



HAL
open science

A high-confidence *Physcomitrium patens* plasmodesmata proteome by iterative scoring and validation reveals diversification of cell wall proteins during evolution

Sven Gombos, Manuel Miras, Vicky Howe, Lin Xi, Mathieu Pottier, Neda S Kazemein Jasemi, Moritz Schladt, J. Obinna Ejike, Ulla Neumann, Sebastian Hänsch, et al.

► To cite this version:

Sven Gombos, Manuel Miras, Vicky Howe, Lin Xi, Mathieu Pottier, et al.. A high-confidence *Physcomitrium patens* plasmodesmata proteome by iterative scoring and validation reveals diversification of cell wall proteins during evolution. *New Phytologist*, 2023, 238 (2), pp.637-653. 10.1111/nph.18730 . hal-04427601

HAL Id: hal-04427601

<https://hal.science/hal-04427601>

Submitted on 30 Jan 2024
















HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

A high-confidence *Physcomitrium patens* plasmodesmata proteome by iterative scoring and validation reveals diversification of cell wall proteins during evolution

Sven Gombos^{1*} , Manuel Miras^{2*} , Vicky Howe^{3*} , Lin Xi¹ , Mathieu Pottier² , Neda S. Kazemineh Jasemi³ , Moritz Schladt² , J. Obinna Ejike² , Ulla Neumann⁴ , Sebastian Hänsch⁵ , Franziska Kuttig³, Zhaoxia Zhang¹, Marcel Dickmanns^{2,6} , Peng Xu⁶, Thorsten Stefan¹, Wolfgang Baumeister⁶ , Wolf B. Frommer^{2,7} , Rüdiger Simon³  and Waltraud X. Schulze¹ 

¹Department of Plant Systems Biology, University of Hohenheim, 70593 Stuttgart, Germany; ²Department of Molecular Physiology, Heinrich Heine University of Düsseldorf, 40225 Düsseldorf, Germany; ³Department of Developmental Genetics, Heinrich Heine University of Düsseldorf, 40225 Düsseldorf, Germany; ⁴Central Microscopy, Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany; ⁵Center for Advanced Imaging, Heinrich Heine University of Düsseldorf, 40225 Düsseldorf, Germany; ⁶Department of Molecular Structural Biology, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany; ⁷Institute for Transformative Biomolecules, Nagoya University, Nagoya 464-0813, Japan

Summary

Author for correspondence:
Waltraud X. Schulze
Email: wschulze@uni-hohenheim.de

Received: 4 November 2022
Accepted: 27 December 2022

New Phytologist (2023) 238: 637–653
doi: 10.1111/nph.18730

Key words: cell wall-modifying enzymes, cell-to-cell communication, evolution of multicellularity, plasmodesmata proteome, *Physcomitrium patens*.

- Plasmodesmata (PD) facilitate movement of molecules between plant cells. Regulation of this movement is still not understood. Plasmodesmata are hard to study, being deeply embedded within cell walls and incorporating several membrane types. Thus, structure and protein composition of PD remain enigmatic. Previous studies of PD protein composition identified protein lists with few validations, making functional conclusions difficult.
- We developed a PD scoring approach in iteration with large-scale systematic localization, defining a high-confidence PD proteome of *Physcomitrium patens* (HC300). HC300, together with *bona fide* PD proteins from literature, were placed in PDDB. About 65% of proteins in HC300 were not previously PD-localized.
- Callose-degrading glycolyl hydrolase family 17 (GHL17) is an abundant protein family with representatives across evolutionary scale. Among GHL17s, we exclusively found members of one phylogenetic clade with PD localization and orthologs occur only in species with developed PD. Phylogenetic comparison was expanded to xyloglucan endotransglucosylases/hydrolases and Exordium-like proteins, which also diversified into PD-localized and non-PD-localized members on distinct phylogenetic clades.
- Our high-confidence PD proteome HC300 provides insights into diversification of large protein families. Iterative and systematic large-scale localization across plant species strengthens the reliability of HC300 as basis for exploring structure, function, and evolution of this important organelle.

Introduction

Plasmodesmata (PD) are important and complex structures present in all land plants, providing connections between plant cells that are otherwise separated by a thick cell wall. During the evolution of multicellular life forms, the exchange of nutrients and information was required to coordinate development and distinct cellular differentiation. The cellular structures and mechanisms facilitating intercellular communication evolved along different paths. In animals, different types of gap junctions mediate the exchange of small molecules between cells, while primary cilia coordinate cell-to-cell signaling (Bangs &

Anderson, 2017; Song *et al.*, 2018; Anvarian *et al.*, 2019). In plants, PD evolved to facilitate translocation of both nutrients and signaling molecules in the presence of a cell wall. While the structure and function of gap junctions and primary cilia are rather well understood, PD structure and function remain somewhat enigmatic.

Plasmodesmata were serendipitously discovered in the 1800s by the botanist Eduard Tangl while staining cell wall cellulose using organic pigments. To his surprise, he observed cytoplasmic connections between plant cells in the endosperm (Tangle, 1879). Similar structures were subsequently identified in many other plant species. Notably, analogous structures likely evolved independently at least six times, underlining their necessity for multicellularity (Raven, 2007).

*These authors contributed equally to this work.

Plasmodesmata connect neighboring cells by forming sophisticated channels lined by a contiguous plasma membrane. A special feature of land plant PD is the desmotubule, a strand of apparently tightly appressed ER that extends through the cytoplasmic sleeve (Ding *et al.*, 1992a,b, 1996). Since their discovery, PD are considered as pores between cells, enabling diffusion of small molecules below a certain size (Wolf *et al.*, 1989). They also seem to enable passage of macromolecules, such as RNAs, peptides, and proteins that are vital for plant development and physiological processes (Lucas *et al.*, 1995, 2009; Kim *et al.*, 2005). Recently, tethering between the plasma membrane and the desmotubule was suggested as a mechanism to control PD pore size and multiple C2 domains and transmembrane region proteins (MCTP) were shown to play a major role in this process (Brault *et al.*, 2019). However, mechanisms of transport through PD are still unclear, as is their precise structural composition.

Proteomic analyses of PD enrichments provide a glimpse of the protein composition of these complex channels. To date, PD proteomes have largely been derived through biochemical enrichment of cell wall preparations, for example in *Arabidopsis* (Bayer *et al.*, 2006; Fernandez-Calvino *et al.*, 2011; Kraner *et al.*, 2017), *Populus trichocarpa* (Leijon *et al.*, 2018), or *Nicotiana benthamiana* (Park *et al.*, 2017). Some of these proposed PD proteins were also found enriched in detergent-resistant membrane fractions. This may be due to the plasma membrane in PD being characterized by a distinct lipid and sterol composition (Grison *et al.*, 2015a; Yan *et al.*, 2019). Out of several thousand candidates identified in such proteomics analyses, only a small subset has actually been validated with PD localization *in planta*, and the fraction of false positives and contaminants in such lists is expected rather high.

Biochemical preparations of PD obtained, for example, through cell wall enrichment or differential centrifugations contain a large proportion of copurifying proteins from cellular domains other than PD. Therefore, quantitative methods distinguishing PD proteins from non-PD proteins are required. In previous studies, comparative quantitative assessments of PD proteomes were performed, for example, by determining an enrichment ratio of proteins identified in PD compared with cell walls to define a 'core PD proteome' (Brault *et al.*, 2019). In another study, a reduced abundance of PD proteins was observed in the loss of function mutant of Choline Transporter 1 (*cher1-1*) and this observation was used to rank PD components (Kraner *et al.*, 2017). A further approach characterizing PD proteomes explored protein recruitment to PD in response to viral infection (Park *et al.*, 2017). Viral movement between cells occurs through PD and thus proteins recruited or depleted from PD in response to viral infection were considered as candidate PD proteins. However, due to limited validation and limited number of species that have been used in PD studies, a reliable PD proteome with the probabilistic classification of core PD components and copurifying proteins remains to be resolved.

Here, we used the bryophyte *Physcomitrium patens* as a model organism to establish a PD proteome by an iterative discovery-verification workflow using experimental protein enrichment and a feature scoring algorithm in iteration with *in planta*

localization. Bryophytes were among the first evolving land plants (Coudert *et al.*, 2019), thus requiring transport and communication between cells despite rigid cell walls. Our results give insights into the large PD protein families, such as cell-wall-modifying enzymes and Exordium (EXO) proteins. These protein families are evolutionarily conserved across species, indicating their importance in the evolution of multicellularity.

Materials and Methods

Plant material and culture conditions

Physcomitrium patens (Hedw.) ecotype Grandsden cultures were obtained from the International Moss Stock Center (IMSC, Freiburg, Germany). *Physcomitrium patens* was cultured on modified BCD medium containing 5 mM diammonium tartrate (Cove *et al.*, 2009; Supporting Information Methods S1). *Nicotiana benthamiana* seeds were soaked in water for 24 h at 20°C and then grown on peat moss substrate in a glasshouse for 3 wk under long-day conditions (Methods S1).

Preparation of plasmodesmata

The PD preparation workflow (Fig. S1A) was based on two previously published methods (Faulkner & Bayer, 2017; Kraner *et al.*, 2017) and optimized for *P. patens* tissue. First step in PD (Fig. S1B) preparation workflow was disruption of plant tissue with a Potter-Elvehjem Homogenizer in presence of high detergent concentrations (150 mM NaCl, 10 mM EDTA, 1 mM DTT, 1% (v/v) Triton X-100, 0.5% (v/v) sodium deoxycholate, 0.1% (w/v) SDS, 10% (v/v) glycerol, and 100 mM Tris-HCl, pH 8). This gently perforated cell walls and removed organelles while leaving cell walls mostly intact to preserve PD structures (Fig. S1C). Two subsequent rounds of grinding cell wall fractions in liquid nitrogen efficiently removed additional soluble protein contaminants and prepared cell walls for digestion by breaking up cell clusters (Fig. S1D). Washing with lower concentrations of detergents (150 mM NaCl, 10 mM EDTA, 0.1% (v/v) Triton X-100, 10% (v/v) glycerol, and 100 mM Tris-HCl, pH 8) prevented premature solubilization of PD membranes at this point. Cell wall fragments were digested using driselase. After ultracentrifugation, the resulting PD pellet contained small membranous components and protein aggregates with very little remains of cell walls (Fig. S1E). Along the PD preparation workflow, four protein fractions were collected as input for later PD scoring: cell wall proteins (CW), total of microsomal membranes (Mic), final pellet of PD-enriched proteins (PD), and a total cell extract (TC).

Protein isolation from other sources

Total protein was extracted in a urea thiourea environment (He *et al.*, 2021). Microsomal fraction isolation followed established protocols (Pertl *et al.*, 2001). Cell wall extracts were obtained after cell homogenization with a Potter-Elvehjem PTFE tissue grinder in a high-detergent environment and centrifugation at 1000 g (Kraner *et al.*, 2017).

LC–MS/MS analysis of peptides

Proteins were trypsin-digested and desalted (Hughes *et al.*, 2019), and peptide mixtures were analyzed on a nanoflow UPLC-coupled Orbitrap hybrid mass spectrometer (Q-Exactive HF; Thermo Scientific, Waltham, MA, USA; Methods S1). Spectra were matched against *P. patens* proteome (Ppatens_318_v.3.3.-protein.fasta, 87 533 entries) using MAXQUANT v.2.0.3.0 (Cox & Mann, 2008). Mass spectrometry proteomics data were deposited to ProteomeXchange Consortium via the PRIDE partner repository (Deutsch *et al.*, 2017) with identifier PXD032820.

PD score

The PD score is the sum of two components, enrichment score (Eqn 1) and feature score (Eqns 2, 3; Fig. S2A). LFQ values from the TC, MIC, CW, and PD fractions derived from MAXQUANT (Tyanova *et al.*, 2016) were \log_2 -transformed before further calculations. Fraction means were calculated and normalized using z -score (Fig. S2B):

$$\text{Enrichment Score} : \sum_{x \in X} \frac{z \text{ Score}(\log_2(\text{LFQ intensity PD})) - z \text{ Score}(\log_2(\text{LFQ intensity } x))}{\text{Eqn 1}}$$

$X = \{\text{MIC, TC, CW}\}$.

Feature scores were based on frequency of protein functions, and PFAM domains (<https://pfam.xfam.org/>) found in the reference data set and validated PD candidates. The feature score is the sum of feature frequencies of PD protein-associated features found in the scored protein (Fig. S2C):

$$\text{Feature Frequency} : \sum_{i=1}^n \frac{\text{count}(\text{feature}_i)}{\max(\text{count}(\text{feature}_i))} \quad \text{Eqn 2}$$

$$\text{Feature Score} : \sum_{i=1}^n \text{feature}_i \text{ frequency} \quad \text{Eqn 3}$$

Quantification of PD candidate localization

Confocal images were screened visually for colocalization of proteins of interest and aniline blue. Candidates showing colocalization with aniline blue were quantitatively scored by the so-called PD index (Grison *et al.*, 2019). To minimize bias in quantification, we used a semi-automated approach of an in-house-generated macro for Fiji software package (Schindelin *et al.*, 2012; Methods S1).

Phylogenetic analysis

GHL17s, XTHs, and EXOs protein sequences were retrieved from PHYTOZOME, ENSEMBL, PHYCOCOSM, and <http://genome.microbedb.jp/klebsormidium> databases (Goodstein *et al.*, 2012; Hori *et al.*, 2014; Howe *et al.*, 2020; Grigoriev *et al.*, 2021). All

protein sequences were subjected to protein domain prediction at INTERPRO server (Hunter *et al.*, 2009). Maximum likelihood trees were generated using the NGPHYLOGENY.FR tool (Lemoine *et al.*, 2019) and visualized and edited in iTOL (Letunic & Bork, 2019).

Results

Detection of PD in *P. patens*

To evaluate the presence and type of PD in *Physcomitrium*, transmission electron micrographs of longitudinal and cross sections were generated from protonemata, in which PD were randomly distributed along the cell-to-cell interface (Fig. 1a,b). Simple PD morphotypes were found at interfaces between protonema cells at a relatively high density of 11.3 (± 1.4) PD μm^{-2} (Fig. 1c) compared with PD densities in Arabidopsis roots, which ranged from 2.5 to 12.5 PD μm^{-2} , depending on the cell-type interface (Zhu & Lucas, 1998). Plasmodesmata averaged to 362.6 nm (± 119.2 nm) in length and 29.7 nm (± 4.9) in diameter. *Physcomitrium* protonema PD were clearly longer than what has been published for Arabidopsis suspension cells or roots (Grison *et al.*, 2015b; Nicolas *et al.*, 2017) because of thicker cell walls in measured *Physcomitrium* protonema. The diameter was in a similar range as described previously for Arabidopsis. (Fig. 1d) We achieved an enrichment of common PD proteins from *Physcomitrium* with concurrent depletion of typical contaminants. Proteins expected in PD, such as C2-domain-containing proteins (Brault *et al.*, 2019), were found with increased abundance in PD fractions compared with TC or CW fractions (Fig. 1e). Photosynthetic proteins in PD pellets were found in reduced abundance, while non-PD membrane proteins, such as plasma membrane ATPases, were most abundant in the microsomal fraction (Fig. 1e), and not in PD.

Our PD preparation workflow (Fig. S1) identified a total of 870 proteins in the final PD pellet. A comparison of published PD pellet proteomes from Arabidopsis (Bayer *et al.*, 2006; Fernandez-Calvino *et al.*, 2011; Kraner *et al.*, 2017; Brault *et al.*, 2019), *P. trichocarpa* (Leijon *et al.*, 2018), and *N. benthamiana* (Park *et al.*, 2017) revealed large overlap of 422 Arabidopsis orthologs in PD pellets of *Physcomitrium* and Arabidopsis (Fig. 1f). Besides typical PD protein functions in cell wall modification or C2-domain-containing proteins, there was a high fraction (22%; $n = 38$) of typical contaminants, such as ribosomes and photosynthetic proteins from light-harvesting complexes and Calvin-Benson Cycle among orthologs found in at least three species. The fact that such nontarget proteins made up of a high percentage of the overlap in PD protein lists, suggesting that definition of PD proteins, solely based on discovery in multiple proteomics data sets, lacks specificity and requires additional criteria for a robust definition of PD composition. Therefore, to identify high-confidence PD protein candidates, a data-driven proteomics-based approach was used in iteration with systematic verification of PD localization.

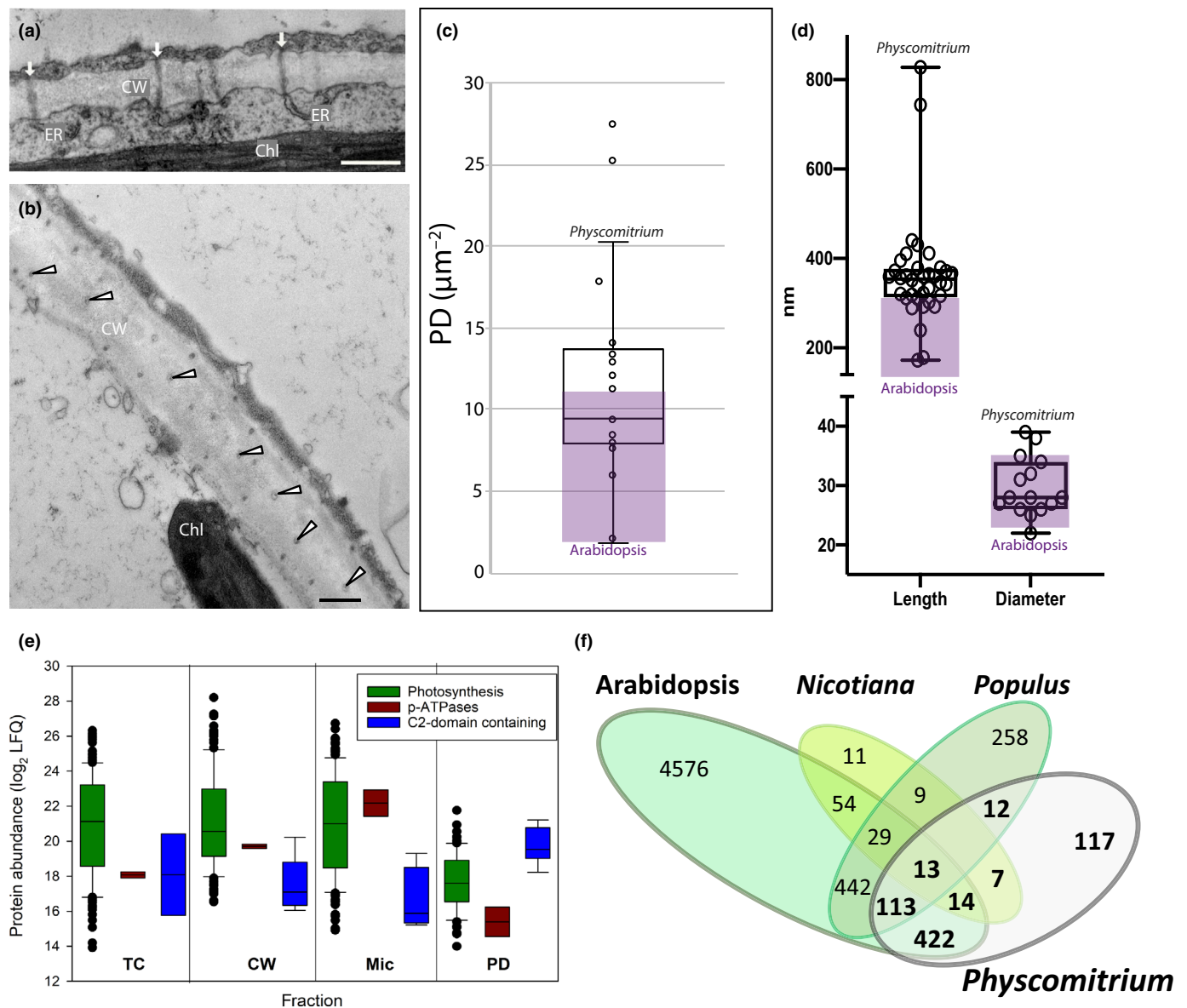


Fig. 1 Plasmodesmata (PD) from *Physcomitrium patens*. (a) Longitudinal section through simple PD between two protonema cells. Desmotubule (pointed with arrows) is visible in the central part of the lumen as a higher electron density. Bar, 500 nm. (b) Cross-section through PD between two protonema cells. Desmotubule (pointed with arrowheads) is visible in the central part of the lumen as a higher electron density. Bar, 500 nm. (c) PD density in protonema cell-to-cell interface as determined by TEM analysis. (d) Measurements of PD length and density based on TEM images. In (c) and (d), purple boxes indicate the ranges of respective values as found in the literature for *Arabidopsis*. The boxes and whiskers represent the 25th to 75th percentiles and minimum–maximum distributions of the data, respectively. Horizontal lines represent the median and open circles represent individual values. (e) Depletion of nontarget proteins (e.g. photosynthetic proteins) and enrichment of target proteins (e.g. C2-domain-containing proteins) in PD preparations. The family of plasma membrane H^+ -ATPases (p-ATPases) was used as control for non-PD plasma membrane. The boxes and whiskers represent the 25th to 75th percentiles and minimum–maximum distributions of the data, respectively. Horizontal lines represent the median and open circles represent individual values. CW, cell wall; Mic, microsomal membranes; PD, plasmodesmata pellet; TC, total cell extract. (f) Overlap of identified protein in PD pellets from *Physcomitrium* and published data sets from *Arabidopsis* (Bayer *et al.*, 2006; Fernandez-Calvino *et al.*, 2011; Kraner *et al.*, 2017; Brault *et al.*, 2019), *Populus trichocarpa* (Leijon *et al.*, 2018) and *Nicotiana benthamiana* (Park *et al.*, 2017). For *Arabidopsis*, the protein lists from four different studies were combined. For this comparison, *Arabidopsis* orthologs from respective species were used based on mappings available from PHYTOZOME (Goodstein *et al.*, 2012).

Definition of a high-confidence PD proteome using an iterative scoring system

To reliably distinguish actual PD proteins from non-PD proteins in biochemical PD preparations, we developed a PD score

(Fig. S2), which is based on protein abundance data along the PD preparation workflow and on annotated features of a reference data set containing confirmed PD proteins from public sources. We compiled published proteome data sets in a relational database (PDDb <http://pddb.uni-hohenheim.de>, Fig. S3)

and complemented this with 70 individual protein entries with literature evidence of experimental PD localization (Table S1). These proteins were used as *bona fide* confirmed PD proteins and were further expanded by 115 proteins published as the PD core proteome (Brault *et al.*, 2019) based on their high enrichment in PD fractions compared with cell wall or membrane fractions. Additional reliable PD proteins from published sources included 152 proteins with greater than twofold depletion in the *cher1-1* mutant (Kraner *et al.*, 2017), and 56 proteins (52 unique Arabidopsis orthologs) with more than fivefold enrichment or depletion in PD in response to viral infection (Park *et al.*, 2017). Respective threshold values were chosen upon re-evaluation and by applying a two-standard deviation width to histograms of published quantitative data (Table S1). The list of 201 proteins published as PD-enriched from *P. trichocarpa* (Leijon *et al.*, 2018) was not included, since no enrichment ratios were published, making a quantitative re-evaluation difficult. Thus, the first PD scoring training data set comprised 338 unique Arabidopsis orthologs (Table S1) coming from complementary data sources (Fig. 2a).

The PD score was then calculated for each identified protein as the sum of an enrichment score and a feature score (Fig. S2). The enrichment score was derived from the abundance of proteins identified in PD fractions compared with the corresponding abundance in CW, MIC, and TC. The feature score was calculated from frequencies of PFAM domains (<http://www.pantherdb.org>) and functional classifications by MAPMAN (Thimm *et al.*, 2004) in the reference data set of 338 *bona fide* PD proteins. Consequently, proteins received higher PD scores when they were enriched in the PD fraction and in addition shared features with reference PD proteins.

For 863 proteins found in the *Physcomitrium* PD pellet, we were able to calculate an enrichment score, and for 200 proteins the enrichment score was > 1 (Fig. 2b). The full PD score was calculated for 2053 out of the 4996 identified proteins (Table S2). There were more proteins with a PD score (Fig. 2c) than with an enrichment score because also proteins not quantified by an enrichment score were awarded a PD score based on their features, and thus potentially qualified as part of the PD proteome. Out of this first PD score distribution, 147 proteins were selected and the corresponding cDNAs were cloned to validate PD localization by transient expression in *N. benthamiana* (Fig. 2d). Proteins which localized to PD were added to the reference data set in PDDB for PD scoring. As a result of an iterative second PD scoring based on the updated PD protein list, most of the experimentally localized PD proteins increased in PD score, as did other proteins with similar features (Fig. 2e). This visualizes the key concept of iterative PD scoring consisting of (1) protein abundance enrichment in PD fractions, (2) identification of proteins with similar features as known reference proteins, and (3) re-evaluating their PD score based on features of experimentally validated new PD-localized proteins.

As a result of an iterative second PD scoring based on the updated PD protein list, most of the experimentally localized PD proteins increased in PD score, as did other proteins with similar features (Fig. 2e). This visualizes the key concept of iterative PD scoring consisting of (1) protein abundance enrichment in PD fractions, (2) identification of proteins with similar features as known reference proteins, and (3) re-evaluating their PD score based on features of experimentally validated new PD-localized proteins.

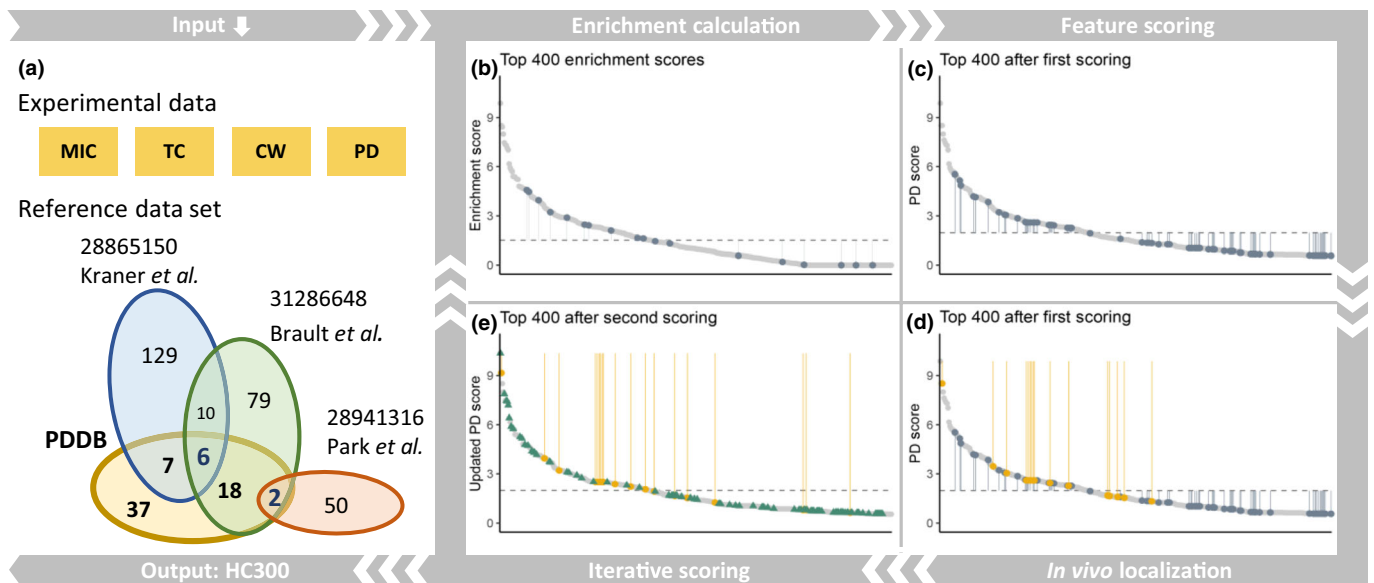


Fig. 2 Plasmodesmata (PD) scoring process visualized with the top 400 best-scoring proteins. (a) Input data consisting of protein abundances in microsomal (MIC), total cell (TC), cell wall (CW), and PD fractions and the *bona fide* PD reference data set of PD proteins characterized by quantitative proteomics (Kraner *et al.*, 2017; Park *et al.*, 2017; Brault *et al.*, 2019) (blue, green, red) and the additional PD-localized proteins in PDB from different literature sources (yellow). This data set was used as a PD positive set in the first round of PD scoring. (b–e) The components of the PD scoring process. The y-axes show enrichment score or PD score. Each of the top 400 scoring proteins is represented by a light gray dot. Proteins are ordered by score in descending order from left to right. Dark gray dots and vertical lines mark the scores of proteins from the reference data set. Yellow dots and vertical lines mark the scores of proteins that have been verified by *in vivo* localization. Gray dashed lines show mean scores. (b) Calculation of the enrichment score that is solely based on proteomics data. (c) PD scores after including a first round of feature scoring. Higher values represent higher likelihood for PD presence. (d) Scores of newly PD-localized proteins are highlighted (yellow dots with vertical yellow lines) before iterative scoring. (e) Iterative second PD scoring including proteins experimentally localized to PD across the full range of PD candidate proteins. Green triangles indicate proteins which were not part of the reference data set but increased in score after iterative scoring.

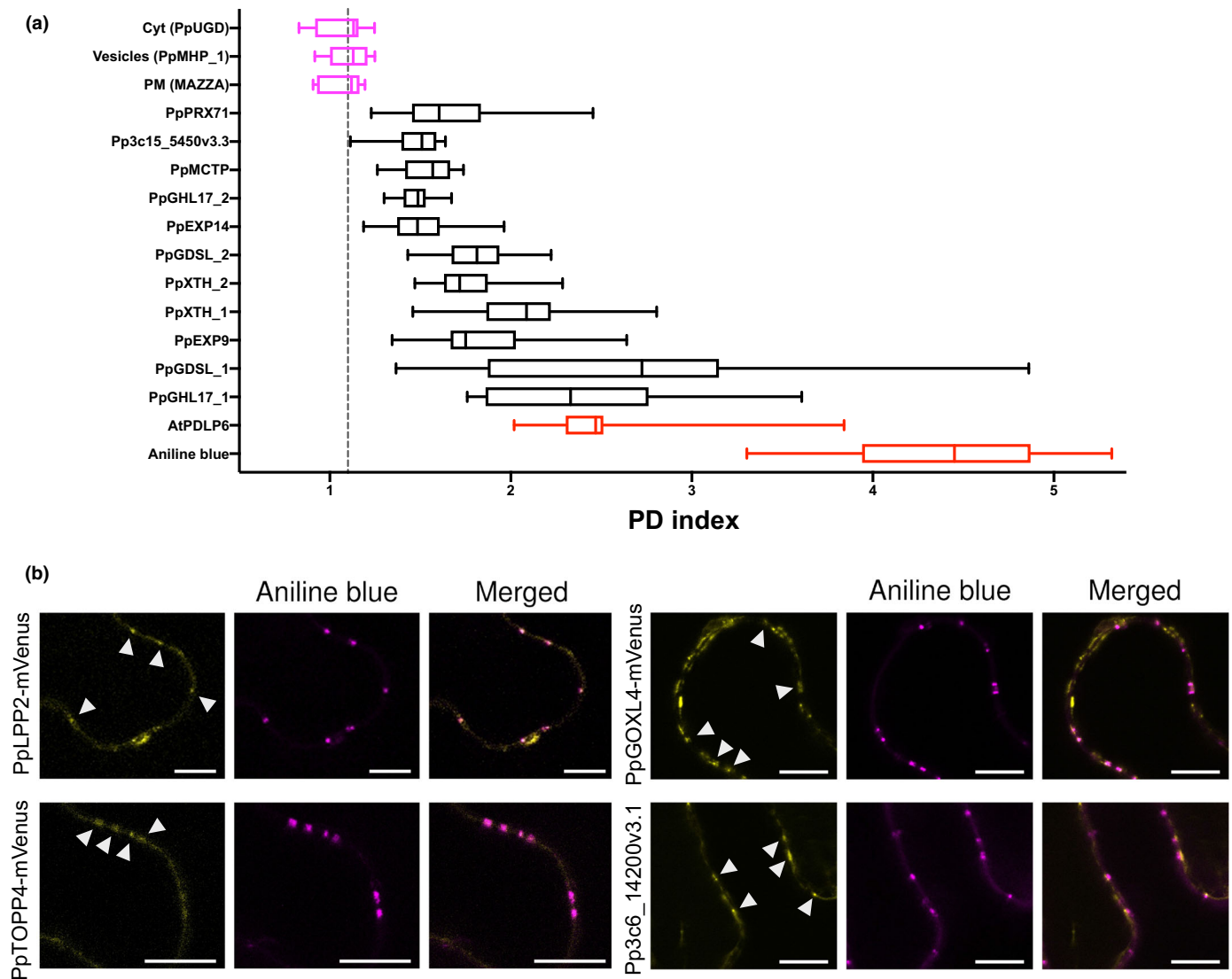


Fig. 3 Selected candidate proteins from *Physcomitrium patens* plasmodesmata (PD) proteome located at PD. Proteins were C-terminally fused to mVenus and transiently expressed under a β -estradiol inducible promoter. Subcellular localization was observed in epidermal cells of *Nicotiana benthamiana* at the confocal microscope. (a) Plasmodesmata index (PD index) of selected candidates, mainly known PD proteins, showed an enrichment at PD similar to well-established PD marker Arabidopsis Plasmodesmata Located Protein 6 (PDL6). Plasmodesmata markers aniline blue and PDL6, and non-PD-localized proteins Arabidopsis MAZZA (PM), PpMHP_1 (Pp3c16_460v3.7; vesicles), and PpUGD (cytoplasm) are shown as red and magenta boxplots, respectively. In the boxplots, the median value is denoted by a line and the boxes and whiskers represent the 25th to 75th percentiles, and minimum to maximum distribution. Dashed lines represent the threshold at which a protein is considered PD-enriched and was set at 1.1 corresponding to the highest mean score of the non-PD-localized proteins PpMHP_1. (b) Examples of single optical sections at cell-to-cell interface showing the colocalization of four novel PD protein candidates with PD marker aniline blue (white arrowheads). Bars, 10 μ m.

Validation of PD localization using confocal microscopy

Putative PD-localized proteins were systematically screened for colocalization with the callose-staining dye aniline blue as a marker of PD pit fields. Candidates were initially screened for PD localization and given a preliminary validation score (Table S3). Proteins not colocalizing with aniline blue (112 candidates) were considered non-PD proteins for the purpose of PD scoring iteration. For the 36 proteins colocalizing with aniline blue, the degree of enrichment at PD compared with PM was calculated using a semi-automated PD indexing script. Proteins with a PD index > 1.1 (the highest mean score of the negative controls,

Figs 3a, S4) were considered enriched at PD compared with bulk PM. For comparison, aniline blue and the Arabidopsis protein Plasmodesmata Located Protein 6 (PDL6), a protein known to localize exclusively to PD (Thomas *et al.*, 2008), were used as positive controls (Fig. S4). Aniline blue had a PD index of 4.4, while PDL6 had a PD index of 2.5. Similarly, three proteins from the initial screen that did not show any obvious colocalization with aniline blue, localizing to either PM, vesicles, or cytoplasm, were used as negative controls and had PD indices of *c.* 1.

Due to deficiencies in image quality, the PD index could not be calculated for all candidates deemed PD-localized by qualitative means. We calculated PD indices (range: 1.4–2.7) for 30 of

the 36 candidates that colocalized with aniline blue (Fig. 3b). For the purposes of iterating the PD scoring, all 36 candidates, including those not assigned a PD index, were treated as PD-localized proteins. Among this set of proteins were 20 proteins (56%) that had not previously been considered as PD proteins in other species (Table S4). Five members of the glyoxal oxidase family (Pp3c12_15000; Pp3c3_25264; Pp3c4_20120; Pp3c2_25840; Pp3c22_8310, and pGGOXL4, Fig. 3b), a serine–threonine phosphatase (Pp3c1_370, TOPP4, Fig. 3b), an uncharacterized transmembrane proteins (e.g. Pp3c6_14200, Fig. 3b), and members of the EXO protein family are examples of novel PD-localized proteins, which ranked among the top 10% proteins in the PD scoring (Table S2). For iterative scoring, the feature score component was modified by the contribution of these PD-localized proteins.

Refinement of a *Physcomitrium* high-confidence PD proteome

The iterative PD scores of proteins confirmed as PD localized in *N. benthamiana* clearly separated from PD scores of proteins that could not be localized to PD (Figs 4a, S5A). The PD score was then used to separate high-confidence PD proteins from other proteins (Figs 4a, S5A). A threshold PD score of 0.7 was chosen by carefully balancing the numbers of proteins not found with PD localization among the top-ranking proteins against proteins found with PD localization among lower ranking proteins. Among those 335 PD protein candidates with PD scores > 0.7 (Table S2) were 16 of the cloned and tested proteins, which did not show PD localization in *N. benthamiana*. Although there may be many valid reasons for failed PD localization of these proteins, we termed these as false positives and they made up 3.9% of the top 335 ranking PD protein subset. In turn, 28 proteins (1.6%) localizing to PD in *N. benthamiana* were found among the 1718 proteins with a calculated PD score of < 0.7, which were considered as false negatives.

The 335 PD protein candidates were further refined by considering the PD score difference between the second and the first round of PD scoring, thus taking advantage of the iterative improvement of PD scores by the contribution of novel PD protein functions after experimental localization. For protein functions expected among PD proteins, such as cell wall modifying proteins and C2-domain-containing proteins, but also for proteins with functions not thus far associated with typical PD proteins, the PD score improved throughout the iterative scoring process leading to high average PD score differences (Fig. S5B). In turn, for typical contaminant proteins, such as ribosomes, catalases, and photosynthetic proteins, the PD score was downgraded by the iterative process resulting in large negative PD score differences (Fig. S5B). By applying the PD score difference as further refinement, we excluded proteins that had a negative PD score difference (PD score difference < -0.13; Fig. 4b), resulting in *P. patens* high-confidence PD Proteome (HC300).

Arabidopsis orthologs were used to compare *Physcomitrium* PD proteomes with previously published PD proteomes of other species (Fig. 4c,d). The whole PD proteome dataset contained

1436 Arabidopsis orthologs (Fig. 4c). Of the 292 proteins contained in the HC300, 273 proteins were annotated with Arabidopsis orthologs. After exclusion of photosynthetic proteins as typical contaminants, HC300 contained 39 proteins annotated to 17 orthologs of confirmed PD proteins in other species and 249 other proteins mapping to 191 orthologous proteins which were previously not considered as high-confidence PD proteins (Fig. 4d,e). Upon stepwise refinement of the *Physcomitrium* HC300 PD proteome based on PD score and PD score difference, the overlap with other quantitative PD proteomes (Kraner *et al.*, 2017; Park *et al.*, 2017; Brault *et al.*, 2019) became smaller at each refinement step (Fig. 4c,d). The refinement steps efficiently removed typical contaminants, such as photosynthetic proteins and ribosomes. The proportion of contaminant proteins was rather high (25%) among the proteins not considered high-confidence PD proteins while in HC300 the fraction of putative contaminants was reduced to 16% (Fig. 4e). Among the novel proteins in HC300 were five out of 17 EXO/Exordium-Like (EXL) proteins, whose Arabidopsis orthologs were previously shown to be located in cell walls (Schröder *et al.*, 2009). HC300 contained five out of seven MCTP protein family members (Pp3c27_520, Pp3c14_25200, and Pp3c10_11080 in Fig. S4, Pp3c16_9250 and Pp3c27_540) as well as two other C2-domain-containing proteins (Pp3c9_16970 and Pp3c9_12510). Three receptor kinases (Pp3c25_12800, Pp3c1_10970, and Pp3c19_8660), a phosphatase (Pp3c1_370), and a transmembrane scaffold protein (Pp3c4_1550) were further top-ranking proteins of HC300 (Table S2).

Within HC300, 62% of the proteins were found to have a signal peptide, 76% had at least one transmembrane domain, and 57% of the proteins were predicted to have intrinsically disordered regions (IDR). Cell wall proteins (84 proteins; bins 10; $P = 5.5 \times 10^{-26}$) were highly over-represented, as were glyoxal oxidases (bin 24; $P = 8.9 \times 10^{-13}$; Fig. S5C). In agreement with the overall high content of cell wall-related proteins, HC300 was significantly enriched with PFAM domains of glycosyl hydrolase family 17 (GHL17) proteins (PF00332 and GHL17s), xyloglucan endotransglucosylases/hydrolases (PF06955, XTHs), and pectin esterase (PF1095). The PFAM X8 domain (PF07983), proposed with a role in protein targeting to PD through its signal sequences for a glycosylphosphatidylinositol (GPI) linkage, was found 14 times (5%; Fig. S5D; Simpson *et al.*, 2009).

Phylogenetic distribution of glycosyl hydrolase family 17

Cell wall proteins made up the largest functional group in HC300. We identified 16 proteins annotated as β -1,3-glucanases containing a GHL17 domain (Table S2). GHL17s are callose-degrading enzymes located at the cell wall and associated with PD (Levy *et al.*, 2007; Zavaliev *et al.*, 2013; Benitez-Alfonso, 2014). Members of the GHL17 family were previously identified in PD proteome lists from Arabidopsis, *P. trichocarpa*, and *N. benthamiana*, and 10 of them overlapped in at least two species. PpGHL17_16 (Pp3c16_16680) was the only member of the GHL17 family, for which orthologs were found in PD proteome lists of all four species. *Physcomitrium patens* HC300

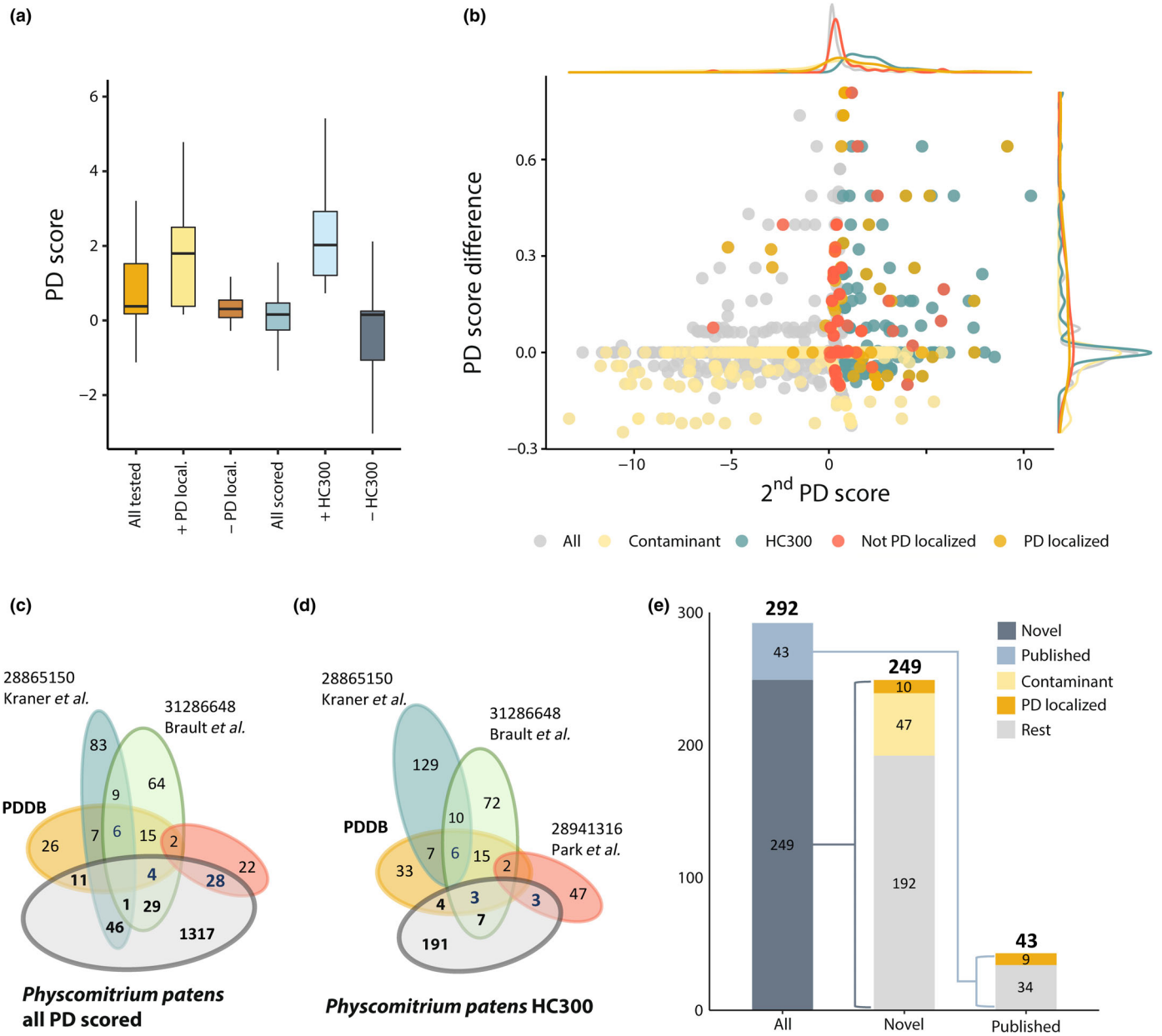


Fig. 4 Characterization of *Physcomitrium* plasmodesmata (PD) scored proteome. (a) Boxplots of PD scores (2nd scoring) of all proteins that were tested for localization at PD in *Nicotiana benthamiana* combined ('All tested') and separated into the proteins localized ('+ PD local.') and nonlocalized to PD ('- PD local.') and boxplots of PD scores (2nd scoring) of all scored proteins combined ('All scored') and separated into proteins that were included into the refined HC300 proteome ('+ HC300') and excluded of the HC300 proteome ('- HC300'). Outliers were removed to improve clarity. See Supporting Information (Fig. S5A) for a comprehensive version of the plot. Horizontal lines represent the median value, whiskers indicate 25th and 75th percentiles. (b) PD score (> 0.7) and PD score difference (> -0.13) as criteria to define the high-confidence *Physcomitrium* PD proteome. Light gray: proteins not included in the refined HC300 proteome. Light yellow: photosynthetic proteins and ribosomal proteins as typical contaminants. Blue, refined HC300 proteome; red, proteins that did not localize at PD in *N. benthamiana*; yellow, proteins with PD localization in *N. benthamiana*. (c) Venn diagram containing the nonrefined set of Arabidopsis orthologs of all PD-scored proteins and the *bona fide* PD reference data set. (d) Venn diagram containing the Arabidopsis orthologs of the refined high-confidence PD proteome HC300 and the *bona fide* PD reference data set. (e) Overview of the HC300 based on *Physcomitrium* protein identifiers classified as 'published' and novel' proteins. For each group, contaminants (e.g. plastidial proteins) and experimentally PD-localized proteins were specifically highlighted.

contained three GHL17 orthologs (PpGHL17_2, PpGHL17_7, and PpGHL17_9; Table S5), which showed no overlap with PD proteomes from angiosperms (Fig. 5a).

Based on sequence similarity, GHL17 proteins of representative species from charophytic algae and embryophytes, including

P. patens, were classified into three major clades (Doxey et al., 2007; Gaudioso-Pedraza & Benitez-Alfonso, 2014). To get insights into their role in the evolution of multicellularity, a new phylogenetic analysis of the GHL17 family was performed including *P. patens*, Arabidopsis, *P. trichocarpa*, *Oryza sativa*, and

representative species from green algae (Figs 5b, S6A). Unexpectedly, no GHL17 orthologs could be identified in the unicellular algae *Mesostigma viride* and *Chlamydomonas reinhardtii*, which belong to the *Streptophyta* and *Chlorophyta* clades and thus are ancestors to embryophytes. This suggests that the GHL17 family emerged in multicellular organisms.

GHL17s from charophytic algae (*Klebsormidium nitens*, *Chara braunii*, and *Nepenthes mirabilis*) were present only near the base of the gamma clade, likely representing the ancestral type of GHL17 proteins (Fig. 5b). The alpha clade only contained GHL17s from embryophytes, suggesting that this group appeared during land colonization. Three of the Arabidopsis GHL17 proteins from the alpha clade were previously described to localize at PD, suggesting that phylogenetic distribution might correlate with the subcellular localization either at PD or apoplasm (Levy *et al.*, 2007; Benitez-Alfonso *et al.*, 2013). To further test this hypothesis, we studied the subcellular localization of six *P. patens* GHL17 representatives, four from the alpha clade and two from the beta clade by transient expression in *N. benthamiana*.

Transient expression of beta clustered PpGHL17_28-mVenus and PpGHL17_30-mVenus showed apoplasmic localization (Fig. S6B), whereas PpGHL17_1-mVenus, PpGHL17_7-mVenus, PpGHL17_8-mVenus, and PpGHL17_15-mVenus from the alpha clades were found with a punctate pattern at the cell periphery and overlaid with aniline blue (Fig. 5c). Since transient expression assays can lead to mislocalization, and PD targeting mechanisms might not be conserved between *N. benthamiana* and *P. patens*, we obtained a *P. patens* stable transgenic line expressing UBQ:PpGHL17_1-mVenus to further validate PD localization. The subcellular localization of PpGHL17_1-mVenus in the stable line confirmed PD localization found in transient expression assays (Fig. 5d,e). In addition, when overexpressing PD-localized PpGHL17_1 in *Physcomitrium*, diffusion of the PD mobile dye carboxyfluorescein diacetate (CFDA) increased compared with wild-type ecotype (Fig. 5f,g) suggesting GHL17s play a role in regulating PD opening through the degradation of callose. Interestingly, expressing beta-clade PpGHL17_30, which did not localize to PD (Fig. S6), did not enhance CFDA diffusion between neighboring protonema cells (Fig. 5f,g). This points to a functional diversification of PD and non-PD-localized members of the GHL17 family or that only PD-localized PpGHL17s can degrade callose at PD. In recent parallel work, GHL17_4 (Pp3c11_25840) and GHL17_18 (Pp3c16_15860), also members of the GHL17 alpha clade, were localized to PD in *P. patens* (Johnston *et al.*, 2023), thus adding further evidence for PD localization of alpha clade GHL17s.

In search of a potential PD localization signal, we modeled structures of alpha-clade PpGHL17s (Fig. 5h). The sequence motif IFALFENE(N) was over-represented among this group of GHL17s and was absent in the members of the beta clade. However, this motif was predicted to be only partially surface-exposed as part of the 40 Å cleft of the glycosyl hydrolase domain and is therefore more likely to be related to enzyme function rather than localization (Fig. 5g). Another interesting observation was that all GHL17s from the alpha clade contained a beta tunnel and cleft

domain, while the distribution of other structural features, such as an X8 domain and GPI-anchor signal, was not equally present (Table S5).

Xyloglucan endotransglucosylases/hydrolases

XTHs are enzymes involved in xyloglucan metabolism and are important for the control of cell wall strength as well as extensibility during growth and development. In Arabidopsis, XTHs were shown to be upregulated in the *ise2* mutants, which displayed an increase in cell-to-cell transport of fluorescent probes (Burch-Smith *et al.*, 2011). We found 12 members of the XTH family in HC300, six of which were confirmed as PD-localized in *N. benthamiana* (Fig. S7A). The molecular phylogeny of XTH genes was previously shown to segregate into three major groups (I, II, and III), and an isolated small group named 'ancestral' (Baumann *et al.*, 2007; Eklöf & Brumer, 2010; Shinohara & Nishitani, 2021). All *Physcomitrium* XTH genes, except Pp3c13_11910 and Pp3c3_9730, clustered in group I. The PpXTHs identified in HC300 were exclusively from group I (Fig. S7B), indicating a correlation between PD localization and phylogenetic distribution to group I.

Exordium and exordium-like proteins

Five members of the Arabidopsis EXO and EXL protein family were identified in HC300: Pp3c19_8770 (PpEXO1), Pp3c10_8680, and Pp3c14_6120 (AtEXO ortholog by protein BLAST), Pp3c18_22500 (AtPHI1/EXL7 ortholog), and Pp3c22_12550 (AtEXL2 ortholog). EXL2 was previously found in Arabidopsis and *P. trichocarpa* PD proteomes, but their PD localization was not confirmed (Fernandez-Calvino *et al.*, 2011; Leijon *et al.*, 2018). Several members of this family were also found in Arabidopsis cell wall proteomes (AtEXO, AtEXL1, AtEXL2, AtEXL3, and AtEXO4) and membrane proteomes (AtEXL4) (Bayer *et al.*, 2006; Feiz *et al.*, 2006; Jamet *et al.*, 2006; Mitra *et al.*, 2007). Here, PpEXO1 and Pp3c22_12550 were found among the top 10 PD score ranks in HC300 (Table S2) and PpEXO1 localized to PD in transient expression assays in *N. benthamiana* (Fig. 6a). The members of the EXO protein family share structural similarity in an N-terminal tail with a transmembrane domain and a globular C-terminal part (Fig. 6b). The globular part of the EXO proteins contains a distinct 'nose' with two clefts, which could serve as scaffolding or binding sites (Fig. 6b).

A phylogenetic analysis was undertaken to study sequence divergence along the EXO family evolution. Exordium proteins from representative species of the embryophytes (17 from *P. patens*, 8 from Arabidopsis, 12 from *O. sativa*, and 17 from *P. trichocarpa*) as well as all EXO proteins identified in algae were therefore included in this analysis (Fig. 6b). We found that EXOs are distributed between three clades. While clade I contained only EXO sequences from algae including both Chlorophyta and Charophyta, clade II contained EXO sequences from all the investigated species, distributed between two subclades. Subclade II-a exclusively included EXOs from the algae *Spiroglaea*

Fig. 5 Plasmodesmata (PD) localization of glycolyl hydrolase 17 (GHL17) proteins correlates with phylogenetic distribution. (a) Members of the GHL17 family identified in PD proteomes from angiosperms (*Arabidopsis*, *Populus trichocarpa*, *Nicotiana benthamiana*) and the moss *Physcomitrium*. (b) Phylogenetic analysis of GHL17 sequences showed three major clades: α (yellow), β (blue), and γ (red), as defined by Gaudioso-Pedraza & Benitez-Alfonso (2014). *Physcomitrium patens* GHL17 proteins identified in the PD proteome are labeled in red and *Arabidopsis* orthologs previously assayed for localization colored in blue (Benitez-Alfonso, 2014; Gaudioso-Pedraza & Benitez-Alfonso, 2014). Red asterisks indicate *P. patens* and *Arabidopsis* PD-localized GHL17 proteins and group in clade α , whereas non-PD-localized proteins are marked with black asterisks and cluster in either β or γ clades. Charophyta green algae grouped in clade β . Fungi GHL17 sequences clustered separately (in green). The phylogenetic tree was generated with the maximum likelihood method implemented in PHML (Lemoine *et al.*, 2019). Support values from 1000 bootstrap samples are shown at the nodes. Clades which do not contain *P. patens* or *Arabidopsis* sequences were collapsed. (c) Subcellular localization in *N. benthamiana* epidermis cells of PpGHL17_1 (Pp3c17_13760V3.2), PpGHL17_7 (Pp3c10_5480V3.1), PpGHL17_8 (Pp3c14_5200V3.2), and PpGHL17_15 (Pp3c12_13470V3.5) mVenus fusion proteins belonging to α clades. Bars, 20 μ m. (d) Localization of PpGHL17_1-mVenus at PD pit fields (white arrowheads) in *P. patens* gametophore cells (Supporting Information Methods S1). (e) Protonema cells of *P. patens* expressing PpGHL17_1-mVenus (Pp3c17_13760V3.2) showed an enrichment at the cell-to-cell interface. Due to the high density of PD at the cell–cell interface, individual PD cannot be resolved in the image. (f) Images of CFDA/FRAP transport assays of *Physcomitrium* ecotype Grandsen chloronema cells for wild-type, overexpressing PpGHL17_1 or PpGHL17_30. Bars, 20 μ m. Asterisks indicate the cell which was bleached. (g) Quantification of CFDA transport rates in *P. patens* ecotype Grandsen wild-type and genotypes overexpressing GHL17_1 or GHL17_30 presented as fraction of CF recovery in the photo-bleached protonema cells within 15 min (Methods S1). Values represent the mean \pm SE of 16 experiments for each line; small letters indicate significant ($P < 0.05$) differences (one-way ANOVA, Holm-Sidak correction). (h) Alpha-Fold model of PpGHL17 α clade glycoside hydrolase domains (cyan) with the conserved IFALFNE(N) motif (yellow).

EXO proteins identified in HC300 were all found in a subclade of clade III which seems to correspond to a cluster of EXO orthologs from seedless plants previously defined by the EXO sequences of *P. patens* and *Selaginella moellendorffii* (Li *et al.*, 2021). Interestingly, EXOs from *Arabidopsis* and *P. trichocarpa* previously identified in PD proteome lists were also found in clade III (Fig. 6b). The EXO protein family is yet the third example of a protein family in which PD-localized members group in a distinct phylogenetic clade.

Discussion

A PD scoring algorithm to define PD proteomes

Plasmodesmata are nanometer-scale structures of high complexity. They contain different types of cellular membranes, and their protein composition depends on environmental cues, tissue, and/or developmental stage. Thus, in order to understand the protein composition of PD, we require *in silico* approaches which tag likely PD-localized proteins and separate them from likely non-PD proteins. No biochemical method yields pure PD. Therefore, highly abundant proteins such as photosynthetic proteins or ribosomes will always be detected in PD preparations. We found that the overlap between unweighted PD proteome lists contains a high proportion of contaminant proteins and mere identification of a protein in biochemically enriched PD preparations is not sufficient to conclude its PD localization without additional information. We demonstrated the improvement of the scoring algorithm after extensive protein localization and used this information to refine a high-confidence PD proteome of *Physcomitrium* (HC300) (Fig. 4). A previous approach (Kirk *et al.*, 2021) also applied a bioinformatic strategy to predict PD proteins based on protein features found in a subset of the published proteomes (Fernandez-Calvino *et al.*, 2011; Park *et al.*, 2017; Brault *et al.*, 2019). However, these authors did not consider biochemical aspects (i.e. enrichment in biochemical PD fractions) and they did not benchmark the predictions against experimental protein localization.

Here, we used strict scoring thresholds to define the high-confidence set of proteins likely present in PD of *P. patens*. However, the drawback of stringent thresholding lies in accepting potential losses of loosely associated proteins or proteins dynamically associated with PD. Similar effects were discussed previously for interactome data: proteins identified with high reproducibility and high-confidence tended to make up stable protein complexes, such as ribosomes, respiratory chain, or photosynthetic complexes (Gilbert *et al.*, 2021). In turn, proteins with higher variability among replicates and lower confidence values tended to contain more protein functions for which dynamic or conditional interactions are expected, such as for example signaling proteins (Gilbert *et al.*, 2021).

A trade-off between accuracy and coverage has also been discussed for yeast interactome data (von Mering *et al.*, 2002). In analogy, HC300 is rich in cell wall proteins and structural elements (e.g. the C2-domain-containing proteins and putative scaffolds) but contains a rather low fraction of signaling elements such as receptor kinases. We found only one member of the GDSL-motif lipase/hydrolase family proteins in HC300, while nine other identified family members were not considered part of the high-confidence proteome based on their PD score rank, although two of them were experimentally localized to PD. Similarly, dynamins and RAN-GTPases were not classified among the HC300 after PD score ranking (Table S2). However, members of these protein families also showed a PD score difference > 0.3 , indicating highly improved ranking based on features of experimental localizations. We therefore considered these protein families part of an extended PD proteome (Table S2).

Besides cell wall modifying proteins, our approach identified several members of the EXO protein family (Fig. 6) and several members of the glyoxal oxidase family (Fig. 3) as prominent novel protein families with PD localization. In general, since protein composition, structure, and molecular functioning of PD still remain largely enigmatic, a high-confidence ‘parts list’ will be a valuable basis for future work on PD structure and function. In that sense, the HC300 defined here presents the community with a set of potentially rewarding targets for in-depth functional studies.

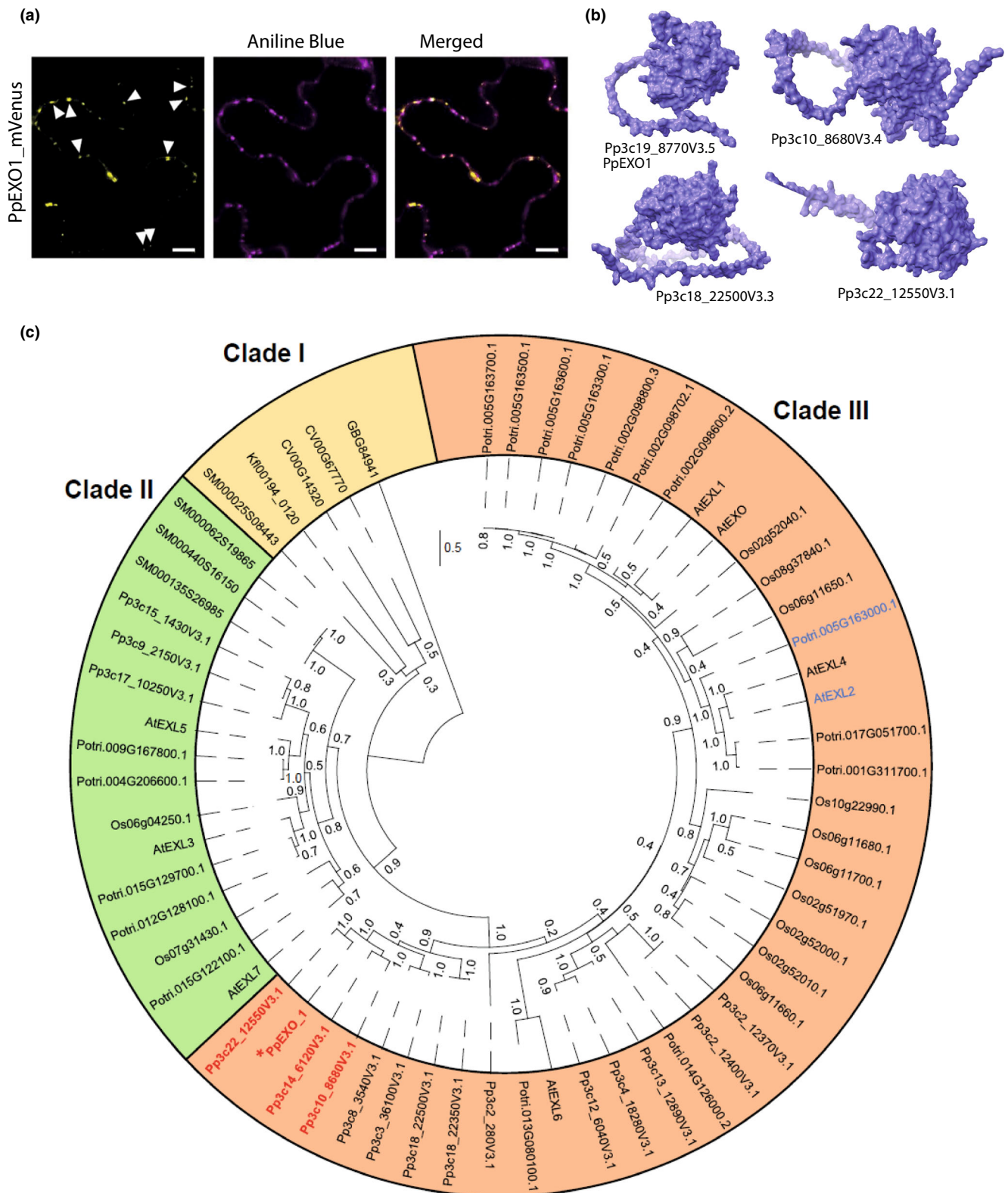


Fig. 6 Exo/Exordium-Like (EXO/EXL) family members. (a) Subcellular localization of PpEXO_1 (Pp3c19_8770V3.5) in *Nicotiana benthamiana* leaf epidermis cells showing overlay with aniline blue. Bars, 10 μm . (b) AlphaFold model of the four top-ranking EXO family members in HC300. (c) Phylogenetic tree of EXO/EXL family members from representative species of the embryophytes (17 from *Physcomitrium patens*, 8 from *Arabidopsis*, 12 from *Oryza sativa*, and 17 from *Populus trichocarpa*) and all algae for which EXO orthologs could be identified (*Coccomyxa subellipsoidea*, *Klebsormidium nitens*, *Chara braunii*, *Spirogloea muscicola*). *Physcomitrium patens* EXO/EXL members identified in the HC300 plasmodesmata (PD) proteome are labeled in red and clustered in clade III in which EXO identified in other PD proteomes are grouped (labeled in blue). Numbers indicate fraction of bootstrapping rounds supporting this node.

Large-scale PD localization in *N. benthamiana*

Despite improvement in distinguishing likely PD proteins from likely non-PD proteins by the PD scoring, experimental localization remains important. For example, we tested five members of the GHL17 family of glycosyl hydrolases from the whole PD proteome for PD localization with positive results for three proteins and no PD localization in *N. benthamiana* for two family members. Overall, the transient localization of *Physcomitrium* proteins in the *N. benthamiana* system proved efficient and applicable also on a larger scale. Most importantly, our approach showed that targeting mechanisms to PD are conserved between bryophytes and vascular plants. However, we cannot exclude the possibility that false negatives occurred, either due to the absence of required chaperones or misinterpretation of sorting signals in this heterologous system. In turn, false positives could have resulted from overexpression or mislocalization. To address such caveats, we used an inducible promoter and appropriate induction times to minimize overexpression artifacts.

Major protein families of *P. patens* PD

Plasmodesmata in land plants evolved to have more complex structures compared with their counterparts found in *Characeae* algae (Brunkard & Zambryski, 2017). This might reflect adaptations to regulate PD conductance and the emergence of specialized metabolic enzymes. Deposition of callose at the neck region of PD is one of the major modes to gate PD (Zavaliev *et al.*, 2011). The maintenance of callose associated with PD is affected mainly by callose synthases and β -1,3-glucanases. Thus, it comes as no surprise that our high-confidence subset of the *P. patens* PD proteome contains a high number ($n = 84$) of cell wall and cell wall-modifying proteins. Among these, the largest group comprised glycosyl hydrolases from the GHL17 subfamily. The GHL17 family is an excellent example of a large protein family which throughout evolution diversified in its subcellular localization, with one clade becoming PD-localized. This subfamily can also be taken as an example highlighting the two components of the PD scoring, namely biochemical enrichment in PD, and a feature scoring component. The feature scoring component alone would not have been able to differentiate proteins from the two clades of GHL17 proteins. However, the whole PD score revealed an important means to differentiate PD-localized and non-PD-localized GHL17s.

GHL17s and XTHs are considered important in controlling PD gating by affecting callose formation. GHL17s as callose-degrading enzymes are considered to be involved in keeping PD in an open state (Levy *et al.*, 2007; Zavaliev *et al.*, 2013; Benitez-Alfonso, 2014; Fig. 5e,f), while for XTHs, a role in cell wall remodeling during formation of secondary PD was proposed (Ehlers & Kollmann, 2001). Similarly, a callose synthase ortholog has been characterized with critical functions in enabling sucrose transport in rice (Wang *et al.*, 2009). The large numbers of proteins involved in cell wall synthesis and remodeling in the *Physcomitrium* PD proteome may reflect the need to dynamically

modify the composition of the PD associated cell wall regions independently of other regions.

We identified four out of the 17 *P. patens* EXO/EXL proteins in HC300. The Arabidopsis ortholog to Pp3c18_22500 is Phosphate induced 1 (PHI-1) and was initially identified in tobacco BY-2 cell suspension cultures among genes induced after recovery from phosphate-induced starvation (Sano *et al.*, 1999). In Arabidopsis, EXO/EXLs were proposed as potential mediators of brassinosteroid-promoted growth (Coll-Garcia *et al.*, 2004; Schröder *et al.*, 2011). Indeed, *AtEXO* and *AtEXL1* expression is induced by brassinosteroids, and *AtEXO* regulates other brassinosteroid-responsive genes (Coll-Garcia *et al.*, 2004; Schröder *et al.*, 2011). While overexpression of *AtEXO* promotes shoot and root growth, its loss of function leads to a reduction in growth resulting from a defect in cell expansion (Schröder *et al.*, 2009). The expression of several *AtEXLs* is also controlled by carbon metabolism and energy status and is required during adaptation to carbon and energy-limiting growth conditions (Schröder *et al.*, 2011). Although physiological functions of EXO proteins are still unclear, it was proposed that *AtEXO* could connect extracellular carbon status to growth responses (Schröder *et al.*, 2009, 2011; Lisso *et al.*, 2013).

Bud dormancy was shown to be associated with a molecular carbon starvation response which includes an increase in *EXL2* and *EXL4* RNA levels (Tarancon *et al.*, 2017). During bud dormancy, meristematic cells are symplasmically isolated through callose-dependent PD obstruction (Rinne *et al.*, 2001; Ruonala *et al.*, 2008; Cooke *et al.*, 2012; Tylewicz *et al.*, 2018). Photoperiod and gibberellic acid mediate opening of PD by inducing the expression of GHL17s. GHL17s then promote callose hydrolysis at PD and thereby contribute to dormancy release (Rinne *et al.*, 2011). Because of increasing evidence of EXO/EXL localization at PD, the involvement of these proteins in sensing of carbon status and regulation of growth response might need to be revisited in the context of symplasmic communication. Indeed, mechanisms by which these proteins mediate their physiological functions remain to be elucidated. Structural features of the EXO proteins suggest they can interact with other proteins through their C-terminal globular domain and thus may have functions in stabilizing protein complexes as scaffolds and/or in presenting signaling components or substrates.

The Arabidopsis homolog to phosphatase Pp3c1_370, TOPP4, was shown to regulate PIN1 polarity and trafficking (Guo *et al.*, 2015). This, together with receptor kinases identified in HC300, indicates that also in *Physcomitrium* PD, posttranslational modification of proteins could be an important mechanism defining protein localization or gating of PD. Indeed, phosphorylation-dependent recruitment to PD was described for coreceptor QSK1 in Arabidopsis (Grison *et al.*, 2019).

Evolution of PD localization within large protein families

Here, we present a thorough characterization of the PD proteome of *P. patens*, a member of the bryophytes, which were among the evolutionary early land plants. It is now possible to compare PD proteomes of an early embryophyte to PD

proteomes of vascular plants to obtain insights into evolution of PD itself and also into evolution of processes that involve PD. Symplasmic transport is highly regulated and involved in numerous processes in plant development, for example, bud dormancy and meristem activity. Thus, some of the PD proteins are involved in transport of signaling molecules and resource sharing and/or regulation thereof. These functions must have evolved and diversified during evolution of multicellularity. The *P. patens* HC300 contains multiple members of larger protein families, exemplified by GHL17s, XTHs, and EXO/EXLs, each of which diverged into several phylogenetic clades. Interestingly, in all cases, PD-localized members grouped in a distinct clade, often – as in the GHL17s – separating PD-localized and non-PD-localized family members. Contrary to what has been previously reported (Li *et al.*, 2021), we also found EXO/EXL orthologs in algae. However, EXO/EXL identified in the PD proteome do not cluster with algae members, suggesting that EXO/EXL PD localization might have emerged later during evolution of embryophytes. The EXLs are not predicted to have GPI anchors, but they contain an N-terminal alpha-helix, which is predicted to contain an ER/secretory pathway targeting signal. We hypothesize that a co-evolution of PD functions and respective protein localization took place, leading to diversification in these protein families.

Acknowledgements

We thank Elmehdi Bahafid for excellent assistance during *Physcomitrium* transformation, Jan Weber for excellent assistance during protein preparations of PD fractions, and Ronja Lange for excellent assistance in cloning of *Physcomitrium* genes. pTH UbiGate plasmid was a kind gift from Prof. Bezanilla, Dartmouth, USA. This work received funding from the European Research Council (ERC) and the Marie Skłodowska-Curie under the European Union's Horizon 2020 research and innovation program (Grant agreements 'SymPore' no. 951292 to WBF, RS, WXS, and WB, 'PDgate' no. 101023981 to MM, and 'CLAVHUB' no. 101023589 to VH, respectively). This research was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2048/1 – project ID 390686111 (RS, SH and WF), and the Alexander von Humboldt Professorship (WBF). Open Access funding enabled and organized by Projekt DEAL.

Author contributions

SG performed proteomics experiments, developed the PD scoring and prepared figures, and wrote the manuscript. MM contributed to cloning and localization of PD candidates in *N. benthamiana*, image analysis, figure preparation, phylogenetic analysis, and wrote the manuscript. VH contributed to cloning and localization of PD candidates in *N. benthamiana*, and image analysis, and contributed to writing of the manuscript. LX contributed to development of the PD scoring and PDDB structure and participated in writing of the manuscript. MP contributed to

cloning and localization of PD candidates in *N. benthamiana*, image analysis, and phylogenetic analysis. NSKJ contributed to cloning and localization of PD candidates in *N. benthamiana*, transformation of PD-localized candidates in *P. patens*, transport assay of CFDA in *P. patens*, and figure preparation. MS contributed to cloning and localization of PD candidates in *N. benthamiana*, and image analysis. JOE contributed to cloning and localization of PD candidates in *N. benthamiana*, image analysis, and transformation of PD-localized candidates in *P. patens*. UN performed transmission electron microscopy. SH wrote PD indexing Fiji macro. FK contributed to cloning and localization of PD candidates in *N. benthamiana*. ZZ contributed to the implementation of PDDB database structure and design. MD performed Cryo transmission electron microscopy of PD fraction preparation and structural analysis of PpGHL17s. PX performed Cryo transmission electron microscopy of PD fraction preparation. TS performed sequence motif analysis of the GHL17 proteins. WB developed the concept and designed experiments. WBF and RS developed the concept, designed experiments, and contributed to writing. WXS developed the concept, designed experiments, performed data analysis, and wrote the manuscript. SG, MM and VH contributed equally to this work.

ORCID

Wolfgang Baumeister  <https://orcid.org/0000-0001-8154-8809>
 Marcel Dickmanns  <https://orcid.org/0000-0003-4001-707X>
 J. Obinna Ejike  <https://orcid.org/0000-0002-4601-8551>
 Wolf B. Frommer  <https://orcid.org/0000-0001-6465-0115>
 Sven Gombos  <https://orcid.org/0000-0003-4177-526X>
 Sebastian Hänsch  <https://orcid.org/0000-0002-7762-2516>
 Vicky Howe  <https://orcid.org/0000-0002-7319-0664>
 Neda S. Kazemine Jasemi  <https://orcid.org/0000-0002-7842-6998>
 Manuel Miras  <https://orcid.org/0000-0002-3636-0265>
 Ulla Neumann  <https://orcid.org/0000-0001-9200-4209>
 Mathieu Pottier  <https://orcid.org/0000-0003-1551-4699>
 Moritz Schladt  <https://orcid.org/0000-0002-9160-4664>
 Waltraud X. Schulze  <https://orcid.org/0000-0001-9957-7245>
 Rüdiger Simon  <https://orcid.org/0000-0002-1317-7716>
 Lin Xi  <https://orcid.org/0000-0002-6811-9505>

Data availability

Proteomics data sets: Mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Deutsch *et al.*, 2017) with the dataset identifier PXD032820.

Image data sets: Images of PD-localized proteins were made public within the PDDB (pddb.uni-hohenheim.de). In addition, confocal images and the TEM micrographs were deposited at Bioimage Archive (<https://www.ebi.ac.uk/bioimage-archive/>) under accession no. S-BIAD466.

PD index quantification scripts: A script was developed aiding assessment of PD localization based on the co-expression of

mVenus fusion proteins with aniline blue. Code of the Fiji macro can be accessed at: https://github.com/SHaensch/2022_PD-index-quantification.

PD scoring: The formulas underlying PD scoring are described in the **Materials and Methods** section of the manuscript.

References

- Anvarian Z, Mykytyn K, Mukhopadhyay S, Pedersen LB, Christensen ST. 2019. Cellular signalling by primary cilia in development, organ function and disease. *Nature Reviews. Nephrology* 15: 199–219.
- Bangs F, Anderson KV. 2017. Primary cilia and mammalian hedgehog signaling. *Cold Spring Harbor Perspectives in Biology* 9: a028175.
- Baumann MJ, Eklöf JM, Michel G, Kallas AM, Teeri TT, Czjzek M, Brumer H 3rd. 2007. Structural evidence for the evolution of xyloglucanase activity from xyloglucan endo-transglycosylases: biological implications for cell wall metabolism. *Plant Cell* 19: 1947–1963.
- Bayer EM, Bottrill AR, Walshaw J, Vigouroux M, Naldrett MJ, Thomas CL, Maule AJ. 2006. Arabidopsis cell wall proteome defined using multidimensional protein identification technology. *Proteomics* 6: 301–311.
- Benitez-Alfonso Y. 2014. Symplastic intercellular transport from a developmental perspective. *Journal of Experimental Botany* 65: 1857–1863.
- Benitez-Alfonso Y, Faulkner C, Pendle A, Miyashima S, Helariutta Y, Maule A. 2013. Symplastic intercellular connectivity regulates lateral root patterning. *Developmental Cell* 26: 136–147.
- Braut ML, Petit JD, Immel F, Nicolas WJ, Glavier M, Brocard L, Gaston A, Fouche M, Hawkins TJ, Crowet JM *et al.* 2019. Multiple C2 domains and transmembrane region proteins (MCTPs) tether membranes at plasmodesmata. *EMBO Reports* 20: e47182.
- Brunark JO, Zambryski PC. 2017. Plasmodesmata enable multicellularity: new insights into their evolution, biogenesis, and functions in development and immunity. *Current Opinion in Plant Biology* 35: 76–83.
- Burch-Smith TM, Brunark JO, Choi YG, Zambryski PC. 2011. Organelle-nucleus cross-talk regulates plant intercellular communication via plasmodesmata. *Proceedings of the National Academy of Sciences, USA* 108: E1451–E1460.
- Coll-Garcia D, Mazuch J, Altmann T, Müssig C. 2004. EXORDIUM regulates brassinosteroid-responsive genes. *FEBS Letters* 563: 82–86.
- Cooke JE, Eriksson ME, Junttila O. 2012. The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. *Plant, Cell & Environment* 35: 1707–1728.
- Coudert Y, Harris S, Charrier B. 2019. Design principles of branching morphogenesis in filamentous organisms. *Current Biology* 29: R1149–R1162.
- Cove JD, Perroud P-F, Charron AJ, McDaniel SF, Khandelwal A, Quatrano RS. 2009. Culturing the moss *Physcomitrella patens*. *Cold Spring Harbor Protocols*. doi: 10.1101/pdb.prot5136.
- Cox J, Mann M. 2008. MAXQUANT enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* 26: 1367–1372.
- Deutsch EW, Csordas A, Sun Z, Jarnuczak A, Perez-Riverol Y, Ternent T, Campbell DS, Bernal-Llinares M, Okuda S, Kawano S *et al.* 2017. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Research* 45: D1100–D1106.
- Ding B, Haudenschild JS, Hull RJ, Wolf S, Beachy RN, Lucas WJ. 1992a. Secondary plasmodesmata are specific sites of localization of the tobacco mosaic virus movement protein in transgenic tobacco plants. *Plant Cell* 4: 915–928.
- Ding B, Kwon MO, Warnberg L. 1996. Evidence that actin filaments are involved in controlling the permeability of plasmodesmata in tobacco mesophyll. *The Plant Journal* 10: 157–164.
- Ding B, Turgeon R, Parthasarathy MV. 1992b. Substructure of freeze-substituted plasmodesmata. *Protoplasma* 169: 28–41.
- Doxey AC, Yaish MW, Moffatt BA, Griffith M, McConkey BJ. 2007. Functional divergence in the *Arabidopsis* β -1,3-glucanase gene family inferred by phylogenetic reconstruction of expression states. *Molecular Biology and Evolution* 24: 1045–1055.
- Ehlers K, Kollmann R. 2001. Primary and secondary plasmodesmata: structure, origin, and functioning. *Protoplasma* 216: 1–30.
- Eklöf JM, Brumer H. 2010. The XTH gene family: an update on enzyme structure, function, and phylogeny in xyloglucan remodeling. *Plant Physiology* 153: 456–466.
- Faulkner C, Bayer EM. 2017. Isolation of plasmodesmata. *Methods in Molecular Biology* 1511: 187–198.
- Feiz L, Irshad M, Pont-Lezica RF, Canut H, Jamet E. 2006. Evaluation of cell wall preparations for proteomics: a new procedure for purifying cell walls from Arabidopsis hypocotyls. *Plant Methods* 2: 10.
- Fernandez-Calvino L, Faulkner C, Walshaw J, Saalbach G, Bayer E, Benitez-Alfonso Y, Maule A. 2011. Arabidopsis plasmodesmal proteome. *PLoS ONE* 6: e18880.
- Gaudio-Pedraza R, Benitez-Alfonso Y. 2014. A phylogenetic approach to study the origin and evolution of plasmodesmata-localized glycosyl hydrolases family 17. *Frontiers in Plant Science* 5: 212.
- Gilbert M, Li Z, Wu XN, Rohr L, Gombos S, Harter K, Schulze WX. 2021. Comparison of path-based centrality measures in protein–protein interaction networks revealed proteins with phenotypic relevance during adaptation to changing nitrogen environments. *Journal of Proteomics* 235: 104114.
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N *et al.* 2012. PHYTOZOME: a comparative platform for green plant genomics. *Nucleic Acids Research* 40: D1178–D1186.
- Grigoriev IV, Hayes RD, Calhoun S, Kamel B, Wang A, Ahrendt S, Dusheyko S, Nikitin R, Mondo SJ, Salamov A *et al.* 2021. PHYCOSSM, a comparative algal genomics resource. *Nucleic Acids Research* 49: D1004–D1011.
- Grison MS, Brocard L, Fouillen L, Nicolas W, Wewer V, Dormann P, Nacir H, Benitez-Alfonso Y, Claverol S, Germain V *et al.* 2015a. Specific membrane lipid composition is important for plasmodesmata function in Arabidopsis. *Plant Cell* 27: 1228–1250.
- Grison MS, Fernandez-Calvino L, Mongrand S, Bayer EM. 2015b. Isolation of plasmodesmata from Arabidopsis suspension culture cells. *Methods in Molecular Biology* 1217: 83–93.
- Grison MS, Kirk P, Braut M, Wu XN, Schulze WX, Benitez-Alfonso Y, Immel F, Bayer EMF. 2019. Plasma membrane-associated receptor like kinases relocate to plasmodesmata in response to osmotic stress. *Plant Physiology* 181: 142–160.
- Guo X, Qin Q, Yan J, Niu Y, Huang B, Guan L, Li Y, Ren D, Li J, Hou S. 2015. TYPE-ONE PROTEIN PHOSPHATASE4 regulates pavement cell interdigitation by modulating PIN-FORMED1 polarity and trafficking in Arabidopsis. *Plant Physiology* 167: 1058–1075.
- He M, Li X, Mang M, Li Z, Ludewig U, Schulze WX. 2021. A systems-biology approach identifies co-expression modules in response to low phosphate supply in maize lines of different breeding history. *The Plant Journal* 109: 1249–1270.
- Hori K, Maruyama F, Fujisawa T, Togashi T, Yamamoto N, Seo M, Sato S, Yamada T, Mori H, Tajima N *et al.* 2014. *Klebsormidium flaccidum* genome reveals primary factors for plant terrestrial adaptation. *Nature Communications* 5: 3978.
- Howe KL, Contreras-Moreira B, De Silva N, 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: D689–D695.
- Hughes CS, Moggridge S, Muller T, Sorensen PH, Morin GB, Krijgsveld J. 2019. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nature Protocols* 14: 68–85.
- Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L *et al.* 2009. INTERPRO: the integrative protein signature database. *Nucleic Acids Research* 37: D211–D215.
- Jamet E, Canut H, Boudart G, Pont-Lezica RF. 2006. Cell wall proteins: a new insight through proteomics. *Trends in Plant Science* 11: 33–39.
- Johnston MG, Breakspear A, Samwald S, Zhang D, Papp D, Faulkner C, de Keijzer J. 2023. Comparative phyloproteomics identifies conserved plasmodesmal proteins. *Journal of Experimental Botany*. doi: 10.1093/jxb/erad022.
- Kim I, Kobayashi K, Cho E, Zambryski PC. 2005. Subdomains for transport via plasmodesmata corresponding to the apical-basal axis are established during

- Arabidopsis embryogenesis. *Proceedings of the National Academy of Sciences, USA* 102: 11945–11950.
- Kirk P, Amsbury S, German L, Gaudioso-Peraza R, Benitez-Alfonso Y. 2021. Comparative meta-proteomic analysis for the identification of novel plasmodesmata proteins and regulatory cues. *bioRxiv*. doi: [10.1101/2021.05.04.442592](https://doi.org/10.1101/2021.05.04.442592).
- Kraner ME, Muller C, Sonnewald U. 2017. Comparative proteomic profiling of the choline transporter-like1 (CHER1) mutant provides insights into plasmodesmata composition of fully developed *Arabidopsis thaliana* leaves. *The Plant Journal* 92: 696–709.
- 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.
- Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S, Gascuel O. 2019. NGPHYLOGENY.FR: new generation phylogenetic services for non-specialists. *Nucleic Acids Research* 47: W260–W265.
- Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v.4: recent updates and new developments. *Nucleic Acids Research* 47: W256–W259.
- Levy A, Erlanger M, Rosenthal M, Epel BL. 2007. A plasmodesmata-associated beta-1,3-glucanase in Arabidopsis. *The Plant Journal* 49: 669–682.
- Li S, Liu Z, Chen G, Qanmber G, Lu L, Zhang J, Ma S, Yang Z, Li F. 2021. Identification and analysis of GhEXO gene family indicated that GhEXO7_At promotes plant growth and development through brassinosteroid signaling in cotton (*Gossypium hirsutum* L.). *Frontiers in Plant Science* 12: 719889.
- Lisso J, Schröder F, Müssig C. 2013. EXO modifies sucrose and trehalose responses and connects the extracellular carbon status to growth. *Frontiers in Plant Science* 4: 219.
- Lucas WJ, Bouche P-S, Jackson DP, Nguyen L, Baker L, Ding B, Hake S. 1995. Selective trafficking of KNOTTED1 homeodomain protein and its mRNA through plasmodesmata. *Science* 270: 1980–1983.
- Lucas WJ, Ham BK, Kim JY. 2009. Plasmodesmata – bridging the gap between neighboring plant cells. *Trends in Cell Biology* 19: 495–503.
- von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P. 2002. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* 417: 339–403.
- Mitra SK, Gantt JA, Ruby JF, Clouse SD, Goshe MB. 2007. Membrane proteomic analysis of *Arabidopsis thaliana* using alternative solubilization techniques. *Journal of Proteome Research* 6: 1933–1950.
- Nicolas WJ, Grison MS, Trepout S, Gaston A, Fouche M, Cordeliers FP, Oparka K, Tilsner J, Brocard L, Bayer EM. 2017. Architecture and permeability of post-cytokinesis plasmodesmata lacking cytoplasmic sleeves. *Nature Plants* 3: 17082.
- Park SH, Li F, Renaud J, Shen W, Li Y, Guo L, Cui H, Sumarah M, Wang A. 2017. NbEXPA1, an alpha-expansin, is plasmodesmata-specific and a novel host factor for potyviral infection. *The Plant Journal* 92: 846–861.
- Pertl H, Himly M, Gehwolf R, Kriechbaumer R, Strasser D, Michalke W, Richter K, Ferreira F, Obermeyer G. 2001. Molecular and physiological characterisation of a 14-3-3 protein from lily pollen grains regulating the activity of the plasma membrane H⁺ ATPase during pollen grain germination and tube growth. *Planta* 213: 132–141.
- Raven JA. 2007. *Evolution of plasmodesmata*. Hoboken, NJ, USA: John Wiley and Sons.
- Rinne PL, Kaikuranta PM, van der Schoot C. 2001. The shoot apical meristem restores its symplasmic organization during chilling-induced release from dormancy. *The Plant Journal* 26: 249–264.
- Rinne PL, Welling A, Vahala J, Ripel L, Ruonala R, Kangasjarvi J, van der Schoot C. 2011. Chilling of dormant buds hyperinduces FLOWERING LOCUS T and recruits GA-inducible 1,3-beta-glucanases to reopen signal conduits and release dormancy in *Populus*. *Plant Cell* 23: 130–146.
- Ruonala R, Rinne PL, Kangasjarvi J, van der Schoot C. 2008. CENL1 expression in the rib meristem affects stem elongation and the transition to dormancy in *Populus*. *Plant Cell* 20: 59–74.
- Sano T, Kuraya Y, Amino S, Nagata T. 1999. Phosphate as a limiting factor for the cell division of tobacco BY-2 cells. *Plant & Cell Physiology* 40: 1–8.
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B *et al.* 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9: 676–682.
- Schröder F, Lisso J, Lange P, Müssig C. 2009. The extracellular EXO protein mediates cell expansion in Arabidopsis leaves. *BMC Plant Biology* 9: 20.
- Schröder F, Lisso J, Müssig C. 2011. EXORDIUM-LIKE1 promotes growth during low carbon availability in Arabidopsis. *Plant Physiology* 156: 1620–1630.
- Shinohara N, Nishitani K. 2021. Cryogenian origin and subsequent diversification of the plant cell-wall enzyme XTH family. *Plant & Cell Physiology* 62: 1874–1889.
- Simpson C, Thomas C, Findlay K, Bayer E, Maule AJ. 2009. An Arabidopsis GPI-anchor plasmodesmal neck protein with callose binding activity and potential to regulate cell-to-cell trafficking. *Plant Cell* 21: 581–594.
- Song DK, Choi JH, Kim MS. 2018. Primary cilia as a signaling platform for control of energy metabolism. *Diabetes and Metabolism Journal* 42: 117–127.
- Tangle E. 1879. Über offene Communicationen zwischen den Zellen des Endosperms einiger Samen. *Jahrbücher für Wissenschaftliche Botanik* 12: 170–190.
- Tarancon C, Gonzalez-Grandio E, Oliveros JC, Nicolas M, Cubas P. 2017. A conserved carbon starvation response underlies bud dormancy in woody and herbaceous species. *Frontiers in Plant Science* 8: 788.
- Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal* 37: 914–939.
- Thomas CL, Bayer E, Ritzenthaler C, Fernandez-Calvino L, Maule AJ. 2008. Specific targeting of a plasmodesmal protein affecting cell-to-cell communication. *PLoS Biology* 6: e7.
- Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. 2016. The Perseus computational platform for comprehensive analysis of (pro)teomics data. *Nature Methods* 13: 731–740.
- Tylewicz S, Petterle A, Marttila S, Miskolczi P, Azeez A, Singh RK, Immanen J, Mahler N, Hvidsten TR, Eklund DM *et al.* 2018. Photoperiodic control of seasonal growth is mediated by ABA acting on cell-cell communication. *Science* 360: 212–215.
- Wang R, Xing X, Wang Y, Tran A, Crawford NM. 2009. A genetic screen for nitrate regulatory mutants captures the nitrate transporter gene NRT1.1. *Plant Physiology* 151: 472–478.
- Wolf S, Deom CM, Beachy RN, Lucas WJ. 1989. Movement protein of tobacco mosaic virus modifies plasmodesmal size exclusion limit. *Science* 246: 377–379.
- Yan D, Yadav SR, Paterlini A, Nicolas WJ, Petit JD, Brocard L, Belevich I, Grison MS, Vaten A, Karami L *et al.* 2019. Sphingolipid biosynthesis modulates plasmodesmal ultrastructure and phloem unloading. *Nature Plants* 5: 604–615.
- Zavaliev R, Levy A, Gera A, Epel BL. 2013. Subcellular dynamics and role of Arabidopsis beta-1,3-glucanases in cell-to-cell movement of tobamoviruses. *Molecular Plant–Microbe Interactions* 26: 1016–1030.
- Zavaliev R, Ueki S, Epel BL, Citovsky V. 2011. Biology of callose (beta-1,3-glucan) turnover at plasmodesmata. *Protoplasma* 248: 117–130.
- Zhu T, Lucas WJ. 1998. Directional cell-to-cell communication in the Arabidopsis root apical meristem. I. An ultrastructural and functional analysis. *Protoplasma* 203: 35–47.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Enrichment protocol of plasmodesmata from *Physcomitrium patens*.

Fig. S2 Details on the plasmodesmata score.

Fig. S3 Structure of the core PDDb database.

Fig. S4 Subcellular localization of control proteins from *Physcomitrium patens* plasmodesmata proteome in *Nicotiana benthamiana* leaf epidermis cells.

Fig. S5 Refinement of the *Physcomitrium* plasmodesmata proteome.

Fig. S6 Plasmodesmata localization of glycosyl hydrolase family 17 (GHL17) proteins correlates with phylogenetic distribution.

Fig. S7 Plasmodesmata-localized *Physcomitrium patens* XTHs grouped in cluster I.

Methods S1 Additional details of experimental procedures.

Table S1 Published plasmodesmata data sets.

Table S2 Plasmodesmata (PD) proteome list with PD scores.

Table S3 Plasmodesmata index validation scale.

Table S4 Verified proteins.

Table S5 List of protein family members GHs, XTHs, and EXOs.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.