pa45LHW
in lhw
P
16/28
CmaLHW
in lhw
Q
2/3
EgLHW
in lhw
R
3/33
SwLHW
in lhw
S
6/6
PhcLHW
in lhw
T
6/6
LduLHW
in lhw
U
10/10
MpLHW
in lhw
V
11/11
KnLHW
in lhw

Fig. S4. Complementation of the *t5t5l1* double and *lhw* single mutant vascular phenotypes by their respective orthologs. (A). Wild type Arabidopsis root with two xylem poles. (B). Arabidopsis *t5t5l1* double mutant root showing only one protoxylem pole. (C-I). Complementation of the *t5t5l1* vascular phenotype by introduction of the TMO5 orthologs from *Populus trichocarpa* (*Pt*), *Oryza sativa* (*Os*), *Picea abies* (*Pa*), *Equisetum diffusum* (*Ed*), *Selaginella moellendorffii* (*Sm*), *Marchantia polymorpha* (*Mp*) and *Klebsormidium nitens* (*Kn*). (J). Arabidopsis *lhw* mutant root showing only one protoxylem pole. (K-V). Complementation of the *lhw* vascular phenotype by introduction of the LHW homologs from *Arabidopsis thaliana* (*At*LHW and *At*LL1) and the orthologs from *Populus trichocarpa* (*Pt*), *Oryza sativa* (*Os*), *Picea abies* (*Pa*), *Cultita macrocarpa* (*Cma*), *Equisetum giganteum* (*Eg*), *Selaginella wallace* (*Sw*), *Phaeoceros carolinianus* (*Phc*), *Leiosporoceros dussii* (*Ldu*), *Marchantia polymorpha* (*Mp*) and *Klebsormidium nitens* (*Kn*). Scale bars represent 20 mm; arrowheads indicate protoxylem poles. In all panels except (D), the number indicated in the right top corner how many show the phenotype out of the total number of individual T1 lines analyzed. In (D), we genotyped 3 independent T2 seedlings and analyzed the phenotype in the *t5t5l1* double mutant and the number indicated how many independent T2 seedlings showed the phenotype.

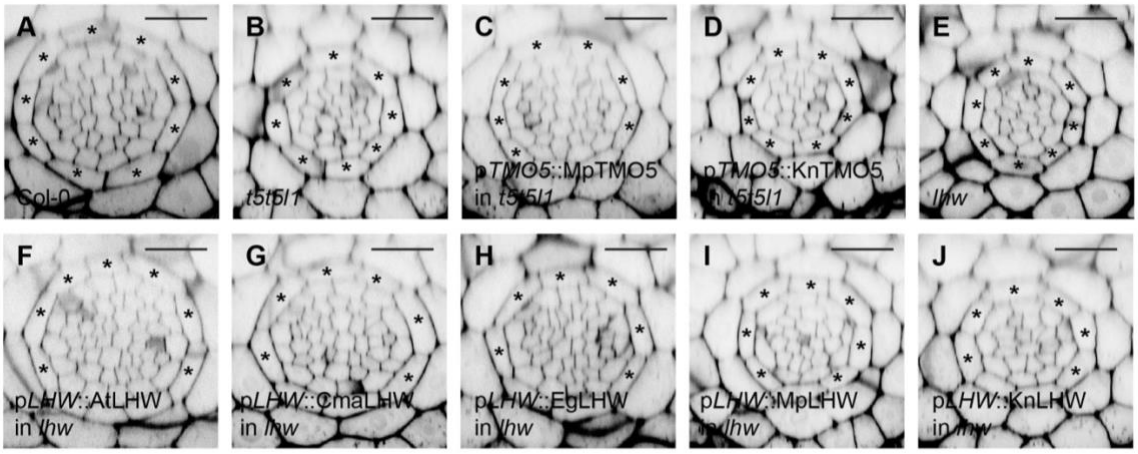

Fig. S5. Representative confocal radial cross-sections of the complementation study. (A). Col-0. (B). The *tmo5 t5l1* double mutant. (C-D). Complementation of the *tmo5 t5l1* double mutant with *Mp*TMO5 (C) and *Kn*TMO5 (D). (E). The *lhw* single mutant. (F-J). Complementation of the *lhw* mutant with *At*LHW (F), *Cma*LHW (G), *Eg*LHW (H), *Mp*LHW (I) and K*n*LHW (J). Scale bars represent 20 mm. Asterisks indicate endodermis cells.

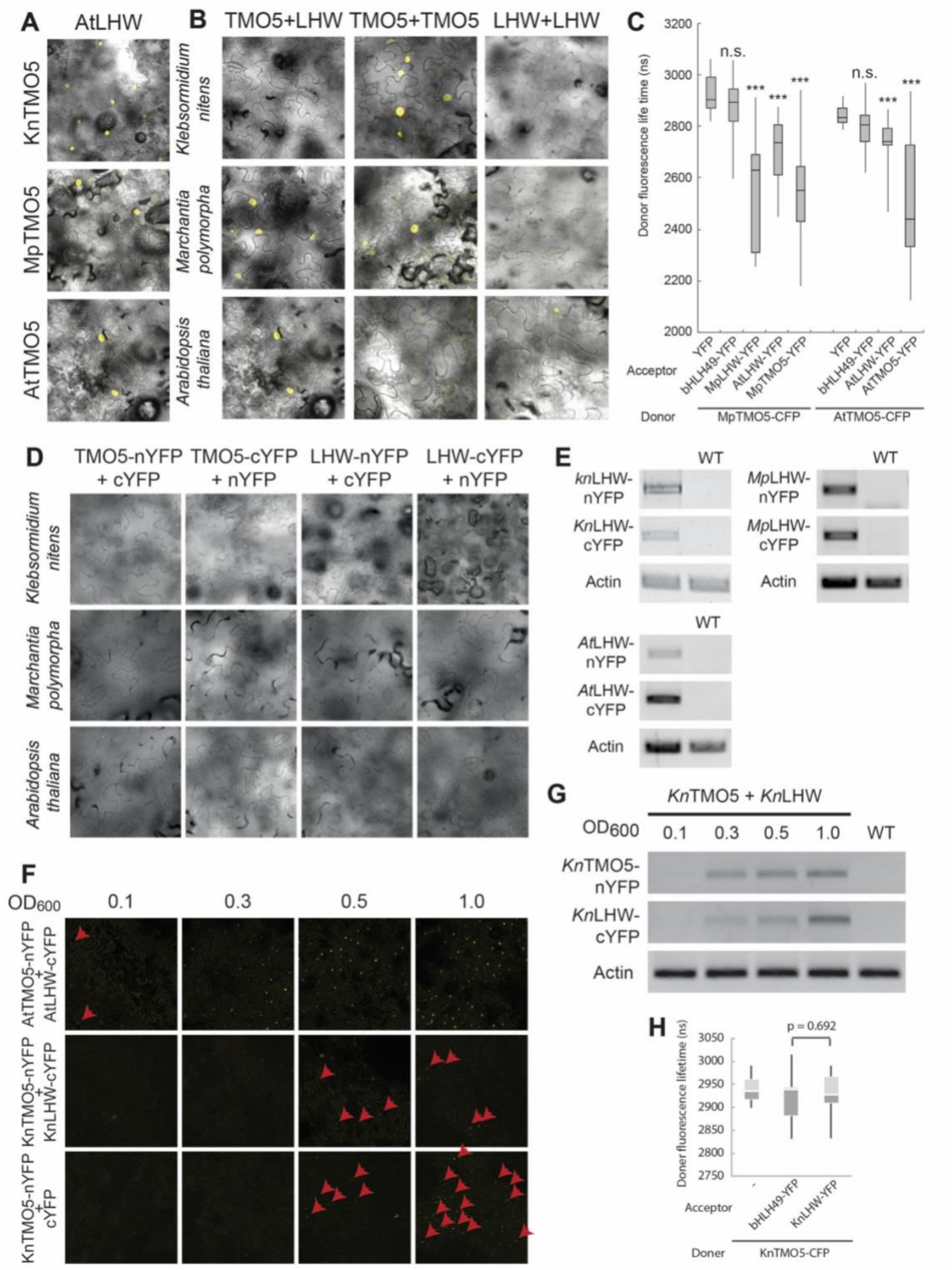

Fig. S6. Homo- and heterodimerization among TMO5 and LHW orthologues. (A) Bimolecular Fluorescence Complementation (BiFC) analyses of heterodimerization of *Klebsormidium nitens* (*Kn*TMO5), *Marchantia polymorpha* (*Mp*TMO5) and *Arabidopsis thaliana* (*At*TMO5) TMO5 orthologues with Arabidopsis LHW (*At*LHW), tested in tobacco leaves. Yellow signals in nuclei

indicate protein interaction. (B) Within-species homo- and heterodimerization of TMO5 and LHW orthologues. In each row, protein pairs from each species indicated on the left were tested in BiFC in all possible combination indicated above each row. (C) Förster Resonance Energy Transfer - Fluorescence Lifetime Imaging (FRET-FLIM) analysis of protein interactions among Marchantia (*Mp*) and Arabidopsis (*At*) TMO5 and LHW orthologues. Data are represented as mean ± SD (n≥15). YFP and the unrelated Arabidopsis bHLH49 were used as negative controls. P-values were calculated by two-tail student t-tests; n.s.: not significant; ***: p-value < 0.01. (D) Negative control of all the constructs used in BiFC. (E) RT-PCR experiments of all the negative combination observed in (B). We detected the expression of all constructs, indicating the negative result was not due to the absence of expression. (F) BiFC of different agrobacterium combinations with different OD_{600} to further confirm the interaction between *Kn*TMO5 and *Kn*LHW. *At*TMO5-nYFP and *At*LHW-cYFP were used as a positive control; *Kn*TMO5-nYFP and cYFP only were used as a negative control. Note that at OD_{600} = 0.3, only *At*TMO5 and *At*LHW showed strong signals. When OD_{600} were over 0.5, both *Kn*TMO5 plus *Kn*LHW and *Kn*TMO5 plus cYFP showed signals. (G) RT-PCR analysis of the indicated constructs as shown in (F). (H) FRET-FLIM analysis of the interaction between KnTMO5 and KnLHW. The Arabidopsis bHLH49-YFP was used as a negative control. Data are represented as mean ± SD (N=15). P-values were calculated by two-tail student t-tests.

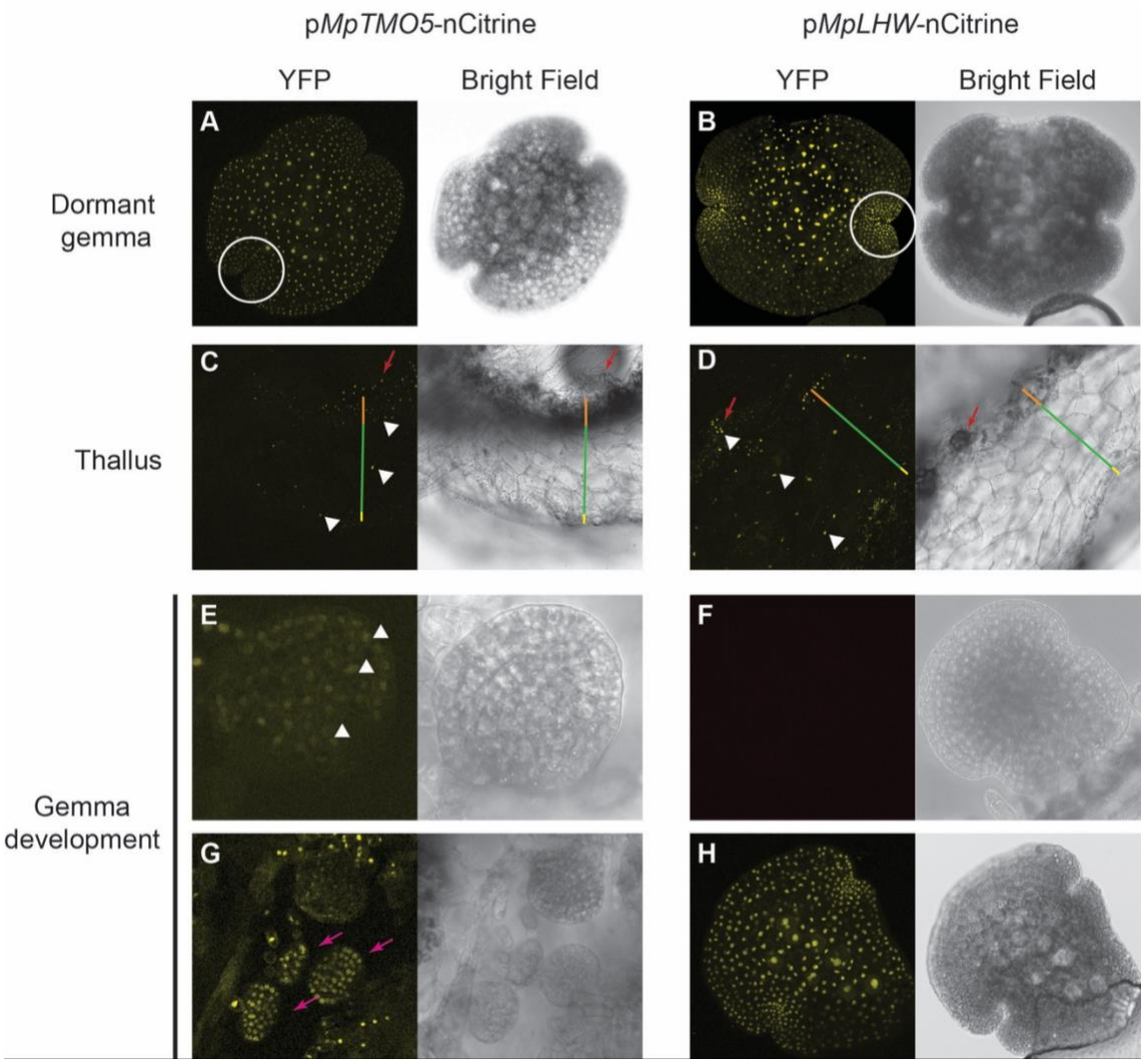

Fig. S7. The expression of p*MpTMO5*-nCitrine (A,C,E,G) and p*Mp*LHW-nCitrine (B,D,F,H) reporter lines at different developmental stages. (A, B) In dormant gemmae, both p*Mp*TMO5 and p*Mp*LHW express ubiquitously. Note that pMpLHW is expressed stronger compare to p*Mp*TMO5 at the apical notch, indicated by the white circle. (C, D) Cross sections of p*Mp*TMO5-nCitrine (C) and p*Mp*LHW-nCitrine (D) mature thalli. In mature thallus, p*Mp*TMO5 expresses mainly in the ventral epidermal cells, including air chamber cells (red arrow) and can only be detected in a few parenchyma cells and dorsal epidermal cells; p*Mp*LHW is expressed ubiquitously in all the cells types in the mature thallus. Orange, green and yellow bars marked ventral epidermal, parenchyma and dorsal epidermal layers, respectively. Red arrows indicated the air chamber. (E-H) p*Mp*TMO5 expression was detected at the early stage of gemmae development (E and G), but not p*Mp*LHW (F); p*Mp*LHW expression was detected at the later stage (H). White arrowheads indicate the expression signals; magenta arrows indicate the early developing gemmae in a gemmae cup.

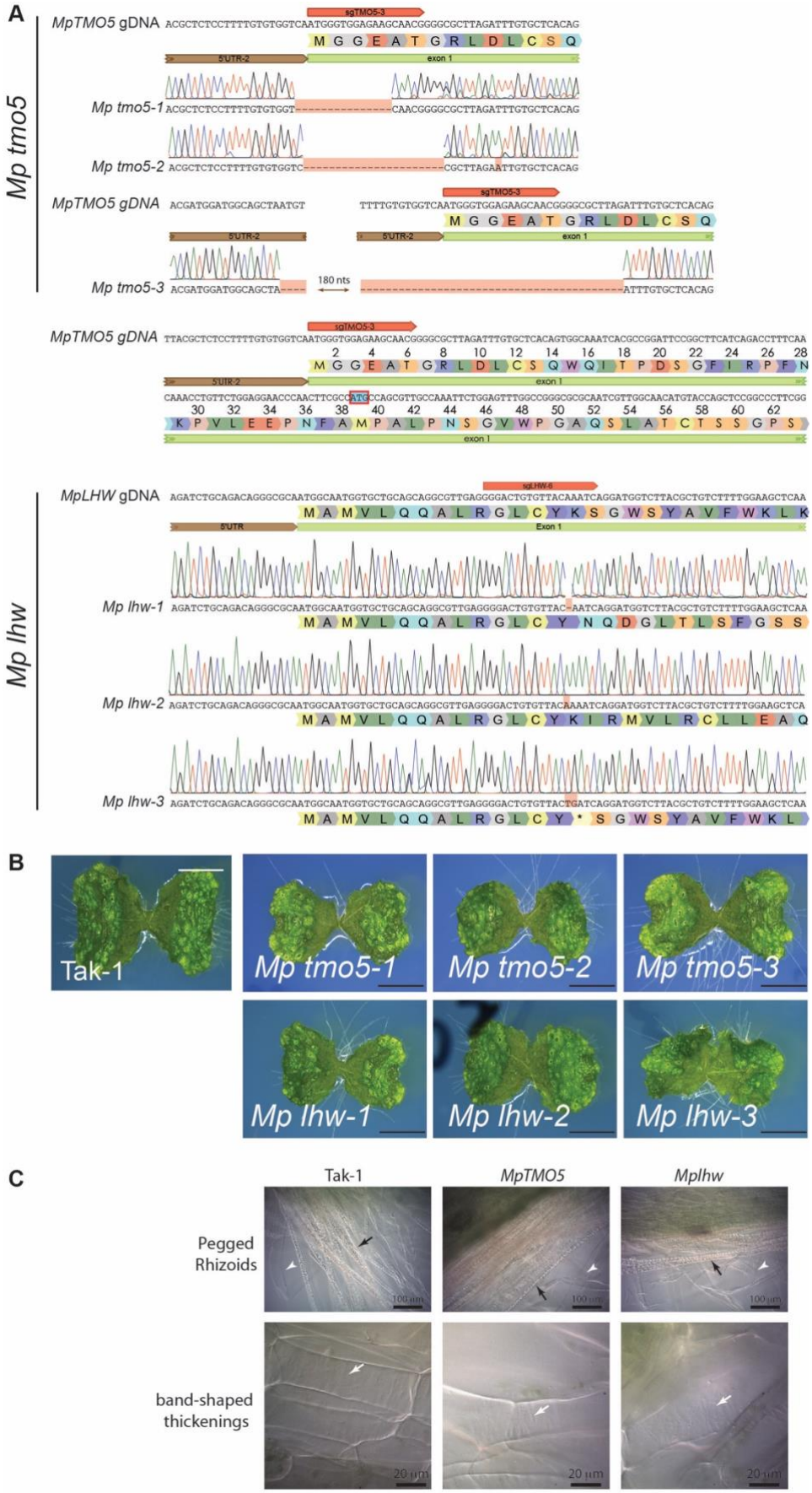
A
Mp tmo5
MpTMO5 gDNA
sgTMO5-3
5'UTR-2
exon 1
Mp tmo5-1
Mp tmo5-2
Mp tmo5-3
180 nts
Mp lhw
MpLHW gDNA
sgLHW-6
5'UTR
Exon 1
Mp lhw-1
Mp lhw-2
Mp lhw-3
B
Tak-1
Mp tmo5-1
Mp tmo5-2
Mp tmo5-3
Mp lhw-1
Mp lhw-2
Mp lhw-3
C
Tak-1
MpTMO5
Mplhw
Pegged Rhizoids
band-shaped thickenings
100 µm
20 µm

Fig. S8. Molecular and phenotypical characterization of multiple independent *Mp tmo5* and *Mp lhw* CRISPR/Cas9 mutant lines. (A). Location of the *Mp tmo5* and *Mp lhw* CRISRP/Cas9 mutations compared to the control genomic DNA sequence. Red tags above the genomic DNA indicate the target region of sgRNAs and mutations are labeled by red colors. The translated amino acid sequences are indicated for each mutant. Note that the *Mp tmo5* mutants all lack the original start codon. The next start codon, highlighted in the sequence below the *Mp tmo5* mutant sequences, generates a truncated protein lacking 38 amino acids at the N-terminal. (B). Resulting young gemmae phenotypes of the CRISPR/Cas9 mutants compared to the control plants (Tak-1). (C). *Mp tmo5* and *Mp lhw* mutants had no defect in pegged rhizoids and the band-shape thickening formation. Black arrows indicate pegged rhizoids, white arrowheads indicate smooth rhizoids, and white arrows indicate the band-shape thickening.

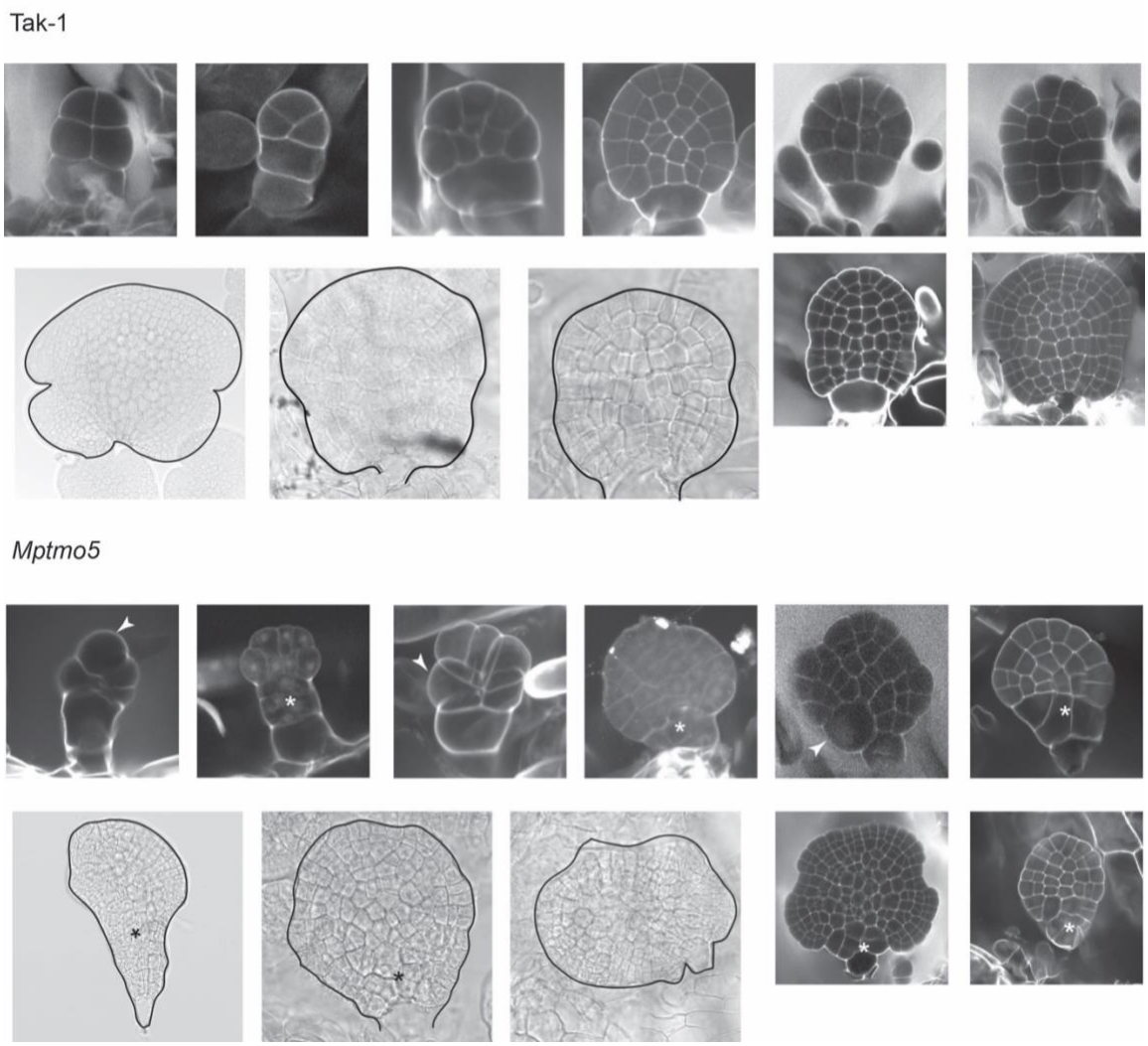

Fig. S9. *Mptmo5* mutants showed occasional division errors. Confocal images of Tak-1 (upper panel) and *Mp tmo5* (lower panel) during early gemmae development. Division defects were occasionally (11 times out of 52 images) found in *Mp tmo5* mutants. No obvious division error was observed in *Mp lhw* mutants. Asterisks indicate abnormal divisions and arrowheads indicate unusual protrusions.

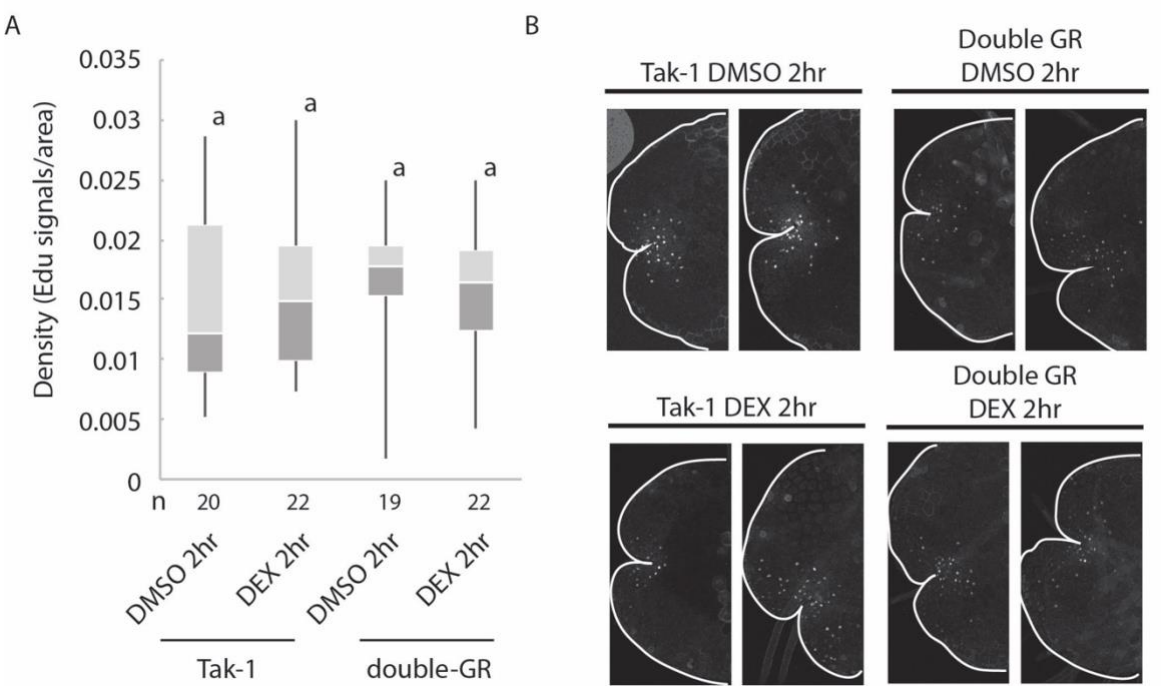

Fig. S10. Both p*EF*::double-GR lines and Tak-1 control lines showed similar division rates after a 10 μm DEX treatment for 2 hours. (A) The Edu density (Edu spot numbers divided by the surface area of half a gemma) of different samples with different treatments. One-way ANOVA analysis indicated no significant difference (p-value = 0.87) between different treatments. (B) The representative confocal images of Edu staining with different treatments.

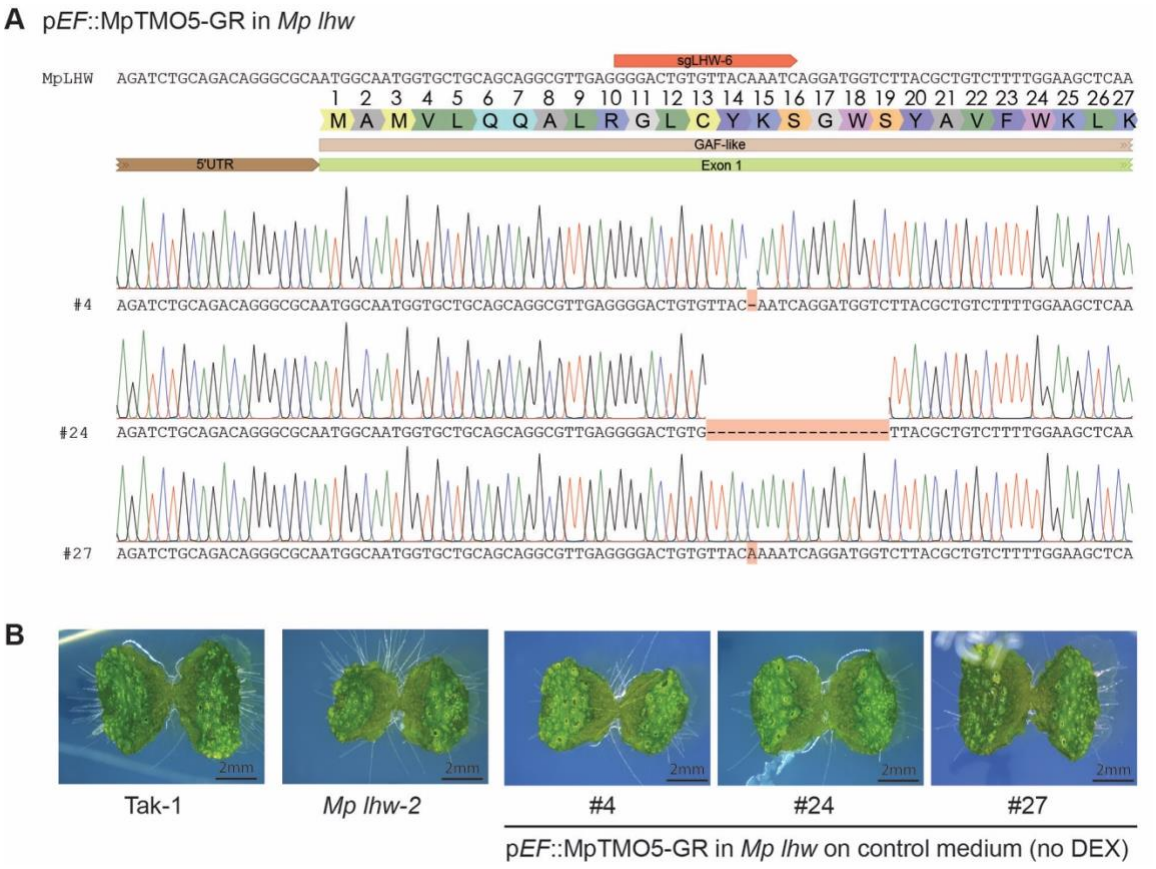

Fig. S11. p*EF*::*Mp*TMO5-GR in *Mp lhw* mutants show similar phenotype to *Mp lhw*. (A) Location of the CRISRP/Cas9 mutations in *Mp*LHW compared to the control genomic DNA sequence. Red tags above the genomic DNA indicate the target region of sgRNAs and mutations are labeled by red colors. The translated amino acid sequences are indicated for each mutant. (B) Gemmaling phenotypes of Tak-1, CRISPR/Cas9 *Mp lhw-2* mutant, and 3 independent *Mp lhw* mutants carrying the p*EF*::*Mp*TMO5-GR construct. Scale bars represent 2 mm.

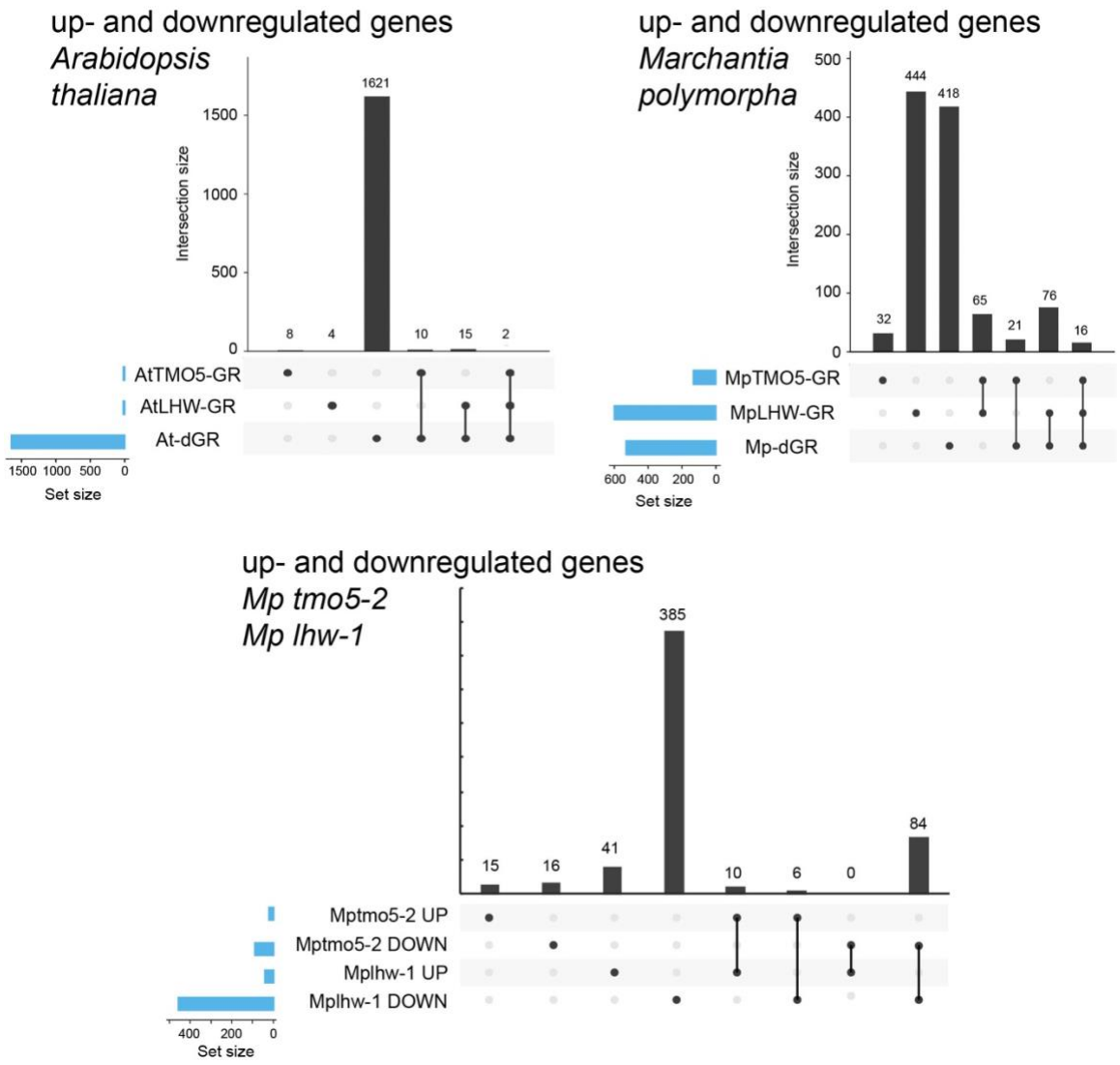

Fig. S12. Independent transcriptional regulation by *Mp*TMO5 and *Mp*LHW. Comparison of genome-wide transcriptional changes (up- and down-regulated genes combined), as measured by RNAseq, upon induction of TMO5-GR, LHW-GR, or both in Arabidopsis (left graph) and Marchantia (right graph), also in *Mp tmo5-2* and *Mp lhw-1* (lower graph). UpSet plots show numbers of differentially expressed genes (DEG) in each DEX-treated transgenic line compared to DEX-treated wild-type (Col-0 in Arabidopsis, Tak-1 in Marchantia). Total DEG numbers in each line are shown in blue bars (set size), while DEG numbers that are either unique to each line (separated block dots), or found in intersections (connected black dots) are shown as black bars (intersection size). See the materials and methods section for more information on the use of upset plots.

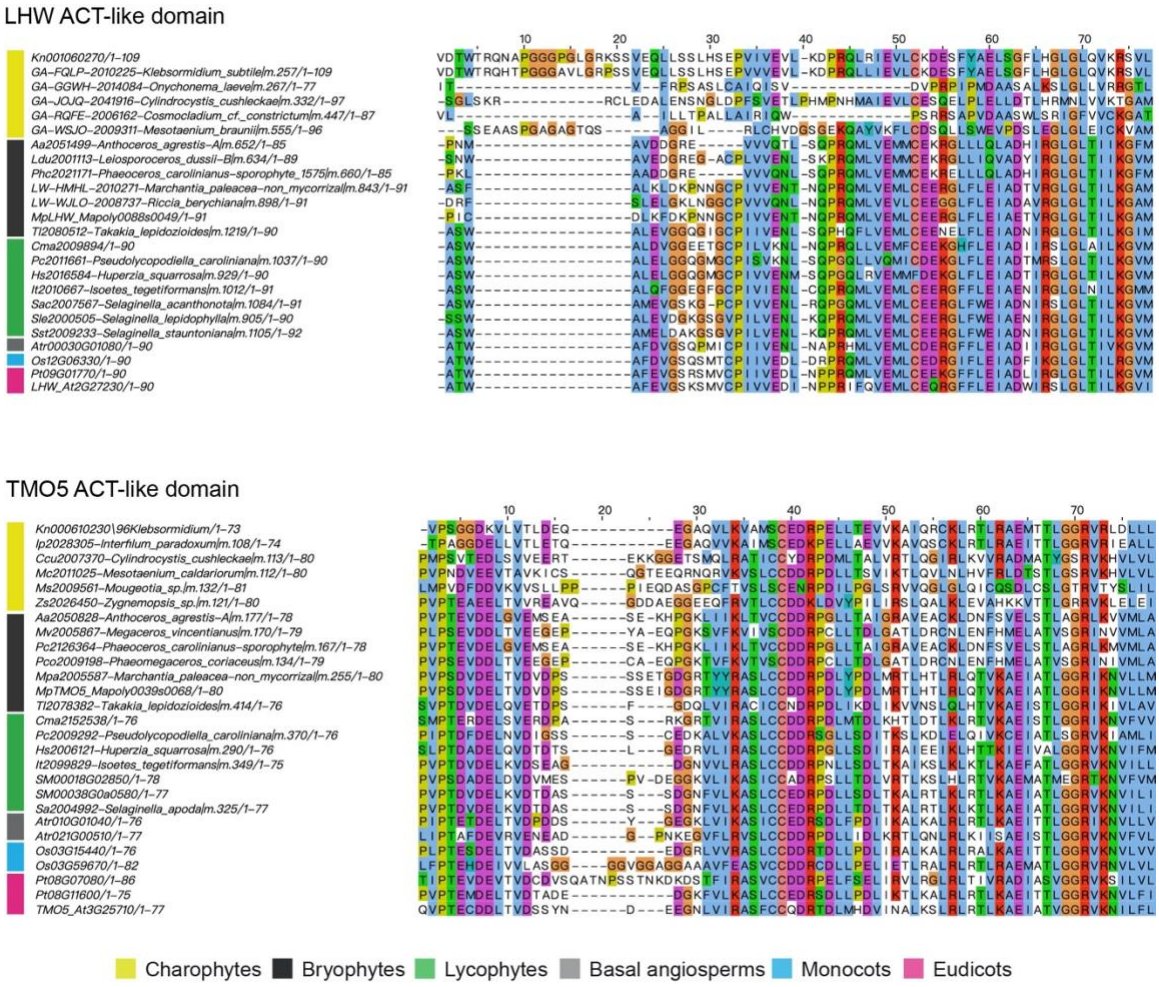

Fig. S13. Multiple sequence alignment (MSA) of the ACT-like domains of LHW and TMO5 orthologs. The color bar on the left indicates different clades which is referred to in the legend.

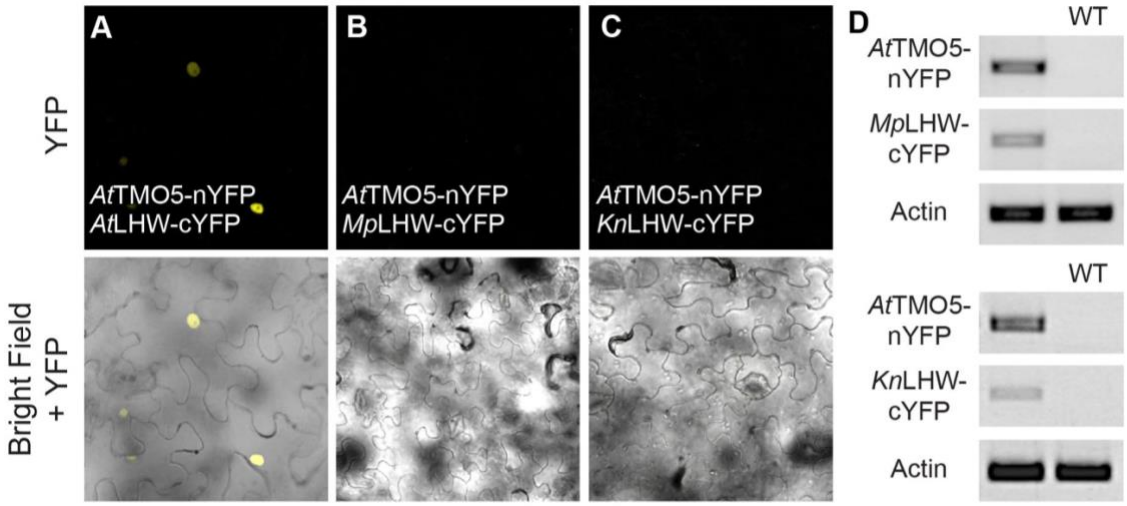

Fig. S14. LHW orthologs do not have conserved dimerization capacities. BiFC results of tested interactions between *At*TMO5 and *At*LHW (A), *Mp*LWH (B) and *Kn*LHW (C). (D) RT-PCR analysis of negative BiFC combinations.

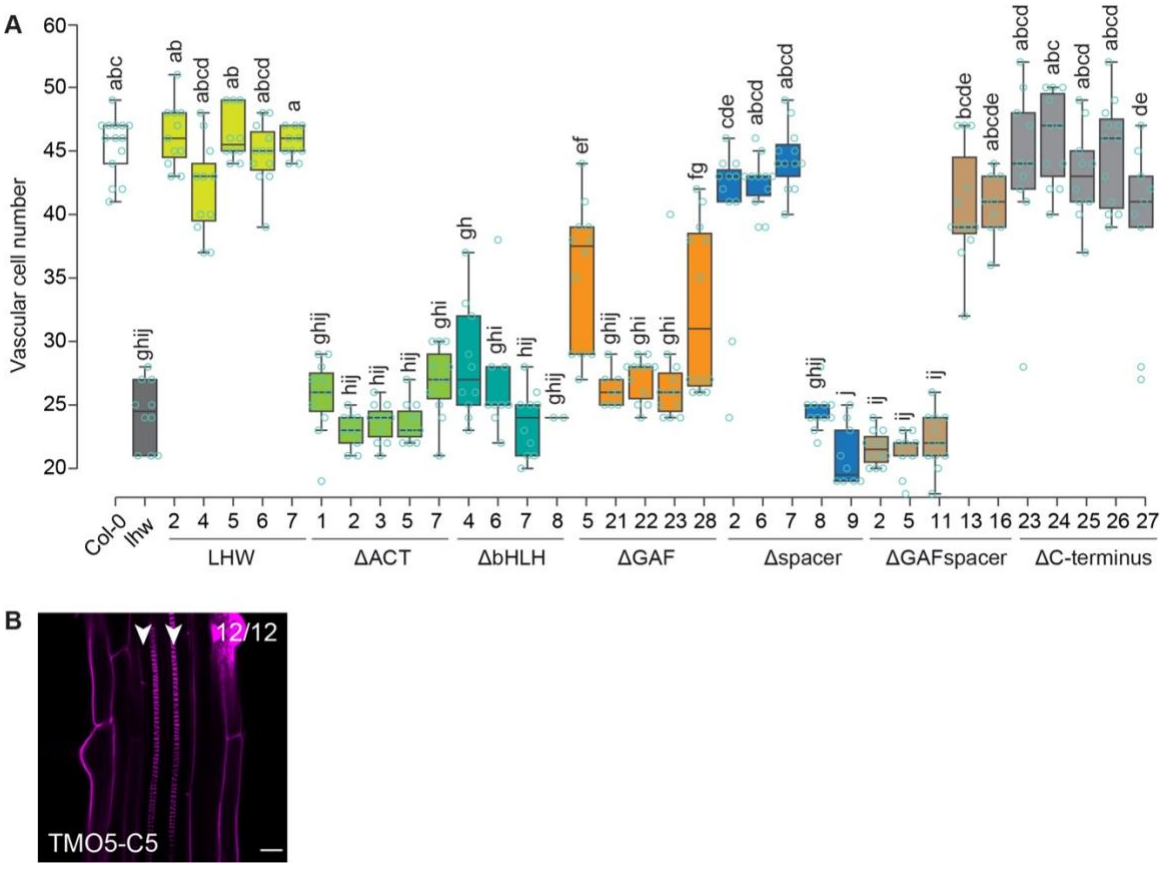

Fig. S15. (A) Quantification of vascular cell file number in radial sections of 5-day-old roots from lines carrying LHW domain deletions in the lhw mutant background. Data are plotted separately for multiple independent transgenic lines. Small case letters indicate significantly different groups as determined using a one-way ANOVA with post-hoc Tukey HSD testing. Blue circles represent all data-points. (B) Domain swaps between *Mp*TMO5 and *Kn*TMO5, and complementation to a diarch vascular pattern phenotype in the Arabidopsis *tmo5 t5l1* mutant using the TMO5-C5 construct. Scale bar represents 20 µm; arrowheads indicate protoxylem poles. The number in the right top corner indicated how many show the phenotype out of the total number of individual T1 lines analyzed.

Table S1: Overview of the complementation experiment showing the percentage of rescue in independent T2 lines of *tmo5 tmo5-like1* and *lhw* mutants in Arabidopsis by introduction of TMO5 and LHW orthologs, respectively.

Table S2: Overview of the complementation experiment showing the percentage of rescue in independent T2 lines of *lhw* mutants in Arabidopsis by introduction of LHW domain deletion constructs (left) and overview of the domain swaps between *Mp*TMO5 and *Kn*TMO5, and their ability to complement (% rescue) the diarch vascular pattern phenotype in the Arabidopsis *tmo5 t5l1* mutant (right).

Table S3: Overview of the species used in the phylogenetic analysis, their abbreviation, clade, order and from which database they were retrieved.

Table S4. Overview of all transcriptomic analyses.

Table S5. DNA alignment of synthetic *MpLHW* CDS and that obtained from Phytozome.

Table S6. Primers used in this research. Different colors indicate the adapter sequences for different cloning methods. Blue, TOPO cloning; Orange, ligation; Green, SLiCE

Dataset S1. FASTA file including all sequences used for the TMO5-related alignments

Dataset S2. FASTA file including all sequences used for the LHW-related alignments

References

33. S. Proost *et al.*, PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* **21**, 3718-3731 (2009).
34. S. Proost *et al.*, PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res* **43**, D974-981 (2015).
35. M. Van Bel *et al.*, Dissecting plant genomes with the PLAZA comparative genomics platform. *Plant Physiol* **158**, 590-600 (2012).
36. M. Van Bel *et al.*, PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic acids research* **46**, D1190-D1196 (2018).
37. D. M. Goodstein *et al.*, Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* **40**, D1178-1186 (2012).
38. K. Hori *et al.*, Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. *Nat Commun* **5**, 3978 (2014).
39. N. Matasci *et al.*, Data access for the 1,000 Plants (1KP) project. *Gigascience* **3**, 17 (2014).
40. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066 (2002).
41. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).
42. R. Lanfear, P. B. Frandsen, A. M. Wright, T. Senfeld, B. Calcott, PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol Biol Evol* **34**, 772-773 (2017).
43. L. T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268-274 (2015).

44. B. De Rybel *et al.*, A bHLH complex controls embryonic vascular tissue establishment and indeterminate growth in Arabidopsis. *Developmental cell* **24**, 426-437 (2013).
45. B. De Rybel *et al.*, Plant development. Integration of growth and patterning during vascular tissue formation in Arabidopsis. *Science* **345**, 1255215 (2014).
46. A. Kubota, K. Ishizaki, M. Hosaka, T. Kohchi, Efficient Agrobacterium-mediated transformation of the liverwort Marchantia polymorpha using regenerating thalli. *Biosci Biotechnol Biochem* **77**, 167-172 (2013).
47. C. R. Barnes, W. Land, Bryological papers. II. The origin of the cupule of Marchantia. *Botanical Gazette* **46**, 401-409 (1908).
48. J. R. Wendrich, C. Y. Liao, W. A. van den Berg, B. De Rybel, D. Weijers, Ligation-independent cloning for plant research. *Methods Mol Biol* **1284**, 421-431 (2015).
49. S. J. Clough, A. F. Bent, Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant Journal* **16**, 735-743 (1998).
50. K. Ishizaki *et al.*, Development of Gateway Binary Vector Series with Four Different Selection Markers for the Liverwort Marchantia polymorpha. *PLoS One* **10**, e0138876 (2015).
51. Y. Zhang, U. Werling, W. Edelmann, Seamless Ligation Cloning Extract (SLiCE) cloning method. *Methods Mol Biol* **1116**, 235-244 (2014).
52. A. F. Rios, T. Radoeva, B. De Rybel, D. Weijers, J. W. Borst, FRET-FLIM for Visualizing and Quantifying Protein Interactions in Live Plant Cells. *Methods in molecular biology (Clifton, N.J.)* **1497**, 135-146 (2017).
53. E. Truernit *et al.*, High-resolution whole-mount imaging of three-dimensional tissue organization and gene expression enables the study of Phloem development and structure in Arabidopsis. *The Plant Cell* **20**, 1494-1503 (2008).
54. I. De Smet *et al.*, An easy and versatile embedding method for transverse sections. *J Microsc* **213**, 76-80 (2004).
55. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360 (2015).
56. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
57. R. Nishihama *et al.*, Phytochrome-mediated regulation of cell division and growth during regeneration and sporeling development in the liverwort Marchantia polymorpha. *J Plant Res* **128**, 407-421 (2015).