Comparative meta-proteomic analysis for the identification of novel plasmodesmata proteins and regulatory cues.

Philip Kirk[1], Sam Amsbury[2], Liam German[1], Rocio Gaudioso-Pedraza[1], Yoselin Benitez-Alfonso[1*]

[1] Centre for Plant Science, School of Biology, University of Leeds, Leeds LS2 9JT, UK.

[2] Department of Animal and Plant Science, University of Sheffield, Sheffield, S10 2TN, UK.

*** Correspondence:** y.benitez-alfonso@leeds.ac.uk; phone: +44 113 343 2811

Summary

A major route for cell-to-cell signaling is via cell wall-embedded pores termed plasmodesmata (PD) forming the symplasm. PD regulate many aspects of plant development and responses to the environment however, our understanding of what factors affect their structure and permeability is limited. In this paper, a meta-analysis is presented as a tool for the identification of conditions affecting PD transport and *in silico* generation of PD proteomes for species of interest. The custom-built pipeline searches the whole genome for protein structural features and conserved domains identified on experimental proteomes and use it to predict PD candidates in 22 compatible plant species. Using the *in silico* proteome and microarray analysis, interactions between PD genes and conditions affecting PD function are identified. High salinity and osmotic stress affect a significant number of PD candidate genes and we provide evidence that these conditions regulate symplasmic transport of GFP. Using the pipeline, the *in silico* PD proteome for *Medicago truncatula* was generated, as an example of a plant in which experimental data is not available. The identification of a candidate receptor like protein was experimentally validated in *M. truncatula* transgenic roots expressing fluorescently tagged protein fusion. Together the results highlight the power of our newly designed tool in the identification of new factors and proteins influencing PD in diverse plant species.

Keywords: *Arabidopsis thaliana*, cell-to-cell communication, *Medicago truncatula*, phloem unloading, plasmodesmata proteome**,** symplasmic transport, transcriptomic analysis.

Introduction

Without careful coordination of the processes occurring in distant cells, plants would be incapable of responding effectively to changes in the environment, to the availability of resources for growth or to defend themselves against predators. Unsurprisingly, plants have evolved a myriad of long- and short- distance signalling pathways underpinning their ability to adapt and thrive in diverse conditions on Earth. A major route for signalling is the symplast, the continuous cytoplasmic connections established by cell walls-embedded pores termed plasmodesmata (PD). PD are dynamic structures tightly controlled to regulate intercellular signalling during development and in response to the environment (Li et al., 2021). Despite their importance, many questions remain unanswered especially regarding the molecular composition and the conditions that affect their function.

35 Over the last decade, proteomic analysis of PD-enriched membrane fractions has greatly
36 improved our understanding of what proteins may localise to and associate with PD
37 (Fernandez-Calvino et al., 2011; Park et al., 2017; Leijon et al., 2018; Brault et al., 2019).
38 Localisation of labelled protein fusions using confocal, FRET-FLIM (Förster resonance energy
39 transfer by fluorescence lifetime imaging) or transmission electron microscopy has
40 experimentally confirmed PD association for around 60 proteins in the model plant *Arabidopsis*
41 *thaliana* (Table S1). The PD proteomes include callose metabolic enzymes, such as callose
42 synthase 3 (CALS3) and the PD-located β-(1,3)–glucanases PdBGs and AtBG_ppap (Sagi *et*
43 *al.*, 2005; Vatén *et al.*, 2011; Benitez-Alfonso *et al.*, 2013). It also comprises signalling proteins
44 and kinases such as the PD-Located Protein family (PDLPs) and the Lysin Motif domain-
45 containing glycosylphosphatidylinositol-anchored protein 2 (LYM2) (Thomas *et al.*, 2008;
46 Faulkner *et al.*, 2013). Proposed functions for many of these factors was recently reviewed in
47 the context of cell-cell connectivity (Amsbury *et al.*, 2018; Dorokhov *et al.*, 2019).
48 Detection in PD proteomes is not proof of PD localisation as preparations of PD domains
49 within cell walls are often contaminated with non-PD structures such as the plasma membrane
50 (PM), cell walls and the endoplasmic reticulum (ER) (Brault *et al.*, 2019). The detection of
51 proteins belonging to the ER is often attributed, not always verified, to the desmotubule (DT),
52 a tubular structure that runs through PD connecting the ER of neighbouring cells. To eliminate
53 contaminants, proteomic data from PM fractions is used but this is far from an ideal
54 solution (Leijon et al., 2018; Brault et al., 2019). Contamination issues aside, the isolation of
55 PD enriched fractions has been optimised for cell cultures which do not reflect the complexity
56 and dynamic composition of PD associated with plant development and responses to the
57 environment. Phylogenetic analysis can be used to identify family members expressed in
58 distinct tissues but this is time consuming and predictions for diverse plant species is not always
59 succesful.
60 In this paper, we present a meta-analysis to overcome some of the limitations imposed by
61 proteomic screens. Comparing four published proteomes, we observed considerable overlap in
62 protein subfamily composition. Based on this, we developed an analysis pipeline (implemented
63 in R and named Plasmodesmata In silico Proteome 1 or PIP1; available as a resource) to rapidly
64 generate candidate PD proteome lists for compatible species (currently 22 species). Overlap
65 between predictions and experimental data demonstrates the validity of this approach. PD
66 proteome lists were obtained *in silico* for the well-studied *A. thaliana* and for *Medicago*
67 *truncatula*, a species for which a PD proteome has not been experimentally determined.
68 Bioinformatic analysis of public transcriptomes allowed us to generate co-expression tables

69 suggesting novel interactions between PD genes and conditions potentially affecting PD
70 function. Using this approach and knowledge, we demonstrate the effect of salinity and high
71 osmotic potential in the symplasmic unloading of GFP and confirm PD localization for a novel
72 receptor like protein in *M. truncatula*. Together, the results highlight the power of our newly
73 designed tool in the identification of new factors affecting PD in diverse plant species.
74
75 **Material and Methods**
76 Plant material and growth conditions
77 Seeds of *A. thaliana* plants expressing p*SUC2*::GFP were surface sterilised with ethanol.
78 Control ATS media was prepared as described by Wilson et al. (1990) with 0.8% (w/v) agar
79 (Type E, Sigma- Aldrich). When required, plates were prepared using ATS-based media
80 supplemented with 3% (w/v) polyethylene glycol (MW 8000) or with 75 mM NaCl as
81 indicated. Seeds were stratified at 4°C for 4 days before being transferred to long day light
82 conditions (22°C, 16 h day, 150 mE/m2/s) for growth.
83 For *p35S::*Medtr1g073320-YFP transgenic roots, Gateway cloning was used to generate the
84 vector following manufacture's instruction (Invitrogen, USA). In brief, primers were designed
85 to amplify Medtr1g073320 with linkers compatible for cloning into pDNR221 by BP reaction
86 (Medtr1g0733201-Attb1: GGG GAC AAG TTT GTA CAA AAA AGC AGG CTC CAT GTT
87 TTG ATT CTC TCT CCA; Medtr1g0733201-Attb2: GGG GAC CAC TTT GTA CAA GAA
88 AGC TGG GTA CCA CAA ATC TCT TTC AGC CAA AA). Positive pDNR clones were
89 confirmed by sequencing and used in LR reaction with the destination vector pB7YWG2
90 (Karimi *et al.*, 2002). The *p35S::*Medtr1g073320-YFP vector was amplified in *E. coli* and
91 expressed in *Agrobacterium rhizogenes* for transformation.
92 Wild-type A17 *M. truncatula* seeds were lightly scarified with sandpaper and sterilised for 3
93 minutes in a 10% sodium hypochlorite solution, washed with water and left undisturbed in
94 water for 4 hours. The seeds were transferred to agar-water plates and left in the dark for 7 days
95 at 4 °C. Plates were transferred to RT overnight and *A. rhizogenes* carrying the *p35S*-
96 Medtr1g073320-YFP was used for root transformation as described by (Boisson-Dernier *et al.*,
97 2001). After two to four weeks post-transformation, transgenic roots expressing YFP fusions
98 were identified using fluorescent microscopy and selected for confocal imaging.
99
100 PD proteome meta-analysis
101 PD proteomic data for *A. thaliana*, *Nicotiana benthamiana* and *Populus trichocarpa* were
102 retrieved from their original publications (Fernandez-Calvino *et al.*, 2011; Park *et al.*, 2017;

103 Leijon *et al.*, 2018; Brault *et al.*, 2019) and subfamilies annotated based on PANTHER16 (Mi
104 *et al.*, 2021). The proteomes were incorporated into a pipeline using a custom-built R script
105 (https://github.com/PhilPlantMan/PIP1). Instructions on how to use the pipeline are included
106 in the GitHub repository along with instructions on how to customise pipeline parameters
107 depending on user requirements. Necessary databases including the PD proteomes used here
108 are packaged with the script.
109 The script dependencies include the R library 'biomartr' (Drost and Paszkowski, 2017).
110 Protein features were predicted using the R library 'ragp' which integrates multiple tools for
111 glycophosphatidylinositol anchor (GPI) and secretory signal peptide (SP) predictions
112 (Emanuelsson *et al.*, 2007; Käll *et al.*, 2007; Pierleoni *et al.*, 2008; Almagro Armenteros *et al.*,
113 2019; Gíslason *et al.*, 2019; Dragićević *et al.*, 2020). Genes were classified as having a GPI
114 and/or SP if at least one tool returned true for that feature. Transmembrane domains (TM)
115 domains prediction was acquired from the ENSEMBL database annotation via the R package
116 'biomartr' using TMHMM (Krogh *et al.*, 2001). Predictions for N-myristoylation, S-
117 farnesylation, S-geranylgeranylation, S-palmitoylation and S-nitrosylation were made using
118 tools available by the Cuckoo workgroup (Xue *et al.*, 2010; Xie *et al.*, 2016; Ning *et al.*, 2020).
119 For protein feature enrichment analysis, the tools described above were used to predict features
120 for the whole *A. thaliana* genome (Araport11). Fisher's exact test was used to determine
121 statistical significance (cutoff provided in figure legends). To be fully pipeline-compatible for
122 proteome input or to generate candidate gene output, the species must be listed in both Ensembl
123 Plant databases (used by biomartr to retrieve sequence information) and in PANTHER16 (used
124 to retrieve subfamily annotation) (Drost & Paszkowski, 2017; Howe *et al.*, 2020; Mi *et al.*,
125 2021). Currently, there are 22 compatible plant species (Table S2). For non compatible species,
126 such as *N. benthamiana*, Arabidopsis orthologues can use. This enable integration of the PD
127 proteome of Park et al. (2017) in the pipeline.
128 Lists of genes were compared by drawing a Venn or Euler diagram using the R library 'eulerr'.
129 The significance of the overlap between candidate lists and proteomes was determined using
130 bootstrap analysis as follows. Sets of genes the same length as a candidate list and a proteome
131 were randomly sampled from Araport11. The overlap in genes between the samples was
132 recorded and repeated for n cycles (n = 10,000). Probability (p) was calculated as the proportion
133 of cycles that attained an overlap at least as large as was observed between the candidate list
134 and the proteome. The size of the overlap by chance was given as the median overlap in random
135 samples over n cycles.
136

137 Expression analysis

138 The gene correlation dataset 'Ath-u.c1-0' was downloaded from the ATTEDII database
139 (Obayashi *et al.*, 2018). Optimal gene order and the corresponding dendrogram were computed
140 using hierarchical clustering. The dendrogram was cut at an optimised height (h = 16) that gave
141 a sufficient number/size of clusters (k = 151). These processes were performed in base R.
142 Enrichment of genes within clusters from candidate lists and verified genes were determined
143 using pairwise comparisons with Fisher's exact test ($p<0.05$, holm-adjusted) via the R library
144 'rcompanion '.

145 Publicly available microarrays for a subset of conditions were independently analysed.
146 Microarray datasets (Affymetrix GeneChip Arabidopsis ATH1-121501) were downloaded
147 from EBI ArrayExpress. For each experiment, expression data were normalised using the
148 robust multi-array average (RMA) method and $\log_2$ transformed with the R package 'oligo'
149 (Carvalho & Irizarry, 2010). Principal component analysis was used to identify and exclude
150 outlier arrays and experiments with insufficient biological replicates. Genes with low levels
151 of expression were filtered out. A design matrix was constructed for each experiment and a
152 linear model applied using the R package 'limma' (Ritchie *et al.*, 2015). Differential expression
153 of genes and a multiple comparison correction were determined using empirical bayes statistics
154 via the package 'limma' and the results filtered by gene IDs. Heatmaps were constructed using
155 the R package 'ComplexHeatmap' (Gu *et al.*, 2016).

156

157 Confocal microscopy imaging for the analysis of GFP diffusion and protein localization

158 Seedlings of *A. thaliana* expressing *pSUC2::*GFP were mounted on glass slides in 10 µg/ml
159 propidium iodide (Sigma- Aldrich). Root tips were imaged using an LSM 800 upright confocal
160 microscope (Zeiss, Germany). Profiles of fluorescence were determined using line and
161 profiling tools in ImageJ. Lateral profiles were taken across the transition zone and rootward
162 profiles started from the basal/apical meristem transition zone ending 150 µm towards the root
163 tip. Fluorescence of lateral and rootward profiles were scaled between 0 and 1. Fluorescence
164 across each lateral profile was binned (bins = 100) to compensate for small differences in root
165 width. Fluorescence profiles of at least 6 plants per treatment were aggregated by calculating
166 the mean (±SD) for position along the profile and plotted.

167 Transgenic roots expressing p*35S*::Medtr1g073320-YFP were stained with aniline blue
168 fluorochrome (Biosupplies, Australia) in 0.1 M K_3PO_4 (pH 12) for callose detection. Roots
169 were imaged by confocal microscopy (LSM-880, Zeiss) using 488 nm excitation laser for YFP

170 and emission at 505–530 nm. Aniline blue was imaged with 405 nm excitation and emission
171 at 463 nm.
172
173 **Results**
174 **Workflow for the prediction of PD proteomes via meta-analysis of conserved subfamilies**
175 **and structural domains.**
176 Proteomic data obtained from *A. thaliana* (Fernandez-Calvino *et al.*, 2011; Brault *et al.*, 2019),
177 *N. benthamiana* (Park *et al.*, 2017) and *P. trichocarpa* (Leijon *et al.*, 2018) PD-enriched
178 fractions were fed into a custom-built R-based pipeline released as a resource with this article
179 (https://github.com/PhilPlantMan/PIP1). The workflow is described in Figure 1 and in the
180 Material and Methods section. First, subfamilies of genes identified in at least one of the PD
181 proteomes mentioned above were annotated using PANTHER16. Subfamilies classifications
182 are used to extract all genes making the *in silico* proteome for a particular plant of interest. For
183 some applications, using families classifications may be more appropriate and PIP1 prompts
184 users whether to generate candidates based on family or subfamily identifiers. Data presented

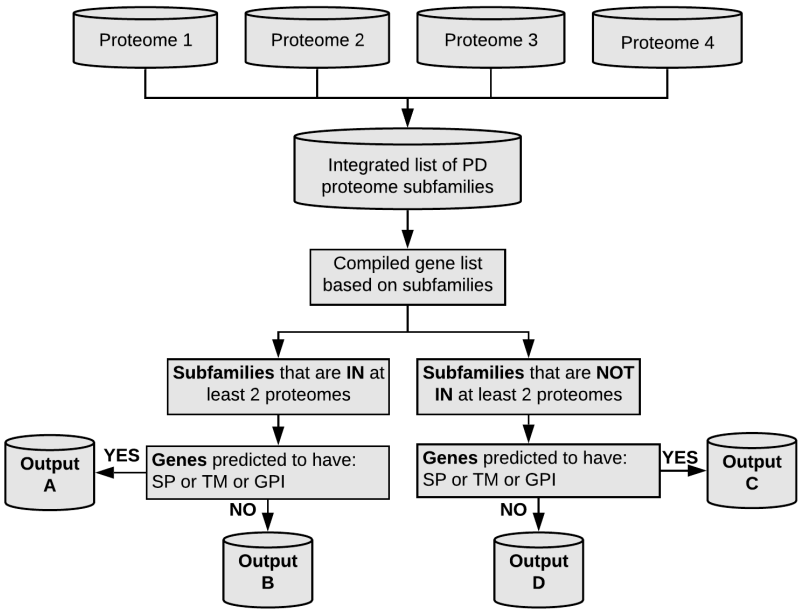

Figure 1. Pipeline to determine candidate plasmodesmata (PD) proteome. 4 proteome databases have been determined experimentally by Fernandez-Calvino et al., 2011; Park et al., 2017; Brault et al., 2019 and Leijon et al., 2018. A list was compiled of subfamilies of proteins identified in the different experimental proteome databases using PANTHER16. Genes belonging to these subfamilies were extracted for the target plant species. These were classified by whether its subfamily was present in at least two PD experimental proteomes or just one. Further classification was based on whether genes were predicted to have a signal peptide (SP), transmembrane domain (TM) or glycosylphosphatidylinositol anchor (GPI) as these features are enriched in verified PD genes (Figure S1). As a result, 4 list of genes were obtained (A-D) representing the predicted PD proteome for the target plant.

185 in this paper are based on candidate generation using subfamily identifiers as a more restrictive
186 approach leading to shorter candidate lists (see example for Arabidopsis below). The pipeline
187 categorises the output by whether the subfamily is present in one or multiple PD proteomes
188 and based on predictions of distinctive features identified in PD verified genes (Figure 1).
189 Features of membrane-bound proteins were predicted for the list of PD verified genes in Table
190 S1 using multiple online platforms as described in Material and Methods. When compared to
191 the Arabidopsis whole proteome, PD-localised proteins are overrepresented in predicted GPI
192 anchors, s-geranylgeranylation, s-palmitoylation, SP and TM domains (Figure S1). The
193 presence of a SP in combination with either a GPI and/or TM domain returned the highest
194 proportion of verified genes thus these features were chosen for gene categorisation.
195 As a result, the pipeline outputs 4 lists of genes: list A are genes of subfamilies present in more
196 than one PD proteome containing either a SP, GPI or TM domains; list B are also present in
197 multiple proteomes but lacking the membrane localising features, list C and D are genes in
198 subfamilies found in a single proteome either with (list C) and without (list D) the predicted
199 SP, GPI or TM features (Figure 1). Using this approach, 1882 gene subfamilies were identified
200 in at least one PD proteome (Table S3). As expected, there is substantial overlap between
201 subfamilies identified in the different proteomes, with 311 subfamilies identified in more than
202 one proteome (Table S3), forming list A and B. As PD verified genes usually display membrane
203 targeting features, list A should be considered the priority list but, since only four proteomes
204 are available, list C cannot be discarded. In support of this prediction, list A contains
205 subfamilies with known PD activities such as enzymes involved in callose metabolism (BG
206 and CALS) and receptors involved in signalling (such as PDLPs).
207
208 **Applying PIP1 to proteome predictions in *Arabidopsis thaliana* and poplar.**
209 For *Arabidopsis*, the *in silico* proteome comprises list A with 206 genes (158 subfamilies), list
210 B: 208 genes (152 subfamilies), list C: 751 genes (597 subfamilies) and D: 1117 genes (802
211 subfamilies) (Table S4). Most known PD genes are isolated from studies in *A. thaliana*. To
212 validate our approach, we determined the number of PD verified genes identified by the
213 pipeline. Generated candidate lists contain 60.7% of the genes known to localise at PD
214 identified either using fluorescent translational fusions and/or isolated in independent, non-
215 proteomic studies (compare Tables S4 and S1). These include the beta-1,3 glucanase
216 AtBG_ppap and C2 calcium/lipid-binding plant phosphoribosyltransferase family protein
217 MTCP6 and MTCP9 identified in the Brault *et al.* (2019) proteome and proteins identified in
218 independent studies such as the leucine-rich repeat receptor-like serine/threonine-protein

219 kinase BAM2 (Rosas-Diaz *et al.*, 2018). We failed to output known PD proteins such as
220 CRINKLY4 (ACR4) or the PD-located protein 5 (PDLP5) likely because these subfamilies are
221 low expressed or absent in cell cultures. Extending the output to family members instead of
222 just subfamily members increased the coverage of known PD genes to 83.6% but consequently
223 increased the total candidate gene size to over 9000 genes (Table 1). Genes in list A and C
224 included in families containing at least one previously verified PD-localised protein were
225 identified as the most likely targets for future experiments (Table S5). In Arabidopsis, this list
226 comprises 55 new proteins.

Table 1. Verified PD genes and candidates identified in the Arabidopsis *in silico* proteome. The number of genes identified when using either subfamilies (subfam) or families (fam) classifications in PIP1 is provided.

***In silico* proteome list**	**Verified PD genes (subfam/fam)[a]**	**Nº of candidates (subfam/fam)[b]**
A	18/41	206/2345
B	1/4	208/1432
C	18/5	751/1961
D	0/1	1117/3935

[a] For a full list of verified PD genes and the publications that report them consult table S1.
[b] The lists of candidates identified using PIP1 and subfamilies classifications is provided is table S4.

227

228 Despite lacking features for PD localisation, list B contains a considerable number of genes
229 identified in more than one PD proteome. The number of known PD proteins identified in list
230 B and D suggest they might contain contaminants but it could also contain PD proteins with
231 no predicted membrane targeting domains such as Remorin 2 REM1.2 (Huang *et al.*, 2019).
232 We explored the hypothesis that list B might contain mobile proteins unintentionally captured
233 within PD fractions. To test this hypothesis, the candidate PD gene lists were compared to the
234 mobile proteome identified in *Cuscuta australis* (dodder) parasitising *A. thaliana* (Liu *et al.*,
235 2020). Candidate mobile proteins were identified in all candidate lists but list B had the largest
236 overrepresentation (20.2%) with an overlap over 10x greater than would be expected by chance
237 (Figure S2).
238 To evaluate the strength of the pipeline further, we generated candidate PD lists for *P.*
239 *trichorcarpa*. 1032 out of 1148 genes/proteins identified in the PD-enriched fraction of poplar
240 cell cultures (Leijon *et al.*, 2018) were predicted in the candidate lists generated with our
241 pipeline. The lists contained 182 (list A), 129 (list B), 397 (list C) and 289 (list D) of the genes
242 identified in the PD-enriched fraction (Leijon *et al.*, 2018). For list A and C the *in silico*

243 proteome is 50% larger than the experimental proteome, reflecting inclusion of proteins not
244 expressed in cell cultures. 116 genes identified in the experimental proteome are excluded
245 because they are not annotated within PANTHER subfamilies. When the poplar proteome is
246 excluded as input in the pipeline, the overlap between the candidates lists and the poplar PD
247 proteome remains high; 20x larger than would be expected by chance significant (Figure S3).
248 In summary, our results validate the use of PIP1 as a new resource to generate candidate PD
249 genes, and potentially mobile proteins, without the cost of time and resources associated with
250 growing and maintaining cell cultures carrying out their biochemical isolations, fractionations
251 and protein sequencing. The *in silico* proteome list is larger than experimental proteomes
252 because there are not constrictions associated with protein expression in cell cultures.
253
254 **Drought and salinity regulate the expression of PD candidates and symplasmic transport.**
255 The generation of PD proteomes not only enables target gene selection but, in combination
256 with transcriptomics, can reveal conditions that regulate PD genes. To establish whether this
257 is the case, co-expression data extracted from ATTED-II database was ordered by hierarchical
258 clustering to identify co-expressed genes in the *in silico* proteomes. For *A. thaliana*, two
259 clusters were found to be significantly over-represented (relative to the wider ATTED-II gene
260 set) in PD-verified genes: cluster 87 and cluster 100 (Figure 2 A, B). 11 genes from list A and
261 18 from list C were found in cluster 87, whereas cluster 100 contained 16 genes of list C (see
262 Table S5 for examples). The expression of PD-candidates in clusters 87 and 100 was also
263 analysed using publicly available microarrays as described in Materials and Methods. Figure
264 2C and Figure S4 show that the expression of the candidate genes overlaps with PD-marker
265 proteins such as AtBG_ppap, PDCB (PD callose binding proteins) and PDLP family members.

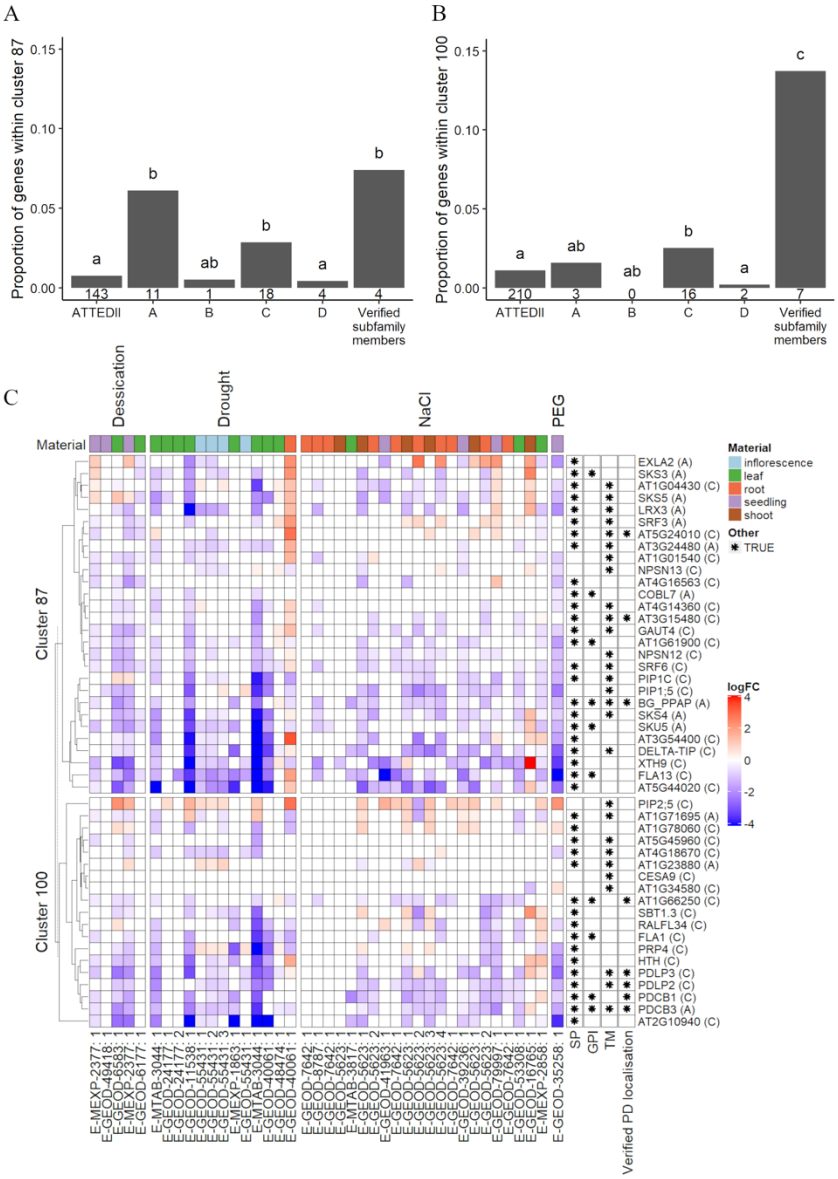

Figure 2: Expression analysis of PD candidates and PD verified genes. *Arabidopsis thaliana* co-expression data from ATTEDII were subset into clusters based on hierarchical clustering (k = 151). Representation analysis of the total ATTEDII database, candidate lists A-D and verified PD subfamilies was performed on each cluster; (A) cluster 87 and (B) cluster 100 show significant overepresentation in verified subfamily members. Number underneath each bar = number of genes within that cluster. Bars with differing letters above are significantly different (Fisher's exact test, Holm corrected, $p \leq 0.05$). (C) Differential expression (log_2FC, ranges from blue: downregulated to red: upregulated in relation to control) of genes in clusters 87 and 100 (only list A, C and verified genes are shown) determined using public microarrays from experiments pertaining to osmotic stress. Other treatments are presented in Supplemental figure 4. Columns labels bottom = ArrayExpress accession codes followed by a reference number. See Supplemental table 6 for further information. (left) Rows are grouped by cluster and then (dendrogram) ordered by hierarchical clustering. (right) Gene ID and properties. Asterisks in cells denote predictions on membrane targeting features (e.g. is predicted to have a SP). SP = Signal peptide, GPI = glycosylphosphatidylinositol anchor, TM = transmembrane domain. Cell colour above each column = plant material sampled.

267 When comparing the expression of PD-verified and PD-candidate genes in a variety of biotic
268 and abiotic stress conditions (Figure S4), osmotic stress (drought, desiccation, elevated NaCl
269 and exogenous treatment with PEG) was found to be a strong regulator (Figure 2C). To
270 determine if these changes in expression reflect conditions affecting symplasmic transport,
271 cell-to-cell connectivity was tested in root tips of transgenic *A. thaliana* seedlings expressing
272 cytosolic GFP, driven by the phloem companion cell specific promoter *Sucrose Symporter 2*
273 *SUC2* (p*SUC2*::GFP) (Imlau *et al.*, 1999).
274 p*SUC2*::GFP seeds were germinated and grown in control conditions, 3% PEG and 75mM
275 NaCl (Hunter *et al.*, 2019). Media water potential (at 25 °C) was estimated by adding the

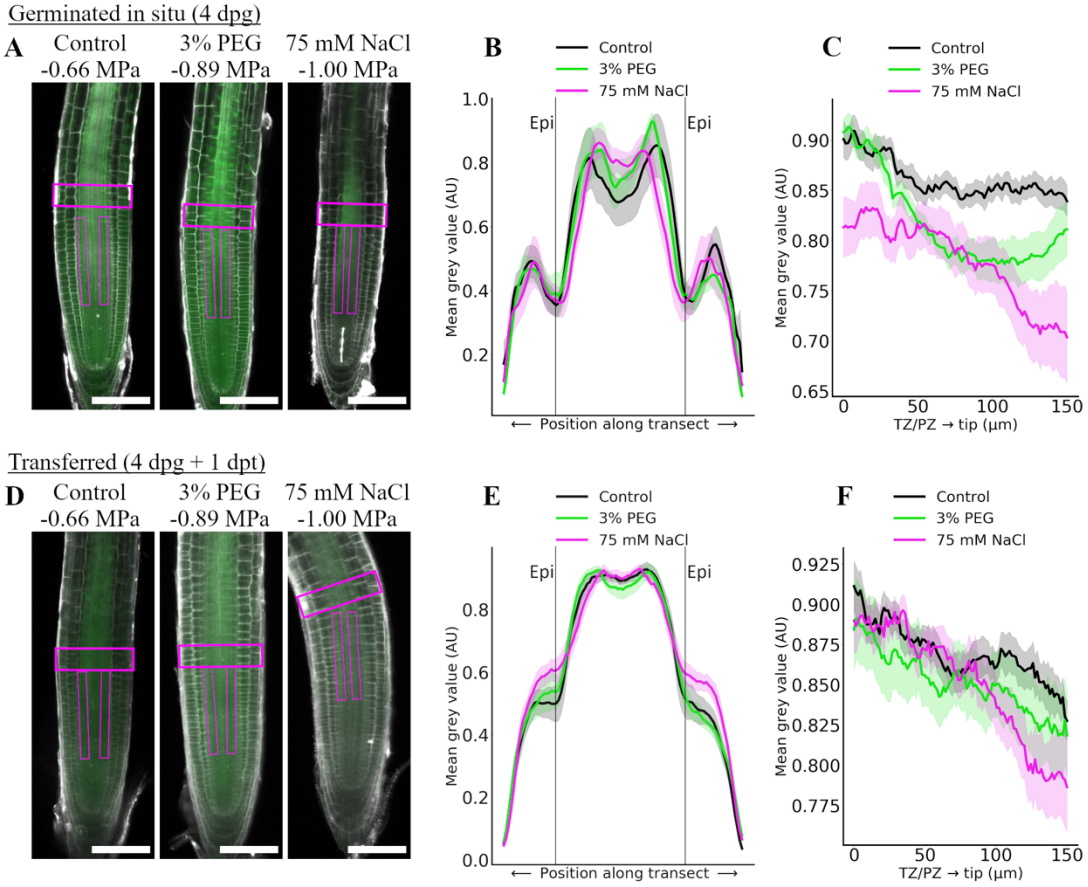

Figure 3: GFP diffusion in the root meristem of *Arabidopsis thaliana* exposed to PEG and sodium chloride. (A) Seeds expressing p*SUC2*::GFP were grown on ATS control media, ATS supplemented with 3% (w/v) polyethylene glycol (PEG) or with 75 mM NaCl. Roots were counterstained with propidium iodide (PI) and images sequentially collected 4 days post-germination (4 dpg) at 561 nm (PI shown in grey) and 488 nm (green: GFP). Estimated water potential (MPa, see main text) is labelled on top of the representative pictures. (B-C) Transects highlighted by magenta boxes in A were used to determine GFP fluorescence profiles using mean grey values (AU: arbitrary units). (B) shows transects across the transition zone (TZ) while (C) shows transects rootward starting from the transition/proliferation zone (TZ/PZ) towards the root tip. (D-F) Similar experiments were carried out with seedlings germinated on ATS media and transferred for 1 day (1dpt) to 3% PEG or 75 mM NaCl. Scale bar = 100 µm. Charts show traces representing min-max normalised GFP fluorescence along the transect for N>6 plants (mean ± SE).

276 estimated solute potentials of individual medium components based on empirical and modelled
277 data (Robinson & Stokes, 1949; Ghashghaie *et al.*, 1991; Gopal & Iwama, 2007). The water
278 potential was reduced in both 3% PEG and 75mM NaCl (Figure 3A). Roots were imaged using
279 confocal microscopy and both lateral and rootward diffusion of GFP from phloem companion
280 cells were measured using mean grey values in Image J. The short and long term effects of
281 changes in water potential were evaluated by comparing changes in GFP distribution in roots
282 grown for 4 days directly in PEG and NaCl media (Figure 3A) and roots grown for 4 days in
283 control media and transferred for 24h to PEG and NaCl media (Figure 3D). There were no
284 noticeable changes in lateral distribution of GFP across primary roots grown permanently in
285 PEG and NaCl relative to control conditions (Figure 3 A, B) but the rootward profiles, show a
286 decrease in fluorescence in both, 3% PEG and 75mM NaCl (Figure 3C). The effect is stronger
287 in NaCl rootward fluorescence profiles, particularly between 50 µm - 150 µm towards the tip
288 where relative fluorescence drops over 13% in 75 mM NaCl and less than 2% in control
289 conditions (Figure 3C).
290 No significant differences in GFP distribution were observed 24 hours after transfer from
291 control to 3% PEG conditions (Figure 3 D-F) but fluorescence was diminished in the rootward
292 profile of roots transferred to 75 mM NaCl (Figure 3F). Over the length of 150 µm from the
293 tip, relative fluorescence decreased 11% in plants transferred to 75 mM NaCl. Lateral profiles
294 in roots transferred to 75 mM NaCl show a higher fluorescence (likely auto-fluorescence) in
295 the epidermis (Figure 3E).
296 Taken together, these data suggest that regulation of PD candidates in osmotic stress is
297 translated into changes in symplasmic connectivity demonstrating the potential for using
298 combined proteomic and transcriptomic meta-analysis as a method to determine conditions
299 regulating PD function and therefore symplasmic communication.
300
301 **A meta-analysis approach predicts plasmodesmata proteins in *Medicago truncatula*.**
302 One of the advantages to the pipeline outlined here is that it can be applied to species for which
303 an experimentally determined PD proteome is not publicly available. To demonstrate this, the
304 pipeline was used to generate candidate PD gene lists for *M. truncatula* (Table S6). List A and
305 C comprised 1018 genes belonging to subfamilies represented in at least one experimental
306 proteome and displaying membrane targeting features (SP, GPI or TM). These lists include *M.*
307 *truncatula* orthologues of the CALS, PDCB and PDLP genes (Table S7).

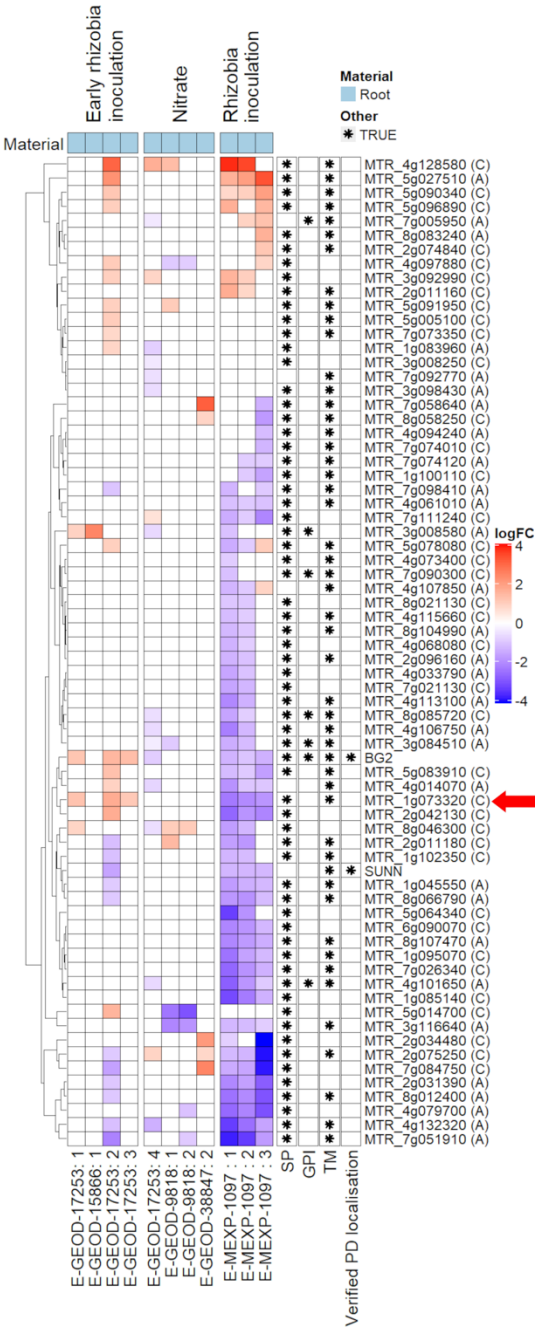

Figure 4: Expression analysis of *Medicago truncatula* PD candidates in rhizobia and nitrate transcriptomes. Gene expression analysis of PD candidates (lists A and C only) and verified PD genes BG2 and SUNN (see table S6). Differential gene expression (log_2FC) was determined using public microarray data of experiments relating to nitrate and rhizobia inoculation in root tissue. Columns labels bottom= ArrayExpress accession codes followed by a reference number. See Supplemental table 8 for further information. (left dendrogram) Rows are ordered by hierarchical clustering. Asterisks in cells denote that gene has a membrane targeting feature or are verified. SP = Signal peptide, GPI = glycophosphatidylinositol anchor, TM = transmembrane domain. Red arrow indicates the position of MTR_1g073320, which is used in further analysis.

Transcriptomic profiling was used to determine clusters of genes co-regulated within the predicted proteome lists A and C. Published microarray data from early rhizobia inoculation, various nitrate concentrations and nodulating roots (Table S8) were re-analysed. These conditions were selected as they have been previously shown to regulate PD genes and symplasmic communication (Haney & Long, 2010; Crook *et al.*, 2016; Gaudioso-Pedraza *et al.*, 2018). The β-(1,3)-glucanase, MtBG2 (Gaudioso-Pedraza et al., 2018), and the Super Numeric Nodule receptor (SUNN, Crook *et al.*, 2016) are, to our knowledge, the only verified PD-localised proteins in *M. truncatula*. A cluster containing MtBG2 and SUNN was identified comprising 70 genes from list A and C (Figure 4). Genes in this cluster share a similar

318 expression pattern, particularly in roots post-inoculated with rhizobia (E-MEXP-1097).
319 Interestingly, four genes encoding receptor-like proteins show a strikingly similar expression
320 profile to MtBG2. These are MTR_5g083910, MTR_4g014070, MTR_1g073320 and
321 MTR_2g011180. MTR_1g073320 (also known as Medtr1g073320) was selected for further
322 studies because it contains sequence domains characteristic of PDLP family members (Table
323 S7).

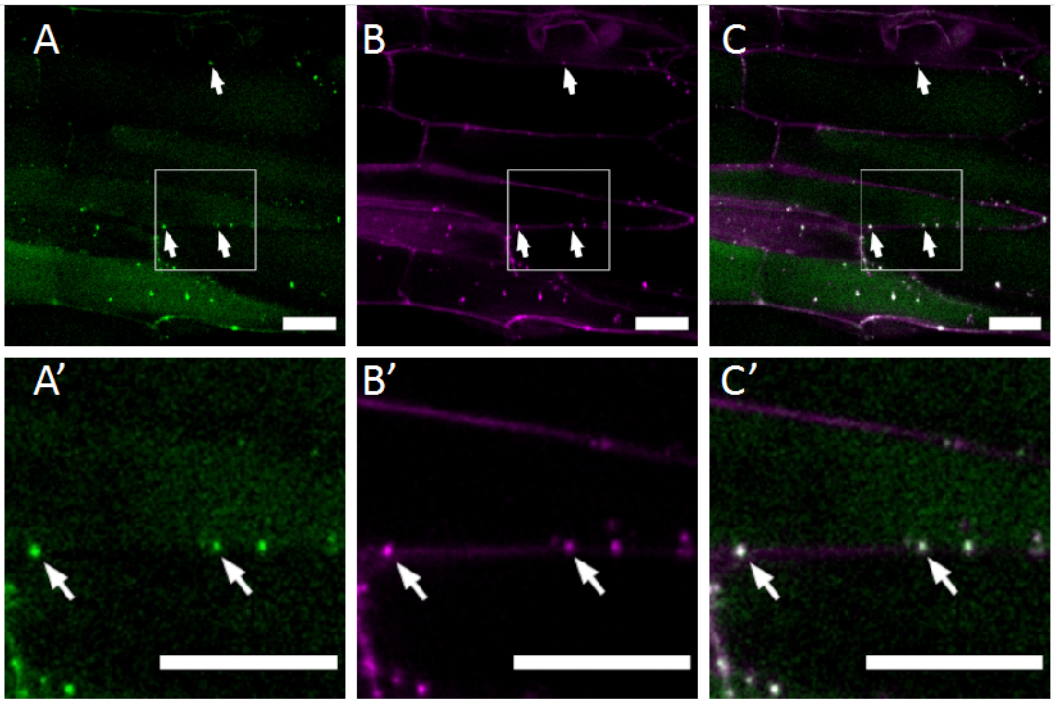

Figure 5: The *Medicago truncatula* protein Medtr1g073320 co-localizes with callose at PD microdomains. *Agrobacterium rhizogenes* transformed roots expressing Medtr1g073320 fused with YFP were counterstained with aniline blue to reveal callose deposition. Roots were imaged with a confocal microscope and (A) excitation laser 561 nm (green: Medtr1g073320-YFP) and (B) 405 nm (shown in magenta: aniline blue). Merged images are shown in C and co-localisation events highlighted with arrows. Bottom panels (A'-C') are magnification of the area highlighted by the white box in the top panels. Scale bar = 20 µm.

324 To established Medtr1g073320 localisation, C-terminal YFP fusions were created and
325 introduced in *M. truncatula* roots using *A. rhizogenes*- mediated transformation. One week old
326 transgenic roots were identified and counterstained with aniline blue which fluoresce in the
327 presence of callose- a cell wall glucan component associated with PD. Confocal microscope
328 images show Medtr1g073320-YFP in a punctate pattern on the cell periphery, co-localising
329 with callose deposits indicating PD targeting (Figure 5).
330 Altogether, the results demonstrate how the pipeline, implemented in our custom build R script,
331 can be used to predict the PD proteome in plant species where this information has not been
332 experimentally derived, opening doors to the identification of novel PD proteins participating
333 in the regulation of cell-to-cell communication.

334

335 **Discussion**

336 Many aspects of PD function and regulation are not well understood despite the increasing
337 evidence of the important role these structures play in plant signalling, organ development and
338 response to physiological and environmental cues (Petit *et al.*, 2020; Vu *et al.*, 2020; Li *et al.*,
339 2021). Proteomic data is lacking for most plant species due to difficulties in isolating clean PD
340 fractions and because association to PD domains can be quite dynamic, varying with
341 environmental conditions and tissue types (Grison *et al.*, 2019; Hunter *et al.*, 2019; Cheval *et*
342 *al.*, 2020). These factors limit our current knowledge of what proteins associate with PD and
343 how their properties and function are controlled.

344 We created a pipeline that exploits the overlap in subfamily composition of existing PD
345 proteomes to generate candidate PD genes (Figure 1). This allows the generation of *in silico*
346 proteomes for 22 plant species which are represented in both PANTHER16 and in Ensembl
347 Plant databases. The gene lists were classified according to their presence in one or more PD
348 proteomes and according to predicted membrane targeting features overrepresented in verified
349 PD proteins. Based on our categorisation, list A (subfamilies present in multiple proteomes and
350 displaying predicted membrane targeting features) is more likely to contain PD localised
351 proteins but genes in list C (subfamilies present in a single proteome) can not be excluded as
352 only four PD proteomes are sequenced so far. Supporting our predictions, list A and C
353 contained 36 out of the 60 proteins reported to target PD in *A. thaliana,* 46 if using families for
354 *in silico* predictions (Table 1). The pipeline demonstrated to be effective in predicting the
355 poplar proteome showing significant overlap with the experimentally determined PD enriched
356 fraction (Fig. S3).

357 Combining proteomic with transcriptomic analysis we identified drought, osmotic and salinity
358 stresses as conditions affecting the expression of clusters of genes identified in the Arabidopsis
359 *in silico* proteomes. The diffusion in root meristems of GFP, expressed in companion cells
360 under the SUC2 promoter, was restricted after continuous (4 days) and short (24h) exposure to
361 3% PEG and 75mM NaCl (Figure 3), suggesting that the expression of PD candidates is linked
362 to the regulation of symplastic transport. We determined for the first time the *in silico* proteome
363 of *M. truncatula* using PIP1 and exploited published microarrays to study gene expression
364 during rhizobia infection and nodulation as conditions reported to affect symplasmic transport
365 (Gaudioso-Pedraza *et al.*, 2018). This strategy led us to identify the receptor protein
366 Medtr1g073320, which we confirmed localises at PD when expressed as fluorescent fusion in
367 transgenic roots (Figure 5). Together our results validate the use of the PIP1 pipeline in

368 combination with transcriptomic analysis for the identification of new PD proteins and
369 conditions influencing symplasmic communication.
370 Our approach relies on subfamily annotation within the PANTHER database. For the 40
371 reference plant genomes included in PANTHER16, family/subfamily annotation coverage is
372 between 60-95% depending on species (Mi *et al.*, 2021). For *A. thaliana* and poplar, the
373 coverage is at 89% and 81% respectively but for *M. truncatula*, coverage is only 67%. Genes
374 without a PANTHER classification will not be included in the output of the pipeline. We expect
375 our approach may have less predictive power in monocots or non-angiosperms, as experimental
376 proteomes are all from dicot species. It is unknown to what extend the composition of PD
377 differs between monocots and dicots, especially as they differ in cell walls composition
378 (Calderan-Rodrigues *et al.*, 2019). The experimental determination of more PD proteomes
379 including representative monocot species and the identification of PD proteins by independent
380 approaches will improve *in silico* predictions as these can be easily added as inputs to PIP1.
381 Adding, as input data, subfamilies of PD proteins identified by different means can get around
382 the limitations associated with the cell culture origin of experimental proteomes. Researchers
383 already use sequence - domain analysis, phylogeny and transcriptomics to study families of PD
384 proteins and to identify members expressed outside cell cultures (Thomas *et al.*, 2008; Simpson
385 *et al.*, 2009; Gaudioso-Pedraza *et al.*, 2018). Our comparative meta-analysis provides a
386 platform to systematically apply this approach enabling *in silico* predictions of whole PD
387 proteome based in multiple instead of single experiments.
388 Applying the pipeline to *A. thaliana* or poplar generates a list of candidates larger than
389 experimentally determined because expression in cell cultures is not a pre-requisite for gene
390 identification. For example, the Arabidopsis *in silico* proteome contains BAM2, the Germin-
391 like protein PDGLP1 and CALNEXIN2 (CNX2), which are not present in any of the
392 experimental proteomes but localised at PD in independent studies (Fernandez-Calvino *et al.*,
393 2011; Liu *et al.*, 2017; Rosas-Diaz *et al.*, 2018; Brault *et al.*, 2019). The candidate lists also
394 includes subfamilies/proteins lacking predicted membrane targeting features (SP, TM and GPI)
395 (list B and D). Overlap between list B (subfamilies present in multiple proteomes) and the
396 mobile proteome, recently reported in dodder parasitising *A. thaliana* (Liu *et al.*, 2020),
397 suggest that these might be subfamilies of proteins identified while in transit via PD. List B
398 might also include proteins with unusual membrane targeting mechanisms or with features
399 poorly predicted by the available platforms. This is the case for REMORIN2, for example,
400 which localises to PD independently of the secretory pathway and lacks a predicted SP, GPI or
401 TM (Reymond *et al.*, 1996; Gronnier *et al.*, 2017; Huang *et al.*, 2019). Based on structural

402 predictions and proteomic count, the lists serve as a tool to prioritise the screening, instead of
403 rejecting, candidate PD proteins.
404 Besides the obvious reasons of obtaining *in silico* proteomes as a tool for the identification of
405 new PD genes, we propose that transcriptomic analysis of the candidate lists can predict
406 conditions affecting PD transport. Here, we showed that osmotic and salinity stresses modify
407 the expression of clusters of PD candidates in Arabidopsis and that symplasmic transport of
408 GFP is affected in these conditions. This aligns well with recent publications reporting changes
409 in symplasmic transport and in PD protein localisation in response to salt and mannitol (Grison
410 *et al.*, 2019; Hunter *et al.*, 2019). It is not clear if PD function is affected by changes in the
411 localisation of PD proteins or by changes in turgor pressure or, more likely, to a combination
412 of these (Hernández-Hernández *et al.*, 2019). The identification and phenotypic
413 characterisation of mutants in PD genes and of relevant symplasmic mobile molecules are
414 necessary to mechanistically understand the biological role of PD in the plant response to these
415 abiotic stresses.
416 Our past research indicates that PD are regulated in response to rhizobia infection, a mechanism
417 mediated by the degradation of the cell wall polysaccharide callose by the MtBG2 protein
418 (Gaudioso-Pedraza *et al.*, 2018). PDLP regulation has been tightly linked to callose deposition
419 in Arabidopsis (Lee *et al.*, 2011; Wang *et al.*, 2013; Caillaud *et al.*, 2014). Here, we identified
420 Medtr1g073320 as a PD-located member of a subfamily of PDLP proteins (Thomas *et al.*,
421 2008) in *M. truncatula*. Expression tables suggest that Medtr1g073320 is co-regulated with
422 SUNN and MtBG2 during rhizobia infection and nodulation. This finding opens doors for
423 research on the mechanisms involving callose and PD regulation in the establishment of
424 nitrogen fixation in legumes, an important process for sustainable agriculture.
425 To summarise, an R-based tool has been made available to integrate the data obtained from PD
426 proteomes and enable the identification of new PD genes in a variety of plant species. Together
427 with transcriptome analysis, the pipeline becomes a useful tool to identify conditions affecting
428 PD composition and function. Our approach works with all plant species and genes with
429 subfamilies annotated in both PANTHER16 and Ensembl Plant databases and will only expand
430 as more new and updated UniProt Reference Proteomes are added (Drost & Paszkowski, 2017;
431 Howe *et al.*, 2020; Mi *et al.*, 2021). The pipeline can help in prioritising the targets for
432 validation from experimental proteomes and can predict PD proteins for species where PD
433 proteomic information is not available. The pipeline is publicly accessible and can be easily
434 modified by the user to add new sequenced proteomes and/or experimentally verified genes
435 thus will serve as a long-lasting resource for the plant community.

Acknowledgements

The authors would like to thank Hendrik Swiegers for his help in testing the pipeline and Dr. Eva Deinum and Dr. Michael Wilson for critical reading and help shaping the manuscript. This work was supported by UKRI Future Leader Fellowship (MR/T04263X/1), Leverhulme Trust Grant RPG-2016-136, a University of Leeds Gosden Studentship and a BBSRC DTP (BB/M011151/1).

Author Contributions

P.K. designed and generated the pipeline, cross-referenced with published proteomics and databases and performed experiments and analysed data for *A. thaliana*. S.A. designed the R script for transcriptomic analysis; L.G. helped acquiring images and analysing data for *M. truncatula*. R.G.P designed and tested the construct for Medtr1g073320, generated transgenic roots and imaged them. Y.B.A. was in charge of conceptualisation, supervision, data curation, and writing- reviewing and editing the manuscript with the help of P.K. All authors discussed the results, edited and commented on the manuscript.

Data Availability Statement

PIP1, the R-based pipeline described here, will be available for download along with associated database from GitHub (https://github.com/PhilPlantMan/PIP1) upon publication. Medtr1g073320 construct will be provided upon request.

References

Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, Von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* **37**(4): 420-423.

Amsbury S, Kirk P, Benitez-Alfonso Y. 2018. Emerging models on the regulation of intercellular transport by plasmodesmata-associated callose. *Journal of Experimental Botany* **69**(1): 105-115.

Benitez-Alfonso Y, Faulkner C, Pendle A, Miyashima S, Helariutta Y, Maule A. 2013. Symplasmic intercellular connectivity regulates lateral root patterning. *Developmental cell* **26**(2): 136-147.

Boisson-Dernier A, Chabaud M, Garcia F, Bécard G, Rosenberg C, Barker DG. 2001. Agrobacterium rhizogenes-transformed roots of Medicago truncatula for the study of nitrogen-fixing and endomycorrhizal symbiotic associations. *Molecular Plant-Microbe Interactions* **14**(6): 695-700.

473 **Brault ML, Petit JD, Immel F, Nicolas WJ, Glavier M, Brocard L, Gaston A, Fouché M,**
474 **Hawkins TJ, Crowet JM, et al. 2019.** Multiple C2 domains and transmembrane region
475 proteins (MCTP s) tether membranes at plasmodesmata. *EMBO reports* **20**(8).
476 **Caillaud M-C, Wirthmueller L, Sklenar J, Findlay K, Piquerez SJM, Jones AME,**
477 **Robatzek S, Jones JDG, Faulkner C. 2014.** The Plasmodesmal Protein PDLP1
478 Localises to Haustoria-Associated Membranes during Downy Mildew Infection and
479 Regulates Callose Deposition. *PLoS Pathogens* **10**(11): e1004496.
480 **Calderan-Rodrigues MJ, Guimarães Fonseca J, De Moraes FE, Vaz Setem L,**
481 **Carmanhanis Begossi A, Labate CA. 2019.** Plant Cell Wall Proteomics: A Focus on
482 Monocot Species, Brachypodium distachyon, Saccharum spp. and Oryza sativa.
483 *International Journal of Molecular Sciences* **20**(8): 1975.
484 **Carvalho BS, Irizarry RA. 2010.** A framework for oligonucleotide microarray preprocessing.
485 *Bioinformatics* **26**(19): 2363-2367.
486 **Cheval C, Samwald S, Johnston MG, De Keijzer J, Breakspear A, Liu X, Bellandi A,**
487 **Kadota Y, Zipfel C, Faulkner C. 2020.** Chitin perception in plasmodesmata
488 characterizes submembrane immune-signaling specificity in plants. *Proceedings of the*
489 *National Academy of Sciences*: 201907799.
490 **Crook AD, Schnabel EL, Frugoli JA. 2016.** The systemic nodule number regulation kinase
491 SUNN inMedicago truncatulainteracts with MtCLV2 and MtCRN. *The Plant Journal*
492 **88**(1): 108-119.
493 **Dorokhov YL, Ershova NM, Sheshukova EV, Komarova TV. 2019.** Plasmodesmata
494 Conductivity Regulation: A Mechanistic Model. *Plants* **8**(12): 595.
495 **Dragićević MB, Paunović DM, Milica, Todorović SI, Simonović AD. 2020.** ragp: Pipeline
496 for mining of plant hydroxyproline-rich glycoproteins with implementation in R.
497 *Glycobiology* **30**(1): 19-35.
498 **Drost H-G, Paszkowski J. 2017.** Biomartr: genomic data retrieval with R. *Bioinformatics*:
499 btw821.
500 **Emanuelsson O, Brunak S, Von Heijne G, Nielsen H. 2007.** Locating proteins in the cell
501 using TargetP, SignalP and related tools. *Nature protocols* **2**(4): 953.
502 **Faulkner C, Petutschnig E, Benitez-Alfonso Y, Beck M, Robatzek S, Lipka V, Maule AJ.**
503 **2013.** LYM2-dependent chitin perception limits molecular flux via plasmodesmata.
504 *Proceedings of the National Academy of Sciences* **110**(22): 9166-9170.
505 **Fernandez-Calvino L, Faulkner C, Walshaw J, Saalbach G, Bayer E, Benitez-Alfonso Y,**
506 **Maule A. 2011.** Arabidopsis plasmodesmal proteome. *PLOS one* **6**(4): e18880.
507 **Gaudioso-Pedraza R, Beck M, Frances L, Kirk P, Ripodas C, Niebel A, Oldroyd GED,**
508 **Benitez-Alfonso Y, de Carvalho-Niebel F. 2018.** Callose-Regulated Symplasmic
509 Communication Coordinates Symbiotic Root Nodule Development. *Curr Biol* **28**(22):
510 3562-3577 e3566.
511 **Ghashghaie J, Brenckmann F, Saugier B. 1991.** Effects of agar concentration on water status
512 and growth of rose plants cultured in vitro. *Physiologia Plantarum* **82**(1): 73-78.
513 **Gíslason MH, Nielsen H, Armenteros JJA, Johansen AR. 2019.** Prediction of GPI-
514 Anchored proteins with pointer neural networks. *bioRxiv*: 838680.
515 **Gopal J, Iwama K. 2007.** In vitro screening of potato against water-stress mediated through
516 sorbitol and polyethylene glycol. *Plant Cell Rep* **26**(5): 693-700.
517 **Grison MS, Kirk P, Brault ML, Wu XN, Schulze WX, Benitez-Alfonso Y, Immel F, Bayer**
518 **EM. 2019.** Plasma Membrane-Associated Receptor-like Kinases Relocalize to
519 Plasmodesmata in Response to Osmotic Stress. *Plant Physiology* **181**(1): 142-160.
520 **Gronnier J, Crowet J-M, Habenstein B, Nasir MN, Bayle V, Hosy E, Platre MP, Gouguet**
521 **P, Raffaele S, Martinez D, et al. 2017.** Structural basis for plant plasma membrane
522 protein dynamics and organization into functional nanodomains. *eLife* **6**.

Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**(18): 2847-2849.

Haney CH, Long SR. 2010. Plant flotillins are required for infection by nitrogen-fixing bacteria. *Proceedings of the National Academy of Sciences* **107**(1): 478-483.

Hernández-Hernández V, Benítez M, Boudaoud A. 2019. Interplay between turgor pressure and plasmodesmata during plant development. *Journal of Experimental Botany*.

Howe KL, Contreras-Moreira B, Nishadi, Maslen G, Akanni W, Allen J, Alvarez-Jarreta J, Barba M, Bolser DM, Cambell L, et al. 2020. Ensembl Genomes 2020—enabling non-vertebrate genomic research. *Nucleic Acids Research* **48**(D1): D689-D695.

Huang D, Sun Y, Ma Z, Ke M, Cui Y, Chen Z, Chen C, Ji C, Tran TM, Yang L, et al. 2019. Salicylic acid-mediated plasmodesmal closure via Remorin-dependent lipid organization. *Proc Natl Acad Sci U S A* **116**(42): 21274-21284.

Hunter K, Kimura S, Rokka A, Tran HC, Toyota M, Kukkonen JP, Wrzaczek M. 2019. CRK2 Enhances Salt Tolerance by Regulating Callose Deposition in Connection with PLDα1. *Plant Physiology* **180**(4): 2004-2021.

Imlau A, Truernit E, Sauer N. 1999. Cell-to-Cell and Long-Distance Trafficking of the Green Fluorescent Protein in the Phloem and Symplasmic Unloading of the Protein into Sink Tissues. *The Plant Cell* **11**(3): 309-322.

Käll L, Krogh A, Sonnhammer EL. 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Research* **35**(suppl_2): W429-W432.

Karimi M, Inzé D, Depicker A. 2002. GATEWAY™ vectors for Agrobacterium-mediated plant transformation. *Trends in plant science* **7**(5): 193-195.

Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *Journal of Molecular Biology* **305**(3): 567-580.

Lee JY, Wang X, Cui W, Sager R, Modla S, Czymmek K, Zybaliov B, van Wijk K, Zhang C, Lu H, et al. 2011. A plasmodesmata-localized protein mediates crosstalk between cell-to-cell communication and innate immunity in Arabidopsis. *Plant Cell* **23**(9): 3353-3373.

Leijon F, Melzer M, Zhou Q, Srivastava V, Bulone V. 2018. Proteomic analysis of plasmodesmata from populus cell suspension cultures in relation with callose biosynthesis. *Frontiers in Plant Science* **9**: 1681.

Li ZP, Paterlini A, Glavier M, Bayer EM. 2021. Intercellular trafficking via plasmodesmata: molecular layers of complexity. *Cellular and Molecular Life Sciences* **78**(3): 799-816.

Liu DY, Smith PM, Barton DA, Day DA, Overall RL. 2017. Characterisation of Arabidopsis calnexin 1 and calnexin 2 in the endoplasmic reticulum and at plasmodesmata. *Protoplasma* **254**(1): 125-136.

Liu N, Shen G, Xu Y, Liu H, Zhang J, Li S, Li J, Zhang C, Qi J, Wang L, et al. 2020. Extensive Inter-plant Protein Transfer between Cuscuta Parasites and Their Host Plants. *Molecular Plant* **13**(4): 573-585.

Mi H, Ebert D, Muruganujan A, Mills C, Albou L-P, Mushayamaha T, Thomas PD. 2021. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research* **49**(D1): D394-D403.

Ning W, Jiang P, Guo Y, Wang C, Tan X, Zhang W, Peng D, Xue Y. 2020. GPS-Palm: a deep learning-based graphic presentation system for the prediction of S-palmitoylation sites in proteins. *Briefings in Bioinformatics*.

Obayashi T, Aoki Y, Tadaka S, Kagaya Y, Kinoshita K. 2018. ATTED-II in 2018: A Plant Coexpression Database Based on Investigation of the Statistical Property of the Mutual Rank Index. *Plant and Cell Physiology* **59**(1): e3-e3.

573 **Park S-H, Li F, Renaud J, Shen W, Li Y, Guo L, Cui H, Sumarah M, Wang A. 2017.**
574 NbEXPA1, an α-expansin, is plasmodesmata-specific and a novel host factor for
575 potyviral infection. *The Plant Journal* **92**(5): 846-861.
576 **Petit JD, Li ZP, Nicolas WJ, Grison MS, Bayer EM. 2020.** Dare to change, the dynamics
577 behind plasmodesmata-mediated cell-to-cell communication. *Current Opinion in Plant*
578 *Biology* **53**: 80-89.
579 **Pierleoni A, Martelli PL, Casadio R. 2008.** PredGPI: a GPI-anchor predictor. *BMC*
580 *bioinformatics* **9**(1): 392.
581 **Reymond P, Kunz B, Paul-Pletzer K, Grimm R, Eckerskorn C, Farmer EE. 1996.** Cloning
582 of a cDNA encoding a plasma membrane-associated, uronide binding phosphoprotein
583 with physical properties similar to viral movement proteins. *The Plant Cell* **8**(12): 2265-
584 2276.
585 **Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015.** limma powers
586 differential expression analyses for RNA-sequencing and microarray studies. *Nucleic*
587 *Acids Research* **43**(7): e47-e47.
588 **Robinson RA, Stokes RH. 1949.** Tables of osmotic and activity coefficients of electrolytes in
589 aqueous solution at 25 C. *Transactions of the Faraday Society* **45**: 612-624.
590 **Rosas-Diaz T, Zhang D, Fan P, Wang L, Ding X, Jiang Y, Jimenez-Gongora T, Medina-**
591 **Puche L, Zhao X, Feng Z. 2018.** A virus-targeted plant receptor-like kinase promotes
592 cell-to-cell spread of RNAi. *Proceedings of the National Academy of Sciences* **115**(6):
593 1388-1393.
594 **Sagi G, Katz A, Guenoune-Gelbart D, Epel BL. 2005.** Class 1 reversibly glycosylated
595 polypeptides are plasmodesmal-associated proteins delivered to plasmodesmata via the
596 Golgi apparatus. *The Plant Cell* **17**(6): 1788-1800.
597 **Simpson C, Thomas C, Findlay K, Bayer E, Maule AJ. 2009.** An Arabidopsis GPI-anchor
598 plasmodesmal neck protein with callose binding activity and potential to regulate cell-
599 to-cell trafficking. *The Plant Cell* **21**(2): 581-594.
600 **Thomas CL, Bayer EM, Ritzenthaler C, Fernandez-Calvino L, Maule AJ. 2008.** Specific
601 targeting of a plasmodesmal protein affecting cell-to-cell communication. *PLoS biol*
602 **6**(1): e7.
603 **Vatén A, Dettmer J, Wu S, Stierhof Y-D, Miyashima S, Yadav SR, Roberts CJ, Campilho**
604 **A, Bulone V, Lichtenberger R. 2011.** Callose biosynthesis regulates symplastic
605 trafficking during root development. *Developmental cell* **21**(6): 1144-1155.
606 **Vu MH, Iswanto ABB, Lee J, Kim J-Y. 2020.** The Role of Plasmodesmata-Associated
607 Receptor in Plant Development and Environmental Response. *Plants* **9**(2): 216.
608 **Wang X, Sager R, Cui W, Zhang C, Lu H, Lee JY. 2013.** Salicylic Acid Regulates
609 Plasmodesmata Closure during Innate Immune Responses in Arabidopsis. *The Plant*
610 *Cell* **25**(6): 2315-2329.
611 **Wilson AK, Pickett FB, Turner JC, Estelle M. 1990.** A dominant mutation inArabidopsis
612 confers resistance to auxin, ethylene and abscisic acid. *Molecular and General Genetics*
613 *MGG* **222**(2-3): 377-383.
614 **Xie Y, Zheng Y, Li H, Luo X, He Z, Cao S, Shi Y, Zhao Q, Xue Y, Zuo Z, et al. 2016.** GPS-
615 Lipid: a robust tool for the prediction of multiple lipid modification sites. *Scientific*
616 *Reports* **6**(1): 28249.
617 **Xue Y, Liu Z, Gao X, Jin C, Wen L, Yao X, Ren J. 2010.** GPS-SNO: Computational
618 Prediction of Protein S-Nitrosylation Sites with a Modified GPS Algorithm. *PLOS one*
619 **5**(6): e11290.

620