



A root phloem pole cell atlas reveals common transcriptional states in protophloem-adjacent cells

Sofia Otero¹, Iris Gildea^{2,5}, Paweł Roszak¹, Yipeng Lu¹, Valerio Di Vittori^{1,3,6}, Matthieu Bourdon¹, Lothar Kalmbach¹, Bernhard Blob¹, Jung-ok Heo¹, Federico Peruzzo^{1,4}, Thomas Laux^{1,4}, Alisdair R. Fernie^{1,3}, Hugo Tavares^{1,7}✉ and Yka Helariutta^{1,2}✉

Single-cell sequencing has recently allowed the generation of exhaustive root cell atlases. However, some cell types are elusive and remain underrepresented. Here we use a second-generation single-cell approach, where we zoom in on the root transcriptome sorting with specific markers to profile the phloem poles at an unprecedented resolution. Our data highlight the similarities among the developmental trajectories and gene regulatory networks common to protophloem sieve element (PSE)-adjacent lineages in relation to PSE enucleation, a key event in phloem biology. As a signature for early PSE-adjacent lineages, we have identified a set of DNA-binding with one finger (DOF) transcription factors, the PINEAPPLES (PAPL), that act downstream of PHLOEM EARLY DOF (PEAR) genes and are important to guarantee a proper root nutrition in the transition to autotrophy. Our data provide a holistic view of the phloem poles that act as a functional unit in root development.

In plants, organs originate from meristems postembryonically and are patterned by mobile signals and the positional information generated in the individual immobile cell types. Determining cell type-specific transcriptional programmes is key to understanding the positional cues guiding plant development¹. However, despite the importance of phloem and radial growth pre-patterning in vascular plants², phloem gene expression is not yet well characterized.

During root development, the term phloem is oftentimes used as a synonym for the protophloem sieve element (PSE), the cell type that undergoes a unique differentiation process to specialize in the transport of sap from source photosynthetic organs to distant sink tissues. This simplification is probably the result of the extensive knowledge we have about PSE specification^{2,3} and differentiation^{4–11}. However, in the *Arabidopsis* primary root, the phloem pole is composed of six cells belonging to four distinct cell types: the central PSE is flanked by two phloem pole pericycle (PPP) cells to the outside and one metaphloem sieve element (MSE) cell to the inside, and both SE cells are in direct contact with the two lateral companion cells (CC)¹² (Fig. 1a).

In the *Arabidopsis* root, both conductive elements (MSE and PSE) derive from the same stem cell¹³ but MSE differentiates later, when PSE cells are no longer functional. Despite having a similar function to PSE, MSE ontogeny is less well characterized¹⁴ and few factors have been directly related to MSE development. Exceptions are the partially redundant homologues *OCTOPUS* (*OPS*, *At3g09070*) and *OCTOPUS-LIKE 2* (*OPL2*, *At2g38070*) identified as important for MSE entry into differentiation¹⁵. Despite some commonalities between PSE and MSE, a recent study highlighted that MSE differentiation is independent of adjacent or preceding PSE¹⁴, underlining

the peculiarities of this cell type. The conducting cell types and CCs originate from different progenitors in the *Arabidopsis* root¹³. CCs are believed to be essential to support enucleated PSE function¹⁶ and their intimate relationship has been evidenced by a common molecular switch controlling SE/CC fate in vitro and in hypocotyls¹⁷, while in the primary root undifferentiated CC and MSE can transdifferentiate to PSE cells if these are misspecified¹⁸. The CC function in leaves consists of loading nutrients into the SE, but their role in the root remains elusive¹⁹. Traditionally, it was thought that they were involved in phloem unloading²⁰, that is, the exit of the nutrients from the sieve element pipe so that they reach meristematic cells for food. However, it was recently demonstrated that this process happens through funnel plasmodesmata connecting PSE to PPP¹².

Despite being considered a non-vascular tissue, PPP and the associated vasculature share a high overlap in gene expression²¹ and are different in size and ultrastructure from the xylem pole pericycle (XPP) population²², exhibiting specific gene expression²³ from early stages, mirroring the diarch pattern in the *Arabidopsis* vasculature²⁴.

In the last 15 years, transcriptomics has been the stepping stone to learn about plant organogenesis. However, even if markers for mature CC and PPP were used for transcriptomics^{1,25,26}, the lack of specific markers for early phloem, combined with the difficulties in accessing phloem cells that are deeply embedded in the root cylinder, has hampered the study of these populations, oftentimes masked under the concept ‘stele’ that groups together the pericycle and vasculature^{27–29}. The more recent root single-cell atlases confer a detailed root panorama but even here, phloem cells remain under-represented compared with more accessible root layers^{30–32}.

¹The Sainsbury Laboratory, University of Cambridge, Cambridge, UK. ²Institute of Biotechnology, HiLIFE/Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland. ³Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm, Germany. ⁴Signalling Research Centres BIOSS and CIBSS, Freiburg, Germany. ⁵Present address: Department of Plant Biology and Genome Center, University of California, Davis, CA, USA. ⁶Present address: Dipartimento di Scienze Agrarie, Alimentari ed Ambientali, Università Politecnica delle Marche, Ancona, Italy. ⁷Present address: Department of Genetics, University of Cambridge, Cambridge, UK.
✉e-mail: hm533@cam.ac.uk; yrjo.helariutta@slcu.cam.ac.uk

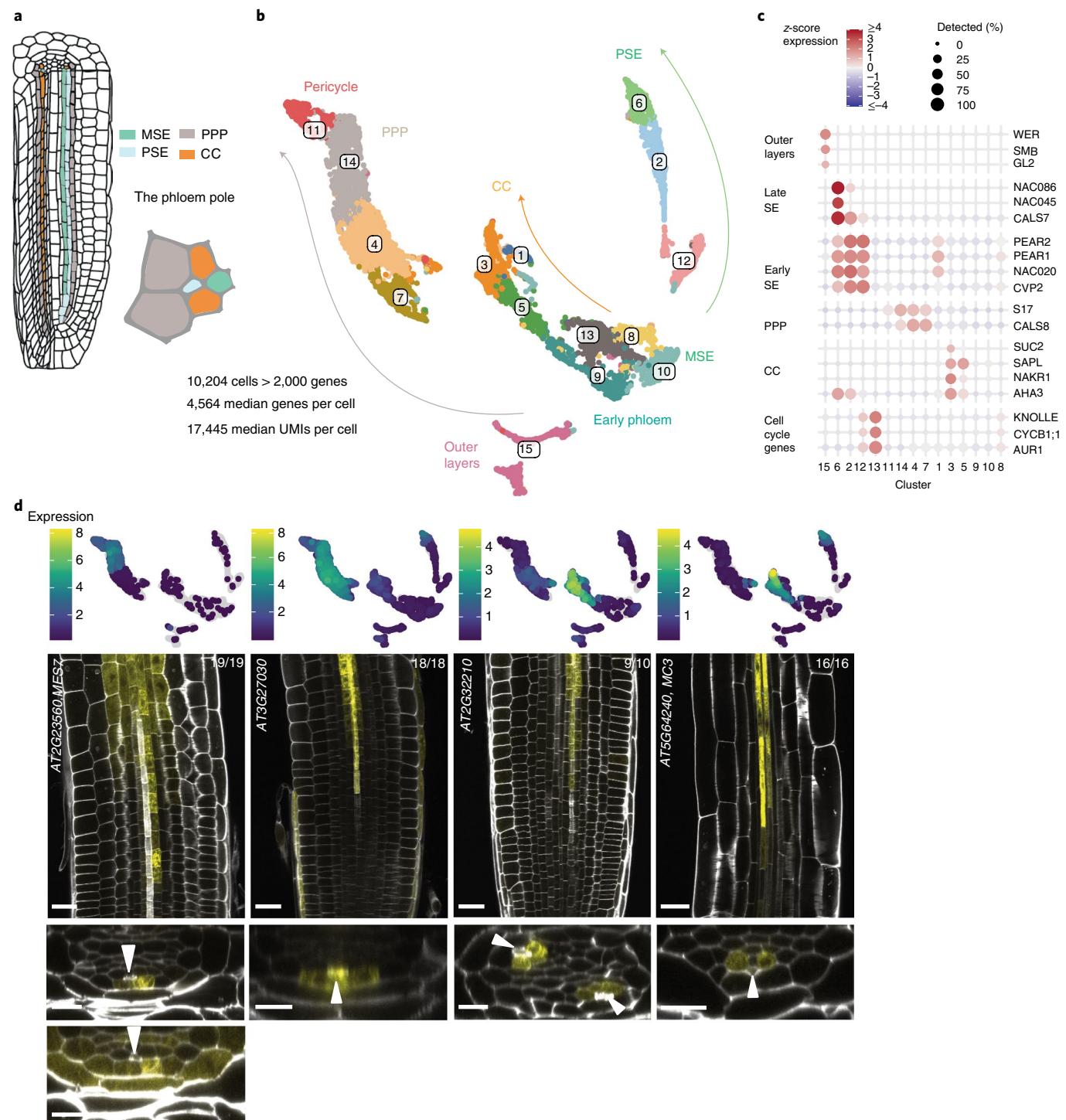


Fig. 1 | A root phloem pole cell atlas containing PSE, MSE, CC and PPP cells. **a**, Root schematic highlighting the cells in the phloem pole coloured by identity with a close-up of the phloem pole (**a'**). One of the radial cuts shows the middle part of the phloem pole (right) and the other shows the side view of the phloem pole (left). **b**, UMAP plot showing the classification of 10,204 cells clustered by cell identity and developmental stage (colours indicate clusters, labelled with a number). The sample of cells has a median of 4,564 detected genes (10%–90% percentiles: 2,600–6,780) and a median of 17,445 total UMIs per cell (10%–90% percentiles: 5,941–52,689). **c**, Cluster annotation based on markers with known tissue- or cell-specific expression. The size of the points represents the percentage of cells in a cluster where the gene was detected (that is, at least 1 UMI). The colour shows the scaled average expression of the gene (z-score, that is, number of standard deviations above/below the gene's mean across all cells). **d**, Newly identified genes significantly enriched in PPP (*At2g23560*, *At3g27030*) and CC (*At2g32210*, *At5g64240*). UMAPs show the particular cluster-weighted normalized expression of each gene in the phloem pole cell atlas. UMAPs and microscopy images are representative of the transcriptional reporter lines, where the gene promoter is fused to VENUSer. Scale bar, 25 µm in the longitudinal sections, 10 µm in the cross sections. White arrowheads point to PSE cells as a reference point. The numbers in each panel indicate samples with similar results, of the total independent biological samples observed.

Combining fluorescent activated cell sorting (FACS) and SMART-seq single-cell technologies allowed the profiling of 758 PSE cells at an unprecedented resolution, identifying the bifurcation of MSE and procambium lineages³³. In this study, we have generated a phloem pole cell atlas of 10,204 cells by sorting phloem marker lines combined with single-cell sequencing. This allowed us to gain resolution not just in the PSE lineage but also in all the surrounding cells (CC, PPP and MSE) in the phloem poles, all of which are underrepresented in general root cell atlases. We investigated not only the specificities of each cell type but also the transcriptional commonalities between them. We additionally identified a second set of DOF transcription factors (TF) expressed in the PSE-adjacent cells, downstream of PEAR TF, that are important in the transition to autotrophy in young seedlings.

Results

To profile phloem cells, we took advantage of new and existing fluorescent markers expressed in SEs, CC and PPP from early meristematic cells until differentiation (Extended Data Fig. 1a). This allowed us to enrich our data with cells of interest, by using FACS and preparing single-cell sequencing libraries using the 10x Chromium droplet-based protocol. This resulted in a total of 10,204 high-quality cells, defined as those having at least 2,000 detected genes and no more than 10% of reads assigned to mitochondrial genes (the resultant sample of cells had a median of 17,455 reads per cell and a median of 4,564 genes per cell). The raw count data were normalized using variance-stabilizing transformation³⁴ and integrated across batches using the mutual nearest neighbours algorithm³⁵, although our main conclusions are robust to normalization and batch effects. These cells were grouped into 15 clusters using the Louvain algorithm on a shared-nearest-neighbour cell graph and visualized using uniform manifold approximation and projection (UMAP)³⁶ (Fig. 1b). Using signature marker genes^{1,2,5,12,17,19,37} (Fig. 1c), we identified all the cell types included in the phloem pole. We manually annotated groups of clusters as: PSE conducting cells (clusters 12, 2, 6), CC (clusters 5, 3) and a third as PPP (clusters 7, 4, 14, 11), all emerging from a central group of less mature cells (clusters 8, 9, 10, 13). Clusters 10 and 1 express MSE genes (Fig. 2b). Clusters 13 and 12 contain G2/M cell-cycle markers, indicating cells undergoing division. While it is usually difficult to infer the identity of cycling cells, in the case of cluster 12 most of the cells express early PSE markers as well as cell division markers, pointing towards PSE-dividing cells. For example, *PEAR1* or *CVP2* are detected in all cells of this cluster and cell-cycle genes such as *KNOLLE*, *AUR1* or *CYCB1* are also detected in over 57% of these cells. Finally, cluster 15 corresponds to the outer layers of the root, as an apparent contamination during cell sorting.

Separated from the rest, clusters 7, 4 and 14 were mainly composed of *pS17::GFP* and *pAPL::3xYFP* markers (Fig. 2a), and expressed genes characteristic of PPP such as *S17* (*At1g022850*) and *GLUCAN SYNTHASE-LIKE 4* (*CALS8*, *At3g14570*) (Fig. 1c). Cluster 11, mainly composed of *pS17::GFP* and the *pMAKR5::MAKR5-3xYFP* sortings, represents mature pericycle cells, since in addition to PPP markers it also expresses markers for XPP (*At1g02460*, *At4g30450*³⁸, *At2g36120*; Fig. 2c) and PPP (Fig. 1c). This is probably because *MEMBRANE-ASSOCIATED KINASE REGULATOR 5* (*MAKR5*, *At5g52870*) is expressed in the whole pericycle layer high up in the root and pericycle cells come together with PPP cells for similarity.

Considering genes that were statistically more highly expressed in PPP-specific clusters, we built reporter lines for two genes, which were confirmed to have PPP-specific expression. One of these, *At3g27030*, was expressed in PPP and late PSE, while the other, *METHYL ESTERASE 7* (*MES7*, *At2g23560*), was expressed in early PPP and soon afterwards becomes more broadly expressed in the vasculature and endodermis (Fig. 1d).

The known CC genes are expressed in cluster 5 (*SISTER OF APL*, (*SAPL*, *At3g12730*³²)), with cluster 3 expressing mature CC genes (*ATPase3* (*AHA3*, *At5g57350*²), *SODIUM POTASSIUM ROOT DEFECTIVE 1* (*NAKRI*)³⁹, *SUCROSE PROTON SYMPORTER 2* (*SUC2*, *At1g22710*⁴⁰)). *AHA3* in particular was statistically more highly expressed in this cluster and allowed the discovery of new CC genes by correlation, which were validated by building reporter lines (Fig. 1d). One of these was *At2g32210*, which is expressed first in PSE and then switches to a strong CC-MSE expression, with a weak expression in the epidermis. *METACASPASE 3* (*MC3*, *At5g64240*) was expressed in late PSE and started being expressed in CC after enucleation, first in a patchy way and then becoming continuous and mostly CC-exclusive. Cloning reporter lines for other genes expressed in these clusters, we found a gene expressed in PSE and CC (*PHOSPHATIDYLINOSITOL-SPECIWC PHOSPHOLIPASE C5* (*PLC5*, *At5g58690*)), previously described to be expressed in vascular tissues³⁷ and *At2g38640*⁴¹ mostly specific to mature CC (Extended Data Fig. 1b). Therefore, we have been able to validate our cell annotation (shown on the UMAP in Fig. 1b) *in vivo* by using new genes highly expressed in these clusters.

Spatiotemporal patterns of differentiation in the atlas

From our initial cell annotation, it seemed clear that our data also captured the temporal aspect of cell differentiation in the phloem. For example, marker genes usually expressed in more differentiated cells showed higher expression at the terminal clusters of our UMAP projection (3, 6, 11 and 14; Extended Data Fig. 1c), while those closer to the cycling cells are less mature. To validate this hypothesis, we compared our data with a microarray dataset¹ of manually microdissected root longitudinal sections (3 to 5 cells thick), assigning each of our cells to the longitudinal section with which they had the highest Spearman correlation (Extended Data Fig. 2a). Using this strategy, we observed that the cells towards the centre of our UMAP matched with the meristematic sections of Brady et al.¹, with a temporal progression towards the terminal clusters of our UMAP, until the more mature cells cap each trajectory. This analysis validates our hypothesis of a temporal trajectory that is well captured by our UMAP projection and cell clustering.

To further infer developmental trajectories and order our cells along a continuous pseudotime, we used Slingshot⁴² (Fig. 3a). Setting a unique origin for all cells in cluster 13 (cycling cells), we obtained 5 different trajectories (Fig. 3a), reflecting the known developmental trajectories in the root. Furthermore, these trajectories agreed with RNA velocity analysis using scVelo⁴³, with velocity vectors aligning towards the end of these trajectories (Extended Data Fig. 2c).

Trajectories 1–3 account for PPP, CC and PSE, respectively. Trajectory 5 is for outer layers and we will not focus on it.

While the PSE trajectory is independent from all others, PPP and CC have cluster 5 in common. While other clusters were unequivocally assigned to a single trajectory (see for instance, cluster 3, with ‘Slingshot’ assigning it a probability close to 1 of belonging to the CC trajectory, or cluster 4, with a probability close to 1 of belonging to the PPP trajectory; Extended Data Fig. 2e) or shared by all trajectories (such as early phloem cells in cluster 8; Extended Data Fig. 2e), cluster 5 was not clear cut, with a probability of ~0.75% of belonging to trajectory 2 (CC) and ~0.25 of belonging to trajectory 1 (PPP) (Extended Data Fig. 2e).

Regarding gene expression, cluster 5 does not express any canonical CC or PPP marker strongly. However, these markers (*SUC2*, *NAKRI*, *AHA3* for CC or *S17* for PPP) are only highly expressed in more mature cells. Cluster 5 has 64% of cells expressing the CC marker *SAPL* (409/638 cells) and 20% expressing the PPP marker *S17* (127/638 cells), with 12% of the cells in this cluster expressing both genes simultaneously (76/638 cells). This indicates that more cells in cluster 5 express CC markers than PPP markers.

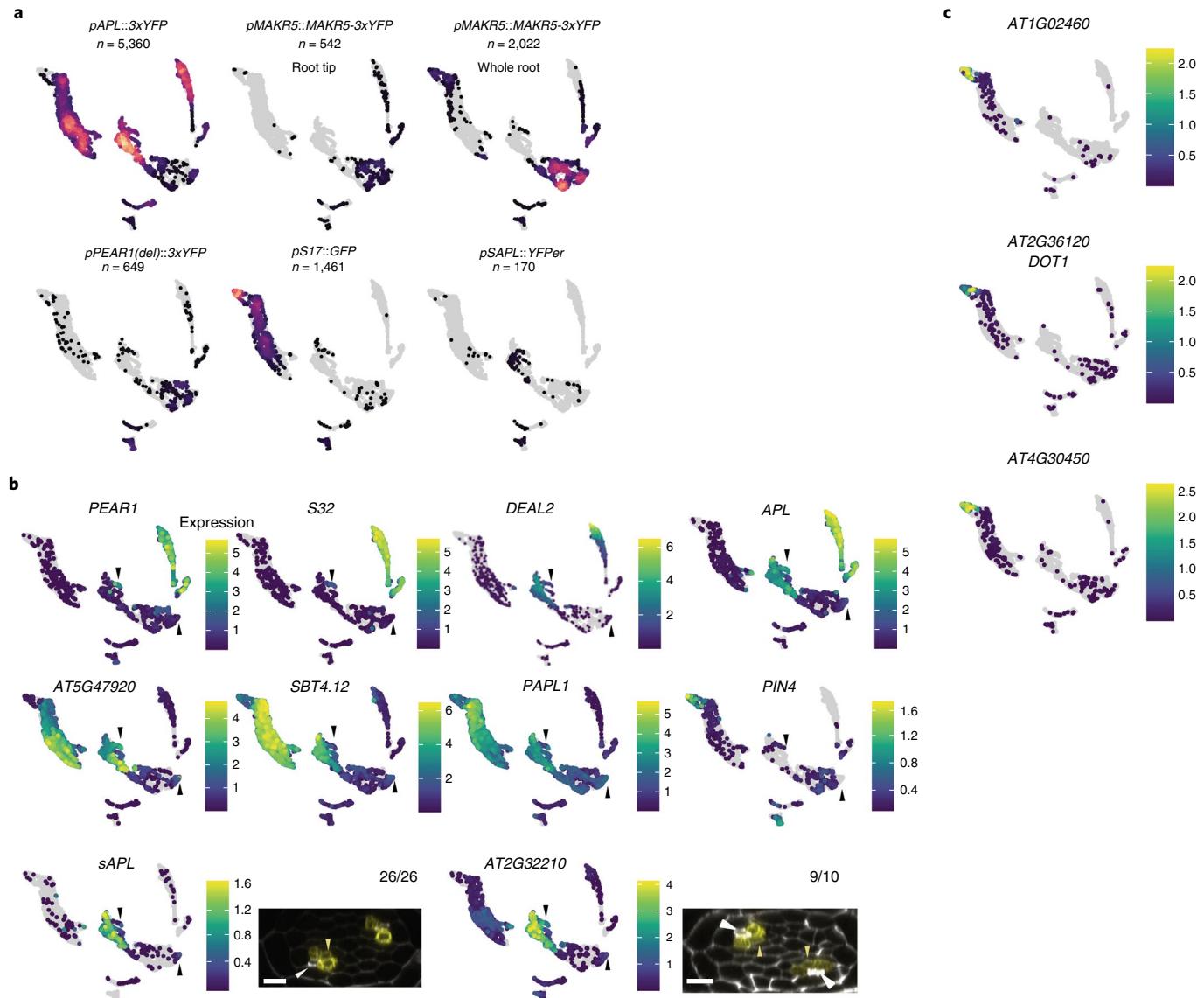


Fig. 2 | MSE cells identification and identity of cluster 11. **a**, Cells were plotted in the UMAP separated by sorting experiment, as indicated in each panel, to show which sorting experiment provided every cell. Colour indicates point density (lighter colour indicates higher density of points), with grey areas meaning an absence of cells. Numbers in each panel indicate the number of filtered cells contributed by that sorting experiment. We sorted *MAKR5* twice, one enriching in root tips (*MAKR5*) and the other sorting in the usual one-third of the root (*MAKR5* differentiated). **b**, UMAPs showing the cluster-weighted normalized expression of marker genes used to determine MSE identity. *PEAR1*, *S32* and *DEAL2* are expressed in sieve elements. *APL* is genuinely expressed in PSE, CC and MSE. *At5g47920*, *SBT4.12* and *PAPL1* are expressed in a ring expression pattern, including MSE and other cell types. *PIN4* is used as a negative control since it is excluded from sieve elements early in development³. *SAPL* and *At2g32210* are expressed in CC and MSE. Black arrowheads point to clusters 1 and 10. Scale bars, 10 µm in the confocal cross sections. White arrowheads point to PSE as a reference point and yellow arrowheads point to MSE. The numbers above each image indicate samples with similar results, of the total independent biological samples observed. **c**, UMAPs for xylem pole pericycle markers, which are found in cluster 11 together with other PPP markers, indicating that this is a late pericycle cluster.

This matches our observations in the root, where *SAPL* starts to be expressed earlier in development than *S17* (Extended Data Fig. 2g). The fact that a small percentage of cells express both markers at the same time despite being specific for different cell types indicates that transcriptional reporters are not always highlighting weak gene expression, so it is possible that our transcriptional data paint broader expression domains than the ones visible with the specific marker lines (see for example, the broader *SAPL* expression domain compared with the cells sorted using *pSAPL::VENUSer* reporter line; Fig. 2a,b).

We tried to distinguish incipient PPP from early CC in cluster 5, but there is no known PPP-specific marker expressed earlier in

development than *S17*. However, these intermediate PPP cells should have been collected in the sorting experiments '*pMAKR5:MAKR5-3xYFP whole root*' and '*pAPL:3xYFP*' (Fig. 2a), and should be present in the UMAP. These cells would sit in between the early PPP cells, sorted using '*pMAKR5:MAKR5-3xYFP root tip*' enriched in root tips, and those expressing *S17*, sorted using *pAPL::3xYFP* and *pS17::GFP* markers.

Therefore, cluster 5 gathers CC and PPP cells that exist in the same transcriptional state but are fated to differentiate into different cell types.

The developmental trajectories obtained reinforce clusters 8, 9 and 10 as early CC, PPP and SE cells. Given that these populations

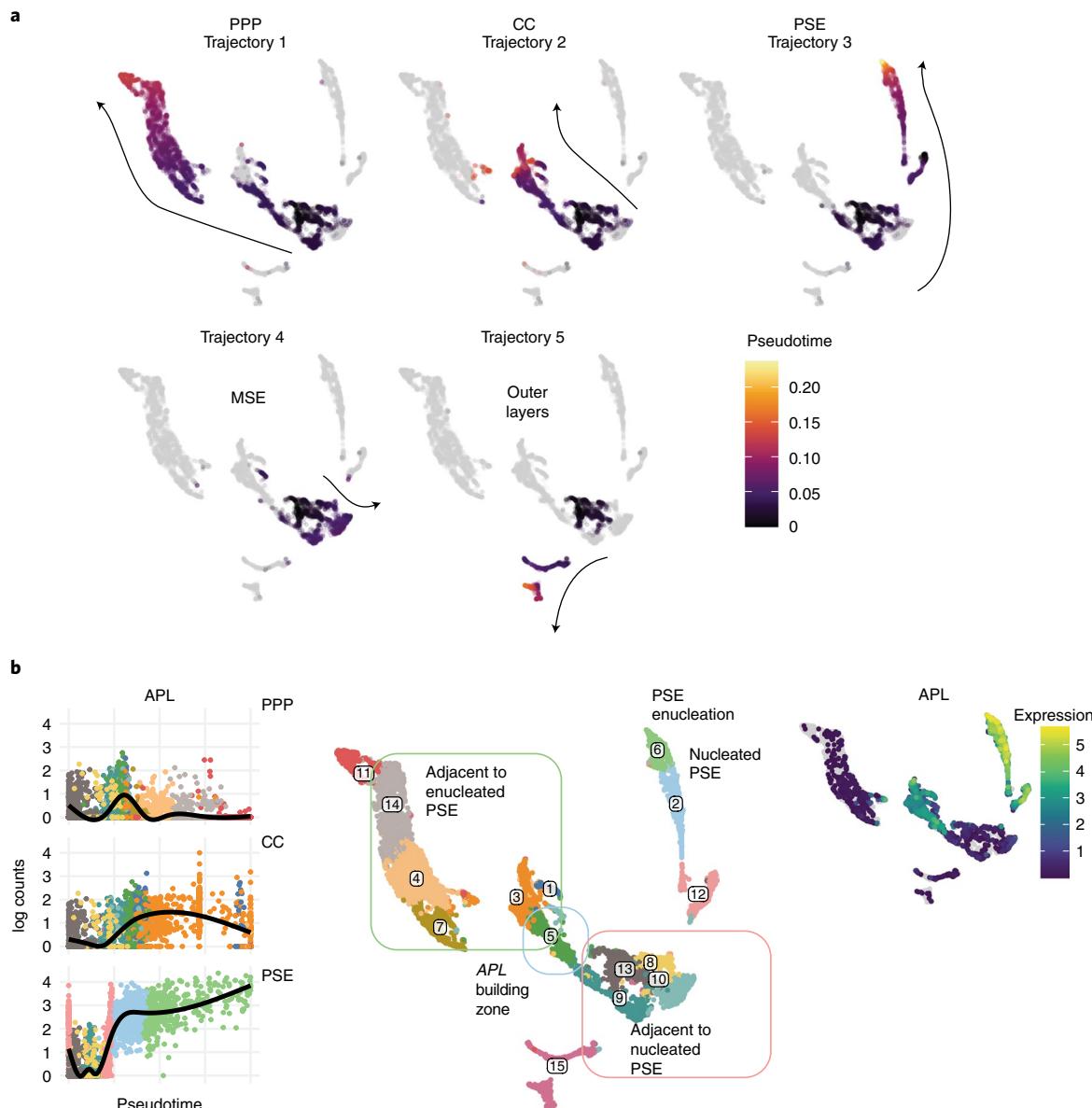


Fig. 3 | Developmental trajectories and mapping of the PSE enucleation point. **a**, Developmental trajectories inferred using Slingshot coloured according to pseudotime, with more mature cells in yellow. The origin for all trajectories was set in the clusters containing cycling cells. **b**, APL expression is plotted along the PPP, CC and PSE trajectories, with the cells coloured by cluster number in the UMAP. The black line is a smoothed trend estimated from a non-parametric generalized additive model. **b'**, APL is used as a standard to coordinate the three trajectories. Cluster 5 groups the cells with an increasing expression of APL in the PPP and CC trajectories, mapping the enucleation point in the adjacent cells. The position of each cell type is indicated in the UMAP in relation to PSE enucleation. **b''**, APL expression plotted in a UMAP of the phloem pole cell atlas.

are composed mainly of cells sorted using *MAKR5* and *PEAR1del* (Fig. 2a), we can conclude that these clusters correspond to early phloem cells containing three different identities (MSE, PPP and CC) but still undifferentiated.

There is no statistically enriched gene in cluster 9 and the few enriched in cluster 8 (Supplementary Table 1) are broadly expressed in whole-root single-cell data. Except for PSE, when we detect cycling cells expressing PSE markers in cluster 12, it is hard to distinguish an early identity in the other trajectories. However, when early phloem cells are compared to the early cells in general root cell atlases, early phloem cells cluster together more than expected by chance compared with other early cells, suggesting that early phloem cells have a specific signature (Fig. 4f,g).

An important event in phloem development is the enucleation of PSE, since at that moment this cell type loses the nucleus and

stops directing phloem progression, becoming dependent on neighbouring cells for survival and probably triggering changes in their transcriptomes. To map the enucleation point in the UMAP and determine which cells are neighbouring PSE before and after enucleation, we needed to coordinate trajectories since each trajectory has a different pseudotime. To coordinate them, we used our knowledge of *ALTERED PHLOEM DEVELOPMENT* (*APL*) expression, which occurs at different times in all three trajectories, combined with the enucleation markers *NAC DOMAIN CONTAINING PROTEIN 86* (*NAC086*, *At5g17260*) and *NAC45/86-DEPENDENT EXONUCLEASE-DOMAIN PROTEIN 4* (*NEN4*, *At4g39810*) (Fig. 3b and Extended Data Fig. 2f). *APL* is first expressed in PSE and at the time of enucleation is transcriptionally activated in CC and MSE⁴. In reporter lines such as *pAPL::3xYFP*, we perceive a strong signal in PPP as well (Extended Data Fig. 1a), a circumstance that

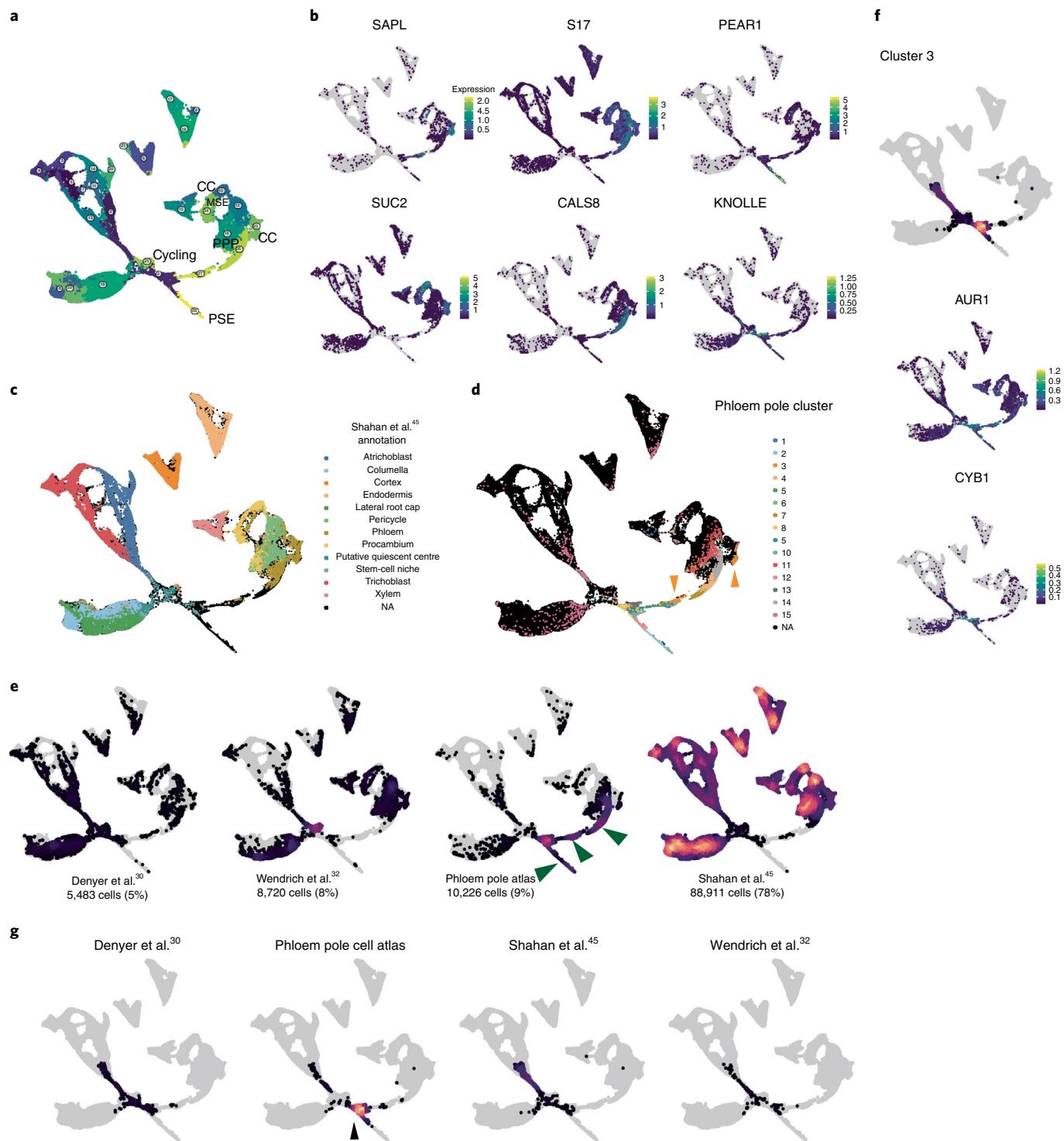


Fig. 4 | Phloem cell types in the integrated UMAP. **a**, A new UMAP containing 113,340 cells was generated by integrating cells from Denyer et al.³⁰, Wendrich et al.³² and Shahan et al.⁴⁴. Colours are used to differentiate cell clusters. **b**, Different markers were plotted in the UMAP to identify the phloem pole cell types: SAPL (CC and MSE), S17 (PPP), PEAR1 (PSE, MSE), SUC2 (mature CC), CALS8 (PPP and CC), KNOLLE (cycling cells). **c**, Integrated UMAP showing cells coloured according to the annotation from Shahan et al. **d**, Integrated UMAP coloured by the original clusters from the phloem pole atlas. Orange arrowheads point to the two parts of cluster 3, split in the integrated dataset. **e**, Integrated UMAP with the cells contributed by each individual project plotted (number indicated below, percentage of the total in brackets), using a coloured scale to indicate cell density (brighter more dense). Green arrowheads point to the clusters mostly contributed by our dataset. **f**, Cluster 3 of the integrated dataset containing root early cells and dividing cells. Cells in cluster 3 (early cells) are indicated in the top panel while the other panels in the column show the expression of G2/M genes (AUR1 and CYCB1) in the integrated dataset, marking dividing cells. **g**, Contribution of each single-cell project to cluster 3. Observe the grouping of early phloem cells (black arrowhead) compared to the higher dispersion of early cells in other datasets.

we took advantage of for sorting, although the phloem pole cell atlas transcriptomics data do not reflect a strong *APL* expression in mature PPP (Fig. 3b' and Extended Data Fig. 2d). In another reporter line, *pAPL::YFP*, we also observe a signal in PPP that gets weaker going shootward (Extended Data Fig. 2d). Based on reporters and transcriptomics data, the signal in PPP is probably not the product of gene expression in this cell type but probably caused by direct unloading from PSE that lags and gets diluted in successive cell divisions.

While *APL* expression increases in the PSE trajectory until enucleation, it starts being detected in the PPP and CC trajectories in the common cluster 5 (Fig. 3b), indicating that this is the transition zone when *APL* starts building up in the neighbouring cell types before enucleation. Therefore, the PSE trajectory is contemporary to the early phloem cells, cluster 5 coincides with PSE enucleation preparation, and clusters for mature PPP and CC contain cells that are neighbouring an enucleated PSE (Fig. 3b').

First stages of MSE development identified

MSE is difficult to identify since there are no specific markers available for this cell type. However, by reducing the diversity of cells in our sample using cell sorting, we were able to gain some insights about this elusive population of cells. Slingshot identified a trajectory (trajectory 4, Fig. 3a) that is mainly formed by cluster 10 (Extended Data Fig. 2b), which is mostly contributed by *MAKR5* sortings, indicating that this could be early MSE cells (Fig. 2a). In cluster 10, we find cells expressing MSE markers such as *sAPL* and *APL* (Fig. 2b), and other genes expressed in MSE and other cell types but excluded from PSE (*At5g47920*, *PAPL1*; Figs. 2b and 7).

In addition, we know that procambial markers such as *PIN-FORMED 4* (*PIN4*, *At2g01420*) become excluded from MSE cells early in development³³ and we find that this marker is absent from cluster 10 (Fig. 2b). Out of 31 genes identified as highly expressed in cluster 10 compared with others (FDR < 1%), 29 were S-phase genes (histones), indicating that these are still early cells and therefore harder to characterize further.

We also know that MSE cells should not display PSE markers in early stages, since these are no longer expressed in MSE after lineage bifurcation but will express these signature genes later in development. For this reason, we interpret that cluster 1 is a more developed MSE, since we find early SE genes such as *PEARI* and *S32* (*At2g18380*) expressed in this cluster, which is mainly contributed by the *APL* sorting. Cluster 1 belongs to the CC trajectory, possibly because CC and MSE at this stage share some transcriptional expression, as evidenced by reporters such as *At2g32210* (Fig. 2b), *SAPL* and the cases shown in Fig. 6b, highlighting how phloem pole cell fates are intertwined along development.

While we have end points for our PSE, CC and PPP trajectories, we don't expect to have an endpoint for MSE, since this cell type differentiates further away from the meristem¹⁴. Out of the 7 genes identified to be expressed in both sieve elements¹⁴, we detected all the genes in PSE clusters but only *DESIGUAL2* (*DEAL2*, *At4g21310*) in cluster 1 and CC, confirming that we have not sampled mature MSE cells. However, we are convinced that we have identified the first stages of MSE development in clusters 10 and 1.

The atlas represents a continuum of phloem development

To explore the depth of our data, we integrated our phloem atlas with existing root single-cell datasets^{30,32,44} (Fig. 4a). After filtering, this process rendered a UMAP with 113,340 reclustered cells, of which 9% belonged to our project, 7.69% to Wendrich et al.³², 4.84% to Denyer et al.³⁰ and 78.4% to Shahan et al.⁴⁴. We used markers to identify cell types (Fig. 4b,d) and projected the clusters of the phloem pole cell atlas in the integrated dataset to confirm our trajectories (Fig. 4d). The relative position of the original clusters is similar in the integrated data as it was in our analysis (Fig. 1b),

except for the cells which we named as 'outer layers', which are dispersed in different parts of the integrated UMAP.

When projecting the cells of each project in the UMAP, a continuity can be observed in the cells contributed by our atlas covering the gaps in the other data (Fig. 4e). Indeed, most PSE cells (cluster 28), a majority of the intermediate PPP and CC cells (cluster 27) and the early cells in cluster 26 (PPP) were provided by our dataset (Extended Data Fig. 3), demonstrating the difficulty in sampling phloem without using an active strategy to enrich this population. Most of the cells classified as 'phloem' by Shahan et al. coincide with our clusters 4 and 14, which we annotated more specifically as PPP cells (Fig. 4c). There are also a few cells near our cluster 3, which we annotated as CC cells. We also noticed that cluster 3 (companion cells) is split in this integrated dataset (Fig. 4d, orange arrowheads) between mature CC (orange dots on the right of the plot, that is, cells expressing mature CC markers such as *SUC2* and *NAK1*; see Extended Data Fig. 3a) and the orange cells in the less mature CC expressing *SAPL* (Extended Data Fig. 3a). Perhaps the more mature CC have higher overall similarity to more mature cells of other phloem cells (such as PPP for example), which are more represented in the integrated dataset than in the original atlas.

While other atlases, in particular Shahan et al., excelled in collecting mature cells (including mature MSE cells; Extended Data Fig. 3a,b), the continuity observed in the UMAP allowed us to track phloem developmental trajectories more accurately (Fig. 3a) and enrich populations that were underrepresented in other general root atlases (Fig. 4e and Extended Data Fig. 3a,b).

We also wanted to compare root phloem with a recently published single-cell dataset on leaf, containing 478 vascular cells⁴⁵. In *Arabidopsis* leaves (Fig. 5a), veins are often formed by multiple sieve elements usually surrounded by at least two CC and one phloem parenchyma cell. In turn, phloem parenchyma cells, which are more irregular and have a much less dense cytoplasm compared with CC, are often in contact with one or more CC, sharing comparatively many more connections than other interfaces⁴⁶.

When the root and leaf data were integrated and clusters were annotated using marker genes (Fig. 5d), we noticed that PPP and phloem parenchyma cells blended in two clusters (Fig. 5b,c). Cluster 9 of the integrated data was formed by CC, which are present in both leaf veins and roots. However, cluster 6 of the integrated data contained a mixture of cells annotated as mature root pericycle cells and phloem parenchyma cells from leaves, showing expression of signature PPP (Fig. 5e), phloem parenchyma genes (Fig. 5f) and XPP (Fig. 5g) in both datasets (see Methods on how we assessed the degree of mixing of the cells from the two datasets in cluster 6). Phloem parenchyma leaf genes are expressed in clusters 11 and 14 in the phloem pole cell atlas, corresponding to pericycle and PPP, respectively (Fig. 5f). Pericycle tissue is present in roots and stems but not in leaves, and phloem parenchyma cells are found in aerial tissue and root secondary phloem but not in primary root. Despite being different cell types with different origins, the transcriptional overlap between phloem parenchyma and mature pericycle is another indication of the importance of positional information for cell function in plants, reinforcing the role of PSE as phloem organizer. These data also suggest that parenchymatous cells share similarities across different organs and underscore their relevance for phloem.

Phloem pole cells share transcriptional programmes

To identify groups of genes showing distinct expression patterns in the phloem poles, we built a gene co-expression network from our single-cell RNA-seq data using the algorithm implemented in 'bigScale2'⁴⁷, which uses a gene-gene correlation metric specifically tailored for sparse single-cell data. This resulted in a gene-gene network containing 5,238 vertices (genes) and 370,794 edges (connecting two genes if their correlation was above 0.9). The biological

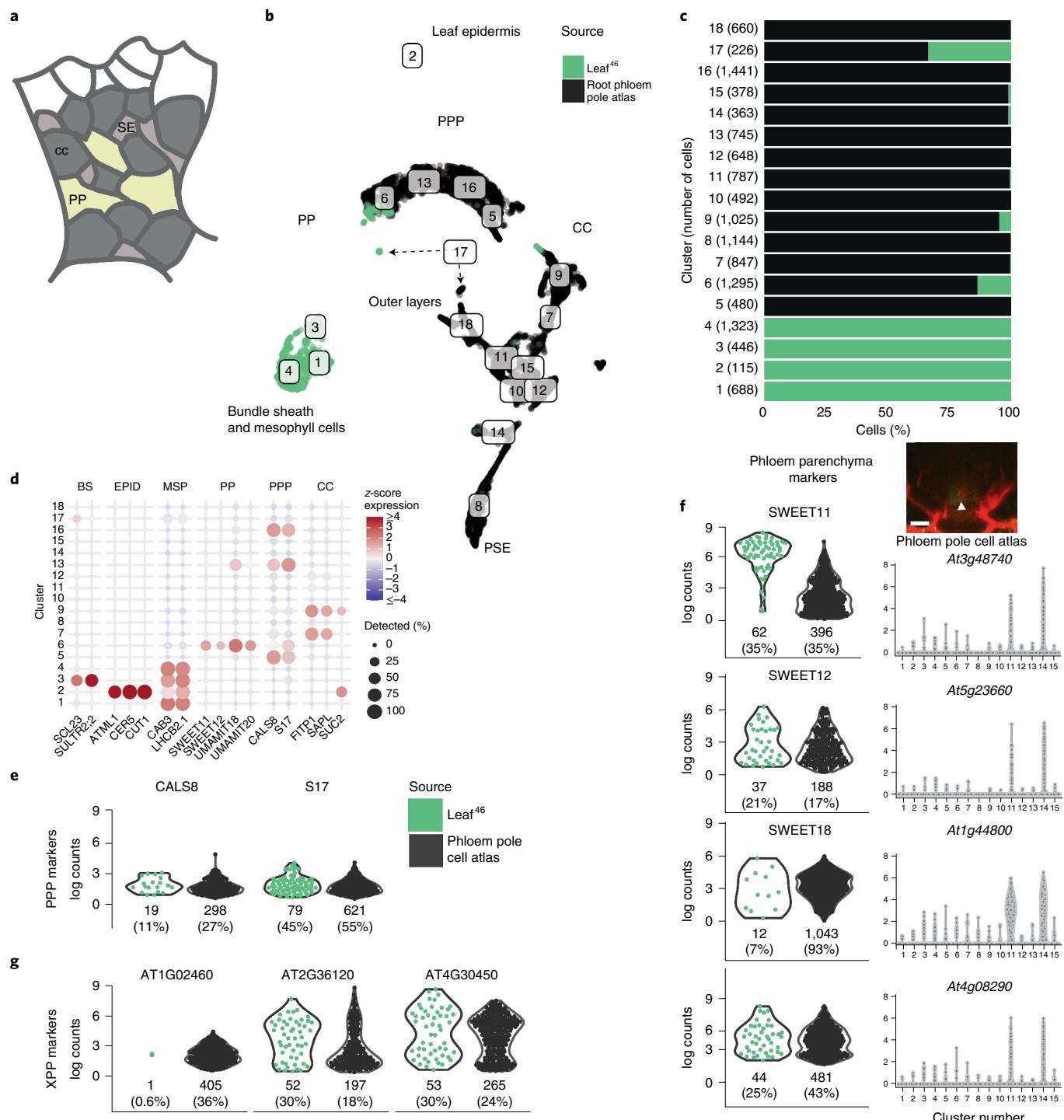


Fig. 5 | Similarities in gene expression between leaf phloem parenchyma and root pericycle. **a**, Schematic of the leaf minor vein showing phloem anatomy. Notice the different composition in terms of cell identities, cell number and organization compared with the root. **b**, UMAP integrating the phloem pole cell atlas with the leaf single-cell dataset. Cells were combined and reclustered, coloured by source (leaf in green, root in black). Notice the separation of the leaf-specific clusters (bundle sheath and mesophyll cells) and overlap in clusters 6 (PPP/phloem parenchyma) and 9 (CC). **c**, Percentage of cells contributed by each dataset in each cluster. The y axis shows the cluster number with the number of cells in it enclosed in brackets. Root cells are coloured black, leaf cells are in green. **d**, Cluster annotation of the root-leaf UMAP based on markers with known tissue-specific expression. The size of the points represents the percentage of cells in a cluster where the gene was detected (that is, at least 1 UMI). The colour shows the scaled average expression of the gene (z-score, that is, number of standard deviations above/below the gene's mean across all cells). **e–g**, Violin plots showing the expression of PPP markers in leaf (green) and root (black) cells for phloem pole pericyclic markers (**e**), phloem parenchyma markers (**f**, left column) and XPP markers (**g**). The numbers under the black/green violin plots indicate the number of cells in cluster 6 of the leaf/root UMAP expressing each gene, with the percentage enclosed in brackets. In **f** (right column), the gene expression of the respective genes is shown in the phloem pole cell atlas. The confocal image in **f** (top right) shows the expression of *pSWEET11::SWEET11-2A-GFP* in PPP in roots. Scale bar, 10 μm. The white arrowhead points to PSE as a reference point. Panel **a** adapted with permission from ref. ⁴⁶, Springer Nature Limited.

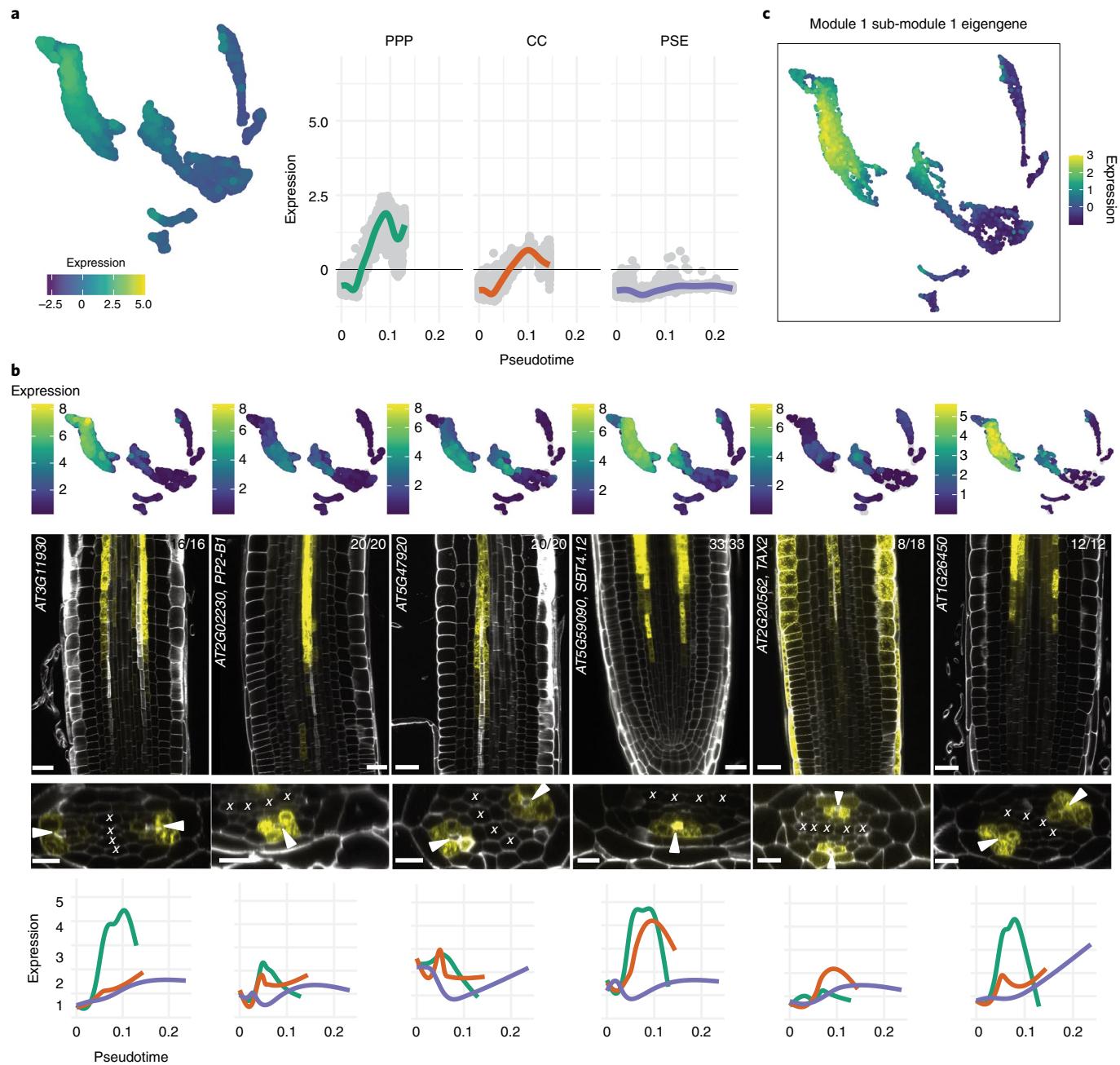


Fig. 6 | Identification of a gene expression pattern common to non-PSE cells frequent after PSE enucleation. **a**, The module 1 eigengene profile with its expression along PPP, CC and PSE trajectories. **b**, New genes with an expression pattern validating the gene profiles grouped in module 1. All genes presented in this panel are expressed forming a ring around PSE at the time of PSE enucleation. *SBT4.12* and *TAX2* are also expressed in late PSE, with *TAX2* also showing expression in the epidermis. Top row: UMAPs show the particular cluster-weighted normalized expression of each gene in the phloem pole cell atlas. Middle row: microscopy images are representative images of the transcriptional reporter lines where the gene promoter is fused to *VENUSer*. Scale bars, 25 μ m in the longitudinal sections, 10 μ m in the cross sections. White arrowheads point to PSE cells as a reference point. 'X' marks xylem cells. Bottom row: each gene has also been plotted in PPP (green), CC (orange) and PSE (purple) Slingshot trajectories, showing average expression values on the y axis and pseudotime on the x axis. **c**, Expression profile of sub-module 1 eigengene, the sub-module of module 1 that is enriched for genes with ring-specific expression. This sub-module contains all the genes in the panel except for *TAX2*, which was not present in our network. In **b**, the numbers on the top right corner in each longitudinal section indicate samples with similar results, of the total independent biological samples observed.

validity of this network was confirmed by the fact that out of 59,545 edges containing genes both present in our network and in known TF-target lists (*Arabidopsis* Gene Regulatory Information Server, AGRIS⁴⁸), 51,658 (~86%) were preserved as linked pairs in our network. To identify groups of genes with correlated expression profiles, we used the Louvain algorithm and obtained a total of 16 gene

modules (Extended Data Fig. 4 and Supplementary Table 2), and summarized their expression as the first principal component of a principal components analysis (PCA), which we refer to as an eigengene¹⁹. Among them, most of the modules were broad in all the trajectories with different temporal patterns. Module 6 seems to represent genes with high expression in PSE (Extended Data Fig. 4).

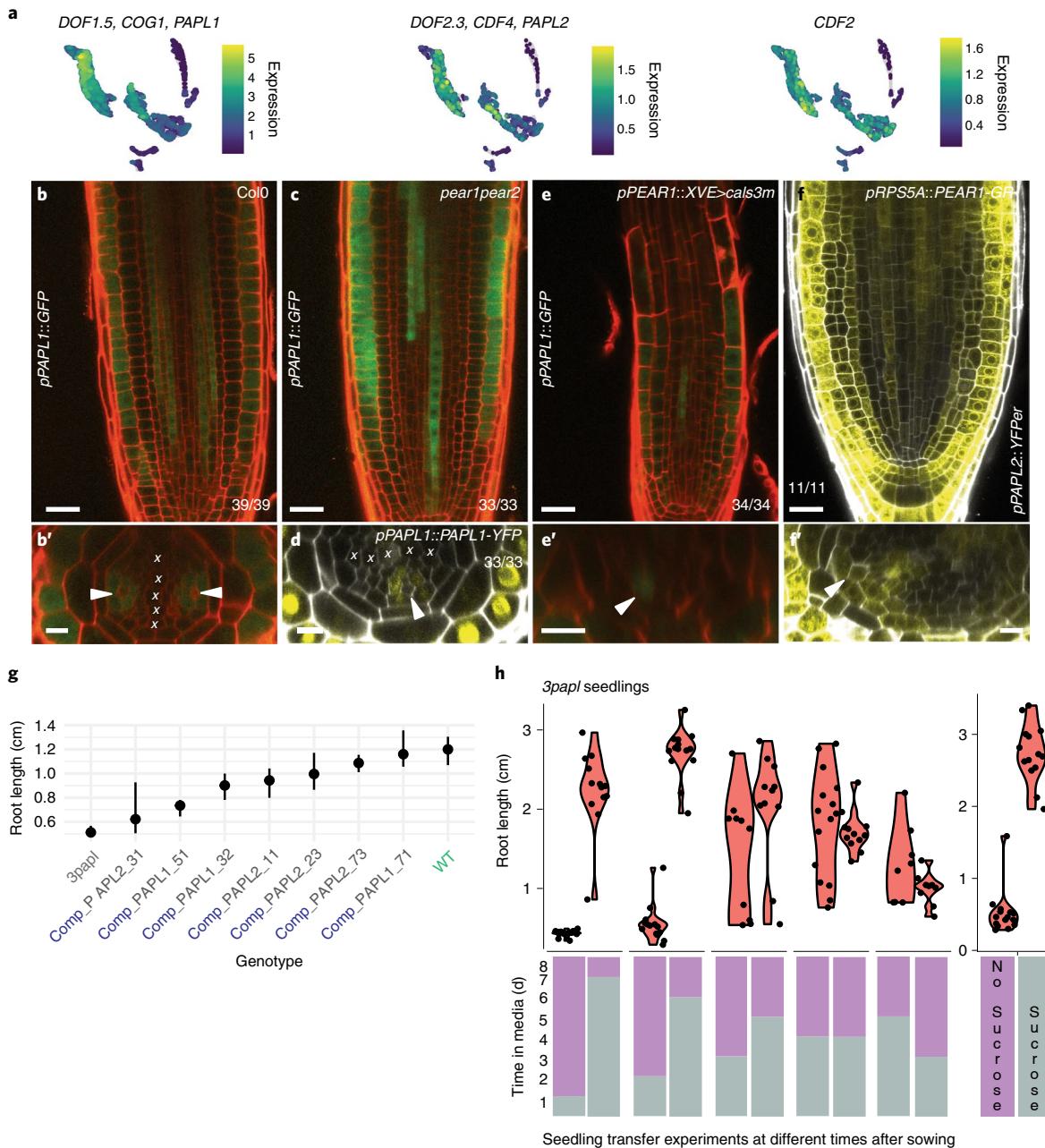


Fig. 7 | PAPL genes are PEAR targets that influence root nutritional status. **a**, UMAPs showing the expression of *PAPL1*, *PAPL2* and *CDF2* in the phloem pole cell atlas. Note that these genes are also expressed in the early phloem cell clusters. **b**, *pPAPL1::GFP/GUS* expression domain, showing expression in the cells around PSE from 40 μ m from the QC and in the epidermis. **c**, Phloem expression of *pPAPL1::GFP/GUS* is delayed until 120 μ m in *pear1pear2* mutant background. **d**, *pPAPL1::PAPL1-YFP* (Col-0) translational domain coincides with the transcriptional domain. **e**, The ring pattern of *pPAPL1::GFP/GUS* gets distorted upon PSE plasmodesmata closure using the ‘cals3m’ tool (*pPEAR1::XVE>cals3m*). **f**, *PAPL2* (*pPAPL2::VENUSer*) becomes ectopically expressed upon *PEAR1* overexpression in the meristem (*pRPS5A::PEAR1-GR*). **g**, Average root length of 6 dps seedlings in 3papl, WT and complementation lines in 3papl background with genomic constructs for *PAPL1* (3 lines) or *PAPL2* (4 lines) in sucrose-depleted media. The median and 95% confidence interval are shown (see Methods). Number of seedlings measured: 488 3papl; 273 *PAPL1-32*; 37 *PAPL1-51*; 33 *PAPL1-71*; 343 *PAPL2-11*; 79 *PAPL2-23*; 37 *PAPL2-31*; 314 *PAPL2-73*; 382 WT. The same data is also shown in Extended Data Fig. 9h, separately for each experimental batch and seed stock (see ‘Experimental design’ in Methods). Statistical analysis comparing each mutant genotype to the WT is in Supplementary Table 8. **h**, Transfer experiment between sucrose and sucrose-depleted plates of 3papl seedlings. Top, root length for seedlings transferred between media containing/depleted of sucrose at different times after sowing. Bottom, amount of time the seedlings in the top panel spent with (grey) or without sucrose (purple). Days spent with and without sucrose are represented by grey and purple bars, respectively, and roots were measured at 8 dps. Number of seedlings: 131 3papl control; 24 3papl seedlings on average per transfer experiment. Statistical analysis comparing each pair of conditions is in Supplementary Table 8. In confocal images **b**, **e** and **f**, primed letters show the respective cross sections. Scale bars, 25 μ m in longitudinal sections, 10 μ m in cross sections. White arrowheads point to PSE cells and ‘X’ marks xylem cells. The numbers at the bottom of each confocal image indicate samples with similar results of the total independent biological samples analysed.

In contrast, module 1, which contains 1,367 genes (Supplementary Table 3), displays an increasing expression in both PPP and CC trajectories and a lower-than-average expression in the PSE trajectory (Fig. 6a). Reporter lines for genes in this module followed these predictions: in addition to genes with broader expression (such as *MES7*; Fig. 1d), we identified genes showing a ‘ring’ pattern, expressed specifically in all the cells around PSE (Fig. 6b and Extended Data Fig. 5). While *At3g11930*, *At2g02230* (*PHLOEM PROTEIN 2-B1*, *PP2-B1*), *At5g47920*, *At3g16330*² and its sister gene *At1g52140*, and *At4g27435* do not show a strong expression in PSE, *At5g59090* (*SUBTILASE 4.12*, *SBT4.12*), *At2g20562* (*TAXIMIN 2*, *TAX2*) and *At1g26450* are expressed in late PSE in addition to being expressed in a ring pattern. Some of the genes found are expressed in some of the cells around PSE (incomplete ring) and other cell types (Extended Data Fig. 5). For instance, *At4g27435* is expressed in CC and occasionally in PPP and protoxylem plus lateral root cap. *At3g21770* (*PER30*) and *At3g11930* are found in the ring around PSE but extend to procambium higher up (Extended Data Figs. 5 and 6b).

Out of the nine genes with a ring expression pattern as observed with reporter lines (see above), seven were found in module 1, with *TAX2* (*At2g20562*) not included in our network and *At4g27435* found in module 4, which includes genes expressed in all trajectories. Despite module 1 being the largest on our network, this result is more than would be expected by chance (hypergeometric test, $P=0.0005$).

Because of the large size of module 1, we tried to refine our analysis by sub-clustering the genes within this module, to identify a more specific group of candidate ‘ring genes’ as defined by the reporter analysis above. This resulted in 15 sub-modules, with 5 of them containing over 100 genes (Extended Data Fig. 6). Six of the 7 ‘ring genes’ from module 1 fell within the same sub-module 1 (except for *At3g16330*), which again is more than would be expected by chance (hypergeometric test, $P=0.0009$). While we do not expect all 326 genes in this sub-module to have a ring expression pattern, this analysis highlights that this pattern is widespread for a variety of phloem genes, which group together by similarity in expression pattern. On the other hand, a gene such as *MES7*, which we saw was not entirely ring-specific, fell in a different sub-module. Therefore, our network analysis suggests that there is a complex ring-specific pattern of expression shared across several genes in the phloem pole.

The complex patterns in the cells around PSE point out that PSE-adjacent cells share some common developmental programmes that are maintained even when cells differentiate into their specific identities, suggesting that the transcriptional signature of phloem cells is influenced by multiple positional cues.

This set of genes could be important in understanding how PSE relates to its neighbouring cell types before and after enucleation. Indeed, as observed in the UMAPs, the ring pattern is frequent right after PSE enucleation, suggesting a shift in the phloem pole governance after PSE enucleation.

PINEAPPLE ring genes are expressed in early phloem

Among the genes in module 1 sub-module 1 that also extend their expression into the less mature clusters, we found a DOF transcription factor, *DOF1.5* (*COGWHEEL1*, *COG1*). This gene and its sister gene *DOF2.3* (*CYCLING DOF 4*, *CDF4*) are expressed in early phloem cells (Fig. 7a). *CDF4* encodes a differentiation factor in columella cells that is repressed by *WOX5*⁵⁰. The role of *COG1* in roots is unknown but this transcription factor is a negative regulator of phytochrome signalling⁵¹ and promotes brassinosteroid biosynthesis by upregulating *PIF4* and *PIF5*, leading to hypocotyl elongation⁵². Both genes have been involved in regulating tolerance to seed deterioration^{53,54} as well as flowering time⁵⁵.

Transcriptional fusions of both *DOF1.5* and *DOF2.3* confirmed the expression of both TF in PPP, CC and MSE from 40 µm from

the quiescent centre (QC), remarkably earlier than the other ring genes described above. While both genes form a ring around PSE reminiscent of a pineapple slice (the expression is weaker or absent in PSE; Extended Data Fig. 7i), *DOF1.5* (henceforth *PINEAPPLE1*, *PAPL1*) is also expressed in the epidermis (Fig. 7b) and *DOF2.3* (*PAPL2*) is found in columella cells with a broader domain towards the QC (Extended Data Fig. 7f,f'). The ring pattern observed with the green fluorescent protein (GFP) fusion construct extends one layer towards procambium when fused to three units of yellow fluorescent protein (3xYFP) (Extended Data Fig. 7a), indicating a weaker expression in this layer. Translational fusions show that these transcription factors are nuclear localized and not mobile (Fig. 7d and Extended Data Fig. 7a–c,e), since transcriptional and translational patterns are coincident. This indicates that PAPL transcription factors act cell-autonomously. Together with the translational domain of *MAKR5*, the expression domain of *PAPL* genes indicates that complex expression patterns in the phloem are relevant from an early stage.

PAPL genes were predicted to be downstream of *PEAR* in microarray data, as were other genes in module 1 (ref. ²). *PEAR* transcription factors move to PSE-adjacent cells to control periclinal cell divisions and other transcriptional programmes non-cell-autonomously. This is evidenced by markers such as *SAPL* and *At3g16330* becoming ectopically expressed after broad *PEAR* overexpression, or *SAPL* being expressed in PSE upon PSE plasmodesmata closure².

To validate that *PAPL* genes are downstream of *PEAR*, *PAPL* reporter lines were expressed in *pear1pear2* double mutants, which resulted in a delay in *PAPL* expression, from 40 to 120 µm from the QC (Fig. 7c and Extended Data Fig. 7g). Since *PEAR* genes are highly redundant, we also introduced *PAPL1* constructs in the *pear sextuple* mutant, *pear sext*³³, where we observed a loss in its usual meristematic expression (Extended Data Fig. 7d). In parallel, closing PSE plasmodesmata connections to the neighbouring cell types using ‘icals3m’ tool alters the ring expression of *PAPL1* (Fig. 7e) and overexpressing *PEAR1* leads to ectopic expression of *PAPL2* (Fig. 7f). These results validate that *PAPL* genes are downstream of *PEAR* and indicate that *PEARS* are needed and sufficient to express *PAPL* genes in the early phloem.

In addition to the *PAPL* genes, we validated that some of the genes in module 1 act downstream of *PEAR* TF. Indeed, *PEARS* are sufficient to induce *SBT4.12*, *At3g11930*, *MES7* and *PER30*, since these genes become ectopically expressed upon induction of *PEAR2* expressed under a ubiquitous promoter (*pRPS5A*) (Extended Data Fig. 7j). In *pear sext*, the expression pattern of *PER30* and *MES7* was modified, while *SBT4.12* expression was decreased and *At3g11930* spread towards the meristem (Extended Data Fig. 7j).

PAPL proteins link PEAR genes to root physiology

Next, we decided to check whether *PAPL* genes were downstream of *PEAR* genes to control periclinal cell divisions. Since *PAPL* expression is delayed in *pear1pear2* double mutants and absent from early phloem, we chose this mutant as a background to express *PAPL1* under the *WOODEN LEG* (*WOL*) promoter. When inducing *PAPL1* expression (20 h treatment or germinated directly in beta estradiol and grown for 5 d), we did not observe a phenotype similar to *PEAR1* overexpression with increased periclinal cell divisions in the root² (Extended Data Fig. 8a–h). A similar result was observed when *PAPL1* was overexpressed in the stele in wild-type (WT) background (Extended Data Fig. 8i–p). These observations indicate that *PAPL* genes do not control periclinal cell divisions downstream of *PEARS*.

To gain insight into the function of *PAPL* genes, and after checking that *papl* single mutants did not show any obvious root phenotype, we generated double mutants (*papl1-1* *papl2* and *papl1-2* *papl2*). Bulk RNA sequencing identified *CYCLING DOF 2*, *CDF2*,

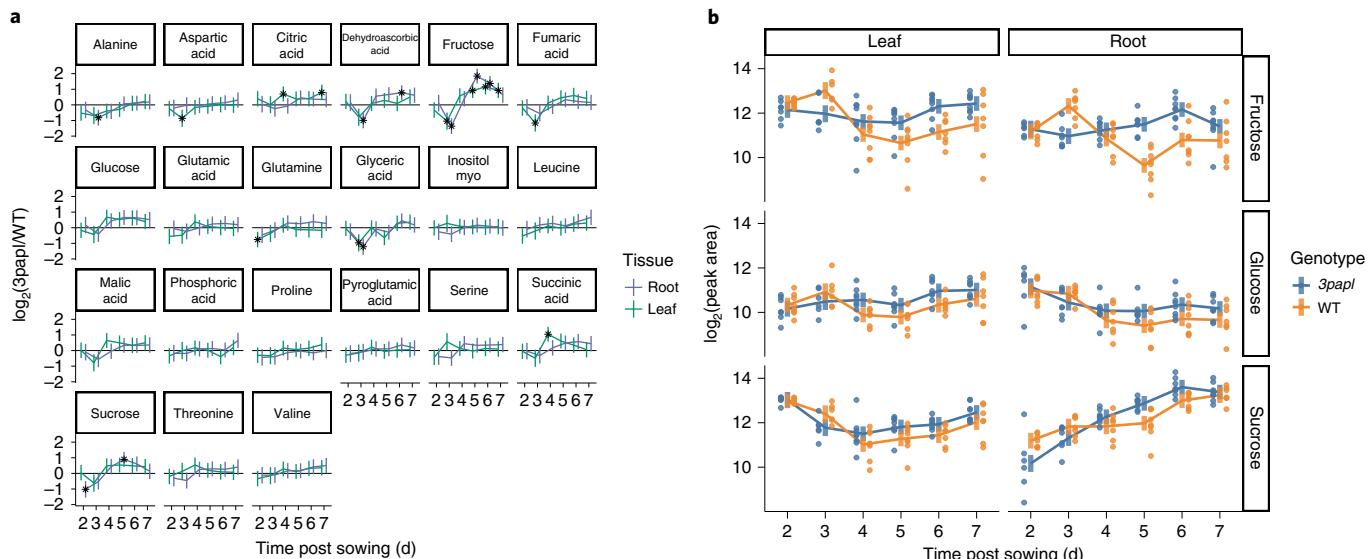


Fig. 8 | Difference in WT and 3papl metabolite levels in leaves and roots. **a**, Overview of the different metabolites with significant differences between WT and mutant seedlings in at least one of the time points. Points and error bars show the mean $\pm 2 \times \text{s.e.m.}$ (that is, an approximate 95% confidence interval) of the log(fold change) between WT and mutant metabolite levels, estimated from our linear model (Methods). Asterisks highlight points that are statistically significant after adjusting for multiple testing across all the tests (FDR of 5%). **b**, Average metabolite levels for sucrose in mutant and WT. The bars denote the 95% confidence intervals estimated from our linear model (Methods). The points show the raw data for individual samples. $N=6-8$ for each time point/tissue/genotype combination (3 of them had 6 replicates, 8 had 7 replicates and 13 had 8 replicates).

as upregulated in *papl1-1papl2* (Supplementary Table 4). This gene encodes another DOF transcription factor expressed in the cortex, pericycle and procambium, partially overlapping with *PAPL* expression (Extended Data Fig. 7h). Presuming this gene was upregulated to compensate for the lack of *PAPL* genes, we generated a triple mutant using a *cdf2* transfer-DNA (T-DNA) allele⁵⁵ (*papl1-1papl2cdf2-1, 3papl*).

The triple-mutant root was shorter than wild type in several conditions (Extended Data Fig. 9a), but the effect was more pronounced when seedlings were grown in media without sucrose (Fig. 7g and Extended Data Fig. 9a,b). A triple mutant with a new allele for *CDF2* generated using CRISPR/Cas9 technology showed similar results (3papl-2; Extended Data Fig. 9h). While wild-type plants grown in media without sucrose often showed a bimodal distribution in terms of root growth (Extended Data Fig. 9a,b), the proportion of roots arresting growth in 3papl was higher (Fig. 7g and Extended Data Fig. 9f). Despite high variation between seed batches, the average root length of the mutant is lower than that of the wild type (Extended Data Fig. 9h). Contrary to other phloem development mutants *apl* and *pearl*, the mutant phenotype of 3papl was mostly suppressed by the addition of 1% sucrose to the media (Fig. 7h and Extended Data Fig. 9e). In this scheme, compared with other mutants, root length in *pearlpear2* mutant was not so affected by the absence of sucrose in the media. When grown with sucrose, it was rescued to wild-type levels.

Since the mutants could be rescued by transferring them to sucrose, we aimed to identify the time point at which sucrose is needed for 3papl. For this experiment, we transferred plants from sucrose-supplemented to sucrose-depleted media and vice versa. The more time the mutant seedlings spent without sucrose, the more difficult it was for them to recover root growth (Fig. 7h and Extended Data Fig. 9c). Those recovering managed to grow well (Extended Data Fig. 9f). Spending at least 3 d in sucrose was necessary for the mutant seedlings to grow normally, while spending only 2 d in sucrose was not enough for root growth recovery (Fig. 7h and Extended Data Fig. 9c). This phenomenon was not observed in wild-type roots grown and transferred in parallel

(Extended Data Fig. 9g). In the confocal imaging, the root meristem of seedlings that got arrested looked shorter and stunted (Extended Data Fig. 9d). *PAPL* genes were expressed at this stage in both sucrose and non-sucrose conditions, showing similar patterns as observed in more mature seedlings (Extended Data Fig. 10a). Other phloem marker genes, such as *MAKR5*, *APL* and *ring gene SBT4.12*, were expressed in 3papl mutant background similar to wild type, suggesting that there are no defects in phloem development in 3papl (Extended Data Fig. 10b-d). In contrast, *MAKR5* expression was delayed in *pearlpear2* mutant background (Extended Data Fig. 10b) and *APL* expression was highly reduced in *pearl* (Fig. 3b), suggesting that *PAPL* genes do not fulfill the same roles as *PEAR* genes. These markers and *SUC2* are expressed similarly when the plants are grown in media with or without sucrose (Extended Data Fig. 10e-h).

To better understand the 3papl phenotype, we carried out metabolic profiling of leaves and roots of seedlings grown in sucrose-depleted media across 6 developmental stages (2–7 days post sowing, dps) (Supplementary Table 5). We identified 7 and 5 metabolites in leaves and roots, respectively, with significant differences between WT and mutant in at least one of the time points (<5% false discovery rate (FDR) from a linear mixed model fit to the whole data; see Methods and Fig. 8a). One of these metabolites was sucrose, with a significant difference only in the roots, where it started at lower levels in the mutant (days 2 and 3) and then continued to increase to reach levels comparable to the WT at the end of the experiment at day 7 (Fig. 8b). A similar pattern, with more significant points, was observed in fructose—a component of sucrose—and to a lesser extent in glucose—the other monosaccharide forming sucrose (Fig. 8b). It has been described that by the time the radicle emerges, all the sugars stored in the *Arabidopsis* seed have been consumed. Within 48 h after germination (approximately at day 3 after sowing), lipid and protein reserves are exhausted and seedlings need to switch to autotrophic growth^{56,57}. The data suggest that *PAPL* genes could be important after the seedling has transitioned to autotrophic growth, facilitating sugar transport to sink tissues such as roots. The continued increase in sucrose in the mutants

could be due to the, on average, smaller size of *3papl* seedlings and stunted growth, which could have led to reduced sucrose consumption and therefore its observed continued accumulation.

Discussion

Our Article demonstrates the power of tissue-specific transcriptomics combining FACS and single-cell sequencing to study elusive cell populations underrepresented in organ general cell atlases. The use of droplet-based technologies also allowed us to gather more cells and a higher resolution than plate-associated methods.

The phloem pole cell atlas allows a holistic understanding of phloem. While there are specific genes for PPP and CC, these cell types share the first stages of their developmental trajectory. Trajectory analysis also revealed the connection between CC and early MSE, providing new insights on early stages of MSE development. The commonalities among the different cell types were validated by gene regulatory network analysis and reporter lines confirmed the relevance of the ring expression pattern in all the cells around PSE.

PSE differentiation involves enucleation and becoming dependent on adjacent cells for survival. Using *APL* expression as a standard, we mapped the enucleation point in the atlas. While PSE organizes the phloem pole in the meristem neighboured by unspecialized cells, PSE enucleation marks the onset of cell differentiation for adjacent cells and switches on similar gene regulatory networks in PSE-surrounding lineages, as evidenced by the ring pattern shown by many genes right after PSE enucleation.

The coordinated expression in the cells of the phloem pole highlights the importance of positional information and cell to cell communication to preserve phloem function when PSE delegates control in the adjacent cells. They also underpin the relevance of PPP cells, which we believe should be considered a built-in part of phloem.

A phloem plasticity zone was recently described in the root meristem, where CC and MSE cells could act as a reservoir for PSE identity¹⁸. This further supports the coordination between the pole identities to ensure correct phloem morphogenesis. It would be interesting to investigate whether PPP can also transdifferentiate to PSE if required.

The similarities between root pericycle cells and phloem parenchyma cells in leaves suggest that parenchymatic cells share characteristics despite being present in different organs with variable anatomic configurations and reinforces PSE as the phloem pole organizer.

The modular analysis of the atlas identified the DOF *PAPL* genes, characterized by early expression in the ring domain and the inability to introduce new periclinal cell divisions when overexpressed in procambium.

Contrary to other phloem mutants, such as *apl*, the presence of sucrose in the media almost completely suppresses the root growth phenotype of *3papl*. Regarding PEARs, the root length in the *pear1pear2* mutant was not so affected by the absence of sucrose in the media and it was rescued to WT levels when grown with sucrose. The fact that the subtle root length phenotype is rescued by sucrose leaves open the possibility that different doses rather than types of the phloem-related DOF genes are responsible for the phenotype. However, we do not favour this scenario because of the functional differences of the DOF genes based on the overexpression phenotypes.

Since phloem transports nutrients, and smaller amounts of sucrose and its component fructose are detected in both mutant leaves and roots at 3 dps when root anatomy is comparable between WT and mutant seedlings, we interpret that *PAPL* genes regulate nutrient allocation between the leaf source organs and the root sink in young seedlings, when embryo reserves are scarce. *PAPL* genes could either regulate phloem loading, long-distance transport or

phloem function, and more studies are required to determine their precise role.

Methods

Plant growth conditions. All *Arabidopsis thaliana* lines used in this study were in Col-0 background except for the *pear1* mutant allele, which is in Ler background, conferring the *pear1pear2* mutant a mixed Ler appearance. Plants were grown in ½ MS basal salts media (0.5 MS salts, 1% Difco agar, with or without 1% sucrose) at 23 °C and long-day conditions, except for sorting experiments where they were grown using 1x MS basal salts at 23 °C with 30% humidity and 188 µM of light, long-day conditions, for comparability with other transcriptomic data.

papl1-1 (*cog1-6*, from gene *At1g29160*, *PAPL1*, *DOF1.5*, *COG1*) has a single nucleotide deletion (G) at position +85, which generates a premature stop codon. This mutant was identified as a *cog1-D* suppressor³². *papl2* (*At2g34140*, *PAPL2*, *DOF2.3*, *CDF4*) has a 4 bp deletion (CAAG) at position +99, creating a premature stop codon. The *cdf2* T-DNA allele (GK782H09) is a knockdown allele⁵⁵. Triple mutant was obtained by crossing the double mutant *papl1-1papl2* to *cdf2-2*⁵⁵, selecting for mutant *3papl* and homozygous wild type alleles for all other genes. The second triple mutant (*3papl-2*) was obtained by generating a new *cdf2* allele by using CRISPR/Cas9 technology directly on the double mutant *papl1-1papl2*. The process rendered a 5 bp deletion (CCCGG) at position +953 (*cdf2-2*), which generated a premature stop codon shortly afterwards.

Beta estradiol (5 µM) or 10 µM dexamethasone was used in the inducible constructs for the indicated times. Plants induced with dexamethasone were treated for 24 h.

Sorting and single-cell sequencing. Seedlings from the different marker lines were grown vertically over mesh (Normesh, 100 µm) for 5 d in the conditions specified above. Approximately one-third of the root including the root tip was chopped with razor blades and the tissue transferred to a 70 µm strainer submerged in 7 ml of the protoplasting solution for 1 h with gentle shaking at room temperature⁵⁸. In the case of the sample 'MAKR5 enriched in root tips', we submerged the root tips of intact roots in Eppendorf tubes containing the protoplasting solution for 15 min, which is enough time for the meristems to be enzymatically cut from the rest of the root. Then the separated root tips were transferred to 70 µm strainers, incubated with 7 ml of protoplasting solution in 4-cm-radius Petri dishes at room temperature for 45 min and from then onwards were treated as the other samples. Washed protoplasts suspended in solution A were taken at room temperature to the sorting facilities and the process from chopping to sorting took approximately 2–2.5 h. For the gating, a wild-type Col-0 sample was first run to establish the fluorescent negative gate. Then this sample was subsequently stained with DAPI and DRAQ5 to gate for intact cells that contained DNA. The corresponding sample containing fluorescent protoplasts was then stained subsequently with DAPI and DRAQ5 and underwent FACS. Gating helped enrich intact (DAPI negative), YFP/GFP positive, DNA containing cells (DRAQ5 positive) that were sorted with a 130 µm nozzle using a high-speed influx cell sorter (BD Biosciences). Sorted protoplasts were collected in W5 solution (154 mM NaCl, 125 mM CaCl₂, 5 mM KCl, 5 mM MES (2-(N-morpholino)ethanesulfonic acid)) in BSA-coated 1.5 ml Eppendorf tubes. Cells were centrifuged for 12 min at 200 g to eliminate excess supernatant. Immediately, single-cell RNA-seq libraries were prepared in the Cancer Research UK Cambridge Institute Genomics Core Facility using the following: Chromium Single Cell 3' Library & Gel Bead Kit v3, Chromium Chip B Kit and Chromium Single Cell 3' Reagent Kits v3 User Guide (Manual Part CG000183 Rev C; 10X Genomics). Cell suspensions were loaded on the Chromium instrument with the expectation of collecting gel-bead emulsions containing single cells. RNA from the barcoded cells for each sample was subsequently reverse-transcribed in a C1000 Touch thermal cycler (Bio-Rad) and all subsequent steps to generate single-cell libraries were performed according to the manufacturer's protocol with no modifications. Complementary DNA quality and quantity were measured with Agilent TapeStation 4200 (High Sensitivity 5000 ScreenTape) after which 25% of material was used for gene expression library preparation.

Library quality was confirmed with Agilent TapeStation 4200 (High Sensitivity D1000 ScreenTape to evaluate library sizes) and Qubit 4.0 fluorometer (ThermoFisher Qubit dsDNA HS assay kit to evaluate double-stranded DNA quantity). Each sample was normalized and pooled in equal molar concentration. The concentration of the pool was confirmed with quantitative PCR using KAPA library quantification kit on QuantStudio 6 Flex before sequencing. All samples were sequenced using Illumina NovaSeq6000 sequencer with the following parameters: 28 bp, read 1; 8 bp, i7 index; and 91 bp, read 2.

Analysis of single-cell RNA-seq. Here we give a briefer description and overview of our analysis steps, but the full details of our analysis pipeline (for example, specific package functions and options used) can be seen in our code repository at https://github.com/tavareshugo/publication_Otero2021_PhloemPoleAtlas.

To obtain unique molecular identifier (UMI) counts for each gene, the raw sequencing reads were aligned to the reference *Arabidopsis* TAIR10 genome using the Araport11 gene annotation (both downloaded from Ensembl release 45) using 10x Genomics Cell Ranger v3.1.0⁵⁹. The data were processed and quality-filtered using several Bioconductor packages⁶⁰. Empty droplets were inferred and removed

using dropletutils v1.8.0⁶¹, and data normalization was done using both the pooling method implemented in scran v1.16.0⁶² and the variance-stabilizing transformation from scctransform v0.2³⁴. To adjust for potential batch effects, data from the different samples (that is, sorted with different GFP fusion markers and/or from different public datasets) were integrated using the mutual nearest neighbours (MNN) algorithm implemented in batchelor v1.4.0³⁵. After initial data exploration and quality checks, we retained cells with at least 2,000 detected genes and genes detected in at least 100 cells (a gene was considered to be detected if it had at least 1 UMI count). Downstream analysis was done on these filtered data, which were then batch-normalized using MNN and variance-stabilized transformed values. However, our conclusions were qualitatively robust to the specific choice of normalization methods. For data visualization purposes, we projected the data to two dimensions using uniform manifold approximation and projection (UMAP), using a neighbourhood size of 30 cells (sizes of 7, 15 and 100 were also explored and gave comparable results). We also visualized the UMAP in three dimensions, which did not provide further insights into the data compared with the two-dimensional projection.

Cell clustering was performed by first defining a ‘shared nearest neighbours’ graph and then identifying modules in the graph using the Louvain algorithm (using scran v1.16.0⁶² and igraph v1.2.6⁶³). To annotate our cells, we used a set of genes with known expression patterns (from promoter fusion microscopy experiments) and calculated, for each cluster, the percentage of cells where each marker gene was detected as well as the (*z*-score-scaled) average expression of the gene in that cluster.

To identify cluster-specific genes, we used pairwise Wilcoxon rank-sum tests between a given target cluster and all others using the ‘findMarkers()’ function in the ‘R/Bioconductor’ package scran v1.16.0⁶¹. We specifically tested for genes upregulated in the target cluster to identify highly expressed genes specific to each cluster (rather than also including genes that are specifically absent from the cluster). The results of the pairwise tests for a given target cluster were then consolidated to obtain a summary *P* value (and corrected FDR) for how enriched each gene is in a given cluster. We summarized the pairwise *P* values for a null hypothesis that the gene is not differentially expressed in at least 8 out of the 15 clusters, allowing us to flexibly identify genes that were highly expressed across multiple cell types (for example, mature ring cells such as PPP and CC) but not others. We also did a more stringent summary of *P* values (null of no differential expression in 12/15 clusters) to obtain genes more specific to particular clusters of interest (namely cluster 10, which was a candidate for early MSE cells).

The same pipeline was applied to the public datasets, also integrated using MNN. The quality of this data integration was confirmed by checking that the majority of our annotated cells were clustering together with the same cell types in other datasets. We produced two sets of data integration, one with root data and another with leaf data. Details of the public datasets used are given in Supplementary Table 6.

To explore how well cells from leaf and root datasets mixed in clusters where they co-occurred (namely cluster 6, which contained both leaf phloem parenchyma and root phloem pole pericycle cells), we used the same shared-nearest-neighbours cell graph used for clustering and calculated the proportion of edges between root-leaf cells (the vertices of the graph). This value was then compared with a null expectation, obtained by randomly shuffling the cell tissue labels 1,000 times and calculating this proportion each time. The 95% inter-percentile range of this null distribution was then used to compare with the observed value. The graph had 19.2% leaf-root edges in this cluster, which is only slightly lower than expected by chance (median 23.5%, 95% CI 22.8%–24.2%, obtained from 1,000 random shuffles of the cell labels). This result suggests that the cells from the two datasets are well mixed. This contrasts with cluster 17, for example, which consists of poorly clustered cells that occur separated in the UMAP. In this case, there were 26.7% leaf-root edges, almost half of the null expectation for that cluster (44.9%).

To further temporally annotate our phloem pole atlas dataset, we used several approaches. Early dividing cells were identified by checking the expression of all annotated cyclins and other cell-cycle markers such as *AUR1* (AT4G32830) and *KNOLLE* (AT1G08560). We also cross-referenced our data with a published dataset that profiled the transcriptome of longitudinal root sections using microarray technology¹. On the basis of 9,674 common genes between the two datasets, we assigned each of our cells to the longitudinal section of Brady et al.¹ so that they had the highest Spearman correlation. We also used the RNA velocity method implemented in scVelo v0.2.2 to infer developmental dynamics in our data¹³. Finally, cells were assigned to lineages and ordered by pseudotime using Slingshot v1.6.0³². We first reduced the dimensionality of the (batch-normalized) counts to 10 components using diffusion maps, which is a dimensionality reduction method suited to capture developmental transitions in the data^{64,65}. For the slingshot analysis, we used a semi-supervised approach where the starting point for the inferred trajectories was set to the cluster highly expressing cell-cycle markers and identified as the earliest cluster when cross-referencing with the Brady et al.¹ dataset. In this manner, we obtained biologically meaningful trajectories (without setting this constraint, several more trajectories were obtained but with an ordering of cells that was the reverse of what was expected from our other analyses). We obtained smooth gene expression patterns for each trajectory using generalized

additive models, as implemented in tradeSeq v1.2.0, which were then used to explore gene expression patterns along the Slingshot trajectories.

To cluster genes on the basis of their similarity of expression across the cells, we built a co-expression network using a modified version of bigScale⁴⁷ adapted to work on any species (rather than the original version suited only for mouse and human). The modified package is available from <https://github.com/tavareshugo/bigScale2/tree/support-any-species>. Briefly, bigScale builds a gene correlation matrix not from the original count data (which in scRNA-seq is too noisy and sparse) but from a *z*-score statistic calculated between pairs of cell clusters. These clusters are iteratively generated to ensure that the *z*-scores capture as much diversity in gene expression patterns across the cells as possible. In this way, correlations between genes are more robust to the noisy and sparse nature of single-cell RNA-seq data. This correlation matrix was then thresholded at 0.9 to obtain a gene-by-gene adjacency matrix, resulting in a network with 5,238 vertices (genes) and 370,794 edges. We identified gene modules using the Louvain algorithm, resulting in 16 modules. From each module, we calculated an eigengene following the procedure in WGCNA vX⁴⁹, which essentially summarizes the expression of all genes of a module as the first principal component score from a PCA done on those genes. The largest of these modules—module 1 containing 1,367 genes—contained several genes of interest for our analysis and was therefore reclustered with Louvain to generate 15 sub-modules. This was further justified by the fact that the variance explained by this module’s eigengene was relatively low (21.44%), suggesting some heterogeneity in expression patterns within the module. To further interpret these results, the eigengenes from these sub-modules were combined with the pseudotime trajectories from Slingshot, although we note that no information about trajectories was used to build the network itself. Therefore, the fact that the different approaches (gene network and pseudotime analysis) reveal groups of genes with similar patterns of expression strengthens our analysis.

Generation of reporter lines and confocal images. Promoter::VENUSer fusions were generated for the genes At1g27030, At2g23560 (*MES7*), At1g32210, At5g64240 (*MC3*), At5g58690 (*PLC5*), At1g38640, At3g11930, At2g02230 (*PP2-B1*), At5g47920, At4g27435, At5g59090 (*SBT4.12*), At3g21770 (*PER30*), At1g26450, At1g29160 (*PAPL1*, *DOF1.5*, *COG1*), At2g34140 (*PAPL2*, *DOF2.3*, *CDF4*) and At5g39660 (*DOF5.2*, *CDF2*).

Translational fusions were also generated for At2g20562 (*TAX2*), *PAPL1* and *PAPL2*. 3xYFP constructs were also generated for transcriptional fusions of *PAPL1* and translational fusions of *PAPL1* and *MAKR5*.

Promoter fragments between 622–4879 bp were amplified by PCR and cloned using MultiSite-Gateway (Supplementary Table 7). Transcriptional fusions to VENUS with an ER tag or translational fusions to YFP were generated in vectors with either resistance to Basta or Hygromycin or a Fast Green/Fast Red selection system. All the constructs were transformed in Col-0 background and at least 2 independent lines were analysed for each.

Roots from 5–7-day-old seedlings were either imaged in the confocal microscope directly after mounting them in 50 µg ml⁻¹ propidium iodide or fixed for 45 min in 4% paraformaldehyde in PBS and cleared using ClearSee solution (10% (w/v) Xylitol, 15% (w/v) sodium deoxycolate, 25% (w/v) urea, water to the final volume)⁶⁶. Cleared roots were then stained with SCRI Renaissance 2200 and observed under the confocal microscope. Images were acquired at 512 × 512 resolution using the confocal microscope Leica SP8. Images were analysed in ImageJ v2.1.0/1.53c.

Bulk RNA-seq transcriptomics. Wild-type and *papl1-1papl2* seedlings were grown on mesh in ½ MS media with sucrose in the above-mentioned conditions for 5 d. Root meristems from wild type and mutant were manually and individually dissected in parallel under a stereomicroscope using 18G needles. Meristems were preserved in RNAlater RNA stabilization reagent (Qiagen) until 120 meristems per replicate were gathered. Three replicates for each mutant and wild type were used for RNA extraction.

RNA was extracted using the RNeasy plant mini kit from Qiagen, and RNA integrity and concentration were checked using TapeStation and Qubit 2.0 fluorometer (Life Technologies), respectively. After quality control in Novogene, the best 3 replicates for mutant and wild type were used for library construction and sequencing following the Novogene pipeline.

Briefly, mRNA was enriched using oligo dT beads and fragmented randomly. cDNA synthesis was performed using random hexamers and reverse transcriptase. After first-strand synthesis, the second strand was synthesized by nick-translation. Library was ready after a round of purification, terminal repair, A-tailing, ligation of sequencing adapters, size selection and PCR enrichment. Library concentration was quantified using a Qubit 2.0 fluorometer (Life Technologies), insert size was checked on Agilent 2100 and quantified more accurately by qPCR. Libraries were fed into the HiSeq XTEN platform for sequencing. Original raw data were transformed to sequence reads by base calling and raw data recorded in FastQ files. Low-quality reads or reads containing adapters were filtered out. TopHat2⁶⁷ v2.0.12 was used to map the reads to the reference genome (TAIR10). HTSeq⁶⁸ v0.6.1 software was used to analyse the gene expression level using the union mode. Fragments per kilobase of transcript sequence per Millions base pair sequenced (FPKM) value of 0.1 or 1 was set as the threshold to determine whether a gene was

expressed or not. To compare gene expression levels under different conditions, FPKM distribution diagrams and violin plots were used. For biological replicates, the final FPKM would be the mean value. The differential gene expression analysis consisted of read-count normalization, model-dependent mean value estimation and FDR value estimation based on multiple hypotheses testing. DESeq⁶⁹ v1.10.1 software was used for these steps. The results of this analysis are given in Supplementary Table 4.

Root length measurements and statistical analysis. To quantify root growth, an EPSON Perfection V700 photo scanner was used to obtain images of the seedlings in plates. A ruler was also scanned to calibrate the images. Roots were measured manually one by one using ImageJ v2.1.0/1.53c.

Because of the nature of the data, which often had a bimodal distribution, we opted to use a non-parametric bootstrap approach for our statistical analysis. This was done by resampling the data 500 times and estimating the difference between groups of interest (either WT vs mutants or between sucrose treatments, as detailed in the respective figure legends). We thus obtained distributions of root length differences, which we used to obtain confidence intervals (based on a 95% inter-percentile range) and a bootstrap *P* value calculated as the number of samples with absolute difference less than a ‘null’ distribution centred on zero. Our *P* values therefore have a lower bound of 1/501~0.002, which we deemed to be of sufficient statistical resolution for our analyses (we added an offset of 1 to both the numerator and the denominator to avoid *P*=0, which would mis-represent the precision of our analysis). Whenever relevant, the bootstrap analysis took into account experimental and seed stock batches by summarizing the results at those levels first, before comparing the groups of interest; this ensured that the uncertainty in our estimates captured these different levels of potential variation. The results of these analyses are provided in Supplementary Table 8.

Experimental design. Experiments were repeated independently for the following number of times. In Fig. 1d, MES7 reporter was imaged 3 times and reporters for MC3, At3g27030 and At2g32210 (also in Fig. 2b) were imaged twice.

For Fig. 6b, reporters for At3g11930, At5g47920 and At1g26450 were imaged twice, PP2-B1, SBT4.12 and TAX2 reporters were imaged 3 times. For Fig. 7, pPAPL1::GFP was imaged 6 times, pPAPL1::GFP in pear1pear2 was imaged 3 times, pPAPL1::PAPL1-YFP (also in Extended Data Fig. 7b) was imaged twice, pPAPL1::GFP in pPEAR1::icasl3m was imaged twice and pPAPL2::YFPer in PRPS5A::PEAR1-GR was imaged once.

For Extended Data Fig. 1a, pMAKR5::MAKR5-3xYFP was imaged 7 times, pS17::GFP was imaged twice, pAPL::3xYFP was imaged 3 times and pSAPL::YFPer (also in Fig. 2b) was imaged 4 times. For Extended Data Fig. 1b, the reporters for PLC5 and At2g38640 were imaged 3 times.

For Extended Data Fig. 2d, pAPL::YFPer was imaged once. For Extended Data Fig. 5, reporters for At3g16330 and PER30 were imaged 3 times, while the reporter for At4g27435 was imaged 4 times.

For Extended Data Fig. 7, pPAPL::3xYFP was imaged 6 times, pPAPL::PAPL1-3xYFP was imaged 3 times and this construct in *pear sext* background was imaged 3 times. pPAPL2::PAPL2-YFPer was imaged 3 times, pPAPL2::VENUSer was imaged 3 times and this construct in *pear1pear2* mutant background was imaged twice. pCDF2::VENUSer was imaged 3 times. For constructs in Extended Data Fig. 7j, reporters in the overexpressor background, induced and control, were imaged once, while reporters were imaged twice in *pear sext* background. For Extended Data Fig. 8, each line was observed independently twice. For Extended Data Fig. 9d, roots were imaged for this figure once, but these two backgrounds were imaged many times with reporter lines in them.

For Extended Data Fig. 10a, reporters with and without sucrose were imaged twice. For Fig 10b, pMAKR5::MAKR5-3xYFP in 3papl mutant was imaged 3 times and it was also imaged 3 times in *pear1pear2* mutant background. The reporter in Extended Data Fig. 10c was imaged 3 times, while the reporter in Extended Data Fig. 10d was imaged once. Reporters in Extended Data Fig. 10e-g were imaged twice and reporters for Extended Data Fig. 10h were imaged once (Col-0) or three times (mutant background).

For the experiment shown in Fig. 7g and Extended Data Fig. 9e, the total number of seedlings measured for each genotype was: 488 3papl; 273 PAPL1-32; 37 PAPL1-51; 33 PAPL1-71; 343 PAPL2-11; 79 PAPL2-23; 37 PAPL2-31; 314 PAPL2-73; and 382 WT. Seedlings were split across 5 experimental batches and came from different seed stocks ($N=24\text{--}46$, with a median of 36 seedlings per experimental batch and seed stock combination). Both seed stock and experimental batch were taken into account in the statistical analysis.

Metabolic profiling. *Arabidopsis* plants were grown across 6 developmental stages (from days 2 to 7) on mesh in solid media with or without sucrose. Each day of the time course, leaves and roots were collected separately and snap frozen in liquid nitrogen. Leaves (50 mg) and roots (20 mg) were ground using a tissue lyser. Extraction was performed as previously described⁷⁰, with modifications. In detail, 750 µl per 300 µl of extraction buffer (100% methanol plus the internal standard adonitol; Sigma) was added to root and leaf samples. Samples were vortexed and transferred to a shaker at 70 °C for 15 min. Then, 375 µl per 200 µl of chloroform and 750 µl per 350 µl of water were added to the tubes for leaves and

roots, respectively, and samples were centrifuged for 10 min at maximum speed. Supernatants (400 µl for roots and 200 µl for leaves) were dried for each sample using a speedvac. Samples were kept at -80 °C until processing.

The dried samples were derivatized for 2 h at 37 °C in 50 µl of 20 mg ml⁻¹ methoxyamine hydrochloride (Sigma-Aldrich, 593-56-6) in pyridine, followed by a 30 min treatment at 37 °C with 100 µl of N-methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA reagent; Macherey-Nagel, 24589-78-4). For each sample, 1 µl was injected in splitless mode to a chromatograph coupled to a time-of-flight mass spectrometer system (Leco Pegasus HT TOF-MS; Leco), using an autosampler Gerstel multi-purpose system (Gerstel). Chromatograms and mass spectra evaluation, as well primary metabolites identification based on the expected retention time and mass fragmentation were performed using the Xcalibur software (Thermo Fisher). The software ChromaTof (Leco) was used to confirm the peaks and retention times for expected metabolite fragments.

To estimate differences between WT and 3papl metabolite levels, we fit a joint hierarchical model to the peak areas of all metabolites, including terms for genotype, stage (dps), tissue and their interactions. The advantage of using this model is that we could include a random effect term to account for multiple measurements per sample (each sample contributed 21 data points, one for each metabolite). Due to the skewed distribution of peak areas, the data were modelled on a log-scale, which produced well-behaved normally distributed residuals. The model was fit with the ‘lme4’⁷¹ v1.1-27.1 R package and we obtained estimates of the difference between the two genotypes for each metabolite and tissue using ‘emmeans’ v1.6.2-1 R package. The *P* values from the ‘emmeans’ contrasts were corrected for multiple testing using the FDR method.

Additional information on metabolomics analysis and metabolites annotation are reported in Supplementary Table 9 (sheets checklist and overview) according to previously published guidelines⁷².

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Sequencing data from 10x Chromium single-cell RNA-seq are available from NCBI’s Gene Expression Omnibus through GEO accession number GSE181999. Sequencing data from bulk RNA-seq are available from NCBI’s GEO accession number GSE182672. All other data (phenotypic scoring, microscopy imaging and plasmid maps) are available from the Cambridge Apollo Repository (<https://doi.org/10.17863/CAM.74836>).

Code availability

Analysis code, with instructions on how to run it, is available from: https://github.com/tavareshugo/publication_Otero2022_PhloemPoleAtlas.

Received: 7 July 2021; Accepted: 24 May 2022;

Published online: 4 August 2022

References

- Brady, S. M. et al. A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science* **318**, 801–806 (2007).
- Miyashima, S. et al. Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature* **565**, 490–494 (2019).
- Depuydt, S. et al. Suppression of *Arabidopsis* protophloem differentiation and root meristem growth by CLE45 requires the receptor-like kinase BAM3. *Proc. Natl Acad. Sci. USA* **110**, 7074–7079 (2013).
- Bonke, M., Thitamadee, S., Mähönen, A. P., Hauser, M.-T. & Helariutta, Y. APL regulates vascular tissue identity in *Arabidopsis*. *Nature* **426**, 181–186 (2003).
- Furuta, K. M. et al. *Arabidopsis* NAC45/86 direct sieve element morphogenesis culminating in enculation. *Science* **345**, 933–937 (2014).
- Dettmer, J. et al. CHOLINE TRANSPORTER-LIKE1 is required for sieve plate development to mediate long-distance cell-to-cell communication. *Nat. Commun.* **5**, 4276 (2014).
- Truernit, E., Bauby, H., Belcram, K., Barthelemy, J. & Palauqui, J.-C. OCTOPUS, a polarly localised membrane-associated protein, regulates phloem differentiation entry in *Arabidopsis thaliana*. *Development* **139**, 1306–1315 (2012).
- Rodríguez-Villalon, A. et al. Molecular genetic framework for protophloem formation. *Proc. Natl Acad. Sci. USA* **111**, 11551–11556 (2014).
- Scacchi, E. et al. Spatio-temporal sequence of cross-regulatory events in root meristem growth. *Proc. Natl Acad. Sci. USA* **107**, 22734–22739 (2010).
- Xiao, Y. & Offringa, R. PDK1 regulates auxin transport and *Arabidopsis* vascular development through AGC1 kinase PAX. *Nat. Plants* <https://doi.org/10.1038/s41477-020-0650-2> (2020).
- Marhava, P. et al. A molecular rheostat adjusts auxin flux to promote root protophloem differentiation. *Nature* **558**, 297–300 (2018).
- Ross-Elliott, T. J. et al. Phloem unloading in *Arabidopsis* roots is convective and regulated by the phloem-pole pericycle. *eLife* **6**, e24125 (2017).

13. Mähönen, A. P. et al. A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root. *Genes Dev.* **14**, 2938–2943 (2000).
14. Graeff, M. & Hardtke, C. S. Metaphloem development in the *Arabidopsis* root tip. *Development* **18**, dev199766 (2021).
15. Sola, M. A. R. et al. OCTOPUS-LIKE 2, a novel player in *Arabidopsis* root and vascular development, reveals a key role for OCTOPUS family genes in root metaphloem sieve tube differentiation. *New Phytol.* **216**, 1191–1204 (2017).
16. van Bel, A. J. E., Ehlers, K. & Knoblauch, M. Sieve elements caught in the act. *Trends Plant Sci.* **7**, 126–132 (2002).
17. Tamaki, T. et al. VISUAL-CC system uncovers the role of GSK3 as an orchestrator of vascular cell type ratio in plants. *Commun. Biol.* **3**, 184 (2020).
18. Gujas, B. et al. A reservoir of pluripotent phloem cells safeguards the linear developmental trajectory of protophloem sieve elements. *Curr. Biol.* **30**, 755–766.e4 (2020).
19. Otero, S. & Helariutta, Y. Companion cells: a diamond in the rough. *J. Exp. Bot.* **68**, 71–78 (2017).
20. Oparka, K. J. & Turgeon, R. Sieve elements and companion cells: traffic control centers of the phloem. *Plant Cell* **11**, 739–750 (1999).
21. Parizot, B., Roberts, I., Raes, J., Beeckman, T. & De Smet, I. In silico analyses of pericycle cell populations reinforce their relation with associated vasculature in *Arabidopsis*. *Phil. Trans. R. Soc. B* **367**, 1479–1488 (2012).
22. Beeckman, T. & Smet, I. D. Pericycle. *Curr. Biol.* **24**, R378–R379 (2014).
23. Lee, J.-Y. et al. Transcriptional and posttranscriptional regulation of transcription factor expression in *Arabidopsis* roots. *Proc. Natl Acad. Sci. USA* **103**, 6055–6060 (2006).
24. Parizot, B. et al. Diarch symmetry of the vascular bundle in *Arabidopsis* root encompasses the pericycle and is reflected in distich lateral root initiation. *Plant Physiol.* **146**, 140–148 (2008).
25. Li, S., Yamada, M., Han, X., Ohler, U. & Benfey, P. N. High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. *Dev. Cell* **39**, 508–522 (2016).
26. Zhang, C., Barthelson, R. A., Lambert, G. M. & Galbraith, D. W. Global characterization of cell-specific gene expression through fluorescence-activated sorting of nuclei. *Plant Physiol.* **147**, 30–40 (2008).
27. Birnbaum, K. et al. A gene expression map of the *Arabidopsis* root. *Science* **302**, 1956–1960 (2003).
28. Ryu, K. H., Huang, L., Kang, H. M. & Schiefelbein, J. Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol.* **179**, 1444–1456 (2019).
29. Shulze, C. N. et al. High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* **27**, 2241–2247.e4 (2019).
30. Denyer, T. et al. Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell* **48**, 840–852.e5 (2019).
31. Zhang, T.-Q., Xu, Z.-G., Shang, G.-D. & Wang, J.-W. A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. *Mol. Plant* **12**, 648–660 (2019).
32. Wendrich, J. R. et al. Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science* <https://doi.org/10.1126/science.aaq4970> (2020).
33. Roszak, P. et al. Cell-by-cell dissection of phloem development links a maturation gradient to cell specialization. *Science* **374**, eaba5531 (2021).
34. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
35. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018).
36. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* <https://f1000research.com/articles/5-2122/v2> (2016).
37. Desvoyes, B. et al. Looking at plant cell cycle from the chromatin window. *Front. Plant Sci.* **5**, 369 (2014).
38. Schürholz, A.-K. et al. A comprehensive toolkit for inducible, cell type-specific gene expression in *Arabidopsis*. *Plant Physiol.* **178**, 40–53 (2018).
39. Tian, H. et al. *Arabidopsis* NPCC6/NaKR1 is a phloem mobile metal binding protein necessary for phloem function and root meristem maintenance. *Plant Cell* **22**, 3963–3979 (2010).
40. Truernit, E. & Sauer, N. The promoter of the *Arabidopsis thaliana* SUC2 sucrose-H⁺ symporter gene directs expression of β-glucuronidase to the phloem: evidence for phloem loading and unloading by SUC2. *Planta* **196**, 564–570 (1995).
41. Zhang, Q. et al. Knock-dDown of *Arabidopsis* PLC5 reduces primary root growth and secondary root formation while overexpression improves drought tolerance and causes stunted root hair growth. *Plant Cell Physiol.* **59**, 2004–2019 (2018).
42. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
43. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
44. Shaham, R. et al. A single cell *Arabidopsis* root atlas reveals developmental trajectories in wild type and cell identity mutants. *Dev. Cell* **57**, 4 (2022).
45. Kim, J.-Y. et al. Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* <https://doi.org/10.1093/plcell/koaa060> (2021).
46. Haritatos, E., Medville, R. & Turgeon, R. Minor vein structure and sugar transport in *Arabidopsis thaliana*. *Planta* **211**, 105–111 (2000).
47. Iacono, G., Massoni-Badosa, R. & Heyn, H. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol.* **20**, 110 (2019).
48. Davuluri, R. V. et al. AGRIS: *Arabidopsis* Gene Regulatory Information Server, an information resource of *Arabidopsis* cis-regulatory elements and transcription factors. *BMC Bioinformatics* **4**, 25 (2003).
49. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
50. Pi, L. et al. Organizer-derived WOX5 signal maintains root columella stem cells through chromatin-mediated repression of CDF4 expression. *Dev. Cell* **33**, 576–588 (2015).
51. Park, D. H. et al. The *Arabidopsis* COG1 gene encodes a Dof domain transcription factor and negatively regulates phytochrome signaling. *Plant J.* **34**, 161–171 (2003).
52. Wei, Z. et al. Brassinosteroid biosynthesis is modulated via a transcription factor cascade of COG1, PIF4, and PIF5. *Plant Physiol.* **174**, 1260–1273 (2017).
53. Bueso, E. et al. *Arabidopsis* COGWHEEL1 links light perception and gibberellins with seed tolerance to deterioration. *Plant J.* **87**, 583–596 (2016).
54. Renard, J. et al. PRX2 and PRX25, peroxidases regulated by COG1, are involved in seed longevity in *Arabidopsis*. *Plant Cell Environ.* **43**, 315–326 (2020).
55. Fornara, F. et al. *Arabidopsis* DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Dev. Cell* **17**, 75–86 (2009).
56. Allen, E. et al. Correlation Network Analysis reveals a sequential reorganization of metabolic and transcriptional states during germination and gene-metabolite relationships in developing seedlings of *Arabidopsis*. *BMC Syst. Biol.* **4**, 62 (2010).
57. Vidal, E. A., Moyano, T. C., Canales, J. & Gutiérrez, R. A. Nitrogen control of developmental phase transitions in *Arabidopsis thaliana*. *J. Exp. Bot.* **65**, 5611–5618 (2014).
58. Birnbaum, K. et al. Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nat. Methods* **2**, 615–619 (2005).
59. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
60. Amezquita, R. A. et al. Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* **17**, 137–145 (2020).
61. Griffiths, J. A., Richard, A. C., Bach, K., Lun, A. T. L. & Marioni, J. C. Detection and removal of barcode swapping in single-cell RNA-seq data. *Nat. Commun.* **9**, 2667 (2018).
62. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* **5**, 2122 (2016).
63. Csardi, G. & Nepusz, T. The igraph software package for complex network research (InterJournal Complex Systems, 2006).
64. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
65. Angerer, P. et al. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **8**, 1241–1243 (2016).
66. Ursache, R., Andersen, T. G., Marhavý, P. & Geldner, N. A protocol for combining fluorescent proteins with histological stains for diverse cell wall components. *Plant J.* **93**, 399–412 (2018).
67. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
68. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
69. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
70. Liseć, J., Schauer, N., Kopka, J., Willmitzer, L. & Fernie, A. R. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.* **1**, 387–396 (2006).
71. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
72. Alsekh, S. et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nat. Methods* <https://doi.org/10.1038/s41592-021-01197-1> (2021).
73. Kang, Y. H. & Hardtke, C. S. *Arabidopsis* MAKR5 is a positive effector of BAM3-dependent CLE45 signaling. *EMBO Rep.* **17**, 1145–1154 (2016).

Acknowledgements

We thank W. Frommer and J.-Y. Kim for providing *pSWEET11:SWEET11-2A-GFP* seeds; C. Hardtke for providing *pMAKR5:MAKR5-GFP* seeds and plasmids; C. Cossetti, R. Schulte and all the staff from Flow Cytometry Core Facility at CIMR for their technical support with cell sorting; K. Kania from Genomics Unit at CRUK for preparing single-cell RNA-seq libraries; B. Guillotin and K. Birnbaum for helpful insights on single-cell analysis; R. Wightman and G. Evans for technical support with microscopy experiments; G. Hindle, J. Salmon and S. Ward for media preparation; K. Blajecka for technical assistance; K. Petkovic and R. Alcaina for technical support; and S. Schornack for helpful comments. S.O. was supported by a Herchel Smith postdoctoral fellowship from the University of Cambridge (2017–2020). L.K. received funding from the SNSF (P2LAP3_178062) and a Marie Curie IEF (No. 795250). This work was supported by the Finnish CoE in Molecular Biology of Primary Producers (Academy of Finland CoE programme 2014–2019) decision no. 271832, the Gatsby Foundation (GAT3395/PR3), the University of Helsinki (award 799992091) and the ERC Advanced Investigator Grant SYMDEV (No. 323052). T.L. was supported by the German Research foundation (DFG) under Germany's Excellence Strategy (CIBSS-EXC-2189-Project ID 390939984) and by grant La606/18-1. A.R.F and V.D.V. acknowledge support from the Max-Planck Society.

Author contributions

S.O. performed the experiments; I.G. identified *PAPL1* and *At3g16330* expression patterns which appeared as PEAR targets in microarray data; P.R. provided *pear1pear2* double mutant, *pPEAR(del)::3xYFP* and advised on experimental design; Y.L. and H.T. analysed gene regulatory networks; P.R., M.B., L.K., B.B. and J.H. participated in

sample collection for sorting and metabolomic profiling; M.B. imaged *pSUC2:GFP*; J.H. provided the *pSAPL::YFP* line; V.D.V. and A.R.F. carried out metabolic profiling and data analysis; F.P. and T.L. provided the *papl2* and *papl1-2* alleles; H.T. designed and performed the single-cell data and statistical analysis; S.O., H.T. and Y.H. conceptualized and designed the study; S.O. wrote the manuscript with input from Y.H., H.T., P.R. and L.K. All authors read, edited and discussed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41477-022-01178-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41477-022-01178-y>.

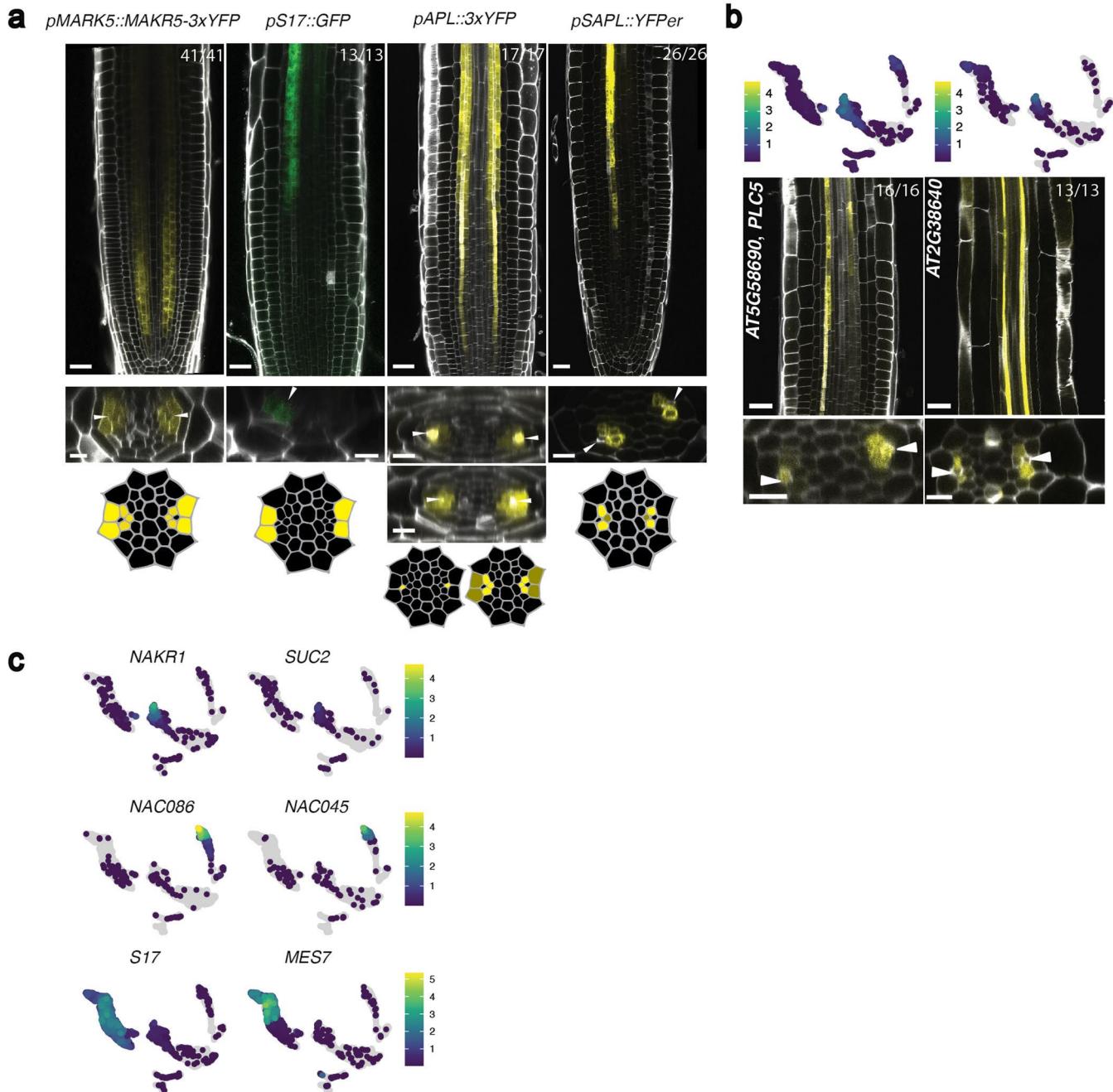
Correspondence and requests for materials should be addressed to Hugo Tavares or Yka Helariutta.

Peer review information *Nature Plants* thanks Jia-Wei Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

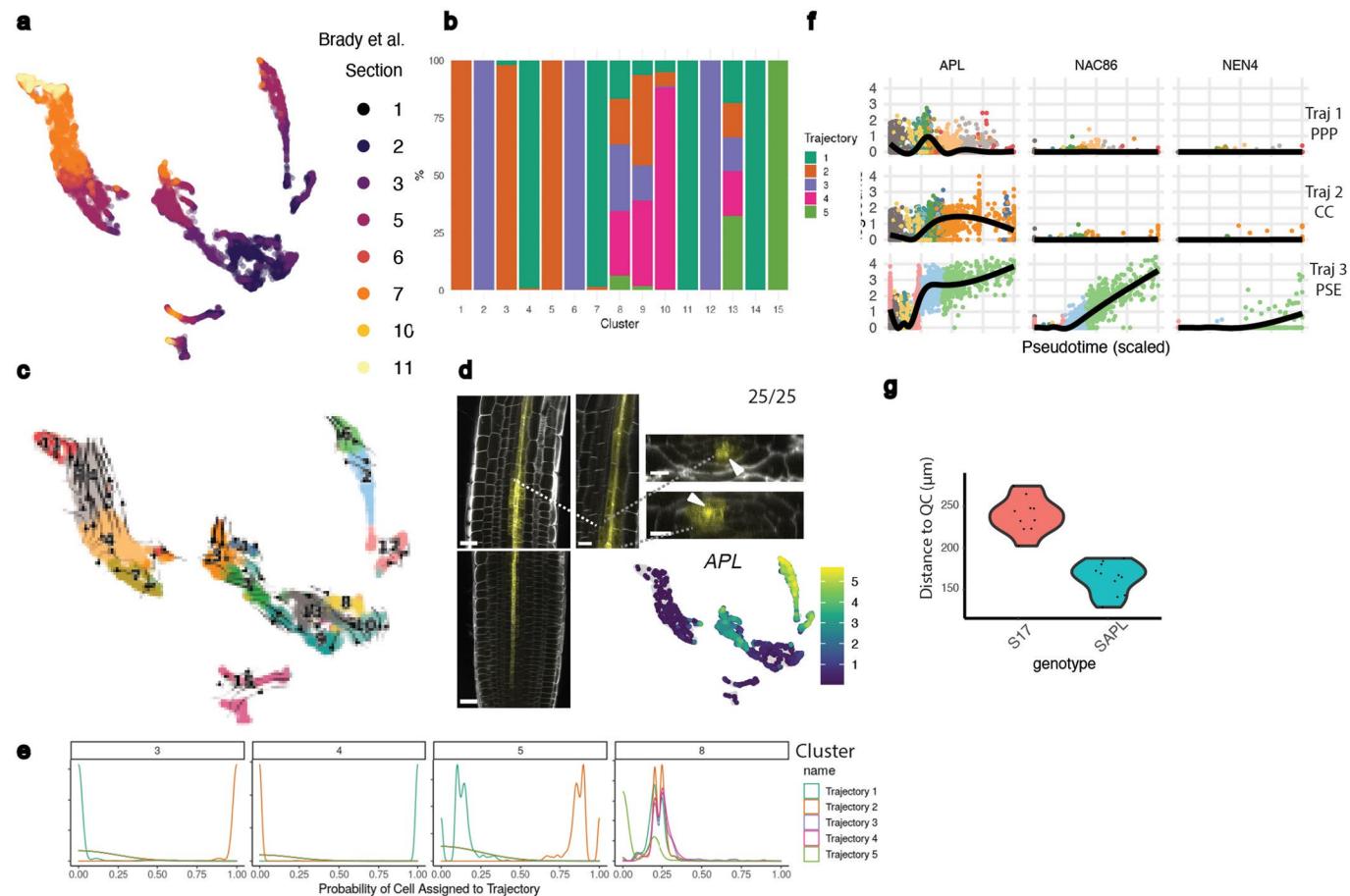
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

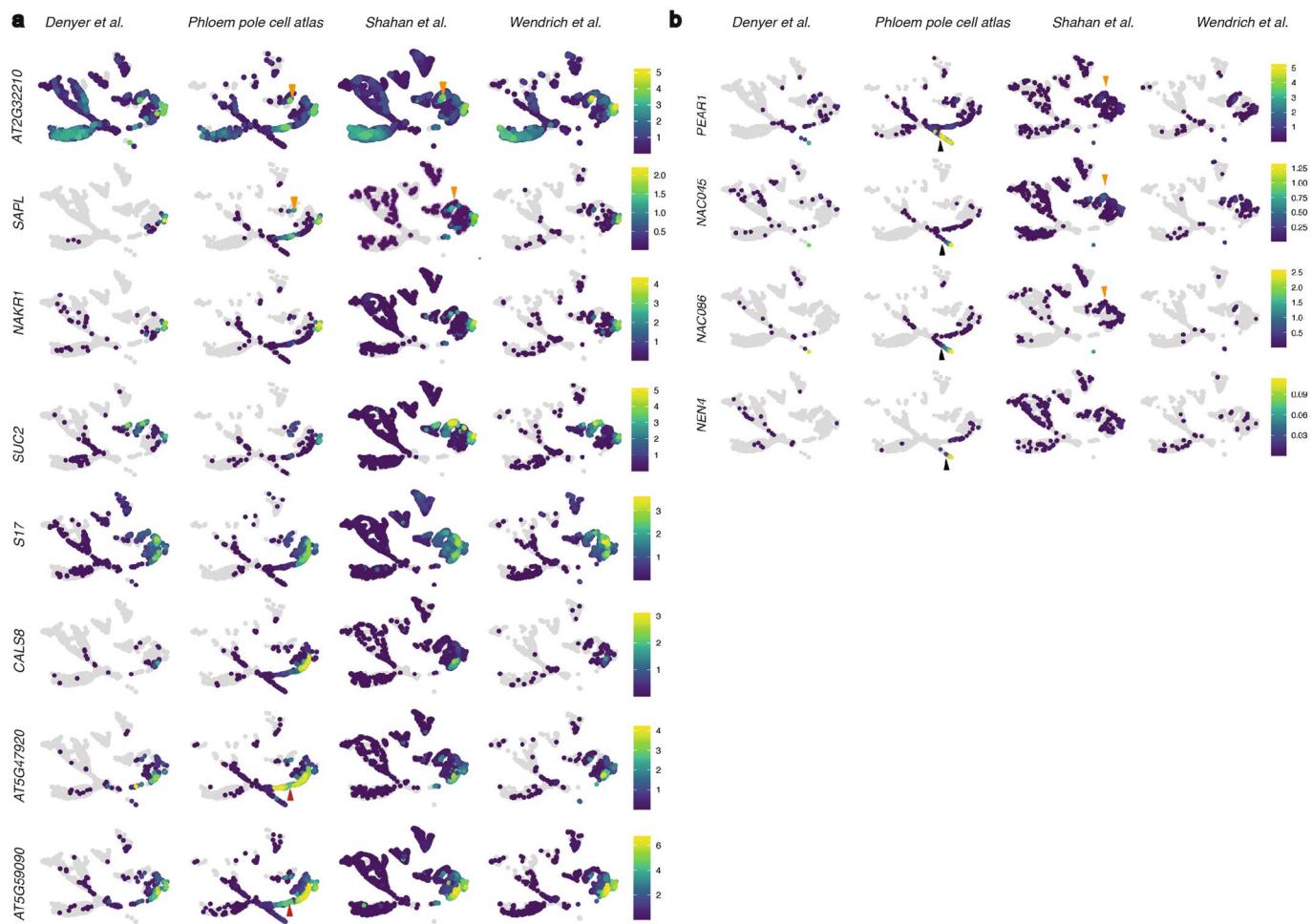
© The Author(s), under exclusive licence to Springer Nature Limited 2022



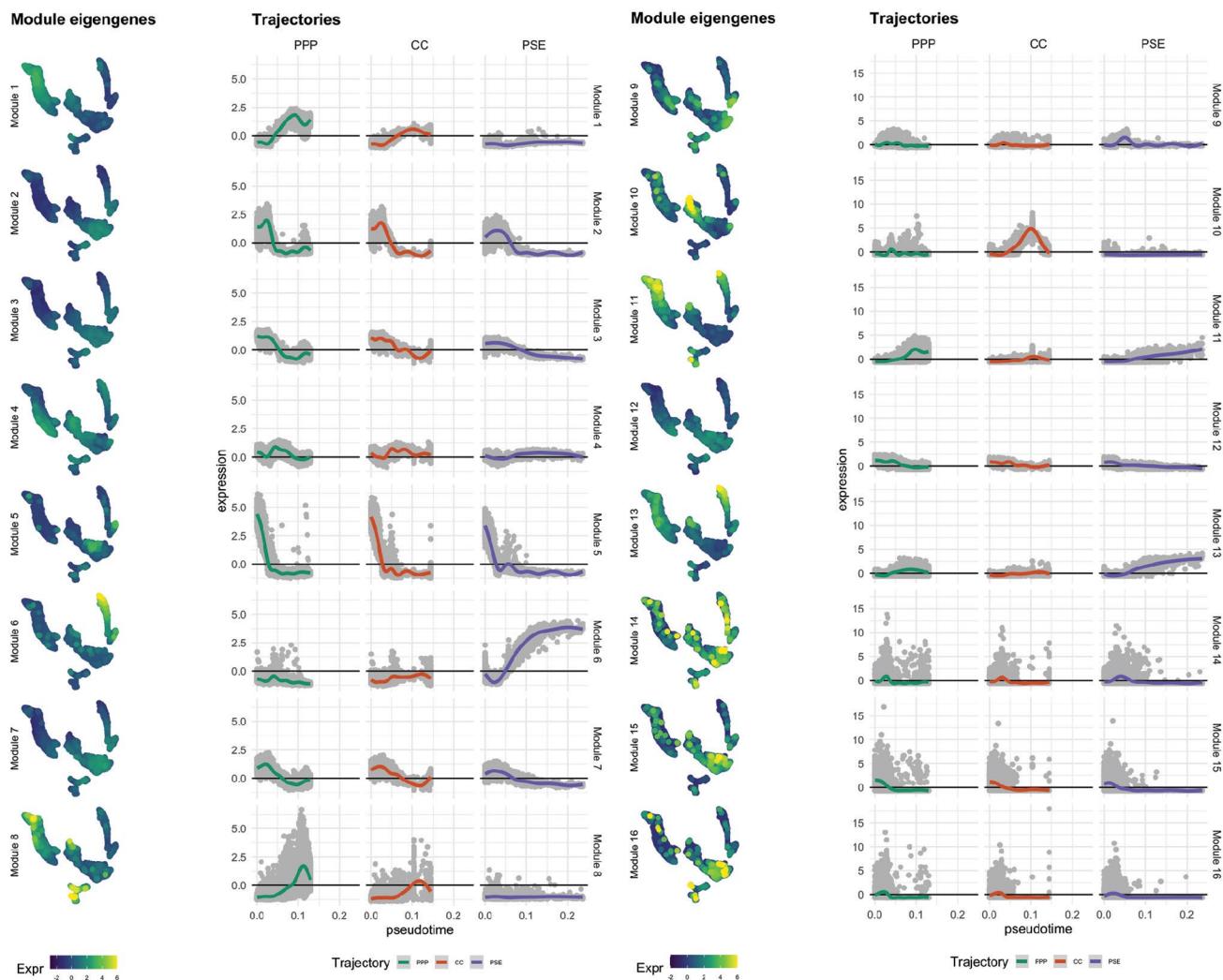
Extended Data Fig. 1 | Markers used for sorting phloem pole cells and new CC genes identified. **a)** *MAKR5* translational fusion highlights all the cells surrounding PSE⁷³ from the quiescent centre (QC) (often stronger from cell number 3) until the differentiation zone, where it becomes weaker. Some weak expression is also found in PSE. In mature parts of the root this marker spreads to the whole pericycle. Published marker was fused to 3xYFP to increase signal. *S17*¹ is expressed in PPP from the unloading zone. *pAPL::3xYFP* is expressed first in PSE and after enucleation switches to all the cells around PSE, stronger in CC. This line is not fully reflecting *ALTERED PHLOEM DEVELOPMENT* (*APL*) endogenous expression, since it has some expression in the outer layers but it is a very strong phloem pole marker. *sAPL* is expressed in CC and MSE (weaker) from 90–120 µm from the QC. *pPEAR1(del)::3xYFP* is a modified version of the *PEAR1* promoter that is expressed in early PSE, MSE and a procambial cell resulting from the same division plus columella cells. See Roszak et al.³³ for the detailed expression pattern **b)** Newly identified genes expressed in CC (*PLCS*, *At5g58690*, and *At2g38640*). **c)** Expression of mature CC (*NAKR1*, *SUC2*), mature PSE (*NAC086*, *NAC045*) and mature PPP (*S17*, *MES7*) genes at the terminal clusters. UMAPs show the particular cluster-weighted normalised expression of each gene in the phloem pole cell atlas and microscopy pictures are representative images of the transcriptional reporter lines where the gene promoter is fused to *VENUSer*. Scale bar in the longitudinal sections is 25 µm while it is 10 µm in the cross sections. White arrowheads point to PSE cells as a reference point. The numbers in each panel indicate samples with similar results, of the total independent biological samples observed.



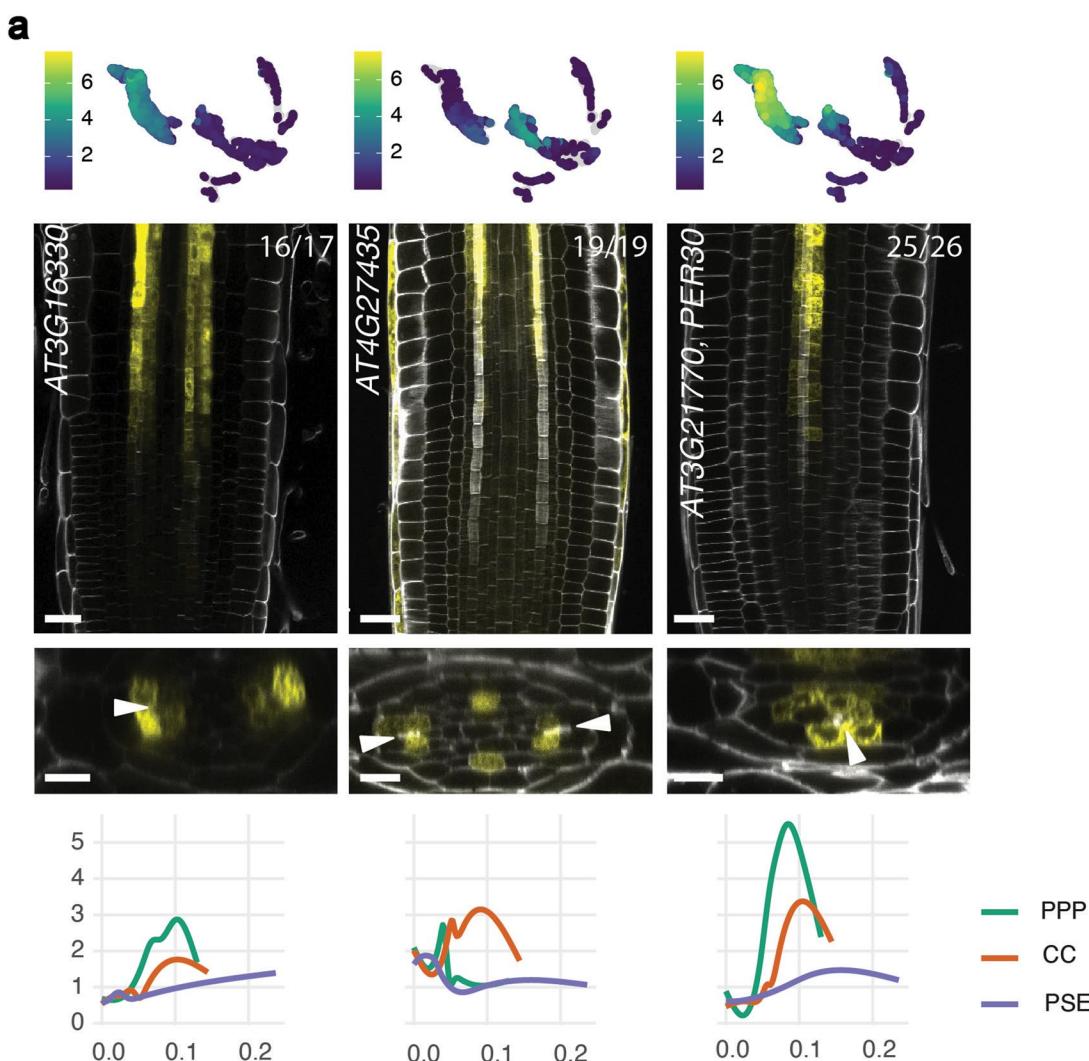
Extended Data Fig. 2 | Validation of the temporal information in the UMAP by using complementary tools. **a)** Color-coded UMAP according to the longitudinal sections in Brady et al. 2007, showing a developmental progression in each cell lineage. **b)** Bar plot indicating the percentage of cells contributed by each cluster to each of the Slingshot trajectories shown in the main text. Bars are coloured by trajectory. **c)** RNA velocity analysis using scVelo, with velocity vectors projected on our UMAP. **d)** Confocal pictures of *pAPL::VENUSer* showing continuous expression in PSE in the early root meristem with a patchy expression in the neighbouring cells that gets stable in CC and MSE shootward. As observed in the zoom picture, the signal in PPP gets weaker after PSE enucleation. The pictures are accompanied by a UMAP showing *APL* cluster-weighted normalised expression in the phloem pole cell atlas. Scale bar in the longitudinal sections is 25 μm while it is 10 μm in the cross sections and zoom. White arrowheads point to PSE cells as a reference point. 25 independent seedlings from three different lines expressing this construct were observed. **e)** Probability of a cell to be assigned to different trajectories, ranging from 0 to 1. In the image we are showing a few clusters as an example. **f)** The expression of *APL* and PSE enucleation markers *NAC086* and *NEN4* was plotted along the PPP, CC and PSE trajectories, with the cells coloured by cluster number in the UMAP. *NAC086*, an *APL* target, appears later than *APL* in the PSE trajectory, and *NEN4*, a *NAC086* target, appears later than *NAC086*, indicating our approach matches the temporal aspects observed in PSE biology in the roots. The dots representing the cells are coloured according to cluster colours in panel c and Fig. 1b. The black line is a smoothed trend estimated from a non-parametric generalised additive model. **g)** Initiation of *S17* and *SAPL* gene expression. Distance from QC (μm) was measured to the first cell expressing each gene in 7 days post sowing (dps) seedlings, $n=10$ for *S17* and $n=11$ for *SAPL*.



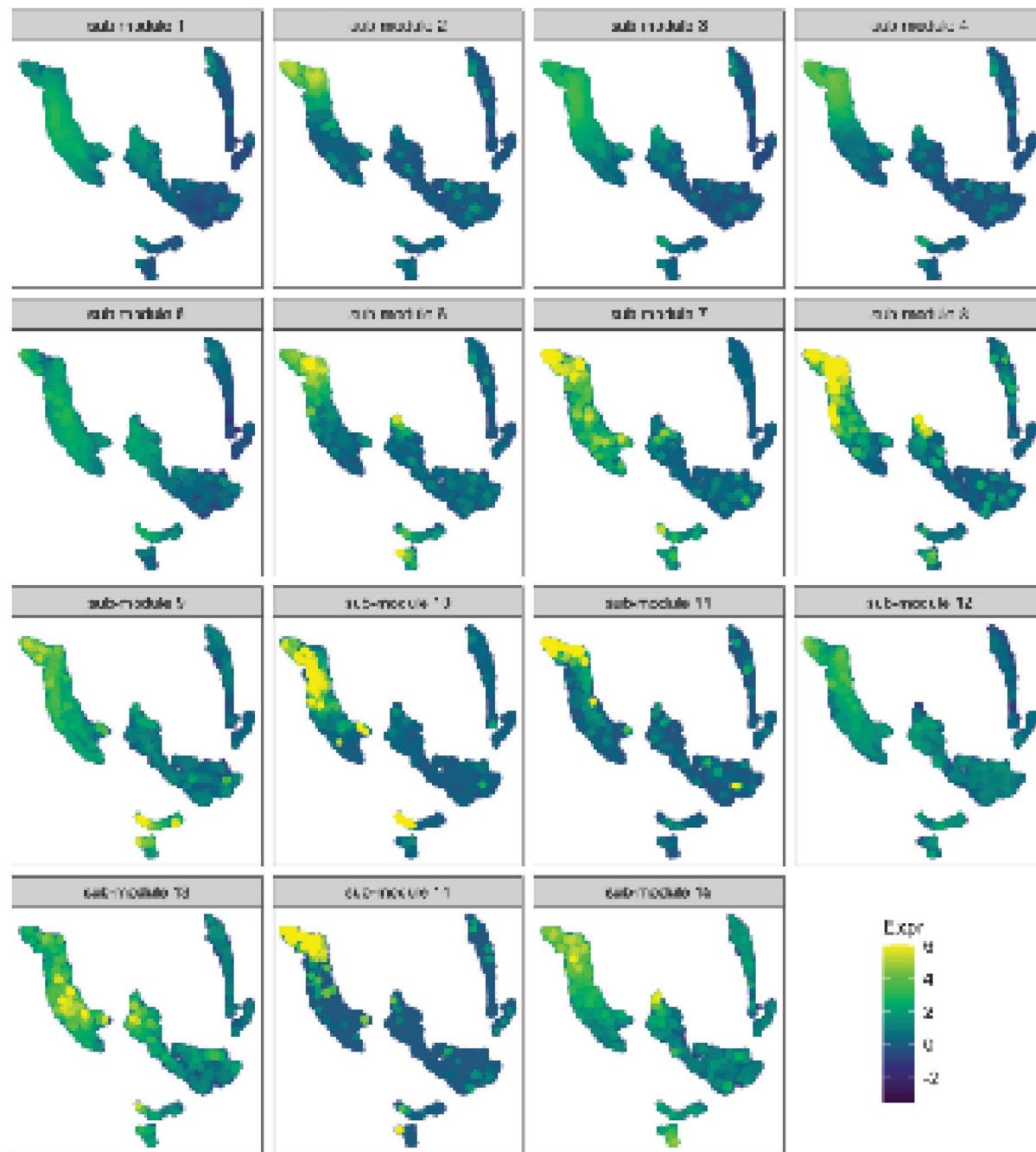
Extended Data Fig. 3 | Expression of CC, SE, PPP and ring genes in the integrated dataset faceted by source. **a)** Orange arrowheads point to MSE cluster 24. Red arrowhead points to cluster 27, cells not present in other dataset that represent cluster 5 of the UMAP and the cells surrounding PSE around enucleation time. **b)** Black arrowheads point to PSE cluster 28. Orange arrowheads point to the blue cells in MSE cluster 24. The phloem pole cell atlas provides a majority of PSE cells.



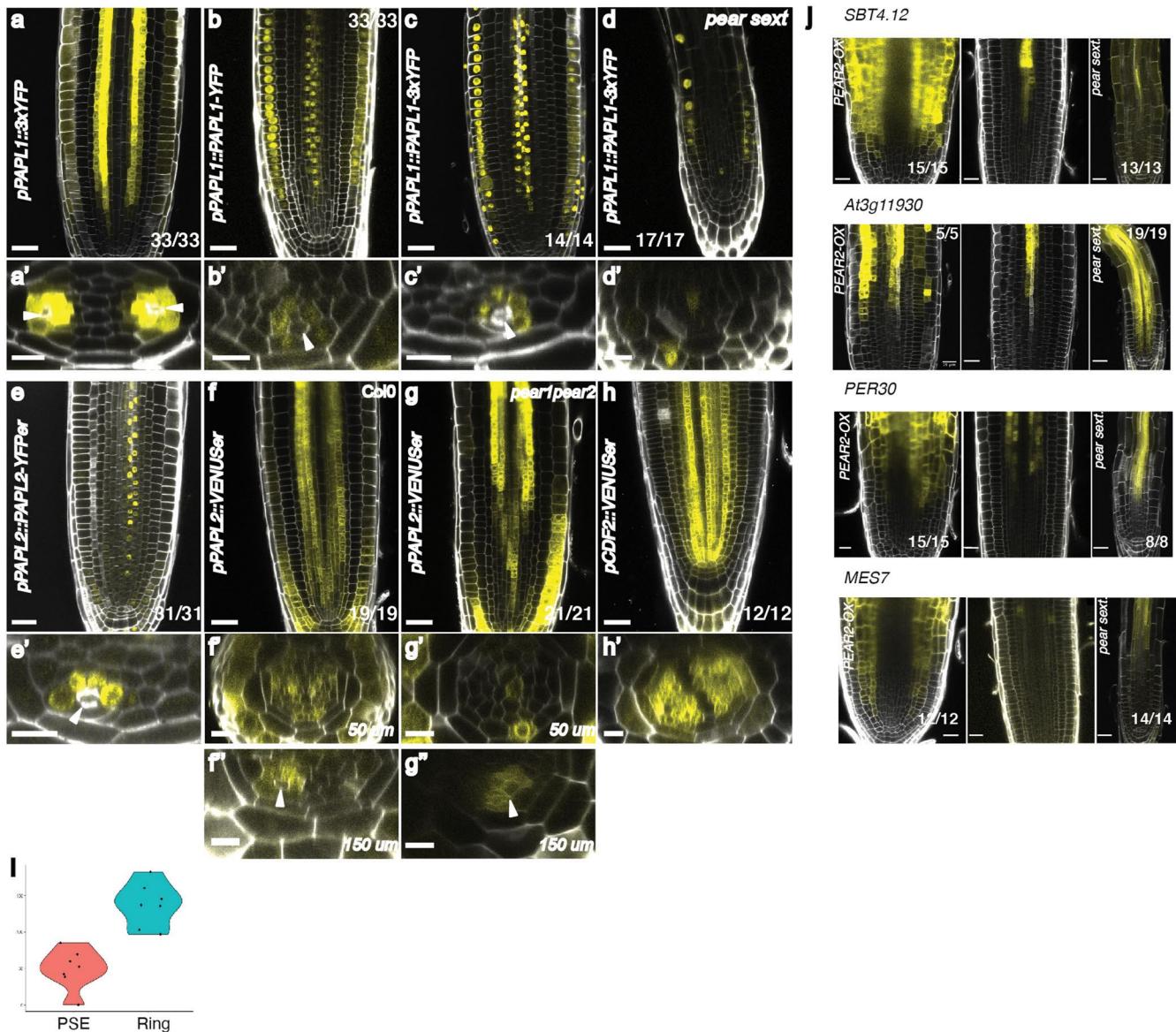
Extended Data Fig. 4 | Gene co-expression network analysis identified 16 gene modules. Gene expression of each module is summarised by the eigengene profile, which is the first principal component score from a PCA done on the expression matrix and are plotted on the UMAP and different trajectories to visualise module expression across various cell types and along developmental trajectories in the phloem pole. Module 1 consists of 136 genes (with variance explained score 21.4%) and shows an increasing expression in both PPP and CC trajectories, while lower than average expression in PSE. Modules 2, 3 and 7 contain 995, 878 and 225 genes respectively, which are highly expressed in the early cells. Module 4 with 778 genes shows broad expression in CC, PPP and PSE, but was lowly expressed on all three trajectories. Module 5, containing 368 genes, shows expression in cell clusters 12 and 13 while the 291 genes in Module 6 are mostly expressed specifically for PSE cells. The 164 genes in Module 8 were mostly expressed in the outer layers, with some lowly expressed in the mature PPP and CC cells. The eigengene profile of Module 9 (134 genes) shows expression in all cell types except CC, but particularly higher in early cells while Module 10 contains 18 genes highly expressed in CC. Modules 11–16 contain no more than 10 genes and the exact sizes are 6, 5, 3, 2, 2, and 2 genes respectively.



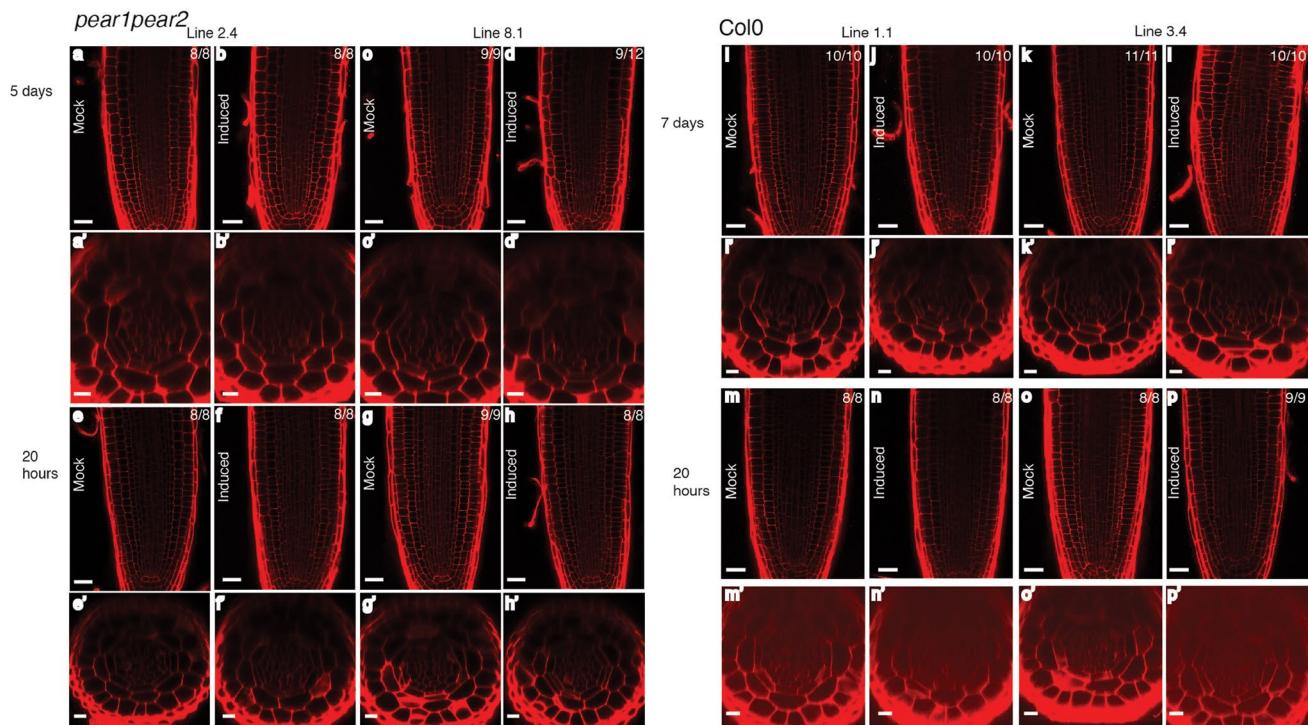
Extended Data Fig. 5 | Module 1 also groups genes with an extended or partial ring expression pattern. **a)** New genes with an expression pattern validating the module eigengene analysis. All the genes presented in this panel are expressed forming a ring pattern (*At3g163380*^a or extended ring (*At4g27435*, note the expression in protoxylem, and *PER30*, note the expression in procambium) at the time of PSE enucleation. They are all grouped in module 1, except *At4g27435*, which belongs to module 4. UMAPs show the particular cluster-weighted normalised expression of each gene in the phloem pole cell atlas and microscopy pictures are representative images of the transcriptional reporter lines where the gene promoter is fused to VENUSer. Scale bar in the longitudinal sections is 25 μ m while it is 10 μ m in the cross sections. White arrowheads point to PSE cells as a reference point. Each gene has also been plotted in PPP (green), CC (orange) and PSE (purple) trajectories, showing average expression values in the Y-axis and pseudotime in the X-axis. The numbers in each panel indicate samples with similar results, of the total independent biological samples observed.

Module 1: sub-module eigengenes

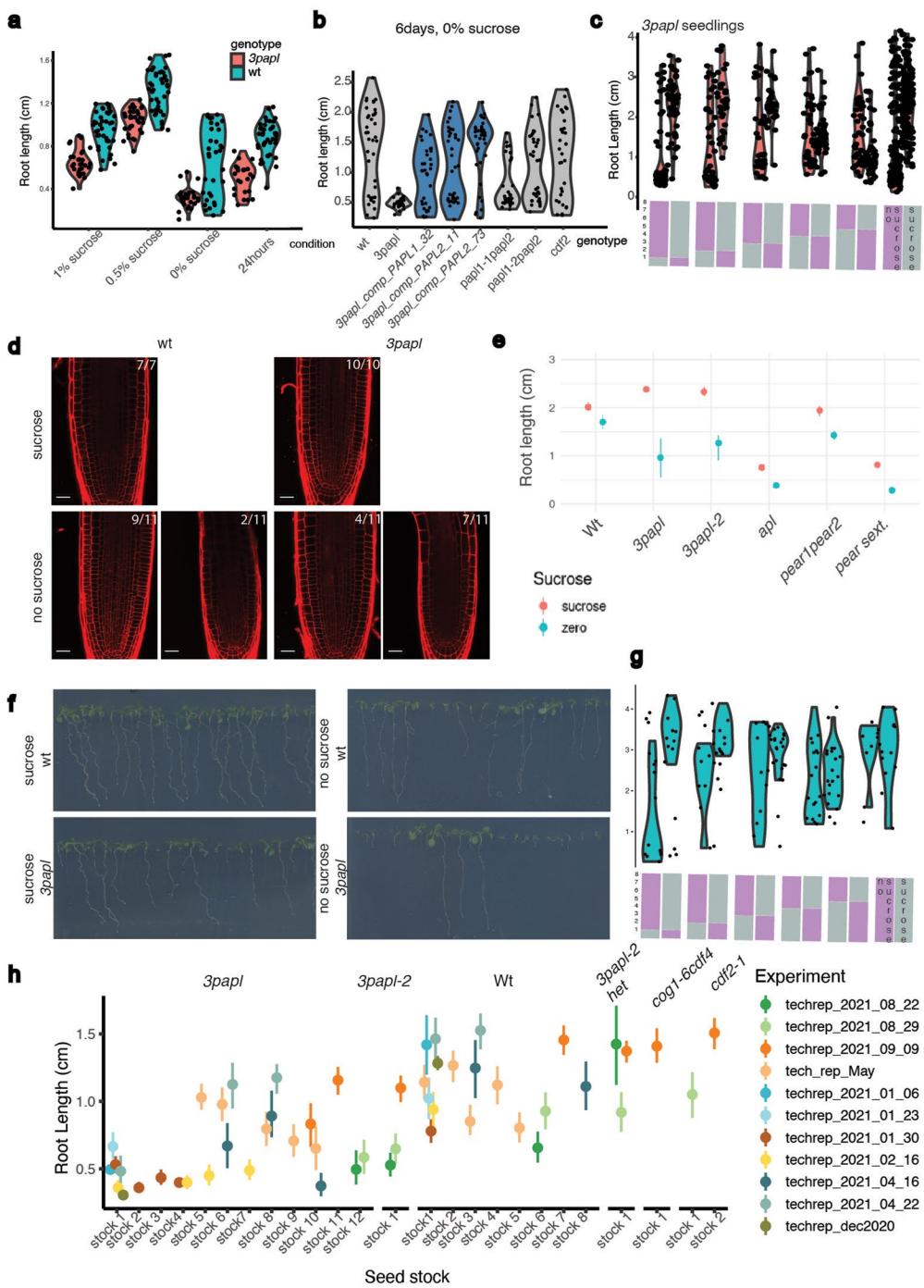
Extended Data Fig. 6 | Analysis on sub-network of Module 1 genes identified 15 sub-modules via Louvain algorithm. The eigengene profile of sub-module 1, containing 326 genes, shows expression in both PPP and CC with relatively low expression in PSE, similar to the pattern revealed by Module 1 eigengene. 8 out of 9 genes with ring-specific expression pattern found in Module 1 fall in this sub-module, while *At3g16330* falls in sub-module 3, which shifts slightly towards mature pericycle cells. Sub-module 2 contains 318 genes specifically for pericycle cells. Genes in sub-module 4 and 6 are highly expressed in PPP cells and some in out layers. For sub-module 5 and 7, the eigengene profiles show relatively broader expression in both PPP and CC, as well as the out layers. The other 8 sub-modules contain no more than 9 genes.



Extended Data Fig. 7 | Detailed analysis of PAPL gene expression. **a)** *pPAPL1::3xYFP* showing a strong expression in all the cells surrounding PSE and a weaker expression in the neighbouring procambial layer. The latter is only observed with the 3xYFP reporter. **b)** Nuclear localization of *pPAPL1::PAPL1-YFP* in epidermis and PSE-adjacent cells which recapitulates the 3xYFP fusion pattern (**c**), indicating that PAPL1 is not mobile. Occasionally some nuclei appear highlighted in the neighbouring procambial layer, where this gene is expressed weakly as shown in **a**. **d)** Phloem meristematic expression of *pPAPL1::PAPL1-3xYFP* disappears in *pear sext.* while the epidermis signal stays. Occasionally the reporter was also observed in a central xylem cell. **e)** *pPAPL2::PAPL2-YFP* recapitulates *PAPL1* translational expression pattern. **f)** *pPAPL2::VENUSer* expression mirrors the *PAPL1* ring transcriptional expression, although it has a broader domain close to QC and it is expressed in columella and epidermis. **g)** Like *PAPL1*, *PAPL2* ring expression domain gets delayed in *pear1/pear2* mutant. **h)** *pCDF2::VENUSer* has a broader expression than *PAPL* genes **i)** Fluorescent signal (mean grey value) in the ring and PSE cells in seedlings of *pPAPL1::3xYFP* **j)** Transcriptional reporter lines where the promoter was fused to *VENUSer* were transformed into *pear sext* mutant background and *pRPSSA::PEAR2-GR*, a line overexpressing ectopically *PEAR2* in the whole meristem. *PEAR2* was sufficient to induce the expression of the different genes in these layers. Primed letters show the cross section of each respective letter. Scale bar in the longitudinal sections is 25 μm while it is 10 μm in the cross sections. White arrowheads point to PSE cells as a reference point. μm in the cross sections indicate the distance from QC. The number in each confocal picture indicates samples with similar results of the total independent biological samples observed.

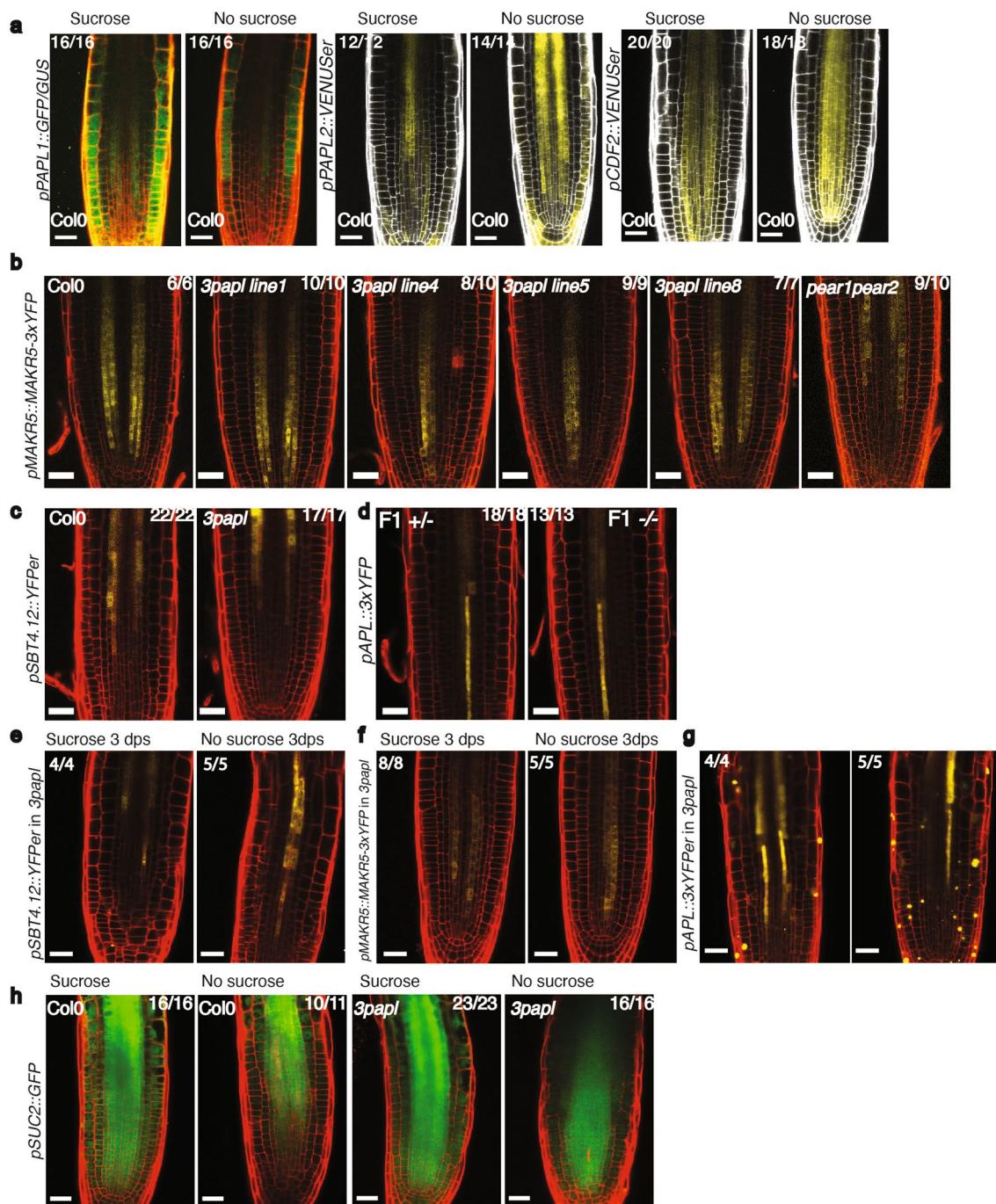


Extended Data Fig. 8 | PAPL genes do not induce periclinal cell divisions. 2 different lines for *pWOL::XVE::PAPL1* in *pear1pear2* mutant background were induced in beta-estradiol for either 5 days (a-d) or 20 h (e-h). The same construct was transformed in the *Col0* background, with 2 lines carried forward. Seedlings were induced for either 7 days (i-l) or 20 h (m-p). In the mock treatment, DMSO was added to the media instead of beta-estradiol. Primed letters show the cross sections of each respective letter. Scale bar in the longitudinal sections is 25 μ m while it is 10 μ m in the cross sections. The number in each panel indicates samples with similar results of the total independent biological samples analysed.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | PAPL genes seem to be important for a correct root nutrition. **a)** Root length in cm for 5dps *3papl* and wt seedlings in different conditions (1% sucrose, 0.5% sucrose, 0% sucrose and 24-hour light regime). For 0.5% sucrose, 42 wt and 36 *3papl* seedlings were used. For 1% sucrose, 39 wt and 29 *3papl* seedlings, for 24 hours, 41 wt and 25 *3papl* and for 0% sucrose, 34 wild type and 30 *3papl* seedlings **b)** Root length in cm for 6 dps seedlings of wt, *3papl* mutant, 3 complementation lines, double mutants and *cdf2* single mutant grown in media depleted of sucrose. 36 seedlings were measured for wt, 37 seedlings for *cdf4cog1-7*, 33 for *cdf4cog1-6*, 32 for *cdf2*, 28 for *3papl*, 38 for *3papl* complementation *PAPL2* line 7.3, 37 for *3papl* complementation *PAPL2* line 1.1 and 34 for *3papl* complementation *PAPL1* line 3.2 **c)** Replicate of the transfer experiment between sucrose and sucrose-depleted plates of *3papl* seedlings (stock 8). Time (in days) spent in sucrose and without sucrose is represented by a grey and purple bar respectively. Transfer was done days 1–5 and all roots were measured at 8 dps. The bars are divided in 8 portions representing the days in each condition. 203 seedlings were grown as a control without sucrose, 156 were grown as a control with sucrose and a total of 99 seedlings were transferred on day 1, 75 on day 2, 70 on day 3, 93 in day 4 and 96 in day 5 **d)** Confocal pictures of 7 dps wt and *3papl* seedlings grown in sucrose containing media or media without sucrose. Scale bar is 25 μ m **e)** Overall root length (cm) of the different phloem mutant phenotypes grown with (red) and without sucrose (blue) at 6 days post germination. The points denote the median and error bars the 95% confidence interval estimated by bootstrap (500 samples; see methods). The number of seedlings measured was: 195 *3papl*; 269 *3papl*-2; 314 *pear* sextuple; 57 *apl*; 272 *pear1pear2*; 311 wt. These were spread across 3 experimental batches with N=11–62 (median = 46) seedlings per combination of batch and treatment. **f)** Scans of 8 dps seedlings (wt, stock 1, and *3papl*, stock 8) grown in 1% sucrose media or without sucrose. **g)** Transfer experiment between sucrose and sucrose-depleted plates of wild type seedlings. Time (in days) spent in sucrose and without sucrose is represented by a grey and purple bar respectively. Transfer was done days 1–5 and all roots were measured at 8 dps. The bars are divided in 8 portions representing the days in each condition. 135 seedlings were measured for wt control (stock1) and an average of 29 wild type total seedlings per transfer experiment. **h)** Same data as in Fig. 4g, but showing the variation across experiment and seed stock batches (only wt and *3papl* are shown for illustration, but similar variation was observed for the complementation lines). Mean and 95% confidence interval per experiment were estimated by bootstrap (500 samples). A total of 1986 seedlings were measured split across 5 experimental batches and, for some genotypes, derived from different seed stocks. The median number of seedlings per experimental batch and seed stock combination was 36 with a range from 24–46. The number in each confocal picture indicates samples with similar results of the total independent biological samples analysed.



Extended Data Fig. 10 | Phloem marker genes are expressed in *3papl* mutant. **a**) Confocal pictures of 3 dps roots expressing reporters for *pPAPL1::GFP/GUS*, *pPAPL2::VENUSer* and *pCDF2::VENUSer* in *Col0* background with and without sucrose. **b**) Confocal pictures of *pMAKR5::MAKR5-3xYFP* in *Col0*, *3papl* and *pear1pear2* mutants. **c**) Confocal pictures of *pSBT4.12::YFPPer* in *Col0* and *3papl* mutant. **d**) Confocal pictures of *pAPL::3xYFP* in *3papl* crossed to *Col0* (*F1 +/-*) or crossed to *3papl* (*F1 -/-*). **e**) Confocal pictures of *pSBT4.12::YFPPer* in *3papl* mutant in the presence and absence of sucrose. **f**) Confocal pictures of *pMAKR5::MAKR5-3xYFP* in *3papl* mutant in the presence and absence of sucrose. **g**) Confocal pictures of *pAPL::3xYFP* in *3papl* mutant in the presence and absence of sucrose. **h)** Confocal pictures of 3dps roots expressing *pSUC2::GFP* in wt and *3papl* in the presence and absence of sucrose. The confocal signal was observed in the number of roots indicated in each picture.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	We used Leica Application Suite to acquire confocal images using mostly Leica SP8. An EPSON Perfection V700 Photo scanner was used to obtain images of the seedlings in plates to quantify root growth.
Data analysis	Raw single-cell sequencing data was processed using l0x Genomics Cell Ranger v31.0 . and downstream analysis done in R v4.0 using several CRAN/Bioconductor packages. Analysis code, with instructions on how to run it, is available from: https://github.com/tavareshugo/publication_Otero2022_PhloemPoleAtlas . Analysis of bulk RNAseq was done with TopHat2 v2..0.12, , HTseq v0.6.1 and DESeq v1.10.1. The metabolites were identified and included in the processing method in Xcalibur by screening the chromatograms and the mass spectra in the same program and comparing with the expected retention time in the literature. Same program was used to quantify the identified metabolites in all the samples. ChromaTof from Leco were used to double check the peaks and the retention time for expected metabolite fragments. Peak areas were analysed jointly for all metabolites, with a hierarchical model in R using the lme4 v1.1-2.7.1 and emmeans v1.6.2 -1 packages to quantify differences between WT and 3papl mutant, within each sampling stage and tissue. The use of a hierarchical model has the advantage that multiple measurements within each sample (21 metabolites measured per sample) are taken into account in the model (this was achieved by adding a random term for "sample ID"). Image J v2.1.0/1.53c was used to process confocal images.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Sequencing data from 10x Chromium single-cell RNA-seq is available from NCBI's Gene Expression Omnibus through GEO accession number GSE181999:
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181999>.

Sequencing data from bulk RNA-seq is available from NCBI's GEO accession number GSE182672:

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE182672>.

All other data will be available from the Cambridge Apollo Repository (doi: <https://doi.org/10.17863/CAM.74836>).

Data are now public.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. Sample size was determined based on preliminary observations which determined how large sample size must be to obtain reproducible results.
Data exclusions	No data were excluded in this study.
Replication	Several batches of single cell data were obtained and data integration methods suggest consistent cell types across batches. Phenotyping experiments were replicated two or more times with similar results. Imaging data was obtained from multiple seedlings (numbers detailed in each figure).
Randomization	For the single-cell experiments, cells were sorted from seedlings grown in separate batches for each reporter line. The data was integrated accounting for these batches as detailed in the methods. For root measurement experiments, individuals from different groups (genotypes or sucrose treatments) were grown simultaneously within each experiment repeat. In cases where the experiment was repeated several times, this was accounted for in the statistical analysis, as detailed in the methods.
Blinding	Blinding was not applied since experiments were carried out without previous knowledge of the results.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging