



UNIVERSITÀ POLITECNICA DELLE MARCHE
Repository ISTITUZIONALE

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

This is the peer reviewed version of the following article:

Original

A root phloem pole cell atlas reveals common transcriptional states in protophloem-adjacent cells / Otero, S., Gildea, I., Roszak, P., Lu, Y., Di Vittori, V., Bourdon, M., Kalmbach, L., Blob, B., Heo, J.-O., Peruzzo, F., Laux, T., Fernie, A.R., Tavares, H., Helariutta, Y.. - In: NATURE PLANTS. - ISSN 2055-0278. - 8:8(2022), pp. 954-970. [10.1038/s41477-022-01178-y]

Availability:

This version is available at: 11566/316272 since: 2024-12-05T13:42:30Z

Publisher:

Published

DOI:10.1038/s41477-022-01178-y

Terms of use:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. The use of copyrighted works requires the consent of the rights' holder (author or publisher). Works made available under a Creative Commons license or a Publisher's custom-made license can be used according to the terms and conditions contained therein. See editor's website for further information and terms and conditions.

This item was downloaded from IRIS Università Politecnica delle Marche (<https://iris.univpm.it>). When citing, please refer to the published version.

(Article begins on next page)

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

Sofia Otero¹, Iris Gildea^{2,3}, Pawel Roszak¹, Yipeng Lu¹, Valerio Di Vittori^{4,5}, Matthieu Bourdon¹, Lothar Kalmbach¹, Bernhard Blob¹, Jung-ok Heo¹, Federico Peruzzo⁶, Thomas Laux⁶, Alisdair R. Fernie⁴, Hugo Tavares^{*1,7}, Yka Helariutta^{*1,2}

Affiliations

¹ The Sainsbury Laboratory, University of Cambridge, Cambridge, United Kingdom.

² Institute of Biotechnology, HiLIFE/Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland.

³ Present address: Department of Plant Biology and Genome Center, University of California, Davis, California 95616

⁴ Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, Potsdam-Golm, 14476 Germany.

⁵ Present address: Dipartimento di Scienze Agrarie, Alimentari ed Ambientali, Università Politecnica delle Marche, Ancona, Italy

⁶ Signalling Research Centres BIOSS and CIBSS, Germany

⁷ Present address: Department of Genetics, University of Cambridge, Cambridge, United Kingdom.

*Corresponding author

Correspondence and requests for materials should be addressed to:

yrjo.helariutta@slcu.cam.ac.uk

Questions related to code and bioinformatic analysis should be addressed to:

hm533@cam.ac.uk

Abstract

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 communal 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.

Main text

INTRODUCTION

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 programs is key to understanding the positional cues guiding plant development¹. However, despite the importance of phloem in vascular plants and radial growth pre-patterning², phloem gene expression is not yet well characterized. During root development, the term phloem is oftentimes used as a synonym of 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

60 tissues. This simplification is probably the result of the extensive knowledge we have about
61 PSE specification^{2,3} and differentiation⁴⁻¹¹. However, in the Arabidopsis primary root, the
62 phloem pole is composed of six cells belonging to four distinct cell types: the central PSE is
63 flanked by two phloem pole pericycle (PPP) cells to the outside and one metaphloem sieve
64 element (MSE) cell to the inside, and both SE cells are in direct contact with the two lateral
65 companion cells (CC)¹² (Fig 1a).

66 In the Arabidopsis root, both conductive elements (MSE and PSE) derive from the same stem
67 cell¹³ but MSE differentiates later, when PSE cells are no longer functional. Despite having a
68 similar function to PSE, MSE ontogeny is less well characterized¹⁴ and few factors have been
69 directly related to MSE development. An exception are the partially redundant homologs
70 *OCTOPUS* (*OPS*, *At3g09070*) and *OCTOPUS-LIKE 2* (*OPL2*, *At2g38070*) identified as
71 important for MSE entry into differentiation¹⁵. Despite some commonalities between PSE and
72 MSE, a recent study highlighted MSE differentiation is independent of adjacent or preceding
73 PSE¹⁴, underlining the peculiarities of this cell type. The conducting cell types and CC originate
74 from different progenitors in the Arabidopsis root¹³. CC are believed to be essential to support
75 enucleated PSE function¹⁶ and their intimate relationship has been evidenced by a common
76 molecular switch controlling SE/CC fate *in vitro* and in hypocotyls¹⁷, while in the primary root
77 undifferentiated CC and MSE can transdifferentiate to PSE cells if these are misspecified¹⁸.

78 The CC function in leaves consists of loading nutrients into the SE but their role in the root
79 remains elusive¹⁹. Traditionally, it was thought they were involved in phloem unloading²⁰, that
80 is, the exit of the nutrients from the sieve element pipe so that they reach meristematic cells
81 for food. However, it was recently demonstrated this process happens through funnel
82 plasmodesmata connecting PSE to PPP¹².

83 Despite being considered a non-vascular tissue, PPP and the associated vasculature share a
84 high overlap in gene expression²¹ and are different in size and ultrastructure to the xylem pole
85 pericycle (XPP) population²², exhibiting specific gene expression²³ from early stages, mirroring
86 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 to access phloem cells, deeply embedded in the root cylinder, have hampered the study of these populations, oftentimes masked under the concept “stele”, that groups pericycle and vasculature^{27–29}. The more recent root single-cell atlases confer a detailed root panoramic but even here phloem cells remain underrepresented compared to more accessible root layers^{30–32}.

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 10204 cells by sorting phloem marker lines combined with single cell sequencing. This allowed us to gain resolution not just in the PSE lineage but in all the surrounding cells (CC, PPP, 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

In order 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 (Fig S1a). 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 2000 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/cell and a median of 4,564 genes/cell). The raw count data was

113 normalised using variance stabilising transformation³⁴ and integrated across batches using
 114 the mutual nearest neighbours algorithm³⁵, although our main conclusions are robust to
 115 normalisation and batch effects. These cells were grouped into 15 clusters using the Louvain
 116 algorithm on a shared-nearest-neighbour cell graph and visualised using uniform manifold
 117 approximation and projection (UMAP)³⁶ (Fig 1b). Using signature marker
 118 genes^{1,2,5,12,17,19,37}(Fig 1c), we identified all the cell types included in the phloem pole. We
 119 manually annotated groups of clusters as: PSE conducting cells (clusters 12, 2, 6), CC
 120 (clusters 5, 3) and a third to PPP (clusters 7, 4, 14, 11), all emerging from a central group of
 121 less mature cells (clusters 8, 9, 10, 13). Clusters 10 and 1 express MSE genes (Fig 2b). In
 122 turn, clusters 13 and 12 contain G2/M cell cycle markers, indicating cells undergoing division.
 123 While it is usually difficult to infer the identity of cycling cells, in the case of cluster 12 most of
 124 the cells express early PSE markers as well as cell division markers, pointing towards PSE-
 125 dividing cells. For example, *PEAR1* or *CVP2* are detected in all of the cells of this cluster and
 126 cell-cycle genes such as *KNOLLE*, *AUR1* or *CYCB1* are also detected in over 57% of those
 127 cells. Finally, cluster 15 corresponds to the outer layers of the root, as an apparent
 128 contamination during cell sorting.

129 Separated from the rest, clusters 7, 4 and 14 were contributed to mainly by *pS17::GFP* and
 130 *pAPL::3xYFP* markers (Fig 2a), and expressed genes characteristic of PPP such as *S17*
 131 (*At2g22850*) and *GLUCAN SYNTHASE-LIKE 4* (*CALS8*, *At3g14570*) (Fig 1c). In turn, cluster
 132 11, mainly contributed to by *pS17::GFP* and the *pMAKR5::MAKR5-3xYFP* sortings,
 133 represents mature pericycle cells, since in addition to PPP markers it also expresses markers
 134 for XPP (*At1g02460*, *At4g30450*³⁸, *At2g36120*, Fig 2c) and PPP (Fig 1c). This is likely because
 135 *MEMBRANE-ASSOCIATED KINASE REGULATOR 5* (*MAKR5*, *At5g52870*) is expressed in the
 136 whole pericycle layer high up in the root and pericycle cells come together with PPP cells for
 137 similarity.

138 Considering genes that were statistically more highly expressed in PPP-specific clusters, we
 139 built reporter lines for two genes, which were confirmed to have PPP-specific expression. One
 140 of these, *At3g27030*, was expressed in PPP and late PSE, while the other, *METHYL*

141 *ESTERASE 7 (MES7, At2g23560)*, was expressed early in PPP and soon afterwards
 142 becomes more broadly expressed in the vasculature and endodermis (Fig 1d).
 143 In turn, the known CC genes are expressed in cluster 5 (*SISTER OF APL*, (*SAPL*,
 144 *At3g12730*¹²)), with cluster 3 expressing mature CC genes (*ATPase3 (AHA3, At5g57350*²),
 145 *SODIUM POTASSIUM ROOT DEFECTIVE 1 (NAKR1)*³⁹, *SUCROSE PROTON*
 146 *SYMPORTER 2 (SUC2, At1g22710*⁴⁰). *AHA3* in particular was statistically more highly
 147 expressed in this cluster and allowed the discovery of new CC genes by correlation, which
 148 were validated building reporter lines (Fig 1d). One of these was *At2g32210*, which is
 149 expressed first in PSE and then switches to a strong CC-MSE expression, with a weak
 150 expression in the epidermis. In turn, *METACASPASE 3 (MC3, At5g64240)*, was expressed in
 151 late PSE and started being expressed in CC after enucleation, first in a patchy way and then
 152 getting continuous and mostly CC-exclusive. Cloning reporter lines for other genes expressed
 153 in these clusters, we found a gene expressed in PSE and CC (*PHOSPHATIDYLINOSITOL-*
 154 *SPECIFIC PHOSPHOLIPASE C5 (PLC5, At5g58690)*), previously described to be expressed
 155 in vascular tissues³⁷ and *At2g38640*⁴¹ mostly specific of mature CC (Fig S1b). Therefore, we
 156 have been able to validate our cell annotation (shown on the UMAP in Fig 1b) *in vivo* by using
 157 new genes highly expressed in these clusters.

158

159 Spatiotemporal patterns of differentiation in the atlas

160 From our initial cell annotation, it seemed clear that our data also captured the temporal aspect
 161 of cell differentiation in the phloem. For example, marker genes usually expressed in more
 162 differentiated cells, showed higher expression at the terminal clusters of our UMAP projection
 163 (3, 6, 11, 14, Fig S1c), while those closer to the cycling cells are less mature. To validate this
 164 hypothesis, we compared our data with a microarray dataset¹ of manually microdissected root
 165 longitudinal sections (3 to 5 cells thick), assigning each of our cells to the longitudinal section
 166 with which they had the highest Spearman correlation (Fig S2a). Using this strategy, we
 167 observed that the cells towards the centre of our UMAP matched with the meristematic

sections of Brady *et al.*, 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 Slingshot⁴³ (Fig 3a). Setting a unique origin for all 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 scVelo⁴⁴, with velocity vectors aligning towards the end of these trajectories (Fig S2c). Trajectories 1-3 account for PPP, CC and PSE respectively. Trajectory 5 is for outer layers and we will not focus on it. While 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 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, Fig S2e) or shared by all trajectories (like early phloem cells in cluster 8, Fig S2e), cluster 5 was not a clear cut, with a probability of around 0.75% of belonging to trajectory 2 (CC) and around 0.25 of belonging to trajectory 1 (PPP), Fig S2e. Regarding gene expression, cluster 5 does not express any canonical CC or PPP marker strongly. However, these markers (*SUC2*, *NAKR1*, *AHA3* for CC or *S17* for PPP) are only highly expressed in more mature cells. Cluster 5 has 64% 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 more cells in cluster 5 express CC markers than PPP markers. This matches our observations in the root, when *SAPL* starts to be expressed earlier in development than *S17* (Fig S2g). The fact that a small percentage of cells express both markers at the same time despite being specific for different cell types would indicate 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 to 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 these populations are contributed mainly by cells sorted using *MAKR5* and *PEAR1del* (Fig 2a), we can conclude that these clusters correspond to the early phloem cells, containing three different identities (MSE, PPP and CC), still undifferentiated.

There is no gene statistically enriched in cluster 9 and those few in cluster 8 (Table S1) 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, phloem early cells cluster together more than expected by chance compared to other early cells, suggesting 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. In order to map the enucleation point in the UMAP and know 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 is expressed 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, Fig S2f). *APL* is first expressed in PSE and at the time of
enucleation is transcriptionally activated in CC and MSE⁴. In reporter lines like *pAPL::3xYFP*
we perceive a strong signal in PPP as well (Fig S1a), a circumstance that we took advantage
of for sorting, but the phloem pole cell atlas transcriptomics data do not reflect a strong *APL*
expression in mature PPP (Fig 3b", Fig S2d). In another reporter line, *pAPL::YFP_{er}* we also
observe a signal in PPP that gets weaker going shootward (Fig S2d). Based on reporters and
transcriptomics data, the signal in PPP is probably not the product of gene expression in this
cell type but likely caused by direct unloading from PSE that lags and gets diluted in
successive cell divisions.

While *APL* expression increases in PSE trajectory until enucleation, it starts being detected in
PPP and CC trajectories in the common cluster 5 (Fig 3b), indicating 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
(trajjectory 4, Fig 3a) that is mainly formed by cluster 10 (Fig S2b), which is mostly contributed
by *MAKR5* sortings, pointing this could be early MSE cells (Fig 2a). In cluster 10, we find cells
expressing MSE markers like *sAPL*, *APL* (Fig 2b), and other genes expressed in MSE and
other cell types but excluded from PSE (*At5g47920*, *PAPL1*, Fig 2b, Fig 7).

In addition, we know procambial markers like *PIN-FORMED 4* (*PIN4*, *At2g01420*) become
excluded from MSE cells early in development⁴² and we find this marker absent from cluster

10 (Fig 2b). Out of 31 genes identified as highly expressed in cluster 10 compared to others (FDR < 1%), 29 were S-phase genes (histones), indicating these are still early cells and therefore harder to characterise 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 like *PEAR1* and *S32 (At2g18380)* expressed in this cluster, which is mainly contributed by the APL sorting. Cluster 1 belongs to CC trajectory, possibly because CC and MSE at this stage share some transcriptional expression, as evidenced by reporters like *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 we have not sampled mature MSE cells. However, we are convinced we have identified the first stages of MSE development in clusters 10 and 1.

The atlas represents a continuum of phloem development

In order to explore the depth of our data, we integrated our phloem atlas with existing root single cell datasets^{30,32,45} (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., 4.84% to Denyer et al. and 78.4% to Shahan et al. 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). The exception are the cells which we named as "outer layers", which are dispersed in different parts of the integrated UMAP.

276 When projecting the cells of each project in the UMAP, a continuity can be observed in the
 277 cells contributed by our atlas covering the gaps in the other data (Fig 4e). Indeed, most of PSE
 278 cells (cluster 28), a majority of the intermediate PPP and CC cells (cluster 27) and the early
 279 cells in cluster 26 (PPP) were provided by our dataset (Fig S3), demonstrating the difficulty to
 280 sample phloem without using an active strategy to enrich this population. Most of the cells
 281 classified as “Phloem” by Shahan et al. coincide with our clusters 4 and 14, which we
 282 annotated more specifically as PPP cells (Fig 4c). There are also a few cells near our cluster
 283 3, which we annotated as CC cells. We also noticed cluster 3 (companion cells) is split in this
 284 integrated dataset (Fig 4d, orange arrowheads), between mature CC (orange dots on the right
 285 of the plot, that is, cells expressing mature CC markers, like *SUC2* and *NAKR1*, see Fig S3a)
 286 and the orange cells in the less mature CC expressing *SAPL* (See Fig S3a). Perhaps the more
 287 mature CC have higher overall similarity to more mature cells of other phloem cells (such as
 288 PPP for example), more represented in the integrated dataset than in the original atlas.
 289 While other atlases, in particular Shahan et al., excelled in harvesting mature cells (including
 290 mature MSE cells, Fig S3a,b), the continuity observed in the UMAP allowed us to track phloem
 291 developmental trajectories more accurately (Fig 3a) and enrich populations that were
 292 underrepresented in other general root atlases (Fig 4e, Fig S3a,b).
 293 We also wanted to compare root phloem with a recently published single cell dataset on leaf,
 294 containing 478 vascular cells⁴⁶. In *Arabidopsis* leaves (Fig 5a), veins are often formed by
 295 multiple sieve elements usually surrounded by at least two CC and one phloem parenchyma
 296 cell. In turn, phloem parenchyma cells, which are more irregular and have a much less dense
 297 cytoplasm compared to CC, are often in contact with one or more CC, sharing comparatively
 298 many more connections than other interfaces⁴⁷.
 299 When the root and leaf data were integrated and clusters were annotated using marker genes
 300 (Fig 5d), we noticed PPP and phloem parenchyma cells blended in two clusters (Fig 5b,c).
 301 Cluster 9 of the integrated data was formed by CC cells, which are present in both leaf veins
 302 and roots. However, cluster 6 of the integrated data contained a mixture of cells annotated as

303 mature root pericycle cells and phloem parenchyma cells from leaves, showing expression of
 304 signature PPP (Fig 5e), phloem parenchyma genes (Fig 5f) and XPP (Fig 5g) in both datasets
 305 (see methods on how we assessed the degree of mixing of the cells from the two datasets in
 306 cluster 6). Phloem parenchyma leaf genes are expressed in clusters 11 and 14 in the phloem
 307 pole cell atlas, corresponding to pericycle and PPP respectively (Fig 5f). Pericycle tissue is
 308 present in roots and stems, but not in leaves, and phloem parenchyma cells are found in the
 309 aerial tissue and root secondary phloem but they are not found in the primary root. Despite
 310 being different cell types with different origins, the transcriptional overlap between phloem
 311 parenchyma and mature pericycle is another indication of the importance of positional
 312 information for cell function in plants, reinforcing the role of PSE as phloem organiser. These
 313 data also suggest parenchymatous cells share similarities across different organs and
 314 underscore their relevance for phloem.

315 Phloem pole cells share transcriptional programmes

316 In order to identify groups of genes showing distinct expression patterns in the phloem poles,
 317 we built a gene co-expression network from our scRNA-seq data using the algorithm
 318 implemented in *bigScale2*⁴⁸, which uses a gene-gene correlation metric specifically tailored
 319 for sparse single-cell data. This resulted in a gene-gene network containing 5,238 vertices
 320 (genes) and 370,794 edges (connecting two genes if their correlation was above 0.9). The
 321 biological validity of this network was confirmed by the fact that out of 59,545 edges containing
 322 genes both present in our network and in known TF-target lists (Arabidopsis Gene Regulatory
 323 Information Server, AGRIS⁴⁹), 51,658 (~86%) were preserved as linked pairs in our network.
 324 To identify groups of genes with correlated expression profiles, we used the Louvain algorithm
 325 and obtained a total of 16 gene modules (Fig S4, Table S2), and summarised their expression
 326 as the first principal component of a PCA, which we refer to as an eigengene⁵⁰. Among them,
 327 most of the modules were broad in all the trajectories with different temporal patterns. Module
 328 6 seems to represent genes with high expression in PSE (Fig S4). In contrast, module 1, which

contains 1,367 genes (Table S3), displays an increasing expression in both PPP and CC trajectories and a lower-than-average expression in PSE (Fig 6a). Reporter lines for genes in this module followed these predictions: in addition to genes with broader expression (like *MES7*, Fig1d), we identified genes showing a “ring” pattern, expressed specifically in all the cells around PSE (Fig 6b, Fig S5). 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 (Fig S5). 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 (Fig S5, Fig 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-value = 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 five of them containing over 100 genes (Fig S6). Six of the seven “ring genes” from module 1 fell within the same sub-module 1 (the exception was *At3g16330*), which again is more than would be expected by chance (hypergeometric test, p-value = 0.0009). While we do not expect that all of the 326 genes in this sub-module 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

356 is a complex ring-specific pattern of expression shared across several genes in the phloem
357 pole.

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

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

366 *PINEAPPLE* ring genes are expressed in early phloem

367 Among the genes in module 1, sub-module 1, that also extend their expression into the less
368 mature clusters, we found a DOF transcription factor, *DOF1.5* (*COGWHEEL1*, *COG1*). This
369 gene and the sister gene *DOF2.3* (*CYCLING DOF 4*, *CDF4*), are expressed in early phloem
370 cells (Fig 7a). *CDF4* encodes a differentiation factor in columella cells, repressed by *WOX5*⁵¹.
371 The role of *COG1* in roots is unknown but this transcription factor is a negative regulator of
372 phytochrome signaling⁵² and promotes brassinosteroid biosynthesis by upregulating *PIF4* and
373 *PIF5*, leading to hypocotyl elongation⁵³. Both genes have been involved in regulating tolerance
374 to seed deterioration^{54,55} as well as flowering time⁵⁶.

375 Transcriptional fusions of both genes confirmed the expression of both TF in PPP, CC and
376 MSE from 40 µm from the QC, remarkably earlier than the other ring genes described above.
377 While both genes form a ring around PSE reminiscent of a pineapple slice (the expression is
378 weaker or absent in PSE, Fig S7i), *DOF1.5* (from now on *PINEAPPLE1*, *PAPL1*) is also
379 expressed in the epidermis (Fig 7b) and *DOF2.3* (*PAPL2*) is found in columella cells with a
380 broader domain towards the QC (Fig S7f, 17f'). The ring pattern observed with the GFP fusion
381 construct extends one layer towards procambium when fused to 3xYFP expression (Fig S7a),

382 indicating a weaker expression in this layer. Translational fusions show these transcription
 383 factors are nuclear localised and not mobile (Fig 7d, Fig S7a-c, S7e), since transcriptional and
 384 translational patterns are coincident. This indicates that PAPL transcription factors act cell-
 385 autonomously. Together with the translational domain of *MAKR5*, the expression domain of
 386 *PAPL* genes indicate complex expression patterns in the phloem are relevant from an early
 387 stage.
 388 *PAPL* genes, as other genes in module 1, were predicted to be downstream of *PEAR* in
 389 microarray data². *PEAR* transcription factors move to PSE-adjacent cells to control periclinal
 390 cell divisions and other transcriptional programs non-cell autonomously. This is evidenced by
 391 markers like *SAPL* and *At3g16330* becoming ectopically expressed after broad *PEAR*
 392 overexpression or *SAPL* being expressed in PSE upon PSE plasmodesmata closure².
 393 To validate *PAPL* genes are downstream of *PEAR*, *PAPL* reporter lines were expressed in
 394 *pear1pear2* double mutant, which resulted in a delay in *PAPL* expression, from 40 to 120 µm
 395 from the QC (Fig 7c, S7g). Since *PEAR* genes are highly redundant, we also introduced
 396 *PAPL1* constructs in the *pear* sextuple mutant, *pear sext*^{t42}, where we observed a loss in its
 397 usual meristematic expression (Fig S7d). In parallel, closing PSE plasmodesmata connections
 398 to the neighbouring cell types using *ica/s3m* tool, the ring expression of *PAPL1* is altered (Fig
 399 7e) and overexpressing *PEAR1* leads to ectopic expression of *PAPL2* (Fig 7f). These results
 400 validate that *PAPL* genes are downstream of *PEAR* and indicate that *PEARs* are needed and
 401 sufficient to express *PAPL* genes in the early phloem.
 402 In addition to the *PAPL* genes, we validated that some of the genes in module 1 act
 403 downstream of *PEAR* TF. Indeed, *PEARs* are sufficient to induce *SBT4.12*, *At3g11930*, *MES7*
 404 and *PER30*, since these genes become ectopically expressed upon induction of *PEAR2*
 405 expressed under a ubiquitous promoter (*pRPS5A*) (Fig S7j). In *pear sext.*, the expression
 406 pattern of *PER30* and *MES7* was modified, while *SBT4.12* expression was decreased and
 407 *At3g11930* spread towards the meristem (Fig S7j).

408 PAPL proteins link *PEAR* genes to root physiology

409 Next, we decided to check if *PAPL* genes were downstream of *PEAR* genes to control
410 periclinal cell divisions. Since *PAPL* expression is delayed in *pear1pear2* double mutant and
411 absent from early phloem, we chose this mutant as a background to express *PAPL1* under
412 the *WOODEN LEG (WOL)* promoter. When inducing *PAPL1* expression (20h treatment or
413 germinated directly in beta estradiol and grown for 5 days), we did not observe a phenotype
414 similar to *PEAR1* overexpression with increased periclinal cell divisions in the root² (Fig S8a-
415 h). A similar result was observed when *PAPL1* was overexpressed in the stele in wild type
416 background (Fig S8i-p). These observations indicate *PAPL* genes do not control periclinal cell
417 divisions downstream of *PEARs*.

418 To gain insight into the function of *PAPL* genes, and after checking *pap1* single mutants didn't
419 show any obvious root phenotype, we generated double mutants (*pap1-1 pap12* and *pap1-2*
420 *pap12*). Bulk RNA sequencing identified *CYCLING DOF 2*, *CDF2*, as upregulated in *pap1-1*-
421 *1pap12* (Table S4). This gene encodes another DOF transcription factor expressed in the
422 cortex, pericycle and procambium, partially overlapping with *PAPL* expression (Fig S7h).
423 Presuming this gene was upregulated to compensate for the lack of *PAPL* genes, we
424 generated a triple mutant using a *cdf2* T-DNA allele⁵⁶ (*pap1-1pap12cdf2-1*, *3pap1*).

425 The triple mutant root was shorter than wild type in several conditions (Fig S9a) but the effect
426 was more pronounced growing the seedlings in media without sucrose (Fig 7g, Fig S9a,b). A
427 triple mutant with a new allele for *CDF2* generated using CRISPR/Cas9 technology showed
428 similar results (*3pap1-2*, Fig S9h). While wild type plants grown in media without sucrose often
429 showed a bimodal distribution in terms of root growth (Fig S9a,b), the proportion of roots
430 arresting growth in *3pap1* was higher (Fig 7g, Fig S9f). Even if there is high variation between
431 seed batches, the average root length of the mutant is lower than that of the wild-type (Fig
432 S9h). Contrary to other phloem development mutants *apl* and *pear* sext., adding 1% sucrose
433 to the media mostly suppressed the mutant phenotype of *3pap1* (Fig 7h, Fig S9e). In this

434 scheme, compared to other mutants, root length in *pear1pear2* mutant was not so affected
 435 by the absence of sucrose in the media. When grown with sucrose, it was rescued to wt levels.
 436 Since the mutants could be rescued by transferring them to sucrose, we aimed to identify the
 437 time point at which sucrose is needed for *3papl*. For this experiment, we transferred plants
 438 from sucrose supplemented to sucrose-depleted media and *vice versa*. The more time the
 439 mutant seedlings spent without sucrose, the more difficult it was for them to recover root
 440 growth (Fig 7h, Fig S9c). Those recovering managed to grow well (Fig S9f). Spending at least
 441 3 days in sucrose was necessary for the mutant seedlings to grow normally while spending
 442 only two days in sucrose was not enough for root growth recovery (Fig 7h, Fig S9c). This
 443 phenomenon was not observed in wild type roots grown and transferred in parallel (Fig S9g).
 444 In the confocal, the root meristem of seedlings that got arrested, looked shorter and stunted
 445 (Fig S9d). *PAPL* genes were expressed at this stage in both sucrose and non-sucrose
 446 conditions showing similar patterns as observed in more mature seedlings (Fig S10a). Other
 447 phloem marker genes, like *MAKR5*, *APL* and *ring gene SBT4.12* were expressed similar to
 448 wild type in *3papl* mutant background, suggesting there are no defects in phloem development
 449 in *3papl* (Fig S10b-d). On the contrary, *MAKR5* expression is delayed in *pear1pear2* mutant
 450 background (Fig S10b) and *APL* expression is highly reduced in *pear sext³³*, suggesting *PAPL*
 451 genes do not fulfil the same roles as *PEAR* genes. These markers and *SUC2* are expressed
 452 similarly when the plants are grown in media containing or depleted of sucrose (Fig S10e-h).
 453 To better understand the *3papl* phenotype, we carried out metabolic profiling of leaves and
 454 roots of seedlings grown in a sucrose-depleted media across six developmental stages (2-7
 455 days post-sowing, dps) (Table S5). We identified 7 and 5 metabolites in leaves and roots,
 456 respectively, with significant differences between WT and mutant in at least one of the time
 457 points (<5% false-discovery rate from a linear mixed model fit to the whole data, see methods;
 458 Fig 8a). One of those metabolites was sucrose, with a significant difference only in the roots,
 459 where it started at lower levels in the mutant (days 2 and 3) and then continued to increase to
 460 reach levels comparable to the WT at the end of the experiment at day 7 (Fig 8b). A similar
 461 pattern, with more significant points, was observed in fructose, which is a component of

462 sucrose, and to a lesser extent in glucose, the other monosaccharide forming sucrose (Fig
463 8b). It has been described that by the time the radicle emerges, all the sugars stored in the
464 Arabidopsis seed have been consumed. Within 48 hours after germination (approximately at
465 day 3 after sowing), lipid and protein reserves are exhausted and seedlings need to switch to
466 autotrophic growth^{57,58}. The data suggest *PAPL* genes could be important after the seedling
467 has transitioned to autotrophic growth, facilitating sugar transport to sink tissues like roots.
468 The continued increase in sucrose in the mutants could be due to the, on average, smaller
469 size of *3pap/* seedlings and stunted growth, which could therefore lead to reduced sucrose
470 consumption and therefore its observed continued accumulation.

471 DISCUSSION

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

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

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

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

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

496 In turn, the similarities between root pericycle cells and phloem parenchyma cells in leaves
 497 suggests parenchymatic cells share characteristics despite being present in different organs
 498 with variable anatomic configurations and reinforces PSE as the phloem pole organiser.

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

502 Contrary to other phloem mutants, like *apl*, the presence of sucrose in the media almost
 503 completely suppresses the root growth phenotype of *3papl*. Regarding PEARs, the root length
 504 in the *pear1pear2* mutant was not so affected by the absence of sucrose in the media and it
 505 was rescued to wt levels when grown with sucrose. The fact that the subtle root length
 506 phenotype is rescued by sucrose leaves the possibility open that different doses of the
 507 phloem-related DOF genes are responsible for the phenotype as opposed to the type of *DOF*
 508 genes. However, we don't favour this scenario because of the functional differences of the
 509 DOF genes based on the overexpression phenotypes.

510 Since phloem is in charge of nutrient transport and a smaller amount of sucrose and its
 511 component fructose is detected in both mutant leaves and roots at 3 dps when root anatomy
 512 is comparable between wt and mutant, we interpret *PAPL* genes regulate nutrient allocation
 513 between the leaf source organs and the root sink in young seedlings, when embryo reserves
 514 are scarce. *PAPL* genes could either regulate phloem loading, long distance transport or
 515 phloem function and more studies are required to determine their precise role.

516 MATERIALS AND METHODS

517 Plant growth conditions

518 All *Arabidopsis thaliana* lines used in this study were in Col-0 background except *pear1* mutant
519 allele, which is in Ler background, conferring *pear1pear2* mutant a mixed Ler appearance.
520 Plants were grown in ½ MS Basal salts media (0.5 MS Salts, 1% Difco agar, with or without
521 1% sucrose) at 23°C and long day conditions, except for sorting experiments, when they were
522 grown using 1x MS Basal salts at 23°C with 30% humidity and 188 µM of light, long day
523 conditions, to be able to compare with other transcriptomic data.

524 *pap1-1* (*cog1-6*, from gene *At1g29160*, *PAPL1*, *DOF1.5*, *COG1*) has a single nucleotide
525 deletion (G) at position +85, which generates a premature stop codon. This mutant was
526 identified as a *cog1-D* suppressor⁵³. *pap2* (*At2g34140*, *PAPL2*, *DOF2.3*, *CDF4*) has a 4 bp
527 deletion (CAAG) at position +99 creating a premature stop codon. The *cdf2* T-DNA allele
528 (GK782H09) is a knockdown allele⁵⁶. Triple mutant was obtained by crossing the double
529 mutant *pap1-1pap2* to *cdf1r235*⁵⁶, selecting for mutant *3pap1* and homozygous wild type
530 alleles for all other genes. The second triple mutant (*3pap1-2*) was obtained by generating a
531 new *cdf2* allele by using CRISPR/Cas9 technology directly on the double mutant *pap1-1pap2*.
532 The process rendered a 5 bp deletion (CCCGG) at position +953 (*cdf2-2*), which generated a
533 premature stop codon shortly afterwards.

534

535 5 µM Beta estradiol or 10 µM DEX were used in the inducible constructs for the indicated
536 times. Plants induced with DEX were treated for 24 hours.

537 Sorting and single cell sequencing

538 Seedlings from the different marker lines were grown vertically over mesh (Normesh, 100 µm)
539 for five days 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 an hour with gentle shaking at room temperature⁵⁹. In the case of the sample "MAKR5 enriched in root tips", we submerged the root tips of intact roots in eppendorfs containing the protoplasting solution for 15 minutes, 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 7ml of protoplasting solution in 4 cm radius petri dishes at room temperature for 45 minutes 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 2-2.5h. For the gating, a wild type Col0 sample was run first to establish the fluorescent negative gate. Then this sample was subsequently stained with DAPI and DRAQ5 to gate for intact cells that contained DNA, respectively. 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 harvested 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 minutes at 200g to eliminate the excess of 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-beads 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. cDNA quality and quantity was 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 dsDNA quantity). Each sample was normalized and pooled in equal molar concentration. To confirm concentration pool was qPCRred using KAPA Library Quantification Kit on QuantStudio 6 Flex before sequencing. All samples were sequenced using Illumina NovaSeq6000 sequencer with 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 (e.g. 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 v3.1.0⁶⁰. The data were processed and quality-filtered using several Bioconductor packages⁶¹. Empty droplets were inferred and removed using dropletutils v1.8.0⁶², and data normalisation was done using both the pooling method implemented in scran v1.16.0⁶³ and the variance-stabilising transformation from sctransform v0.2³⁴. To adjust for potential batch effects, data from the different samples (i.e. 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 2000 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, batch-normalised using MNN and using variance-stabilised transformed

593 values. However, our conclusions were qualitatively robust to the specific choice of
594 normalisation methods. For data visualisation purposes, we have projected the data to two
595 dimensions using uniform manifold approximation and projection (UMAP), using a
596 neighbourhood size of 30 cells (sizes of 7, 15 and 100 were also explored and give comparable
597 results). We have also visualised the UMAP in three dimensions, which did not provide further
598 insights into the data compared to the two-dimensional projection.

599 Cell clustering was performed by first defining a “shared nearest neighbours” graph and then
600 identifying modules in the graph using the Louvain algorithm (using *scrn* v1.16.0⁶³ and *igraph*
601 v1.2.6⁶⁴. To annotate our cells we used a set of genes with known expression patterns (from
602 promoter fusion microscopy experiments) and calculated, for each cluster, the percentage of
603 cells where each marker gene was detected as well as the (z-score scaled) average
604 expression of the gene in that cluster.

605 To identify cluster-specific genes, we used pairwise Wilcoxon rank sum tests between a given
606 target cluster and all others using the *findMarkers()* function in the *R/Bioconductor* package
607 *scrn* v1.16.0⁶². We specifically tested for genes upregulated in the target cluster, to identify
608 highly-expressed genes specific to each cluster (rather than also including genes that are
609 specifically absent from the cluster). The results of the pairwise tests for a given target cluster
610 were then consolidated to obtain a summary p-value (and corrected false-discovery rate) for
611 how enriched each gene is in a given cluster. We summarised the pairwise p-values for a null
612 hypothesis that the gene is not differentially expressed in at least 8 out of the 15 clusters,
613 allowing us to flexibly identify genes that were highly expressed in across multiple cell types
614 (e.g. mature ring cells such as PPP and CC) but not others. We also did a more stringent
615 summary of p-values (null of no differential expression in 12/15 clusters) to obtain genes more
616 specific to particular clusters of interest (namely cluster 10, which was a candidate for early
617 MSE cells).

618 The same pipeline was applied to the public datasets, also integrated using MNN. The quality
619 of this data integration was confirmed by checking that the majority of our annotated cells were
620 clustering together with the same cell types in other datasets. We produced two sets of data

621 integration, one with root data and another with leaf data. Details of the public datasets used
622 are given in (Table S6).

623 To explore how well cells from leaf and root datasets mixed in clusters where they co-occurred
624 (namely cluster 6, which contained both leaf phloem parenchyma and root phloem pole
625 pericycle cells) we used the same shared-nearest-neighbours cell graph used for clustering
626 and calculated the proportion of edges between root-leaf cells (the vertices of the graph). This
627 value was then compared with a null expectation, obtained by randomly shuffling the cell tissue
628 labels 1000 times and calculating this proportion each time. The 95% inter-percentile range of
629 this null distribution was then used to compare with the observed value. The graph had 19.2%
630 leaf-root edges in this cluster, which is only slightly lower than expected by chance (median
631 23.5%, 95% CI 22.8%-24.2%, obtained from 1000 random shuffles of the cell labels). This
632 result suggests that the cells from the two datasets are well mixed. This is in contrast with
633 cluster 17, for example, which consists of poorly clustered cells that occur separated in the
634 UMAP. In this case, there were 26.7% leaf-root edges, almost half of the null expectation for
635 that cluster, which was 44.9%.

636 To further temporally annotate our phloem pole atlas dataset we used several approaches.

637 Early dividing cells were identified by checking the expression of all annotated cyclins and
638 other cell cycle markers such as *AUR1* (*AT4G32830*) and *KNOLLE* (*AT1G08560*). We also
639 cross-referenced our data with a published dataset that profiled the transcriptome of
640 longitudinal root sections using microarray technology¹. Based on 9,674 common genes
641 between the two datasets, we assigned each of our cells to the longitudinal section of Brady
642 et al. that had the highest Spearman correlation with it. We also used the RNA velocity method
643 implemented in scVelo v0.2.2 to infer developmental dynamics in our data⁴⁴. Finally, cells were
644 assigned to lineages and ordered by pseudotime using slingshot v1.6.0⁴³. We first reduced
645 the dimensionality of the (batch-normalised) counts to 10 components using diffusion maps,
646 which is a dimensionality reduction method suited to capture developmental transitions in the
647 data^{65,66}. In this latter case we used a semi-supervised approach, where the starting point for
648 the inferred trajectories was set to the cluster highly expressing cell-cycle markers and

649 identified as the earliest cluster when cross-referencing with the Brady et al. dataset. In this
650 manner we obtained biologically meaningful trajectories (without setting this constraint several
651 more trajectories were obtained but with an ordering of cells which was the reverse of what
652 was expected from our other analyses). We obtained smooth gene expression patterns for
653 each trajectory using generalised additive models, as implemented in tradeSeq v1.2.0, which
654 were then used to explore gene expression patterns along the slingshot trajectories.

655 To cluster genes based on their similarity of expression across the cells, we built a co-
656 expression network using a modified version of bigScale⁴⁸, adapted to work on any species
657 (rather than the original version suited only for mouse and human). The modified package is
658 available from
659 <https://github.com/tavareshugo/bigScale2/tree/support-any-species>.

660 Summarily, bigScale builds a gene correlation matrix not from the original count data (which
661 in scRNA-seq is too noisy and sparse), but from a z-score statistic calculated between pairs
662 of cell clusters. These clusters are iteratively generated to ensure the z-scores capture as
663 much diversity in gene expression patterns across the cells as possible. In this way,
664 correlations between genes are more robust to the noisy and sparse nature of single-cell
665 RNA-seq data. This correlation matrix was then thresholded at 0.9 to obtain a gene-by-gene
666 adjacency matrix, resulting in a network with 5,238 vertices (genes) and 370,794 edges. We
667 identified gene modules using the Louvain algorithm, resulting in 16 modules. From each
668 module, we calculated an eigengene following the procedure in WGCNA vX⁵⁰, which
669 essentially summarises the expression of all genes of a module as the first principal
670 component score from a principal components analysis (PCA) done on those genes. The
671 largest of these modules - module 1 containing 1,367 genes - contained several genes of
672 interest for our analysis, and was therefore re-clustered with Louvain to generate 15 sub-
673 modules. This was further justified by the fact that the variance explained by this module's
674 eigengene was relatively low (21.44%), suggesting some heterogeneity in expression
675 patterns within the module. To further interpret these results, the eigengenes from these
676 sub-modules were joined with the pseudotime trajectories from slingshot, although we note

677 that no information about trajectories was used to build the network itself. Therefore, the fact
678 that the different approaches (gene network and pseudotime analysis) reveal groups of
679 genes with similar patterns of expression is a strengthening point in our analysis.

680 Generation of reporter lines and confocal images

681 *Promoter::VENUS* fusions were generated for the genes *At3g27030*, *At2g23560* (*MES7*),
682 *At2g32210*, *At5g64240* (*MC3*), *At5g58690* (*PLC5*), *At2g38640*, *At3g11930*, *At2g02230* (*PP2-*
683 *B1*), *At5g47920*, *At4g27435*, *At5g59090* (*SBT4.12*), *At3g21770* (*PER30*), *At1g26450*,
684 *At1g29160* (*PAPL1*, *DOF1.5*, *COG1*), *At2g34140* (*PAPL2*, *DOF2.3*, *CDF4*), *At5g39660*
685 (*DOF5.2*, *CDF2*).

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

689 Promoter fragments between 622-4879 bp were amplified by PCR and cloned using MultiSite-
690 Gateway (Table S7). Transcriptional fusions to *VENUS* with an ER tag or translational fusions
691 to YFP were generated in vectors with either resistance to Basta or Hygromycin or a Fast
692 Green/Fast Red selection system. All the constructs were transformed in Col0 background
693 and at least 2 independent lines were analysed for each.

694 Roots from 5-7-day-old seedlings were either imaged in the confocal directly after mounting
695 them in 50 µg/ml propidium iodide or fixed for 45 minutes in 4% paraformaldehyde in PBS and
696 cleared using ClearSee solution (10% (w/v) Xylitol, 15% (w/v) sodium deoxycolate, 25% (w/v)
697 urea, water to the final volume)⁶⁷. Cleared roots were then stained with SCRI Renaissance
698 2200 and observed under the confocal. Images were acquired at 512x512 resolution using
699 the confocal Leica SP8. Images were analysed in ImageJ v2.1.0/1.53c.

700 Bulk RNA-seq transcriptomics

701 Wild type and *pap1-1pap12* seedlings were grown on mesh in ½ MS media with sucrose in
702 the above-mentioned conditions for 5 days. Root meristems from wild type and mutant were
703 manually and individually dissected in parallel under a stereomicroscope using 18G needles.
704 Meristems were preserved in RNAlater RNA stabilisation reagent (Qiagen) until 120
705 meristems per replicate were gathered. 3 replicates for each mutant and wild type were used
706 for RNA extraction.

707 RNA was extracted using the RNeasy Plant Mini kit from Qiagen and RNA integrity and
708 concentration were checked using TapeStation and Qubit 2.0 fluorometer (Life Technologies)
709 respectively. After quality control in Novogene company, the best 3 replicates for mutant and
710 wild type were used for library construction and sequencing following the Novogene pipeline.
711 Briefly, mRNA was enriched by using oligo dT beads and fragmented randomly. cDNA
712 synthesis was performed using random hexamers and reverse transcriptase. After first-strand
713 synthesis, the second strand is synthesized by nick-translation. Library is ready after a round
714 of purification, terminal repair, A-tailing, ligation of sequencing adapters, size selection and
715 PCR enrichment. Library concentration was quantified using a Qubit 2.0 fluorometer (Life
716 Technologies), Insert size was checked on Agilent 2100 and quantified more accurately by
717 quantitative PCR. Libraries were fed into the HiSeq XTEN platform for sequencing. Original
718 raw data were transformed to Sequence Reads by base calling and raw data recorded in
719 FastQ files. Low quality reads or reads containing adaptors were filtered out. TopHat2⁶⁸
720 v2.0.12 was used to map the reads to the reference genome (TAIR10). HTSeq⁶⁹ v0.6.1
721 software was used to analyze the gene expression level using the union mode. Fragments
722 Per Kilobase of transcript sequence per Millions base pair sequenced (FPKM) value of 0.1 or
723 1 was set as the threshold to determine whether a gene is expressed or not. To compare
724 gene expression levels under different conditions, FPKM distribution diagram and violin plot
725 were used. For biological replicates, the final FPKM would be the mean value. The differential
726 gene expression analysis consisted of read-count normalization, model-dependent mean

727 value estimation and FDR value estimation based on multiple hypotheses testing. DESeq⁷⁰
728 v1.10.1 software was used for these steps. The results of this analysis are given in Table S4.

729 **Root length measurements and statistical analysis**

730 To quantify root growth, an EPSON Perfection V700 Photo scanner was used to obtain images
731 of the seedlings in plates. A ruler was also scanned to calibrate the images. Roots were
732 measured manually one by one using ImageJ v2.1.0/1.53c.
733 Because of the nature of the data, which often had a bimodal distribution, we opted for using
734 a non-parametric bootstrap approach for our statistical analysis. This was done by resampling
735 the data 500 times and estimating the difference between groups of interest (either WT vs
736 mutants or between sucrose treatments, as detailed in the respective figure legends). We thus
737 obtained distributions of root length differences, which we used to obtain confidence intervals
738 (based on a 95% inter-percentile range) and a bootstrap p-value calculated as the number of
739 samples with absolute difference less than a "null" distribution centred on zero. Our p-values
740 therefore have a lower bound of $1/501 \sim 0.002$, which we deemed to be of sufficient statistical
741 resolution for our analyses (we added an offset of 1 to both the numerator and denominator
742 to avoid p-value = 0, which would mis-represent the precision of our analysis). Whenever
743 relevant, the bootstrap analysis took into account experimental and seed stock batches by
744 summarising the results at those levels first, before then comparing the groups of interest; this
745 ensured that the uncertainty in our estimates captures those different levels of potential
746 variation. The results of these analyses are provided in Table S8.

747 **Experimental Design**

748 Experiments were repeated independently for the following number of times. In Fig 1d, *MES7*
749 reporter was imaged 3 times and reporters for *MC3*, *At3g27030* and *At2g32210* (also in Fig
750 2b) were imaged twice.

751 For Fig 6b, reporters for *At3g11930*, *At5g47920* and *At1g26450* were imaged twice, *PP2-B1*,
 752 *SBT4.12* and *TAX2* reporters were imaged 3 times. For figure 7, *pPAPL1::GFP* was imaged
 753 6 times, *pPAPL1::GFP* in *pear1pear2* was imaged 3 times, *pPAPL1::PAPL1-YFP* (also in Fig
 754 S7b) was imaged twice, *pPAPL1::GFP* in *pPEAR1::icasl3m* was imaged twice and
 755 *pPAPL2::YFP* in *pRPS5A::PEAR1-GR* was imaged once.
 756 For Fig S1a, *pMAKR5::MAKR5-3xYFP* was imaged 7 times, *pS17::GFP* was imaged twice,
 757 *pAPL::3xYFP* was imaged 3 times and *pSAPL::YFP* (also in Fig 2b) was imaged four times.
 758 For Fig S1b, the reporter for *PLC5* and *At2g38640* were imaged 3 times.
 759 For Fig S2d, *pAPL::YFP* was imaged once. For Fig S5, reporters for *At3g16330* and *PER30*
 760 were imaged 3 times, while the reporter for *At4g27435* was imaged 4 times.
 761 For figure S7, *pPAPL::3xYFP* was imaged 6 times, *pPAPL::PAPL1-3xYFP* was imaged 3
 762 times and this construct in *pear sext.* background was imaged 3 times. *pPAPL2::PAPL2-YFP*
 763 was imaged 3 times, *pPAPL2::VENUS* was imaged 3 times and this construct in *pear1pear2*
 764 mutant background was imaged twice. *pCDF2::VENUS* was imaged 3 times. For constructs
 765 in Fig S7j, reporters in the overexpressor background, induced and control, were imaged once
 766 while reporters were imaged twice in *pear sext.* background. For Fig S8, each line was
 767 observed independently twice. For Fig S9d, roots were imaged for this figure once but these
 768 two backgrounds were imaged many times with reporter lines in them.
 769 For Fig S10a, reporters with and without sucrose were imaged twice. For Fig 10b,
 770 *pMAKR5::MAKR5-3xYFP* in *3papl* mutant was imaged 3 times and it was also imaged 3 times
 771 in *pear1pear2* mutant background. Reporter in Fig S10c was imaged 3 times while the reporter
 772 in Fig S10d was imaged once. Reporters in Fig S10e-g were imaged twice and reporters for
 773 Fig S10h were imaged once (Col0) or three times (mutant background).
 774 For the experiment shown in Fig 7g and S9e, the total number of seedlings measured for each
 775 genotype was: 488 *3papl*; 273 *PAPL1-32*; 37 *PAPL1-51*; 33 *PAPL1-71*; 343 *PAPL2-11*; 79
 776 *PAPL2-23*; 37 *PAPL2-31*; 314 *PAPL2-73*; 382 wt. Seedlings were split across 5 experimental
 777 batches and came from different seed stocks (N = 24 - 46 with a median of 36 seedlings per

778 experimental batch and seed stock combination). Both seed stock and experimental batch
779 were taken into account in the statistical analysis.

780

781 Metabolic profiling

782 Arabidopsis plants were grown across six developmental stages (from day to day 7) on mesh
783 in solid media containing sucrose or devoid of sucrose. Each day of the time course, leaves
784 and roots were harvested separately and snapped frozen in liquid nitrogen. 50 mg of leaves
785 and 20 mg of roots were ground using a Tissue Lyser. Extraction was performed according to
786 Lisec et al. (2006)⁷¹, with modifications. In detail, 750 µl/300 µl of extraction buffer (100%
787 methanol plus the internal standard adonitol, Sigma) were added to root and leaf samples
788 respectively. Samples were vortexed and transferred to a shaker at 70 °C for 15 minutes.
789 375 µl/200 µl of chloroform and 750 µl/350 µl of water were added to the tubes for leaves and roots
790 respectively, and samples were centrifuged for 10 minutes at maximum speed. 400 µl (roots)
791 and 200 µl (leaves) of supernatant were dried for each sample using the speedvac. Samples
792 were kept at -80°C until processing.

793 The dried samples were derivatized for 2 hours at 37 °C in 50 µl of 20 mg ml⁻¹ methoxyamine
794 hydrochloride (Sigma-Aldrich, cat. no. 593-56-6) in pyridine followed by a 30 min treatment at
795 37 °C with 100 µl of N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA reagent; Macherey-
796 Nagel, cat. no. 24589-78-4). For each sample, 1 µl was injected in splitless mode to a
797 chromatograph coupled to a time-of-flight mass spectrometer system (Leco Pegasus HT TOF-
798 MS; Leco Corp., St Joseph, MI, USA), using an autosampler Gerstel Multi-Purpose system
799 (Gerstel GmbH & Co.KG, Mülheim an der Ruhr, Germany). Chromatograms and mass spectra

800 evaluation, as well primary metabolites identification based on the expected retention time and
801 mass fragmentation were performed using the software Xcalibur software (Thermo Fisher
802 Scientific). The software ChromaTof (Leco) was used to confirm the peaks and retention times for
803 expected metabolite fragments.

804 To estimate differences between WT and *3pap/* metabolite levels, we fit a joint hierarchical
805 model to the peak areas of all metabolites, including terms for genotype, stage (dps), tissue
806 and their interactions. The advantage of using this model is that we could include a random
807 effect term to account for multiple measurements per sample (each sample contributed 21
808 data points, one for each metabolite). Due to the skewed distribution of peak areas, the data
809 were modelled on a log-scale, which produced well-behaved normally distributed residuals.

810 The model was fit with the *lme4*⁷² v1.1-27.1 R package and we obtained estimates of the
811 difference between the two genotypes for each metabolite and tissue using the *emmeans*
812 v1.6.2-1 R package. The p-values from the *emmeans* contrasts were corrected for multiple
813 testing using the false discovery rate method.

814 Additional information on metabolomics analysis and metabolites annotation are reported in
815 table S9 (sheets checklist and overview) according to the guidelines provided in Alseekh et
816 al⁷³.

817
818

819 Data availability

820 Sequencing data from 10x Chromium single-cell RNA-seq is available from NCBI's
821 Gene Expression Omnibus through GEO accession number GSE181999⁷⁴:
822 <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181999>.

823 Sequencing data from bulk RNA-seq is available from NCBI's GEO accession number
824 GSE182672⁷⁵:

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

826 All other data (phenotypic scoring, microscopy imaging, plasmid maps) are available
827 from the Cambridge Apollo Repository (<https://doi.org/10.17863/CAM.74836>). A
828 persistent DOI will be available upon acceptance⁷⁶.

829 **Code availability**

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

832 **Acknowledgements**

833 We thank Wolf Frommer and Ji-Yun Kim for providing *pSWEET11:SWEET11-2A-GFP* seeds,
834 Christian Hardtke for providing *pMAKR5:MAKR5-GFP* seeds and plasmids, Chiara Cossetti,
835 Reiner Schulte and all the staff from Flow Cytometry Core Facility at CIMR for their technical
836 support with cell sorting, Katarzyna Kania from Genomics Unit at CRUK for preparing Single-
837 cell RNA-seq libraries, Bruno Guillotin and Kenneth Birnbaum for helpful insights on single cell
838 analysis, Raymond Wightman and Gareth Evans for technical support with microscopy
839 experiments, Gill Hindle, Jemma Salmon and Sally Ward for media preparation, Karolina
840 Blajacka for technical assistance, Kristina Petkovic and Rut Alcaina for technical support.
841 Sebastian Schornack for helpful comments.

842 S.O. was supported by a Herchel Smith postdoctoral fellowship from the University of
843 Cambridge (2017-2020). L.K. received funding from the SNSF (P2LAP3_178062) and a Marie
844 Curie IEF (No. 795250). This work was supported by Finnish CoE in Molecular Biology of
845 Primary Producers (Academy of Finland CoE program 2014-2019) decision #271832, the
846 Gatsby Foundation (GAT3395/PR3)), University of Helsinki (award 799992091) and the ERC
847 Advanced Investigator Grant SYMDEV (No. 323052). T.L. was supported by the German
848 Research foundation (DFG) under Germany's Excellence Strategy (CIBSS-EXC-2189-Project

849 ID 390939984) and by grant La606/18-1. A.R.F and V.D. acknowledge support from the Max-
850 Planck Society.

851 **Author Contributions**

852 S.O. performed the experiments, I.S. identified *PAPL1* and *At3g16330* expression patterns
853 which appeared as *PEAR* targets in microarray data, P.R. provided *pear1pear2* double
854 mutant, *pPEAR(del)::3xYFP* and advised on experimental design, Y.L. and H.T. analysed
855 gene regulatory networks, P.R., M.B., L.K.,B.B. and J-o.H. participated in sample collection
856 for sorting and metabolomic profiling, M.B. imaged *pSUC2:GFP*, J-o.H. provided
857 *pSAPL::YFP* line, V.D. and A.R.F. carried out the metabolic profiling and data analysis, F.P.
858 and T. L. provided the *papl2* and *papl1-2* alleles, H.T. designed and performed the single cell
859 data and statistical analysis, S.O., H.T. and Y.H. conceptualised and designed the study. S.O.
860 wrote the manuscript with input from Y.H., H.T., P.R. and L.K. All authors read, edited and
861 discussed the manuscript.

862 **Competing Interests statement**

863 The authors declare no competing interests.

864 **Figure legends**

865 **Figure 1. A root phloem pole cell atlas containing PSE, MSE, CC and PPP cells.**
866 a) Root schematic highlighting the cells in the phloem pole coloured by identity (adapted from⁴²) with a
867 close-up of the phloem pole (a'). One of the radial cuts shows the middle part of the phloem pole and
868 the other shows the side view of the phloem pole b) UMAP plot showing the classification of 10,204
869 cells clustered by cell identity and developmental stage (colours indicate clusters, labelled with a
870 number). The sample of cells has a median of 4,564 detected genes (10%-90% percentiles: 2,600-
871 6,780) and a median of 17,445 total UMIs per cell (10%-90% percentiles: 5941-52689). c) Cluster
872 annotation based on markers with known tissue- or cell-specific expression. The size of the points
873 represents the percentage of cells in a cluster where the gene was detected (i.e. at least 1 UMI). The
874 colour shows the scaled average expression of the gene (z-score, i.e. number of standard deviations
875 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 normalised expression of each gene in the phloem pole cell atlas. UMAP 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. The numbers in each panel indicate samples with similar results, of the total independent biological samples observed.

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 another sorting the usual one third of the root (*MAKR5* differentiated). b) UMAPs showing the cluster-weighted normalised expression of marker genes used to identify 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. In the confocal cross sections, the scale is 10 μ m. White arrowheads point to PSE as a reference point and yellow arrowheads point to MSE. The numbers over each picture 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 this is a late pericycle cluster.

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 generalised 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 PPP and CC trajectory, 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

Fig 4. Phloem cell types in the integrated UMAP.

a) A new UMAP containing 113340 cells was generated by integrating cells from Denyer et al. 2019³⁰, Wendrich et al. 2020³² and Shahan et al. 2020⁴⁵. 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 on top (number indicated below, percentage of the total in brackets), using a coloured scale to indicate cell density. 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 first panel while the other panels in the row show the expression of G2/M genes 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.

Fig 5. Similarities in the 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 to the root. Adapted from³⁶ 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. Y axis shows the cluster number with the number of cells in it between brackets. Root cells are coloured in 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 (i.e. at least 1 UMI). The colour shows the scaled average expression of the gene (z-score, i.e. number of standard deviations above/below the gene's mean across all cells) e) Violin plots showing the expression of PPP markers in leaf (green) and root (black) cells for phloem pole pericycle markers (e), phloem parenchyma markers (f) and XPP markers (g) The confocal picture in f shows the expression of *pSWEET11::SWEET11-2A-GFP* in PPP in roots. In f, the gene expression of the respective genes is shown in the phloem pole cell atlas. 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 between brackets. In the confocal picture, the scale is 10 μ m and the white arrowhead points to PSE as a reference point.

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 the 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. 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. "X" marks xylem cells. Each gene has also been plotted in PPP (green), CC (orange) and PSE (purple) Slingshot trajectories, showing average expression values in the Y-axis and pseudotime in the X-axis. c) Expression profile of sub-module 1 eigengene, the sub-module of module 1 which 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. The numbers in each panel indicate samples with similar results, of the total independent biological samples observed.

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 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* (Col0) translational domain coincides with the transcriptional domain. e) The ring pattern of *pPAPL1::GFP/GUS* gets distorted upon PSE plasmodesmata closure using the *cal3m* tool (*pPEAR1::XVE>>cal3m*). f) *PAPL2* (*pPAPL2::VENUSer*) becomes ectopically expressed upon *PEAR1* overexpression in the meristem (*pRPS5A::PEAR1-GR*). g) Average root length of 6 days post-sowing (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 (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 Fig S9h, separately for each experimental batch and seed stock (see "Experimental Design" section in the methods). Statistical analysis comparing each mutant genotype to the WT is in Table S8. h) Transfer experiment between sucrose and sucrose-depleted plates of *3papl* seedlings. 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 Table S8.

In confocal pictures b, e and f, primed letters show the cross sections of each respective letter. Scale bars: 25 µm in longitudinal sections; 10 µm in cross sections. White arrowheads point to PSE cells and "X" marks xylem cells. The number in each confocal picture indicates samples with similar results of the total independent biological samples analysed.

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 in at least one of the time points. The point and error bars show, respectively, the mean and 2 times the standard error (i.e. an approximate 95% confidence interval) of the log-fold-change between WT and mutant metabolite levels, estimated from our linear model (see methods). The asterisk highlights points that were statistically significant after adjusting for multiple testing across all the tests (false discovery rate of 5%). b) Average metabolite levels for sucrose in mutant and wild-type. The bars denote the 95% confidence interval estimated from our linear model (see methods). The points show the raw data for individual samples. N = 6 - 8 for each timepoint/tissue/genotype combination (3 of them had 6 replicates, 8 had 7 replicates and 13 had 8 replicates).

REFERENCES

1. Brady, S. M. *et al.* A High-Resolution Root Spatiotemporal Map Reveals Dominant Expression Patterns. *Science* **318**, 801–806 (2007).
2. Miyashima, S. *et al.* Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature* **565**, 490–494 (2019).
3. 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.* **110**, 7074–7079 (2013).
4. Bonke, M., Thitamadee, S., Mähönen, A. P., Hauser, M.-T. & Helariutta, Y. APL regulates vascular tissue identity in Arabidopsis. *Nature* **426**, 181–186 (2003).
5. Furuta, K. M. *et al.* Arabidopsis NAC45/86 direct sieve element morphogenesis culminating in enucleation. *Science* **345**, 933–937 (2014).
6. Dettmer, J. *et al.* CHOLINE TRANSPORTER-LIKE1 is required for sieve plate development to mediate long-distance cell-to-cell communication. *Nat. Commun.* **5**, 1–11 (2014).
7. 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).

- 1021 8. Rodriguez-Villalon, A. *et al.* Molecular genetic framework for protophloem formation.
1022 *Proc. Natl. Acad. Sci.* 111, 11551–11556 (2014).
- 1023 9. Scacchi, E. *et al.* Spatio-temporal sequence of cross-regulatory events in root meristem
1024 growth. *Proc. Natl. Acad. Sci.* 107, 22734–22739 (2010).
- 1025 10. Xiao, Y. & Offringa, R. PDK1 regulates auxin transport and Arabidopsis vascular
1026 development through AGC1 kinase PAX. *Nat. Plants* 1–12 (2020) doi:10.1038/s41477-
1027 020-0650-2.
- 1028 11. Marhava, P. *et al.* A molecular rheostat adjusts auxin flux to promote root protophloem
1029 differentiation. *Nature* 558, 297–300 (2018).
- 1030 12. Ross-Elliott, T. J. *et al.* Phloem unloading in Arabidopsis roots is convective and
1031 regulated by the phloem-pole pericycle. *eLife* 6, e24125 (2017).
- 1032 13. Mähönen, A. P. *et al.* A novel two-component hybrid molecule regulates vascular
1033 morphogenesis of the Arabidopsis root. *Genes Dev.* 14, 2938–2943 (2000).
- 1034 14. Graeff, M. & Hardtke, C. S. Metaphloem development in the Arabidopsis root tip. *bioRxiv*
1035 2021.04.30.442082 (2021) doi:10.1101/2021.04.30.442082.
- 1036 15. Sola, M. A. R. *et al.* OCTOPUS-LIKE 2, a novel player in Arabidopsis root and vascular
1037 development, reveals a key role for OCTOPUS family genes in root metaphloem sieve
1038 tube differentiation. *New Phytol.* 216, 1191–1204 (2017).
- 1039 16. van Bel, A. J. E., Ehlers, K. & Knoblauch, M. Sieve elements caught in the act. *Trends*
1040 *Plant Sci.* 7, 126–132 (2002).
- 1041 17. Tamaki, T. *et al.* VISUAL-CC system uncovers the role of GSK3 as an orchestrator of
1042 vascular cell type ratio in plants. *Commun. Biol.* 3, 1–9 (2020).
- 1043 18. Gujas, B. *et al.* A Reservoir of Pluripotent Phloem Cells Safeguards the Linear
1044 Developmental Trajectory of Protophloem Sieve Elements. *Curr. Biol.* 30, 755-766.e4
1045 (2020).
- 1046 19. Otero, S. & Helariutta, Y. Companion cells: a diamond in the rough. *J. Exp. Bot.* 68, 71–
1047 78 (2017).
- 1048 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*. *Philos. Trans. R. Soc. B Biol. Sci.* 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.* 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. Shulse, 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

1077 (2019).

1078 32. Wendrich, J. R. *et al.* Vascular transcription factors guide plant epidermal responses to
 1079 limiting phosphate conditions. *Science* (2020) doi:10.1126/science.aay4970.

1080 33. Roszak, P. *et al.* Cell-by-cell dissection of phloem development links a maturation
 1081 gradient to cell specialization. *Science* 374, eaba5531.

1082 34. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-
 1083 seq data using regularized negative binomial regression. *Genome Biol.* 20, 296 (2019).

1084 35. Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell
 1085 RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat.*
 1086 *Biotechnol.* 36, 421–427 (2018).

1087 36. A step-by-step workflow for low-level analysis of... | F1000Research.
 1088 <https://f1000research.com/articles/5-2122/v2>.

1089 37. Desvoyes, B. *et al.* Looking at plant cell cycle from the chromatin window. *Front. Plant*
 1090 *Sci.* 5, 369 (2014).

1091 38. Schürholz, A.-K. *et al.* A Comprehensive Toolkit for Inducible, Cell Type-Specific Gene
 1092 Expression in Arabidopsis. *Plant Physiol.* 178, 40–53 (2018).

1093 39. Tian, H. *et al.* Arabidopsis NPCC6/NaKR1 Is a Phloem Mobile Metal Binding Protein
 1094 Necessary for Phloem Function and Root Meristem Maintenance. *Plant Cell* 22, 3963–
 1095 3979 (2010).

1096 40. Truernit, E. & Sauer, N. The promoter of the Arabidopsis thaliana SUC2 sucrose-H⁺
 1097 symporter gene directs expression of β -glucuronidase to the phloem: Evidence for
 1098 phloem loading and unloading by SUC2. *Planta* 196, 564–570 (1995).

1099 41. Zhang, Q. *et al.* Knock-Down of Arabidopsis PLC5 Reduces Primary Root Growth and
 1100 Secondary Root Formation While Overexpression Improves Drought Tolerance and
 1101 Causes Stunted Root Hair Growth. *Plant Cell Physiol.* 59, 2004–2019 (2018).

1102 42. Roszak, P. *et al.* Analysis of phloem trajectory links tissue maturation to cell
 1103 specialization. *bioRxiv* 2021.01.18.427084 (2021) doi:10.1101/2021.01.18.427084.

1104 43. Street, K. *et al.* Slingshot: cell lineage and pseudotime inference for single-cell

transcriptomics. *BMC Genomics* 19, 477 (2018).

44. 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).

45. Shahan, R. *et al.* A single cell Arabidopsis root atlas reveals developmental trajectories in wild type and cell identity mutants. *Dev. Cell* 57, 4 (2022).

46. Kim, J.-Y. *et al.* Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* (2021) doi:10.1093/plcell/koaa060.

47. Haritatos, E., Medville, R. & Turgeon, R. Minor vein structure and sugar transport in *Arabidopsis thaliana*. *Planta* 211, 105–111 (2000).

48. Iacono, G., Massoni-Badosa, R. & Heyn, H. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol.* 20, 110 (2019).

49. 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).

50. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559 (2008).

51. 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).

52. 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).

53. Wei, Z. *et al.* Brassinosteroid Biosynthesis Is Modulated via a Transcription Factor Cascade of COG1, PIF4, and PIF5. *Plant Physiol.* 174, 1260–1273 (2017).

54. Bueso, E. *et al.* Arabidopsis COGWHEEL1 links light perception and gibberellins with seed tolerance to deterioration. *Plant J.* 87, 583–596 (2016).

55. 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).

1133 56. Fornara, F. *et al.* Arabidopsis DOF Transcription Factors Act Redundantly to Reduce
1134 CONSTANS Expression and Are Essential for a Photoperiodic Flowering Response.
1135 *Dev. Cell* 17, 75–86 (2009).

1136 57. Allen, E. *et al.* Correlation Network Analysis reveals a sequential reorganization of
1137 metabolic and transcriptional states during germination and gene-metabolite
1138 relationships in developing seedlings of Arabidopsis. *BMC Syst. Biol.* 4, 62 (2010).

1139 58. Vidal, E. A., Moyano, T. C., Canales, J. & Gutiérrez, R. A. Nitrogen control of
1140 developmental phase transitions in Arabidopsis thaliana. *J. Exp. Bot.* 65, 5611–5618
1141 (2014).

1142 59. Birnbaum, K. *et al.* Cell type–specific expression profiling in plants via cell sorting of
1143 protoplasts from fluorescent reporter lines. *Nat. Methods* 2, 615–619 (2005).

1144 60. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells.
1145 *Nat. Commun.* 8, 14049 (2017).

1146 61. *Orchestrating Single-Cell Analysis with Bioconductor.*

1147 62. Griffiths, J. A., Richard, A. C., Bach, K., Lun, A. T. L. & Marioni, J. C. Detection and
1148 removal of barcode swapping in single-cell RNA-seq data. *Nat. Commun.* 9, 2667
1149 (2018).

1150 63. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level
1151 analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* 5, 2122 (2016).

1152 64. Csardi, G. & Nepusz, T. The igraph software package for complex network research.,
1153 *InterJournal Complex Syst.* (2006).

1154 65. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* 15,
1155 e8746 (2019).

1156 66. destiny: diffusion maps for large-scale single-cell data in R | Bioinformatics | Oxford
1157 Academic. <https://academic.oup.com/bioinformatics/article/32/8/1241/1744143>.

1158 67. Ursache, R., Andersen, T. G., Marhavý, P. & Geldner, N. A protocol for combining
1159 fluorescent proteins with histological stains for diverse cell wall components. *Plant J.* 93,
1160 399–412 (2018).

- 1161 68. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of
1162 insertions, deletions and gene fusions. *Genome Biol.* 14, R36 (2013).
- 1163 69. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-
1164 throughput sequencing data. *Bioinformatics* 31, 166–169 (2015).
- 1165 70. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
1166 for RNA-seq data with DESeq2. *Genome Biol.* 15, 550 (2014).
- 1167 71. Lisec, J., Schauer, N., Kopka, J., Willmitzer, L. & Fernie, A. R. Gas chromatography
1168 mass spectrometry–based metabolite profiling in plants. *Nat. Protoc.* 1, 387–396 (2006).
- 1169 72. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models
1170 Using lme4. *J. Stat. Softw.* 67, 1–48 (2015).
- 1171 73. Alseekh, S. *et al.* Mass spectrometry-based metabolomics: a guide for annotation,
1172 quantification and best reporting practices. *Nat. Methods* (2021) doi:10.1038/s41592-
1173 021-01197-1.
- 1174 74. Otero, S., Gildea, I., Roszak, P., Lu, Y., Di Vittori, V., Bourdon, M., Kalmbach, L., Blob,
1175 B., Heo, J-O, Peruzzo, F., Laux, T., Fernie, A.R., Tavares, H., Helariutta, Y. A root
1176 phloem pole cell atlas reveals common transcriptional states in protophloem adjacent
1177 cells, *Gene Expression Omnibus*, GSE181999, 2022.
- 1178 75. Otero, S., Gildea, I., Roszak, P., Lu, Y., Di Vittori, V., Bourdon, M., Kalmbach, L., Blob,
1179 B., Heo, J-O, Peruzzo, F., Laux, T., Fernie, A.R., Tavares, H., Helariutta, Y. A root
1180 phloem pole cell atlas reveals common transcriptional states in protophloem adjacent
1181 cells, *Gene Expression Omnibus*, GSE182672, 2022.
- 1182 76. Otero, S., Gildea, I., Roszak, P., Lu, Y., Di Vittori, V., Bourdon, M., Kalmbach, L.,
1183 Blob, B., Heo, J-O, Peruzzo, F., Laux, T., Fernie, A.R., Tavares, H., Helariutta, Y. A root
1184 phloem pole cell atlas reveals common transcriptional states in protophloem adjacent
1185 cells, Cambridge Apollo Repository, 2022. <https://doi.org/10.17863/CAM.74836>

Commented [1]: Hugo, is this the right year?

Commented [2]: Hugo, is this the right year?

1186 **Methods References**

1187 77. Kang, Y. H. & Hardtke, C. S. Arabidopsis MAKR5 is a positive effector of BAM3-
1188 dependent CLE45 signaling. *EMBO Rep.* **17**, 1145–1154 (2016).
1189 78. Brady, S. M. *et al.* A High-Resolution Root Spatiotemporal Map Reveals Dominant
1190 Expression Patterns. *Science* **318**, 801–806 (2007).
1191 79. Roszak, P. *et al.* Analysis of phloem trajectory links tissue maturation to cell
1192 specialization. *bioRxiv* 2021.01.18.427084 (2021) doi:10.1101/2021.01.18.427084.
1193 80. Miyashima, S. *et al.* Mobile PEAR transcription factors integrate positional cues to
1194 prime cambial growth. *Nature* **565**, 490–494 (2019).

1195