Plant Physiology Preview. Published on September 2, 2014, as DOI:10.1104/pp.114.243758

Running title: Genetic analysis of DEK1 Loop function.

Telephone numbers and E-mail addresses of the authors:

Viktor Demko:	+421904312576	viktor.demko@nmbu.no
Pierre-François Perroud:	+1 (314) 935-7593	perroud@biology2.wustl.edu
Wenche Johansen:	+47 62517865	wenchej@hihm.no
Charles F. Delwiche:	+1 301-405-8300	delwiche@umd.edu
Endymion Cooper:	+1 301-405-8300	endymion.dante.cooper@gmail.com
Pål Remme:	+47 48040365	milo83@gmail.com
Ako Eugene Ako:	+47 62517867	ako5552002@gmail.com
Karl G. Kugler:	+49 151 23232769	karl.kugler@helmholtz-muenchen.de
Klaus F.X. Mayer:	+49 89 31873584	k.mayer@helmholtz-muenchen.de
Ralph Quatrano:	+1(314) 935-6350	rsw@seas.wustl.edu
Odd-Arne Olsen:	+47 91581293	odd-arne.olsen@nmbu.no

Research area: Genes, Development and Evolution

Genetic analysis of DEK1 Loop function in three-dimensional body patterning in *Physcomitrella patens*

Viktor Demko[1§], Pierre-François Perroud[2§], Wenche Johansen[3], Charles F. Delwiche[4], Endymion D. Cooper[4], Pål Remme[3], Ako Eugene Ako[3], Karl G. Kugler[5], Klaus F. X. Mayer[5], Ralph Quatrano[2] and Odd-Arne Olsen[1*]

[1)] Norwegian University of Life Sciences, N-1432 Ås, Norway.

[2)] Department of Biology, Campus Box 1137, Washington University in St Louis, St. Louis, MO 63130, USA.

[3)] Department of Natural Science and Technology, Hedmark University College, N-2318 Hamar, Norway.

[4)] Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA

[5)]MIPS/IBIS, Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, 85764 Neuherberg, Germany

* *Corresponding author,* [§] These authors contributed equally to the work

One sentence summary

Proposed regulatory Loop segment of the DEK1 transmembrane domain is required for gametophore patterning in *Physcomitrella patens*

Footnotes:

Financial sources:

This work was supported by research grants from the Norwegian Research Council to the Norwegian University for Life Sciences (to O-A.O.) for personnel support (V.D. and P-F.P.) and to Hedmark University College (to O-A.O.) for personnel support (W.J., A.E.A.). R.S.Q. and P-F.P. were supported in part by funds from Washington University in St. Louis. KFXM acknowledges generous funding by the Deutsche Forschungsgemeinschaft (DFG) to project SFB924 Molecular mechanisms regulating yield and yield stability in plants. CFD and EDC acknowledge financial support from the US National Science Foundation (grant DEB-1036506).

Current address:

Viktor Demko: Department of Plant Physiology, Faculty of Natural Sciences, Comenius University in Bratislava, Mlynska dolina, 08315 Bratislava, Slovakia.

Corresponding author:

Odd-Arne Olsen, odd-arne.olsen@nmbu.no

Abstract

DEK1 of higher plants plays an essential role in position dependent signaling and consists of a large transmembrane domain (MEM) linked to a protease catalytic domain (CysPc) and a regulatory domain (C2L). Here we show that the postulated sensory Loop of the MEM domain plays an important role in the developmental regulation of DEK1 activity in the moss *Physcomitrella patens*. Compared with *P. patens* lacking DEK1 (Δ*dek1*), the *dek1*Δ*loop* mutant correctly positions the division plane in the bud apical cell. In contrast to an early developmental arrest of Δ*dek1* buds, *dek1*Δ*loop* develops aberrant gametophores lacking expanded phyllids resulting from mis-regulation of mitotic activity. In contrast to the highly conserved sequence of the catalytic CysPc domain, the Loop is highly variable in land plants. Functionally, the sequence from *Marchantia polymorpha* fully complements the *dek1*Δ*loop* phenotype, whereas sequences from *Zea mays* and *Arabidopsis thaliana* give phenotypes with retarded growth and affected phyllid development. New bioinformatic analysis identifies MEM as a member of the Major Facilitator Superfamily, membrane transporters reacting to stimuli from the external environment. Transcriptome analysis comparing WT and Δ*dek1* tissues identifies an effect on two groups of transcripts connected to *dek1* mutant phenotypes, i.e. transcripts related to cell wall remodeling and regulation of the APB2 and APB3 transcription factors known to regulate bud initiation. Finally, new sequence data support the hypothesis that the advanced charophyte algae that evolved into ancestral land plants lost cytosolic calpains, retaining DEK1 as the sole calpain in the evolving land plant lineage.

Introduction

A novel principle introduced in body patterning of early land plants, evolving from green algae related to charophytes 470-450 million years ago, was the ability to control growth in three dimensions (Graham et al, 2000; Pires and Dolan, 2012). The earliest multicellular forms of charophytes resembled Klebsormidiales, which form unbranched filaments with cells dividing by centripetal furrowing. Later, within Charales, Coleochaetales and Zygnematales, cell plate expansion was facilitated by phragmoplast (Leliaert et al., 2012). The ability to orient cell division in two cutting faces contributed to the formation of branched filaments and a more complex stem-like or discoid thallus. Finally, three-dimensional body patterning evolved within the early diverging land plants represented by bryophytes, displaying apical meristematic cells capable of dividing in three or more cutting faces (Graham et al., 2000). Subsequent evolution of multicellular apical meristems facilitated an increased morphological complexity in seed plants. Precise determination of division plane became critical for asymmetric cell divisions that drive plant morphogenesis (De Smet and Beeckman, 2011). An asymmetric cell division is tightly linked to differential cell fate establishment and several molecular players have been identified which play a role in these processes in land plants. These include transcription factors (e.g. WOX family, ATML1, PDF2, GRAS family) (reviewed in Lau et al., 2012), microtubule-associated proteins (e.g. TANGLED1, CLASP, MAP65) (reviewed in Müller et al., 2009), protein phosphatases (e.g. PP2A complex) (Spinner et al., 2013), protein kinases (e.g. CLAVATA1, CRINKLY4), proteins involved in vesicular trafficking and hormonal signaling (e.g. GNOM, PIN carriers) and others (reviewed in De Smet et al., 2009; De Smet and Beeckman, 2011). In all current models for the regulation of plant body patterning facilitated by coordinated cell divisions, an upstream-acting mechanism that detects interprets and transmits positional information (external, mechanical and intrinsic) triggering the down-stream events remains unknown.

We previously proposed DEK1 as a candidate protein for sensing and signaling surface cell position in land plants based on its predicted structure as well as genetics and evolutionary data (Tian et al., 2007; Liang et al., 2013). DEK1 consists of a multi-(21)-spanning transmembrane domain, DEK1 MEM (MEM), interrupted

by a suggested "sensor"-DEK1 Loop (Loop) and a C-terminal calpain protease DEK1 CysPc-C2L (CysPc-C2L) connected to MEM by the DEK1 Arm (Arm) segment (Lid et al., 2002). Depending on the computer algorithms used, the Loop is predicted to be extracellular (Lid et al, 2002) or cytosolic (Kumar et al., 2010). Neither the 3D structure of DEK1, nor the specific function of MEM or Arm is known. The emerging model for DEK1 function holds that the CysPc-C2L domains, encoding a calpain-like cysteine proteinase, is released from its inhibitory state by activation of MEM, resulting in autocatalytic cleavage of CysPc-C2L mediated by the Arm (Tian et al., 2007, Johnson et al., 2008). Confirmation of CysPc-C2L as the effector molecule has come from the observation that the *dek1* mutant phenotypes of *A. thaliana* and *P. patens* can be fully complemented by expression of the CysPc-C2L domain alone (Johnson et al., 2008; Perroud et al., 2014) if certain conditions are met. These conditions include expression under a promoter with sufficiently high activity during early embryogenesis and a ubiquitous pattern of expression throughout the development, as *pRPS5A* (Johnson et al., 2008). In addition, when *pRPS5A* is used to express CysPc-C2L, there appears to be a narrow window of transgene expression for full complementation to occur (Johnson et al., 2008). Deviation from these conditions, e.g. expression under the control of the 35S promoter in *A. thaliana* or overexpression of the CysPc-C2L under the control of *pRPS5A* introduces a range of new phenotypes affecting organ development globally (Lid et al., 2005; Johnson et al., 2008). In the *P. patens ⊿dek1* mutant complemented with the CysPc-C2L driven by the native DEK1 promoter or with full-length *DEK1* cDNA driven by 2x35S promoter, phenotypes ranging from WT-like to stunted plants develop (Perroud et al., 2014). These observations all point to an important role for MEM and Arm in the proper regulation of CysPc-C2L activity during plant development. This conclusion is further supported by the observation that overexpression of MEM in *A. thaliana* causes a dominant negative phenotype, mimicking the phenotypes of 35S-DEK1 RNAi lines (Tian et al., 2007).

Requirement of DEK1 for surface position-dependent aleurone cell fate specification and maintenance as well as normal embryogenesis was shown in maize (Lid et al., 2002). The involvement of DEK1 in three-dimensional body plan transition of early land plants is further supported by the phenotype of the DEK1 deletion mutant in *P. patens* (Perroud et al., 2014). In this mutant, the development of

protonemata, filamentous cells growing by polar tip growth, is not affected, whereas the transition to the three-dimensional growth of the gamethophore is severely affected in its ability to reorient the division plane in the bud apical cell perpendicularly to the first asymmetric division of the bud initial (Perroud et al., 2014). This phenotype is highly reminiscent of the *dek1* embryo phenotype of *A. thaliana*, where the first asymmetric division of the zygote, similarly to *P. patens* bud initial, gives rise to the cytoplasm-rich apical cell and a vacuolated basal cell. After the correct zygote division, subsequent cell divisions fail to occur in the oriented manner specified in wild-type embryos and as a consequence, the protoderm (the outermost cell layer of the globular embryo) fails to develop (Johnson et al., 2005; Lid et al., 2005). We interpret these data to suggest that the basic function of DEK1 in cell division plane orientation is conserved between mosses, one of the earliest diverging lineages of land plants, and angiosperms.

DEK1 is a member of one of four ancestral calpain variants that were established in the early evolution of eukaryotes, TML-calpains, in which the CysPc-C2L domains are attached to multi-spanning transmembrane anchors (Zhao et al., 2012). The other ancient calpains consisted either of CysPc alone, or CysPc attached to other domains (Zhao et al., 2012). Among modern calpains, the so-called classical calpains in humans are the most intensively studied, with the domain structure Nter-CysPc-C2L-PEF (Ono and Sorimachi, 2012). Our working hypothesis is that DEK1 assumed a novel role in positional signaling during land plant evolution, contributing to the ability of land plants to develop three-dimensional organs. This hypothesis is supported by several lines of circumstantial evidence, including the highly conserved sequence and function of DEK1 in land plants, forming a separate clade among the land-plants (Zhao et al., 2012, Liang 2013). Recently, we showed that the CysPc-C2L calpain moiety of DEK1 from the moss *P. patens* is capable of complementing the *A. thaliana dek1-3* mutant (Liang et al., 2013), representing a functional conservation that spans about 450 million years of evolutionary time (Kenrick and Crane, 1997). In contrast, the CysPc-C2L domains of the unicellular alga *Mesostigma viride*, the earliest diverging lineage of charophyte algae, do not complement the *A. thaliana dek1* mutant (Liang et al., 2013). Importantly, the appearance of the DEK1-clade during land plant evolution coincides with the establishment of three-dimensional growth habit of land plants, whereas members of the chlorophytes that display

unicellular or planar body plans lost TML-calpains, but retained multiple cytosolic calpains (Zhao et al., 2012; Liang et al., 2013). The exact point in time during charophyte evolution when only DEK1-calpains were retained is unknown.

In this paper we continue to explore the separate DEK1-domains, focusing on the Loop region. First, using homologous recombination, we create a *P. patens* Loop deletion mutant, *dek1Δloop*. Based on phylogenetic analysis of Loop sequences from Charophyta and land plant species we use Loop coding regions from *Marchantia polymorpha*, *Zea mays* and *A. thaliana* to complement the *dek1Δloop* mutant in *P. patens* in order to study the functional conservation of Loop sequences from land plants. Bioinformatics analysis is used to re-examine the structure of DEK1 MEM in order to identify homologous proteins or protein domains that help elucidating the MEM function. Next, in order to develop a better understanding of the global role of DEK1, we use RNAseq differential expression analysis to study the effect of DEK1 on the transcriptome of *P. patens* by comparing WT and *Δdek1* protonemata before and after bud initiation. Finally, we use novel data to identify the last charophycean species of green algae that possessed multiple calpain forms before retention of DEK1 as the single calpain of land plants.

Results

The DEK1 Loop of land plants is highly divergent from its algae counterparts.

The transmembrane domain of *Pp*DEK1 is interrupted by a ~300 amino acid residue Loop segment located between the 9th and 10th transmembrane segment (TMS) (Fig. 1A, Supplemental Fig. S1A). To analyze the degree of sequence conservation, we aligned the Loop sequence of 60 DEK1 proteins, including the sequences from three charophyte algae species (Supplemental Table S1). This analysis revealed that the algae Loop sequences are highly divergent from each other and from the corresponding land plant sequences, preventing meaningful alignments (data not shown). To further investigate whether any local similarity in the Loop exists, we carried out pairwise comparisons between the different algae and *P. patens* Loop sequences using the exact Smith-Waterman algorithm (EMBOSS Water). Only the N-terminal end of the charophyte algae Loop sequences align with significant

expect (E) values to the *P. patens* Loop, but with low similarity scores (< 20%) (data not shown).

Next, we investigated the phylogenetic relationship among land plant Loop sequences. This analysis grouped the sequences into the major clusters corresponding to bryophytes, lycophytes, monocots and dicots (Fig. 1B). A pairwise sequence comparison of the Loop sequences shows that the amino acid identity decreases with evolutionary distance; the sequence identity between *A. thaliana* and *P. patens* being 38% (Supplemental Fig. S1B). As portrayed by the sequence logo representation of the alignment (Supplemental Fig. S1C), the N- and C-terminal ends of Loop are highly conserved with a more divergent middle part. In the conserved regions, blocks of conserved amino acids, single amino acid positions with conservative substitutions within the Loop are identified. Using the consensus sequences from each group alignment we performed a new alignment identifying the fully conserved positions in the Loop sequences (Fig. 1C).

The DEK1 MEM shows homology to Major Facilitator Superfamily of membrane transporters.

In spite of the fact that DEK1 was discovered more than a decade ago, similarity of the MEM domain to proteins in existing databases that could hint to a function has remained elusive. In addition, computer modeling of MEM has given inconsistent results both with respect to the number of TMS (21 *vs*. 23) and the position of the Loop (intracellular *vs*. extracellular) (Lid et al., 2002; Kumar et al., 2010). Here, we present a reanalysis of MEM from *A. thaliana*, *Z. mays* and *P. patens* using TMHMM (Krogh et al., 2001), SPOCTOPUS (Viklund et. al., 2008), TOPCONS (Bernsel et al., 2009), PHOBIUS (Käll et al., 2004) and HMMTOP (Tusnády & Simon, 1998). This analysis consistently locates the Loop intracellularly in all species (Fig. 1D) and predicts 23 TMSs located N-terminally to the predicted Arm segment (Supplemental Fig. S1A). Furthermore, the MEM topography of charophyte algae, as represented by *Klebsormidium flaccidum* (Klebsormidiales), *Nitella mirabilis* (Charales) and *Mougeotia scalaris* (Zygnematales) is highly similar to land plants, showing overall structural conservation of MEM in Streptophyta, representing an evolutionary time of ~1000 million years (Zimmer et al., 2007; Pires and Dolan, 2012), (Supplemental Fig. S1A). New homology searches in recent

protein databases also detect similarity between the Major Facilitator Superfamily (MFS) domain of secondary transporters (cd:06174) and TMSs 16-22 of DEK1-MEM (Supplemental Fig. S2). MFS proteins includes uni-, sym- and anti-porters and are a large and diverse group of proteins facilitating transport of various solutes across the membranes in response to chemiosmotic gradients, including ions, sugars, phosphates, drugs, neurotransmitters, nucleosides, amino acids and peptides (for review see Yan, 2013).

Deletion of the DEK1 Loop severely affects *P. patens* gametophore body patterning.

Our strategy for assessing the function of the Loop in *P. patens* is first to utilize homologous recombination to create a Loop deletion mutant (*dek1Δloop*) and then re-target Loop-coding sequences from representatives of the bryophytes, monocots and dicots grouped in the phylogenetic studies above (Fig. 1B). In order to create the *dek1Δloop* mutant we first transformed *P. patens* protoplasts using the *pBHRF-DEK1-ΔLoop* construct (Supplemental Fig. S3A). Out of the 65 stable transformants, 47 displayed the Δ*dek1* phenotype lacking gametophores as described in Perroud et al., 2014. The PCR genotyping of the *dek1Δloop* locus was performed in two steps. First, we assessed the loss of Loop ORF by attempting to PCR-amplify a sequence targeted for removal using the primer pair *delta loop fra-fw* and *delta loop fra-rv* (Supplemental Table S2 for primer sequence). Twenty-one transformants lacked WT bands and were analyzed further. Second, we PCR amplified the targeted locus using the primer pair *delta loop diag fw* and *delta loop diag rv* designed outside the genomic fragment used to build the *pBHRF-DEK1-ΔLoop* vector to select single insertion events. Four transformants showed a signal corresponding to a single replacement event. All of these events displayed the Δ*dek1* phenotype lacking gametophores (Perroud et al., 2014) (Supplemental Fig. S4A and B). Potentially, an insertion of a resistance marker in an intron may cause a null mutant phenotype by interfering with posttranscriptional modifications of the transcript thus preventing expression of an active protein. To test this we generated transformants showing loss of hygromycin resistance using the Cre recombinase approach (Trouiller et al. 2006). Interestingly, the resistance marker-free mutants we obtained displayed a distinct new phenotype, different from WT or previously described Δ*dek1*, carrying gametophores with altered morphology (Supplemental Fig. S4C shows the line designated

dek1Δloop selected for further description). The new locus was cloned and sequenced, confirming loss of the resistance marker and Southern blot analysis confirmed that the Loop was eliminated from the genome (Supplemental Fig. S5A and B). To confirm proper splicing of the *DEK1* transcript in *dek1Δloop* line, the cDNA region overlapping the deleted Loop-coding sequence was amplified and sequenced using the *RT-Loop-F* and *RT-Loop-R* primers (Supplemental Fig. S3A). Sequencing confirmed in-frame removal of the Loop-coding sequence from exon 7 and proper splicing at the locus. Transcription of the truncated gene was also confirmed by RT-PCR using the primers from DEK1 CysPc-C2L coding regions and judged from this semi quantitative RT-PCR, the level of *dek1Δloop* transcript is not changed compared to WT (Supplemental Fig. S5D).

In contrast to emerging phyllids of WT plants (Fig. 2A), the most pronounced phenotype of the *dek1Δloop* mutant gametophores is retarded growth and lack of expanded phyllids (Fig. 2B). Instead, short filamentous protrusions form on the mutant gametophore stem (Fig. 2B and E). Neither phyllids nor gametangia are formed on the mutant gametophore even after two months of cultivation under sporophyte-inducing conditions (Fig. 2E). Thus, we conclude that *dek1Δloop* mutant is capable of forming gametophore apical stem cells giving rise to a phyllid-less stem, but that cell division and differentiation activities from the lateral domains which normally shape a leafy-shoot gametophore are blocked.

Gametophore pattern formation depends on asymmetric cell divisions coordinated by local cues within the developing body as described in details by Harrison et al. (2009). In order to characterize morphological changes in *dek1Δloop* gametophores we studied the pattern of cell divisions in early buds and juvenile gametophores (Fig. 3). In WT, the bud initial cell divides first asymmetrically (Fig. 3A) giving rise to the bud apical and basal cells, respectively, which in a few hours undergo additional asymmetric divisions. The apical cell divides perpendicular to the first asymmetric division of the bud initial (Fig. 3A). As previously described by Perroud et al. (2014), positioning of the division plane in bud apical cells depends on DEK1 activity, misorientation of cell divisions in *Δdek1* mutants preventing establishment of the stem cell. Subsequently, the *Δdek1* mutant fails to undergo transition to three-dimensional growth and further development of the gametophore is arrested at the early bud stage (Fig, 2C) (Perroud et al., 2014). As shown in Fig. 3E,

the apical cell of the two-cell stage bud in *dek1Δloop* mutant divides similar to WT, although with a slight bending of the cell wall. Unlike the *Δdek1* mutant, the *dek1Δloop* mutant buds continue to grow. However, the pattern of cell division become irregular when compared to WT (Fig. 3G and H). In WT, phyllid initial cells emerge from the lateral domains early during juvenile gametophore development (Fig. 3D). Phyllid primordials then start to expand in medio-lateral and proximo-distal dimensions (two-cell file phyllid shown on Fig. 3I). The *dek1Δloop* mutant initiates formation of the phyllid progenitor cell-like structures (Fig. 3H), however they fail to expand laterally and, instead, continue to proliferate as curved filamentous structures (Fig. 3J). Further proliferation of the filamentous protrusions from defective gametophores stops when they reach the 3-6 cell state (Supplemental Fig. S6). Based on the reduced growth and lack of expanded phyllids in the *dek1Δloop* mutant we infer that an intact Loop is critical for DEK1 calpain activity required for locally coordinated asymmetric divisions, especially at the lateral domains which shape gametophore organs.

Deletion of the entire PpDEK1 coding sequence causes an over-budding phenotype where the number of buds per 15 protonemal filaments (counted from the apical cell) is increased to approximately four in comparison to one bud in WT (Perroud et al., 2014). As shown on Fig. 4, the bud induction in *dek1Δloop* mutant reaches an intermediate level between the WT and *Δdek1* mutant with an average number of buds close to two.

The Loop from Liverworts, but not dicots or monocots, fully complements the *dek1Δloop* mutant of *P. patens*.

To investigate whether Loops from the phylogenetically separate groups of Loop sequences from bryophytes, monocots, and angiosperms (Fig. 1A) are functionally conserved, we introduced the Loop coding sequences from the liverwort *M. polymorpha*, the monocot *Z. mays*, and the dicot *A. thaliana* into the original locus of the *P. patens dek1Δloop* mutant line. In order to verify the functionality of retargeted sequences we first retargeted the WT Loop coding sequence to the *dek1Δloop* locus as described in Material and Methods (see also Supplemental Fig. S3 B). Four independent lines were obtained in this experiment, all of which reverted to WT phenotype (Fig. 5A and C; Supplemental Fig. S3A; Supplemental Fig. S7A-C).

The Loop coding sequences from *M. polymorpha*, *Z. mays* and *A. thaliana* were then targeted to the *dek1Δloop* locus. The constructs used to transform the *dek1Δloop* line are depicted in Supplemental Fig. S3C and lines carrying the heterologous *Loop* sequences were selected as described in Material and Methods.

In the experiment where the Loop coding sequence from *M. polymorpha* was introduced (Supplemental Fig. S3C), 15 transformants with no PCR-signal for the original *dek1Δloop* locus out of the 31 analyzed lines displayed the *Δdek1* phenotype. In a PCR screen for single insertion events, three lines with a positive signal were detected. One of these lines (*MpLoop#29*, for Southern analysis see Supplemental Fig. S5B and C) was subjected to the Cre recombinase-mediated elimination of the resistance cassette. Three lines with no resistance to the G418 were obtained, all fully reverted to the WT phenotype a shown in more details for one of the lines designed *MpLoop* (Fig. 5). In-frame insertion of the *Mp*Loop coding sequence and removal of the resistance cassette was confirmed by sequencing the Loop overleaping genomic DNA region. Sequencing of the Loop overlapping cDNA regions confirmed proper splicing of the *PpDEK1* transcript containing heterologous *MpLoop* sequence (data not shown). Semiquantitative RT-PCR using the primers from the CysPc coding region showed that the transcript abundance in the *MpLoop* line is the same as in WT (Supplemental Fig. S5D). Fully developed gametophore of the *MpLoop* line is shown in Fig. 5D. The size and overall morphology of the *MpLoop* gametophore is indistinguishable from the WT with fully expanded phyllids, differentiated marginal serrated cells and midrib (Fig. 5H). The bud onset on protonemata in the *MpLoop* line shows the same rate as in the WT (Fig. 4). After cultivation of this line under sporophyte-production conditions, we observed fully developed sporophytes indistinguishable from WT (Supplemental Fig. S7A and D). These results show that the heterologous DEK1-Loop segment from the liverwort, sharing approximately 43% amino acid identity with *Pp*Loop, is fully functional in moss demonstrating functional conservation of the DEK1-Loop within the groups of early diverging land plants (Fig. 1A).

Next, we investigated whether the monocot and dicot Loops that form separate clusters in the phylogenetic tree (Fig. 1B) are functional in *P. patens*. For the *A. thaliana* Loop sequence, six *P. patens* lines were identified that contained the *At*Loop out of the 95 transformants obtained. All of these lines showed proper targeting of the

construct from both 5`and 3`end and they all contained multiple copies of the targeted sequence at the locus. The line *AtLoop#14* was next subjected to the Cre recombinase-mediated elimination of the resistance cassette. Three *AtLoop* lines were obtained with no resistance to G418, all showing the same phenotype as described in more details for one of the lines below (for construct design and Southern blots see Supplemental Figs. S3C and S5B and C, respectively). Correct in-frame insertion of the *AtLoop* was confirmed by sequencing of the *Loop* overlapping genomic region. Semi quantitative RT-PCR using the primers from the DEK1 CysPc-coding region showed that the transcript abundance in the *AtLoop* line is the same as in WT (Supplemental Fig. S5D). Replacement of the *Pp*Loop coding sequence with its *A. thaliana* counterpart causes reduced growth of gametophores and morphogenetic changes affecting phyllid development (Fig. 5E). *AtLoop* phyllids are narrow with blade composed of 3 to 8 files of cells with variable size (Fig. 5I). No marginal serrated cells are differentiated. The midrib is formed in phyllids composed of more than 3 blade cell files, but never differentiates through the entire phyllid axis (Fig. 5I). Onset of buds on protonemata in *AtLoop* line shows the same rate as in WT (Fig. 4). After prolonged cultivation of the *AtLoop* line under the sporophyte development-promoting conditions, gametangia were formed however, we were not able to detect any sporophytes (Supplemental Fig. S7E).

A similar result as for *A. thaliana* was achieved when the *Loop* coding sequence from amonocot *Z. mays* was used to replace the *Pp*Loop (for construct design see Supplemental Fig. S3). Out of the 86 genotyped transformants, 9 lines showed targeting of the *ZmLoop* to the *dek1Δloop* locus, all showing the *Δdek1* mutant phenotype. Based on PCR-genotyping, all these lines showed proper targeting of the construct from both 5`and 3`end and they all contained multiple copies of the targeted sequence at the locus. The line *ZmLoop#5* (for Southern see Supplemental Fig. S5B and C) was then subjected to the Cre recombinase-mediated elimination of the resistance cassette. Fifteen lines were selected with no resistance to the G418, all showing a phenotype similar to that of the *ZmLoop* line described below (Fig. 5F). In-frame insertion of the *Zm*Loop coding sequence and removal of the resistance cassette was confirmed by sequencing of the *Loop* overlapping genomic region. Semi quantitative RT-PCR using the primers from the DEK1 CysPc-C2L coding region showed that transcript abundance in the *ZmLoop* line is the same as in WT

(Supplemental Fig. S5D). Similar to the *AtLoop* line, replacement of the *Pp*Loop with its maize homolog caused reduced gametophore growth and aberrant phyllid development with narrow phyllids and blades composed of 3 to 7 cell files (Fig. 5J). The size and morphology of the phyllid blade cells are variable, effecting phyllid morphology. All phyllids lack differentiated marginal serrated cells. The midrib-like structures are formed only in phyllids with more than 3 blade cell files, but never reach the phyllid tip (Fig. 5J). Morphology of such midribs in the *ZmLoop* line appears more affected compared to the *AtLoop* line as depicted in representative examples of isolated phyllids in Fig. 5I and J. After cultivation of the *ZmLoop* line under sporophyte development-promoting conditions, gametangia were formed. However, no sporophytes were detected (Supplemental Fig. S7F). Despite the morphological abnormalities in *ZmLoop* gametophores, the number of buds formed per 15 filament cells is the same as in WT (Fig. 4).

PpDEK1 deletion alters gene expression prior to bud formation.

In order to detect genes and pathways regulated by DEK1 we performed a transcriptome analysis of WT and *Δdek1* tissues at 6 and 14 days after culture initiation. At the first time point, the samples consisted of protonemata cells in both strains (Supplemental Fig. S8A and B). During the next 8 days, budding occurred in both WT and in *Δdek1*, but only gametophores developed in WT (Supplemental Fig. S8C and D). Three independent culture sets were used, giving a total of 12 data points. cDNA library building and Illumina RNA-seq were performed at BGI (http://www.genomics.cn/en/index) as described in Material and Methods. RNA-seq data from this article can be found in the ArrayExpress database (www.ebi.ac.uk/arrayexpress/) under accession number E-MTAB-2588. After sequence read mapping against the *P. patens* genome, transcript abundance (FPKM) and differential expression were calculated using uniquely mapped reads only as described in Material and Methods (see full dataset in Supplemental Table S3 and Supplemental Protocol 1 for the datatset validation details).

As expected, the majority of genes are expressed (FPKM > 1) under all four conditions. However, there is also a substantial number of genes that are only expressed under certain conditions (Fig. 6A). Of the total number of transcripts assembled, 17506 (85,2%) were present under all conditions, an unsurprising result

since both strains are viable and they do not show morphological difference at the early time point. A bird's view of the GO term annotation shows that the percentage of transcript annotation for the common pool (57% of annotated transcript) is slightly increased compared with the total published *P. patens* gene model annotation (41%). At the other end of the scale *Δdek1* and WT specific transcript pools for each time point show a significant drop in the frequency of gene annotations with values between 8% to 30%. This drop may reflect an understudy of multicellular gametophytic development since this stage occurs in mosses and ferns to an extent not present in the more well studied angiosperms. Next we focused the analysis on transcript variation associated both with the *dek1* mutation as well as developmental stages by looking at the interaction between the two factors using DESeq2 (Supplemental Table S4). In this analysis we found 380 genes for which the change between the time points was significantly different between the wild type and the *Δdek1* mutant (BH adjusted p-value < 0.05, absolute log2 fold change > 1). In this set of 380 genes there are 179 genes with a positive log2 fold change in the interaction contrast and 201 with a negative log2 fold change. A GOSlim term analysis performed on this gene subset (Supplemental Fig. S9) showed a clear enrichment in term function associated with cell wall and cell periphery term. For example, the absence of gametophore in *Δdek1* at 14 days reflects the reduction of transcript coding for *Pp*-ABCG28 (Pp1s198_152V6), a gene associated with cuticle transport present only in gametophores (Buda et al., 2014). Similarly, NAC transcription factor *PpVNS* 1 and 5 (Pp1s182_37V6.1 and Pp1s223_12V6.1 respectively, see Supplemental Fig. S10) recently shown to control part of phyllid mid rib development (Xu et al., 2014), remains at a protonemal accumulation level in *Δdek1*. In addition, the *PpDEK1* deletion also affects the transcriptome before onset of budding. More specifically, we identified a gene set that can be directly linked to the *Δdek1* phenotype, namely over-budding, that also occur at a lower level in *dek1Δoop* (Fig. 4). Two AP2 containing transcription factors (Pp1s131_139V6.1 and Pp1s131_131V6.1 named ABP2 and ABP3, respectively by Aoyama *et al.* (2012) are significantly up regulated in *Δdek1* at both time points (Fig. 6B). These two genes belong to a small homologous gene set analyzed by Aoyama et al. (2012), which are necessary for the budding process since the quadruple knockout led to bud-less protonemal tissue. Conditional overexpression of one of this gene, ABP4, in a KO background not only restored the budding phenotype but also led to over-budding, a similar phenotype observed in the *Δdek1*

line of *P. patens* (Perroud et al., 2014). ABP upregulation in *Δdek1* suggests the existence of a regulatory function of DEK1 in protonema independent of three-dimensional development.

Charophycean algae were the last group of land plant predecessors with multiple forms of calpains.

Previous analysis concluded that the predecessor of Chlorophyta and Charophyta possessed multiple members of the calpain superfamily (Liang et al., 2013). In land plants, calpains are represented only by the DEK1-clade of TML-calpain, whereas only cytosolic calpains, but not DEK1, has been identified in the chlorophyte genome sequences available at that time (Liang et al., 2013). Using novel RNAseq data from organisms which represent the five major lineages of charophyte algae we identified 17 different CysPc transcripts within Mesostigmatales (*Mesostigma viride*), Klebsormidiales (*Klebsormidium flaccidum*), Charales (*Nitella mirabilis*), Coleochaetales (*Coleochaete orbicularis*) and Zygnematales (*Spyrogyra pratensis* and *Mougeotia scalaris*) (Supplemental Table S5). In all species examined, except in *S. pratensis, Dek1*-like transcripts were detected that encod proteins with identical modular composition to land plant DEK1 proteins. In addition, we identified several CysPc sequences in transcripts encoding proteins without the unique DEK1-domains (MEM and Arm), including transcripts containing single or multiple CysPc domains, with or without the conserved catalytic triad (Cys, His, Asn) and with or without a C-terminal C2L domain (Supplemental Table S5). These cytosolic non-DEK1-like calpains were detected in *M. viride*, *K. flaccidum*, *N. mirabilis*, and *S. pratensis*, but not in transcriptome data available for *C. orbicularis*, possibly representing the stage at which cytosolic calpains were lost during land plant evolution. Phylogenetic reconstruction was performed to infer the evolutionary relationship between the various charophyte CysPc sequences, together with representative land plant and chlorophyte sequences. The resulting phylogeny (Fig. 7) confirmed that *Mesostigmatales*, *Klebsormidiales* and *Zygnematales* species, early diverging charophytes, harbor both DEK1-clade TML-calpains (clustering with land plant calpains) and cytosolic calpains (clustering with cytosolic chlorophyte and animal calpains). The cytosolic calpains where subsequently lost within the

evolutionary advanced charophytes leaving the DEK1 protein as the single calpain of land plants.

Discussion

This study expands our knowledge of DEK1 evolution by showing the distribution of calpain family in chlorophyte and charophyte algae as well as land plants, all together representing an evolutionary time span of about 1000 million years (Zimmer et al., 2007; Pires and Dolan, 2012). As reported earlier, TML-calpains, calpains with more than 15 transmembrane segments linked to the protease CysPc-C2L originated approximately 1.5 billion years ago as a result of a fusion between the CysPc-C2L domains and the TML domain, forming a monophyletic group (Zhao et al., 2012). Reanalysis of the TML domain presented here indicate that segments of the TML domain belong to the large family of Major Facilitator Superfamily (MFS) domains of secondary transporters, which, similar to CysPc, are also of prokaryotic origin (Pao et al., 1998). Currently, we have no information about the function of TML in the regulation of CysPc activity in groups other than plants, including members of the Excavata, SAR supergroups, as well as in *Thecamonas trahens*, an ancient eukaryote sister to Opisthokonta (animals, protists and fungi) (Zhao et al., 2012). Previously, we have shown that the last common ancestor of the chlorophyte and charophyte algae likely contained both TML-calpains and cytosolic calpains, but that we were unable to detect TML-calpains in species belonging to the Chlorophyta, including *Chlamydomonas reinhradtii, Volvox carteri, Micromonas pusilla*. Here we show that early diverging charophytes harbored both the cytosolic calpains and DEK1-like calpains. The cytosolic calpains were subsequently lost within the evolutionary advanced charophytes, leaving the TML-calpain as the single calpain giving rise to the DEK1 clade of land plants. This supports our hypothesis that DEK1 evolved from TML-calpains by assuming a novel positional sensing function and thereby enabling the critical ability of land plants to direct 3D growth and development of complex organs. One possible explanation for the loss of cytosolic calpains is that cytosolic calpains interfered with DEK1 action and was therefore selected against during the transition from charophyte algae to land plants. Methods for genetic transformation of charophyte algae that could aid in elucidating calpain function in land plant predecessors are currently underway (Sørensen et al., 2014).

A vital role of DEK1 in controlling 3D growth and development is also supported by the fact that DEK1 exists as a single gene in the overwhelming majority of plants examined to date. During land plant evolution, a single TML-calpain evolved into the DEK1 clade, in which the calpain catalytic core domain CysPc is highly conserved (Liang et al., 2013). As shown here, the Loop sequence is much more divergent than the sequence of the calpain moiety domains, indicating a role for evolutionary pressure on the regulatory function of the Loop in driving morphological and functional divergence between bryophytes and early tracheophytes, and later dicots and monocots. Retention of a single *DEK1* gene in the vast majority of land plants implies selection pressure to maintain DEK1 calpain and its indispensable role in the developmental control of land plants. Elimination of additional copies following multiple genome-wide duplication events has been reported for genes with essential physiological and developmental roles (De Smet et al., 2013).

Previous studies have shown that complementation of *dek1* mutants in both *A. thaliana* and *P. patens* can be achieved by expression of the native CysPc-C2L domains, and in the case of *A. thaliana dek1*, also by the CysPc-C2L from *P. patens* (Johnson et al., 2008; Liang et al., 2013). However, the high frequency of abnormal phenotypes, and the requirement for a promoter with specific spatio-temporal activity, strongly point to an important role for MEM domain in regulating CysPc enzyme activity. An overview of *dek1* mutants, *DEK1* down-regulation and over-expression lines as well as genetic complementation experiments in *P. patens* and representative angiosperm species can be seen in Supplemental Table S6. Although a complete understanding of the mechanism of DEK1 activation in surface cells or gametophores via the MEM domain can only be achieved after determination of its 3D structure, some progress towards a better understanding of how this works is provided in this study by the identification of homology between subdomains of MEM and the Major Facilitator Superfamily (MFS) domains of secondary transporters (Supplemental Fig. S2). We propose that the function of MFS proteins in facilitating transport of various solutes across the membranes in response to chemiosmotic gradients is compatible with evolution of a functional role for MEM in sensing the difference between the surface membrane of a neighboring cell and the external environment. It is likely that this relationship was not discovered earlier due to the large sequence divergence between current day MFSs and MEM. The significance of the positioning of the Loop

on the cytosolic side can only be fully appreciated after the 3D structure has been solved. The importance of the Loop in modulating DEK1 calpain activity shown in this study is in line with a dominant negative effect reported in *A. thaliana* lines on which the MEM containing the Loop were overexpressed, while the lines overexpressing MEM without the Loop appeared WT (Tian et al., 2007). Removing the Loop in *P. patens dek1Δloop* mutant has significant effect on gametophore development. However, the effect is less severe than in the *Δdek1* mutant, since *dek1Δloop* buds are able to form and orient the wall in the bud apical cell perpendicularly to the first asymmetric division of the bud initial, and perpetuate cell proliferation (Fig. 3). In our interpretation, this suggests that the CysPc-C2L domain of *dek1Δloop* plants undergoes a basal level of activation under less stringent control leading to proteolytic activity, albeit at a lower level than in WT. In the subsequent development of the gametophore, the effect of removing the Loop becomes evident, causing complete blocking of phyllid development. Interestingly, phyllid progenitor-like cells are formed on *dek1Δloop* buds and later on an aberrant gametophore stem (Fig. 2 and Fig. 3), which shows that an intact Loop is required for the CysPc-C2L activity necessary for the asymmetric cell divisions that drive phyllid expansion. This assumption is further supported by the phenotypes of lines with introduced Loop coding sequences from *A. thaliana* and *Z. mays*, where the phyllids are formed but their proximo-distal and medio-lateral expansion is greatly affected (Fig. 5). One possible explanation for the phenotype of *dek1Δloop* that cannot presently be excluded is destabilization of the molecular structure of MEM as a result of removal of the Loop. However, the observation of near normal division plane positioning in the *dek1Δloop* early bud cells and progressive stem growth suggests that the MEM structure is not completely disturbed, but rather improperly regulated without the Loop. The independent folding of separate domains within a multi-domain protein has been reported for several proteins as a mechanism preventing separate domains from engaging in aberrant interactions with one another (Netzer and Hartl, 1997; Rüßmann et al., 2012). This fact also speaks against the severe disruption of MEM in the *dek1Δloop* mutant. The 3D structure determination and knowledge of intramolecular interactions within the MEM would shed more light to the structure-function relationships between the Loop and the rest of the MEM domain. In the future, we hope to be able to measure the *in vivo* activity of calpains in different mutants and in different cell types in which the effects on division plane determination is affected.

This has met with considerable difficulties even in the best studied cases with animal classical calpains (Zadran et al., 2010), however it represents a powerful future tool to understand the spatio-temporal control of calpain action.

Interestingly, the Loop segment from the liverwort *M. polymorpha* is fully functional in *P. patens* even if the amino acid identity between the Loops from these species is only 43%. It is therefore interesting to note that the *P. patens* Loop shares 38% amino acid identity to *A. thaliana* and 35% identity to *Z. mays* Loop, respectively. Recent phylogenetic studies resolved the liverworts as the earliest-divergent clade of land plants and mosses as the sister group to hornworts plus tracheophytes (Ligrone et al., 2012). According to a number of studies, liverworts, mosses, and hornworts diverged sequentially and form a paraphyletic group with the hornworts sister to the tracheophytes (Karol et al., 2001; Qiu et al., 2006). On the other hand, according to other recent analyses, bryophytes represent a monophyletic group (Cox et al., 2014). Our functional analysis presented in this work shows that despite the low sequence identity and uncertain evolutionary distance, Loops from the moss and the liverwort are inter-functional in the gametophytic and sporophytic phase and that a likely shift in Loop function occurred in angiosperms which displayed a different body plan organization. Another possible explanation for the lack of full complementation is failure of the postranscriptional prosessing of the *PpDEK1* transcripts containing the Loop-coding sequences from *A. thaliana* and maize. This possibility is raised by the presence of alternative splicing variants in the *AtLoop* and *ZmLoop* lines in addition to normal splice variants (data not shown).

Analysis of the transcriptome of WT and *Δdek1* protonemata and early gamethophores identifies DEK1 as a regulator of the initiation of gametophore buds in protonemata by suppressing ABP2 and ABP3 transcription factors, thereby keeping the number of buds to only one per 15 filaments in WT. In *Δdek1*, with an increased level of these transcription factors, four buds per filament are formed, whereas in *dek1Δloop* the level is intermediate, consistent with the observation of less severe phenotypes for this mutant in other traits as well. Recently, it was shown that the AP2-type transcription factors (APB1-4) are positively regulated by auxin and restrict the fate of the bud initial cell in *P. patens* protonema tissue (Aoyama et al., 2012). Here we show that a lack of DEK1 function in the *Δdek1* mutant causes an overall up regulation of APB2 and APB3 in protonemata (Fig. 6), which is accompanied by a

significant increase in bud formation (Perroud et al., 2014 and Fig. 4 in this work). This indicates that a physiological role of DEK1 in protonemata involves sensing of cues defined by local gradients of signaling molecules in the growing protonemal tissue, triggering the downstream events, which restrict the fate of side branch initial cells.

The negative control of bud initiation is relaxed in the *Δdek1* mutant. However, the buds fail to establish a functional meristematic stem cell and their further development is arrested at an early stage (Perroud et al., 2014). Tight control of almost invariant cell division plane positioning typical for WT buds is completely lost in *Δdek1* mutant, and, in addition to their misorientation, the newly formed cell walls are often bent and wrinkled (Perroud et al., 2014). A growing body of evidence shows that physical properties of the cell wall and mechanical forces between the neighboring cells constrain genetic regulation of cell proliferation and specification (Murray et al., 2012). However, the integrating mechanisms remains largely unknown. Recently, the functional interplay between *WUSCHEL*-related genes, which control stem cells in the meristems of flowering plants, and cell wall modifying enzymes has been reported in *P. patens* by Sakibara et al. (2014). In their work, the authors show that *PpWOX13L* activity is required for the upregulation of cell wall loosening enzymes which appear to be involved in stem cells formation and growth in *P. patens*. Potential role(s) of DEK1 in the pathways involving *WUSCHEL*-related genes has been hypothesized earlier (reviewed in Lau et al., 2012). Cell division plane orientation, cell wall expansion and fate specification are uncoupled in the *Δdek1* mutant (Perroud et al., 2014). As we show in this work, a lack of DEK1 function disturbs transcriptional regulation of the genes associated with cell wall modification and morphogenesis (Supplemental Fig. S9; Supplemental Table S4). This again may reflect an inability of the *Δdek1* mutant to sense and respond to the local cues generated by mechanical forces or gradients of signaling molecules both in protonemata and buds.

Conclusions

Here we show that *P. patens* strains in which the DEK1 Loop is deleted from the transmembrane domain (*dek1Δloop)* retain sufficient activity to allow cell divisions during early bud development. Subsequent leafy gametophore development

is, however, compromised. Although *dek1Δloop* strains form phyllid primordial cells on gametophore stems, they are incapable of further expansion. These results support our postulated sensory/regulatory role of the Loop segment in the spatio-temporal control of DEK1 activity. *In silico* re-examination of the Loop topology in DEK1 proteins from algae and land plants shows that the Loop is oriented towards the cytosol. Furthermore, we identify for the first time significant homology between DEK1 MEM and a known membrane protein family, namely Major Facilitator Superfamily. We believe that the function of MSF proteins give important hints as to how DEK1 may function in positional signaling to be explored in future experiments. DEK1 evolved from an ancient form of calpains containing large transmembrane domain (TML-calpains) some 1,5 billion years ago. Here we show that in contrast to the chlorphyte algae, which retained only cytosolic calpains, charophyte algae harbor both DEK1-like calpains and cytosolic calpains. During the evolutionary transition from advanced charophyte algae to land plants, the cytosolic calpins were lost, and the DEK1-calpain clade evolved, supporting our postulated key role for DEK1 during land plant evolution. In contrast to the highly conserved CysPc-C2L protease domains of land plants, we show that the Loop segment is more variable, both between representatives of charophyte algae and between charophytes and within DEK1 of land plants. A functional differentiation of the loop is also supported by the observation that the Loop sequence of the liverwort *M. polymorpha* genetically complement the *P. patens dek1Δloop* phenotype, whereas Loop sequences from *A. thaliana* and *Z. mays* only partially complement the same mutant. We interpret this finding to show that in order to stay functional as the land plant morphology evolved, the Loop also evolved. Finally, we show that DEK1 activity is required for controlled expression of genes involved in cell wall remodeling and developmental transition in side branch initials from secondary protonema to bud initial cells. These results indicate that DEK1, although not essential for protonemata cell division and differentiation, may play a role in modulating growth responses globally, likely via the sensing of local cues which determine bud initiation and cell wall expansion. Later, in developing buds and gametophores, DEK1 plays essential role in body patterning, were various activities are likely needed to control different morphogenetic programs.

Material and Methods

Plant material and growth conditions

In this study, we used *Physcomitrella patens* Gransden strain. Tissue maintenance and production was performed on BCDA media as described in Cove et al. (2009). *P. patens* tissue and protoplasts were grown under long day conditions (16 hours light [70 to 80 $\mu molm^{-2}s^{-1}$]/8 hours dark) at 25 °C. Medium was supplemented with 30 μgl^{-1} of Hygromycin B or 50 μgl^{-1} of G418 for selection of transformed cells. All phenotypic characterizations were performed on BCD medium unless specifically mentioned (Cove et al., 2009). Culture for bud count was established as follow. BCD containing Petri dish was inoculated with 16 equally spaced spot inoculums consisting of 10 µl of protonemal tissue suspension. Bud count was performed after 14 days of growth on 15 cells caulonemal filament. Standardly, 100 filaments were randomly picked from each plate to establish a budding pattern. Tissue for sporophyte production was grown on sterile Jiffy7 soil blocks placed in the glass flasks under short day conditions (8 hours light [70 to 80 $\mu molm^{-2}s^{-1}$]/16 hours dark) at 15 °C and manipulated as described by Perroud et al. (Perroud et al., 2011). Tissue for RNAseq analysis was grown and harvested as follows: tissue of protonema cells grown on BCDA medium under the long day regime (16 hours light [70 to 80 $\mu molm^{-2}s^{-1}$]/8 hours dark) at 25 °C was collected, homogenized in sterile water and inoculated on BCD medium overlaid with cellophane discs (2 ml of homogenized tissue per plate) as described in Cove et al., (2009). Then, the tissue was collected after 6 and 14 days of growth under the long day regime (16 hours light [70 to 80 $\mu molm^{-2}s^{-1}$]/8 hours dark) at 25 °C, frozen in liquid nitrogen and stored at -80 °C. The procedure was repeated three times to obtain replicates for further analysis.

In silico analyses

The DEK1 sequences used in this study are listed in Supplemental Table S1. The topography of the transmembrane domain of DEK1 was analyzed with HMMTOP2.0 (Tusnády and Simon, 2001), using the image creator MyDomain tool (http://prosite.expasy.org/mydomains/) to visualize the result. DEK1 sequences were submitted to the SMART server (http://smart.embl-heidelberg.de/) to identify and extract the amino acids corresponding to the DEK1-Loop segment. The DEK1-Loop sequences were aligned using MAFFT v.7.058b and the L-INS-i algorithm. To

calculate the pairwise sequence identity between the DEK1-Loop sequences, the MAFFT alignment was submitted to SIAS at http://imed.med.ucm.es/Tools/. This tool calculates the identity as the number of identical residues divided by the length of the shortest sequence. To generate the DEK1-Loop sequence logo, the WebLogo tool (http://weblogo.berkeley.edu/logo.cgi) was used. We performed phylogenetic analyses using the maximum likelihood method as implemented in RAxML version 7.2.6 (Stamatakis, 2006). We used the WAG model of protein evolution (Whelan and Goldman, 2001) with gamma distributed rate heterogeneity. Branch support was assessed by running 1000 bootstrap replicates.

PpDEK1_ΔLoop and Loop complementation vectors construction

Primers used for vector construction are listed in Supplemental Table S2. The schematics for the gene deletion and knock-in complementation constructs are shown in Supplemental Fig. S3. All nucleotide numberings are relative to the A[1]TG start site in the *P. patens DEK1* gene sequence (Pp1s173_19V6.1; www.phytozome.net) unless otherwise stated. All generated plasmid vectors were verified by restriction digestion analysis and sequencing.

The Loop deletion vector was designed to remove the Loop coding sequence and insert a hygromycin resistance cassette in a single intron. After sequences and assembly design, fragment syntheses and cloning were ordered and performed by Genscript. Inc. USA. Shortly, 5' targeting sequence spans nucleotides 2561-3563 and is flanked by 5' *Hind*III and 3' *Nru*I restriction enzymes. This fragment ends in intron 7. In order to avoid any splicing conflict with heterologous splicing border pair, the 3' targeting sequence starts with the end of intron 6 exon 7 border, nucleotides 3562-3609, and continue with exon 8 sequence, nucleotides 5298-6301 (Supplemental Fig. S3A). Additionally, 5' *Kas*I and 3' *Nsi*I restrictions enzymes flank the 3' fragment. Both fragments were cloned sequentially into *pBHRH* (Schaefer et al. 2010) using *Hind*III/*Nru*I and *Kas*I/*Nsi*I restriction enzyme pairs, respectively creating the *pBHRF-DEK1-Loop* vector (Supplemental Fig. S3A). Prior to transformation, *pBHRF-DEK1-Loop* was digested with the restriction enzymes *Hind*III and *Nsi*I.

Complementation constructs have been assembled to re-insert the Loop sequences from different species into the deleted locus. First, to re-insert the *P. patens* Loop sequence into its native locus, a DNA fragment, spanning the Loop sequence,

and 5' and 3' flanking regions (nucleotides 2563-5719), was PCR amplified from genomic *P. patens* DNA using primers V1/SP and V1/ASP. The resulting 3156 bp PCR fragment was cloned into the *Zero Blunt* PCR cloning vector (Invitrogen, Carlsbaden, USA), giving plasmid *pCR_PpLOOP_V1* (Supplemental Fig. S3B). *pCR_PpLOOP_V1* was digested with *Eco*RI restriction enzyme prior protoplast transformation. Secondly, we built three constructs aimed to test heterologous Loop sequences from *A. thaliana (At), Z. mays (Zm)* and *M. polymorpha (Mp)*. In these constructs chimeric Loop from *At* (nucleotide 2028-2855 relative to the ATG start site; AT1G55350), *Zm* (nucleotide 4181-5029 relative to the ATG start site; A4061804.1) or *Mp* (see below) flanked by *P. patens* DEK1 5' and 3' targeting sequence were constructed as follow: Two DNA fragments were synthetized *de novo* by GeneScript (http://www.genscript.com/): 1) The Loop 5' targeting fragment was a 2174 bp chimeric sequence composed of *P. patens DEK1* nucleotides 2565-3681/*M. polymorpha DEK1* nucleotides 1036-1935 (Liang et al., 2013)/*P. patens DEK1* nucleotides 4528-4655, and flanked with 5' and 3' *Pml*I and *Xho*I restriction sites, respectively; 2) The Loop 3' targeting fragment was a 968 bp sequence of *P. patens* DEK1 nucleotides 2179-3136, and flanked with 5' and 3' *Mlu*I and *Cla*I restriction sites, respectively. The Loop 5' and 3' targeting fragments were inserted into the vector *pBNRF* using *Pml*I/*Xho*I and *Mlu*I/*Cla*I restriction sites, respectively, resulting in plasmid *pBNRF-MpDEK1-Loop-Comp* (Supplemental Fig. S3C). Before transformation, *pBNRF-MpDEK1-Loop-Comp* was digested with *Pml*I and *Cla*I restriction enzymes. To make the corresponding *At* and *Zm* Loop complementation constructs, the In-Fusion Cloning Strategy was used (Clontech Laboratory). To generate the *At* and *ZmLoop* In-Fusion inserts, forward and reverse gene specific primers containing 5' and 3' 15 bp extensions complementary to the *P. patens Loop* 3' flanking sequences in the *pBNRF-MpDEK1-Loop-Comp* plasmid was used to PCR amplify *At* (primers *At_Loop_ifc_SP* and *At_Loop_ifc_ASP*) and *Zm* (primers *Zm_Loop_ifc_SP* and *Zm_Loop_ifc_ASP*) from genomic DNA. These inserts were each mixed with linearized vector *pBNRF-MpDEK1-Loop-Comp*, produced by PCR amplification using primers *SP_Loop_Comp* and *ASP_Loop_Comp* to exclude the *Mp*Loop coding sequence. The inserts and linearized vector were ligated using the In-Fusion Cloning strategy according to the manufacturer's instructions, resulting in plasmids *pBNRF-AtDEK1-Loop-Comp* and *pBNRF-ZmDEK1-Loop-Comp*

(Supplemental Fig. S3C). Prior transformation, these two plasmids were digested using *Sal*I/*Swa*I restriction enzymes.

Transformation procedure

P. patens protoplast production and stable transformation was performed according Schaefer and Zryd 1997 modified by Cove et al. 2009 with 15 µg of linearized plasmid DNA used per transformation. Shortly, transformed protoplast regeneration and selection was performed by transferring the culture to different media according to the following sequence: 6 days of protoplast regeneration on PRMB medium, 6 days of selection on BCDA medium supplemented with the appropriate antibiotic, 14 day of growth on BCDA medium and 7 days on BCDA supplemented with the appropriate antibiotic. Resistant plant were then picked individually on fresh BCDA medium and used for genotyping and phenotype analysis after sufficient growth.

Cre recombinase procedure to remove resistance marker from primary transformant was performed as previously described (Trouiller et al 2006) with minor modifications. Transformed tissue was grown as wild type and protoplast production and transformation carried out using 20 µ pAct-Cre (Trouiller et al 2006) using regular procedure. Protoplasts were plated on cellophane diluted (25 000 counted protoplast per 9 cm Petri dish) to avoid picking mixed regenerated plant. Protoplast regeneration and test procedure were performed as follow: 1) four days protoplasts regeneration on PRMB medium; 2) four days protoplasts growth on BCDA medium; 3) individual plant picking on fresh BCDA plate and growth for eight days; and 4) replica plating of each individual plant unto BCDA and BCDA with the appropriate antibiotic. Strain showing loss of antibiotic resistance were selected and grown until sufficient tissue was available for genotyping and phenotype analysis.

Molecular characterization of transformants

Genomic DNA for Southern Blot analysis was extracted using the Nucleon™ PhytoPure™ Genomic DNA Extraction Kit (GE Healthcare). Southern Blot was performed as described by Perroud and Quatrano (2006) using 1 µg DNA per digestion. Probes were DIG-labelled using the DIG Probe PCR synthesis kit (Roche, Indianapolis, IN, USA) according to the manufacturer's instructions. Templates for

PCR amplification of probes were genomic DNA extracted from *P. patens* and *pBHRF* vector. Four different probes (Supplemental Fig. S5) were made to hybridize to the 5' and 3' targeting sequences, the kanamycin resistant gene (G418) and the Loop coding sequence using primer pairs *PpL5_S_Sp/PpL5_S_Asp*, *PpL3_S_Sp/PpL3_S_Asp*, *G418-F/G418-R* and *PpLL_S_Sp/ PpLL_SAsp*, respectively.

RT-PCR and sequencing was used to analyze *P. patens* DEK1 transcripts to ensure proper deletion/insertions and splicing at the *DEK1 Loop* locus. Total RNA was isolated from *P. patens* protonemata using the Plant RNeasy Kit (Qiagen). 500 ng *DNase*I-treated total RNA was reverse-transcribed by 200 U Superscript III Reverse Transcriptase (Invitrogen) primed with random hexamers (50 µM) at 55 °C for 60 min. Phusion High-Fidelity DNA polymerase was used to amplify the target sequence spanning a region from exon 6 to exon 9 (genomic nucleotides 3190-5358) as follows: 1 µL undiluted cDNA template was PCR amplified with primers *RT-Loop-F* and *RT-Loop-R* (Supplemental Fig. S3) using the following cycling conditions: 98 °C for 30 sec, 35 cycles of 98 °C for 10 sec, 56 °C for 30 sec and 72 °C for 30 sec, and a final elongation step of 72 °C for 5 min. The PCR product was treated with exonuclease I (Fermentas) to remove excess primers, and then cycling sequencing reactions were performed using primers *PpL_5_Tar-Fw* and *PpL_3_Tar-Rv*, and the ABI BigDye v.3.1 chemistry according to the St_eP method (Platt et al., 2007). DNA fragments were precipitated using sodium-acetate/ethanol and finally sequenced by Capillary Electrophoresis using the 3130xL Genetic Analyzer. The Genomic Workbench Software was used to analyze the sequences.

Genotyping of the complemented lines

First, we genotyped obtained transformants by PCR, looking for loss of the original *dek1Δloop* locus with *RT-Loop-F, RT_Loop-R* primers. Then, we screened selected lines from the first round of genotyping for single-copy insertion at the locus using *LoopGenot-F* and *LoopGenot-R* primers annealing upstream and downstream of the targeting sequences, respectively (Supplemental Fig. S3). To confirm targeting of the cassette from both 5` and 3`sites, we PCR-genotyped the lines using the *LoopGenot_F* and *35S-R* primers (5`targeting) and *Term-F* and *LoopGento-R* primers (3`targeting) (Supplemental Fig. S3). Southern blot was performed to identify the

lines with random insertion of the targeting construct in the genome (Supplemental Fig. S5). Resistance cassettes were eventually removed from selected lines using Cre recombinase-mediated approach as described above. The lines which showed loss of the resistance against G418 were further genotyped in three steps: 1) by PCR using the *RT-Loop-F*/*RT_Loop-R*, Southern blot and sequencing of the cDNA regions overlapping introduced chimeric Loops as described above.

Molecular procedure for high throughput sequencing

Total RNA was extracted from frozen material using the RNeasy lipid tissue mini kit (Qiagen) with few modifications. Briefly, the frozen tissue was thoroughly homogenized in liquid nitrogen using a mortar and pistil. Approximately 120 mg of powdered tissue was lysed in 1ml QIAzol lysis reagent. Two hundred microliters of chloroform was added and the mixture was centrifuged at 4 °C. The aqueous phase was collected, 1.5 volume of 100% ethanol was added and the mixture was vortexed. After binding of the RNA to the RNeasy mini spin column, on-column DNaseI treatment was performed to remove genomic DNA. The column was washed with the RPE buffer, dried and RNA eluted in 45µl of RNase-free water. The concentration of RNA was measured and RNA integrity was further assessed using an Agilent 2100 Bioanalyzer (DE54704553, Agilent Technologies, Inc.) with a RNA 6000 LabChip kit. The RNA samples were stored at -80^{0}C until sent for sequencing.

The library construction and sequencing was performed at Beijing Genomics Institute (BGI), Hong Kong, China (http://www.genomics.cn/en/index). The total RNA samples were treated with DNase I to degrade any possible DNA contamination. Then the mRNA was enriched by using the oligo(dT) magnetic beads. Mixed with the fragmentation buffer, the mRNA was fragmented into short fragments (about 200 bp). The first strand of cDNA was synthesized by using random hexamer-primers. Buffer, dNTPs, RNase H and DNA polymerase I were added to synthesize the second strand and the double strand cDNA was purified with magnetic beads. End reparation and 3'-end single nucleotide adenine addition was then performed. Finally, sequencing adaptors were ligated to the fragments and the fragments were enriched by PCR amplification. During the QC step, Agilent 2100 Bioanaylzer and ABI StepOnePlus Real-Time PCR System were used to qualify and quantify the sample

library. The library products were sequenced via Illumina HiSeqTM 2000. The read length for all samples was 49 bp.

Bioinformatics analyses of high throughtput dataset

Reads were aligned against the genomic assembly of *P. patens* (Rensing et al., 2008; Zimmer et al., 2013), (http://www.phytozome.net/physcomitrella.php, v1.6; Ppatens_152.fa and Ppatens_152_gene_exons.gff3) using Bowtie (2.1.0) (Langmead et al., 2009) and Tophat (2.0.10) (Trapnell et al., 2009) using default parameters. Sample quality was assessed using FastQC (0.10.1) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), with all samples passing quality control. For downstream analysis only uniquely mapped reads were kept (Supplemental Table S7).

Gene ontology (GO) (Ashburner et al., 2000) terms were taken from the Physcomitrella annotation (http://www.phytozome.net/physcomitrella.php, v1.6; Ppatens_152_annotation_info.txt). GOSlim terms were derived using GOSlimAuto provided by AgBase (McCarthy et al., 2006) based on the automated slimming of GO term annotations (Davis et al., 2010). The significance of over representation was quantified using hypergeometric tests as implemented in the R package GOstats (Falcon and Gentleman, 2007) with parameter conditional set to FALSE and FDR adjustment of p-values (Benjamini and Hochberg, 1995). We applied cuffdiff (2.1.1) (Trapnell et al., 2010) to estimate gene abundance and to test for pair-wise differential expression using the *Physcomitrella* gene models (parameters: -dispersion-method per-condition --library-norm-method quartile --frag-bias-correct Ppatens_152.fa). Genes with an FPKM larger than 1 as reported from cuffdiff were marked as being expressed. We applied a count-based approach for finding genes showing a significant difference between WT and *Δdek1* in the change of expression between 6 days and 14 days. After extracting the gene-wise unique raw counts by using HTSeq (http://www-huber.embl.de/users/anders/HTSeq), we applied DESeq2 (Anders and Huber, 2010) which makes use of negative binomial generalized linear models to test for differentially expressed genes. We tested for an interaction between genotype and time, so that the resulting genes were those for which the amount of change between the two time points was significantly different between WT and *Δdek1*. Genes with an FDR-adjusted p-value below 0.05 and an absolute log2 fold change larger than 1 were kept as being differentially expressed.

Supplemental Material

Supplemental Protocol 1. RNAseq data validation.
Supplemental Table S1. DEK1 Loop sequences from land plants and charophyte algae used in this study.
Supplemental Table S2. Primer sequences used in this study.
Supplemental Table S3. Gene expression as reported by cufflinks.
Supplemental Table S4. Interaction differential expression analysis results.
Supplemental Table S5. Overview of identified CysP transcripts from charophyte algae.
Supplemental Table S6. Overview of phenotypes of the *dek1* muatnts, *DEK1* down-regulation and over-expression lines and genetic complementation experiments in *Physcomitrella patens* and angiosperm species.
Supplemental Tabe S7. Read mapping results.
Supplemental Figure S1. Bioinformatic analyses of Loop sequences.
Supplemental Figure S2. Conserved domains detected in the *Physcomitrella patens* DEK1 protein by RPS-BLAST using the Conserved Domain Architecture Retrieval Tool (CDART) at NCBI.
Supplemental Figure S3. Vector construction for targeted deletion and replacements of the PpLoop.
Supplemental Figure S4. Gametophore morphology in *dek1Δloop* line before and after the Cre recombinase-mediated removal of the resistance cassette.
Supplemental Figure S5. Southern blot genotyping, RT-PCR.
Supplemental Figure S6. Phyllid development failure in the *dek1Δloop* mutant.
Supplemental Figure S7. Sporophyte formation in WT, *dek1Δloop* and Loop complemented lines.
Supplemental Figure S8. Micrographs of the *Physcomitrella patens* tissue used for RNA-seq analysis .
Supplemental Figure S9. Transcriptome comparison between WT and *Δdek1*. GOSlim enrichment for DEGs from interaction.
Supplemental Figure S10. Expression of selected *PpVNS* genes.
Supplemental Figure S11. Corelation between biological replicates.
Supplemental Figure S12. K-means clustering of the dataset.
Supplemental Figure S13. Full dataset principal component analysis (PCA).
Supplemental Figure 14. Comparison of the dataset expressed genes with external datasets.
Supplemental Figure S15. Expression of *PpDEK1* and control genes in the dataset.
Supplemental Figure S16. Track view of *PpDEK1* expression in the dataset.
Supplemental references.

Literature Cited

Anders S,Huber W (2010) Differential expression analysis for sequence count data. Genome Biol **11(10):** R106

Ashburner M,**Ball CA**,**Blake JA**,**Botstein D**,**Butler H**,**Cherry JM**,**Davis AP**,**Dolinski K**, **Dwight SS**,**Eppig JT**,**Harris MA**,**Hill DP**,**Issel-Tarver**

L,**Kasarskis A**,**Lewis S**,**Matese JC**, **Richardson JE**,**Ringwald M**,**Rubin GM**,**Sherlock G** (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet **25(1):** 25-29

Aoyama T,**Hiwatashi Y**,**Shigyo M**,**Kofuji R**,**Kubo M**,**Ito M**,**Hasebe M** (2012) AP2-type transcription factors determine stem cell identity in the moss Physcomitrella patens. Development **139(17):** 3120-3129

Benjamini Y,Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc 57(1): 289-300

Bernsel A,Viklund H,Hennerdal A,Elofsson A (2009) TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res. 37(Web Server issue): W465-468

Buda GJ,Barnes WJ,Fich EA,Park S,Yeats TH,Zhao L,Domozych DS,Rose JKC (2013) An ATP Binding Cassette Transporter Is Required for Cuticular Wax Deposition and Desiccation Tolerance in the Moss Physcomitrella patens. Plant Cell 25: 4000–4013

Cove DJ,Perroud PF,Charron AJ,McDaniel SF,Khandelwal A,Quatrano RS. (2009) The moss Physcomitrella patens. A novel model system for plant development and genomic studies. In: DA Crotty, A Gann, eds, Emerging model organisms, a laboratory manual. New York, NY, USA: Cold Spring Harbor Laboratory Press, p 69–104

Cox CJ,Li B,Foster PG,Embley TM,Civán P (2014) Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst Biol **63(2):** 272-279

Davis MJ,**Sehgal MS**,**Ragan MA** (2010) Automatic, context-specific generation of Gene Ontology slims. BMC Bioinformatics **11:** 498

De Smet I,Voss U,Jürgens G,Beeckman T (2009) Receptor-like kinases shape the plant. Nature Cell Biol **11:** 1166-1173

De Smet I,Beeckman T (2011) Asymmetrcic cell divisions in land plants and algae: the driving force for differentiation. Nat Rev Mol Cell Biol **12(3):** 177-188

De Smet R,Adams KL,Vandepoele K,Van Montagu MC,Maere S,Van de Peer Y (2013) Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants. Proc Natl Acad Sci USA **110(8):** 2898-903

Falcon S,**Gentleman R** (2007) Using GOstats to test gene lists for GO term association. Bioinformatics **23(2):** 257-258

Graham LE,Cook ME,Busse JS (2000) The origin of plants: Body plan changes contributing to a major evolutionary radiation. Proc Nat Acad Sci USA **97(9):** 4535-4540

Harrison CJ,Roeder AH,Meyerowitz EM,Langdale JA (2009) Local cues and asymmetric cell divisions underpin body plan transitions in the moss Physcomitrella patens. Current Biology **19:** 461–471

Johnson KL,**Degnan KA**,**Ross Walker J**,**Ingram GC** (2005) AtDEK1 is essential for specification of embryonic epidermal cell fate. Plant J **44(1):** 114-127

Johnson KL,Faulkner C,Jeffree CE,Ingram GC (2008) The phytocalpain defective kernel 1 is a novel Arabidopsis growth regulator whose activity is regulated by proteolytic processing. Plant Cell **20(10):** 2619-2630

Karol KG,McCourt RM,CiminoMT,Delwiche CF (2001) The closest living relatives of land plants. Science **294:** 2351–2353

Käll L,Krogh A,Sonnhammer EL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol **338(5):** 1027-1036

Kenrick P,Crane PR (1997) The origin and early evolution of plants on land. Nature **389:** 33-39

Krogh A1,Larsson B,von Heijne G,Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305(3):** 567-580

Kumar SB,Venkateshwaran K and Kundu S (2010) Alternative conformational model of a seed protein DeK1 for better understanding of structure-function relationship. J Prot Proteomics **1:** 77-90

Langmead B,**Trapnell C**,**Pop M**,**Salzberg SL** (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol **10(3):** R25

Lau S,Slane D,Herud O,Kong J,Jürgens G (2012) Early embryogenesis in flowering plants: setting up the basic body pattern. Annu Rev Plant Biol **63:** 483-506

Leliaert F,Smith DR,Moreau H,Herron MD,Verbruggen H,Delwiche CG,De Clerck O (2012) Phylogeny and Molecular Evolution of the Green Algae. Crit Rev Plant Sci **31:** 1-46

Liang Z,Demko V,Wilson R,Johnson K,Ahmad R,Perroud P,Quatrano R,Zhao S, Shalchian-Tabrizi K,Otegui M,Olsen OA, Johansen W (2013) The catalytic domain CysPc of the DEK1 calpain is functionally conserved in land plants. Plant J **75(5):** 742-754

Lid SE,Gruis D,Jung R,Lorentzen JA,Ananiev E,Chamberlin M,Niu XM,Meeley R,Nichols S, Olsen OA (2002) The defective kernel 1 (dek1) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily. Proc Nat Acad Sci USA **99(8):** 5460-5465

Lid SE,Olsen L,Nestestog R,Aukerman M,Brown RC,Lemmon B,Mucha M,Opsahl-Sorteberg HG,Olsen OA (2005) Mutation in the Arabidopisis thaliana DEK1 calpain gene perturbs endosperm and embryo development while over-expression affects organ development globally. Planta **221(3):** 339-3515

Ligrone R,Duckett JG,Renzaglia KS (2012) Major transitions in the evolution of early land plants: a bryological perspective. Ann Bot **109(5):** 851-871

McCarthy FM,**Wang N**,**Magee GB**,**Nanduri B**,**Lawrence ML**,**Camon EB**,**Barrell DG**, **Hill DP**,**Dolan ME**,**Williams WP**,**Luthe DS**,**Bridges SM**,**Burgess SC** (2006) AgBase: a functional genomics resource for agriculture. BMC Genomics **7:** 229

Murray JA,**Jones A**,**Godin C**,**Traas J** (2012) Systems analysis of shoot apical meristem growth and development: integrating hormonal and mechanical signaling. Plan Cell **24(10):** 3907-3919

Müller S,Wright AJ,Smith LG (2009) Division plane control in plants: new players in the band. Trends Cell Biol **19(4):** 180-188

Netzer WJ,**Hartl FU** (1998) Protein folding in the cytosol: chaperonin-dependent and -independent mechanisms. Trends Biochem Sci **23(2):** 68-73

Ono Y,Sorimachi H (2012) Calpains: an elaborate proteolytic system. Biochim Biophys Acta **1824(1):** 224-236

Pao SS,Paulsen IT,Saier MH (1998) Major facilitator superfamily. Microbiol Mol Biol Rev **62(1):** 1-34

Perroud PF,Quatrano RS (2006) The role of ARPC4 in tip growth and alignment of the polar axis in filaments of Physcomitrella patens. Cell Motil Cytoskeleton **63(3):** 162-171

Perroud PF,Cove DJ,Quatrano RS,McDaniel SF (2011) An experimental method to facilitate the identification of hybrid sporophytes in the moss Physcomitrella patens using fluorescent tagged lines. New Phytol **191(1):** 301-330

Perroud PF,Demko V,Johansen W,Wilson RC,Olsen OA,Quatrano RS (2014) Defective Kernel 1 (DEK1) is required for three-dimensional growth in Physcomitrella patens. New Phytol **203(3):** 794-804

Pires ND,Dolan L (2012) Morphological evolution in land plants: new designs with old genes. Phil Trans R Soc B **367:** 508-518

Platt AR,Woodhall RW,George AL (2007) Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. BioTechniques **43:** 58-62

Qiu YL,Li L,Wang B,Chen Z,Knoop V,Groth-Malonek M,Dombrovska O,Lee J,Kent L,Rest J,Estabrook G.F,Hendry TA,Taylor DW,Testa CM,Ambros M,Crandall-Stotler B,Duff RJ,Stech M,Frey W,Quandt D,Davis CC (2006) The deepest divergences in land plants inferred from phylogenomic evidence. Proc Natl Acad Sci USA **103:** 15511–15516

Rensing SA,Lang D,Zimmer AD,Terry A,Salamov A,Shapiro H,Nishiyama T, Perroud PF,Lindquist EA,Kamisugi Y,Tanahashi T,Sakakibara K,Fujita T, Oishi K,Shin-I T, Kuroki Y,Toyoda A,Suzuki Y,Hashimoto S,Yamaguchi K, Sugano S,Kohara Y,Fujiyama A,Anterola A,Aoki S,Ashton N,Barbazuk WB, Barker E,Bennetzen JL,Blankenship R,Cho SH,Dutcher SK,Estelle M,Fawcett JA,Gundlach H,Hanada K,Heyl A,Hicks KA,Hughes J,Lohr M,Mayer K, Melkozernov A,Murata T,Nelson DR,Pils B,Prigge M,Reiss B,Renner T, Rombauts S,Rushton PJ,Sanderfoot A,Schween G,Shiu SH,Stueber K, Theodoulou FL,Tu H,Van de Peer Y,Verrier PJ,Waters E,Wood A,Yang L,Cove D,Cuming AC,Hasebe M,Lucas S,Mishler BD,Reski R,Grigoriev IV,Quatrano RS,Boore JL (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science **319(5859):** 64-69

Rüßmann F,**Stemp MJ**,**Mönkemeyer L**,**Etchells SA**,**Bracher A**,**Hartl FU** (2012) Folding of large multidomain proteins by partial encapsulation in the chaperonin TRiC/CCT. Proc Natl Acad Sci USA **109(52):** 21208-21215

Schaefer DG,Zrÿd JP (1997) Efficient gene targeting in the moss Physcomitrella patens. Plant J **11(6):** 1195-1206

Schaefer DG,Delacote F,Charlot F,Vrielynck N,Guyon-Debast A,Le Guin S,Neuhaus JM, Doutriaux MP,Nogue F (2010) Rad51 loss of function abolishes gene targeting and de-represses illegitimate integration in the moss Physcomitrella patens. DNA Repair (Amst) **9:** 526–533

Sørensen I,Fei Z,Andreas A,Willats WG,Domozych DS,Rose JK (2014) Stable transformation and reverse genetic analysis of Penium margaritaceum: a platform for studies of charophyte green algae, the immediate ancestors of land plants. Plant J **77(3):** 339-351

Spinner L,**Gadeyne A**,**Belcram K**,**Goussot M**,**Moison M**,**Duroc Y**,**Eeckhout D**,**De Winne N**, **Schaefer E**,**Van De Slijke E**,**Persiau G**,**Witters E**,**Gevaert K**,**De Jaeger G**,**Bouchez D**, **Van Damme D**,**Pastuglia M** (2013) A protein phosphatase 2A complex spatially controls plant cell division. Nat Commun **4:** 1863

Stamatakis A (2006) RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. Bioinformatics **22(21):** 2688-2690

Tian Q,Olsen L,Sun B,Lid SE,Brown RC,Lemmon BE,Fosnes K,Gruis DF,Opsahl-Sorteberg HG,Otegui MS,Olsen OA (2007) Subcellular localization and functional domain studies of DEFECTIVE KERNEL1 in maize and Arabidopsis suggest a model for aleurone cell fate specification involving CRINKLY4 and SUPERNUMERARY ALEURONE LAYER1. Plant Cell **19(10):** 3127-3145

Trapnell C, **Pachter L**,**Salzberg SL** (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25(9):** 1105-1111

Trapnell C,**Williams BA**,**Pertea G**,**Mortazavi A**,**Kwan G**,**van Baren MJ**,**Salzberg SL**, **Wold BJ**,**Pachter L** (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28(5):** 511-515

Trouiller B,Schaefer DG,Charlot F,Nogue F (2006) MSH2 is essential for the preservation of genome integrity and prevents homeologous recombination in the moss Physcomitrella patens. Nucleic Acids Res **34:** 232-242

Tusnády GE,Simon I (1998) Principles governing amino acid composition of integral membrane proteins: application to topology prediction. J Mol Biol **283(2):** 489-506

Viklund H,Bernsel A,Skwark M,Elofsson A (2008) A SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology. Bioinformatics. **24(24):** 2928-2929

Whelan S,Goldman N (2001) A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. Mol Biol Evol **18(5):** 691-699

Yan N (2013) Structural advances for the major facilitator superfmaily (MFS) transporters (2013) Trends Biochem Sci **38(3):** 151-159

Xu B,**Ohtani M**,**Yamaguchi M**,**Toyooka K**,**Wakazaki M**,**Sato M**,**Kubo M**,**Nakano Y**,**Sano R**, **Hiwatashi Y**,**Murata T**,**Kurata T**,**Yoneda A**,**Kato K**,**Hasebe M**,**Demura T** (2014) Contribution of NAC transcription factors to plant adaptation to land. Science **343(6178):** 1505-1508

Zadran S,Jourdi H,Rostamiani K,Qin Q,Bi X,Baudry M (2010) Brain-derived neurotrophic factor and epidermal growth factor activate neuronal m-calpain via mitogen-activated protein kinase-dependent phosphorylation. J Neurosci **30(3):** 1086-1095

Zhao S, Liang Z, Demko V, Wilson R, Johansen W, Olsen OA, Shalchian-Tabrizi K (2012) Massive expansion of the calpain gene family in unicellular eukaryotes. BMC Evol Biol **12(1):** 193

Zimmer A,Lang D,Richardt S,Frank W,Reski R,Rensing SA (2007) Dating the early evolution of plants: detection and molecular clock analyses of orthologs. Mol Genet Genomics **278(4):** 393-402

Zimmer AD,Lang D,Buchta K,Rombauts S,Nishiyama T,Hasebe M,Van de Peer Y,Rensing SA,Reski R (2013) Reannotation and extended community resources for the genome of the non-seed plant Physcomitrella patens provide insights into the evolution of plant gene structures and functions. BMC Genomics **14:** 498-518

Figure legends

Figure 1. Land plant Loop phylogeny and structure. Bioinformatic analyses of land plant Loop sequences. A, Diagram of the predicted DEK1 structure. Bar represents the length of 100 amino acids. B, Phylogenetic analysis of land plant Loop sequences. The Loop sequences cluster in four major phylogenetic groups corresponding to bryophytes, lycophytes, monocots and dicots. C, Alignments between the bryophyte and lycophyte (group 1), monocot (group 2) and dicot (group 3) Loop consensus sequences showing the absolutely conserved positions and the degree of sequence conservation (red = 100 % conserved positions/regions) using CLC Genomic Workbench to visualizethe result. D, Consensus prediction of *Physcomitrella patens* DEK1 membrane protein topology (TOPCONS) suggests the Loop (black lined box) to be localized at the cytoplasmic side of the plasma membrane.

Figure 2. Gametophore development in WT, *dek1Δloop* and *Δdek1*. A, young WT gametophore grown on BCD medium. B, young *dek1Δloop* gametophore grown on BCD medium. C, arrested *Δdek1* bud grown on BCD medium. D, mature WT gametophore with differentiated sporophyte cultivated under the sporophyte-production conditions on soil block. E, mature *dek1Δloop* gametophore cultivated under the sporophyte-production conditions on soil block (lower left sector shows the same mutant gametophore in the scale comparable to WT gametophore on D). Arrow points to the *Δdek1* bud; arrowheads point to the filamentous protrusions formed on the *dek1Δloop* gametophore stem. Bar: 500 µm.

Figure 3. Bud and early gametophore development in WT and *dek1Δloop* mutant. A-D, bud development in WT. E-H, bud development in *dek1Δloop*. I, juvenile WT gametophore with emerging phylid. J, juvenile *dek1Δloop* gametophore with filamentous protrusion formed from the phyllid progenitor cell. White arrows point to the first asymmetric division of the bud initial. Red arrows point to the first division of the bud apical cell. Arrowheads point to the phyllid primordials. Barr: 50 µm.

Figure 4. Bud initiation in WT, *Δdek1, dek1Δloop, MpLoop, AtLoop* and *ZmLoop* lines. Graph showing average number of buds formed per 15 filament cells in WT, *Δdek1, dek1Δloop, MpLoop, AtLoop* and *ZmLoop* lines. Values shown are means ± s.e. Different letters denote significant differences at P = 0.05 (ANOVA, LSD-test).

Figure 5. Gametophore development in WT and *dek1Δloop* complementation lines. A, WT gametophore. B, *dek1Δloop* gametophore. C, *PpLoop - dek1Δloop* complemented with the Loop coding sequence from *Physcomitrella patens* showing the WT phenotype. D, *MpLoop - dek1Δloop* complemented with the Loop coding sequence from *Marchantia polymorpha*. E, *AtLoop - dek1Δloop* complemented with the Loop coding sequence from *Arabidopsis thaliana*. F, *ZmLoop - dek1Δloop* complemented with the Loop coding sequence from *Zea mays*. G-J, isolated phyllids from apical, sub-apical, middle and basal part of the gametophores. G, WT. H, *MpLoop*. I, *AtLoop*. J, *ZmLoop*. Bar: 1 mm (A-F), 500 µm (G-K).

Figure 6. Transcriptome comparison between WT and *Δdek1*. A, Venn diagram representing the number of detected transcripts with a FPKM>1 in the two strains (WT and *Δdek1*) at the two time points (7 and 14 days). B, Expression of AP2-type transcription factors. Expression of the AP2-type transcription factors, APB1, APB2, APB3, and APB4. The height of the bars corresponds to the reported FPKM, and the error bars represent the standard error (n=3). Asterisks indicate significance based on the adjusted p-value (**: p-value < 0.01; ***: p-value < 0.001).

Figure 7. Phylogenetic analysis of streptophyte and representative chlorophyte CysPc sequences. The CysPc sequences were aligned with MAFFT v.7.058b. The tree was constructed using RAxML with 1000 bootstrap replicates using the WAG model with GAMMA distributed rate heterogeneity. TML-calpains cluster in a separate group

from the cytosolic calpains. Land plant sequences are highlighted in green, charophyte and chlorophyte algae sequences in red and blue, respectively. capn1-3 represent cytosolic rat calpains.

A breakdown of author responsibilities in the work and manuscript preparation

Viktor Demko: design of experiments, DNA constructs design, genetic experiments in *P. patens*, phenotypic characterization of created *P. patens* strains, interpretation of the data, manuscript writing; **Pierre-Francois Perroud:** DNA constructs design, genetic experiments in *P. patens*, consulting phenotypic characterization of created *P. patens* strains, analysis and interpretation of the RNAseq data, manuscript writing; **Wenche Johansen:** DNA constructs design, bioinformatics analysis, molecular characterization of created *P. patens* strains, interpretation of the data, manuscript writing; **Charles F. Delwiche:** providing sequence data for DEK1 analyses from charophyte algae species; **Endymion D. Cooper:** search for calpain sequencing in databases from Delwiche lab, consulting phylogenetic analyses, reading and commenting on the manuscript; **Pål Remme:** molecular characterization of created *P. patens* strains, creating DNA constructs for complementation studies in *P. patens*; **Ako Eugene Ako:** molecular characterization of created *P. patens* strains; **Karl G. Kugler:** RNAseq data analysis and interpretation, writing the manuscript; Klaus F.X. Mayer: providing expertise and infrastructure for RNAseq analysis; **Ralph Quatrano:** providing expertise and resources for experiments in *P. patens*, reading and commenting on the manuscript; **Odd-Arne Olsen:** project PI, experimental design, data interpretation, co-ordination of manuscript writing.

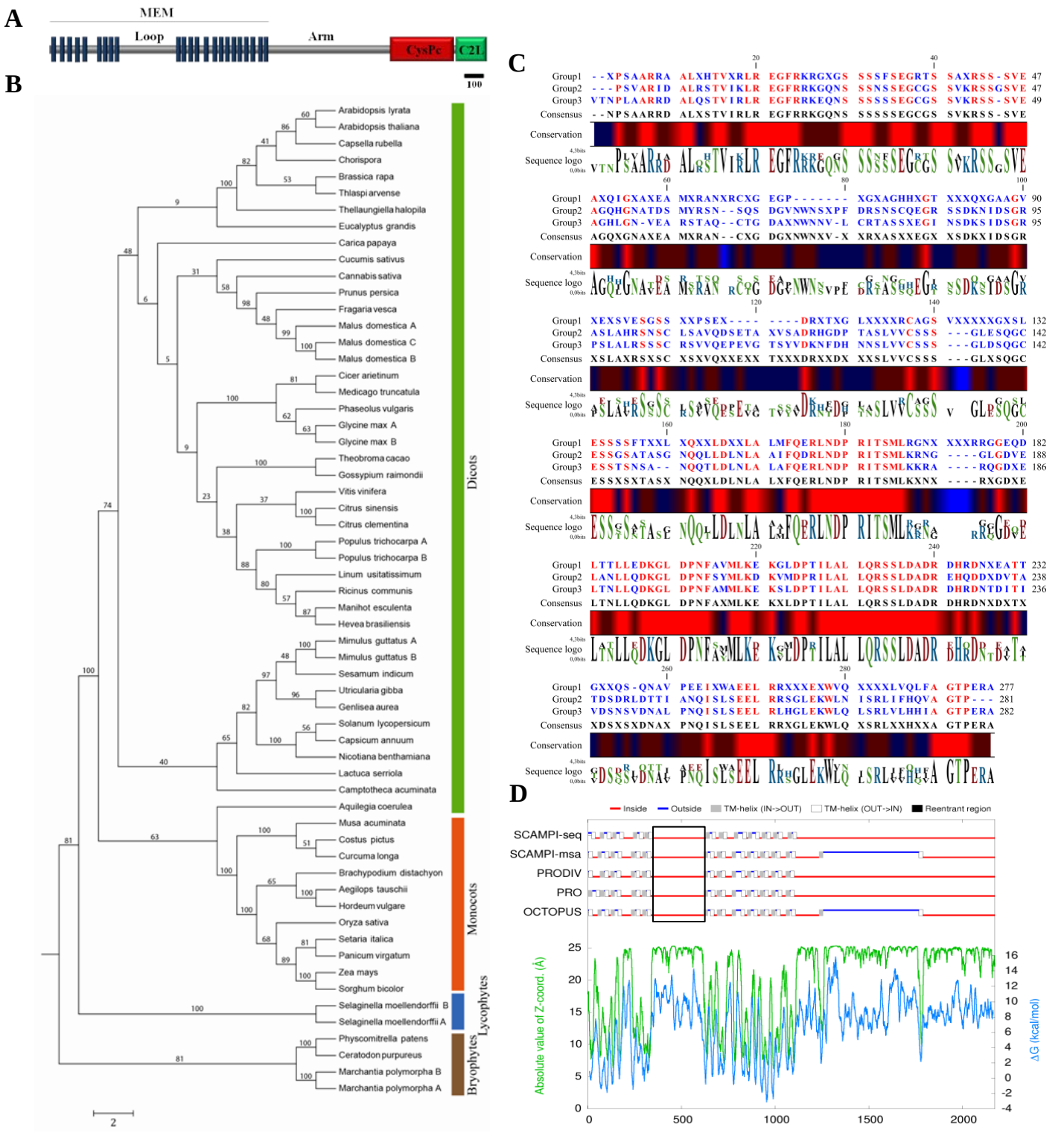

Figure 1. Land plant Loop phylogeny and structure. Bioinformatic analyses of land plant Loop sequences. A, Diagram of the predicted DEK1 structure. Bar represents the length of 100 amino acids. B, Phylogenetic analysis of land plant Loop sequences. The Loop sequences cluster in four major phylogenetic groups corresponding to bryophytes, lycophytes, monocots and dicots. C, Alignments between the bryophyte and lycophyte (group 1), monocot (group 2) and dicot (group 3) Loop consensus sequences showing the absolutely conserved positions and the degree of sequence conservation (red = 100 % conserved positions/regions) using CLC Genomic Workbench to visualizethe result. D, Consensus prediction of *Physcomitrella patens* DEK1 membrane protein topology (TOPCONS) suggests the Loop (black-lined box) to be localized at the cytoplasmic side of the plasma membrane.

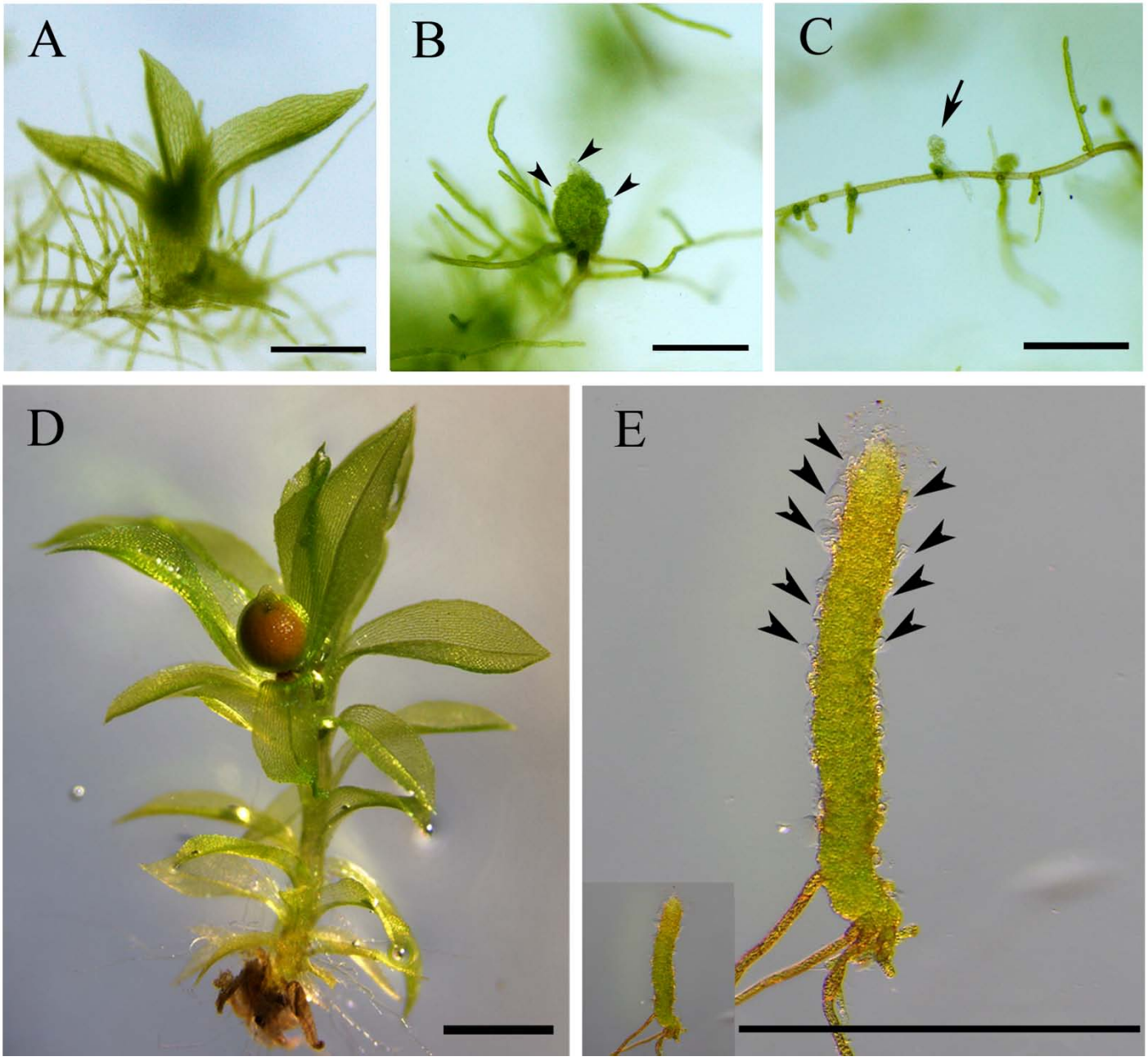

Figure 2. Gametophore development in WT, *dek1Δloop* and *Δdek1*. A, young WT gametophore grown on BCD medium. B, young *dek1Δloop* gametophore grown on BCD medium. C, arrested *Δdek1* bud grown on BCD medium. D, mature WT gametophore with differentiated sporophyte grown on soil block. E, mature *dek1Δloop* gametophore cultivated under the sporophyte-production conditions on soil block (lower left sector shows the same mutant gametophore in the scale comparable to WT gametophore on D). Arrow points to the *Δdek1* bud ; arrowheads point to the filamentous protrusions formed on the *dek1Δloop* gametophore stem. Bar: 500 μm.

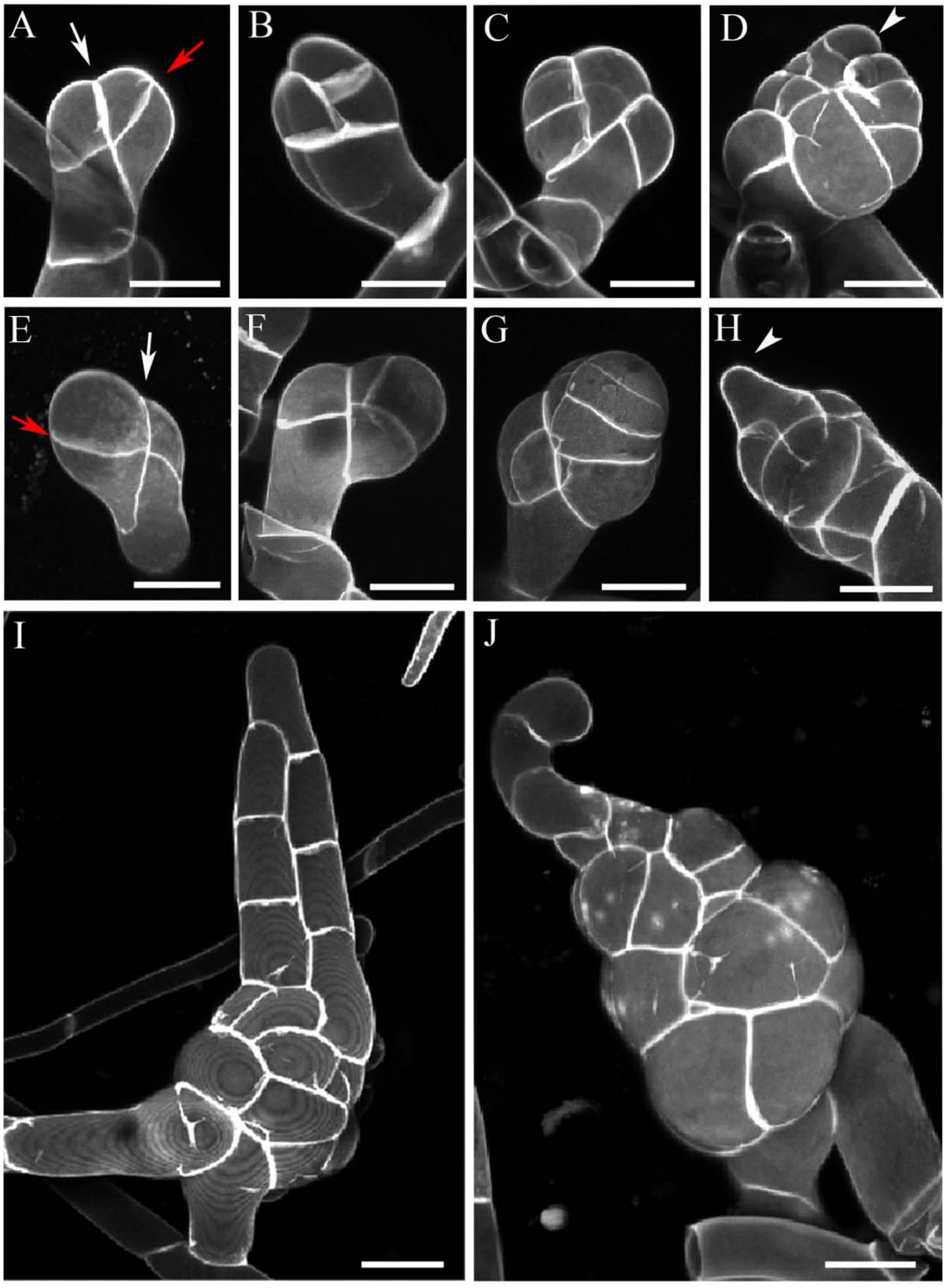

Figure 3. Bud and early gametophore development in WT and *dek1Δloop* mutant. A-D, bud development in WT. E-H, bud development in *dek1Δloop*. I, juvenile WT gametophore with emerging phylid. J, juvenile *dek1Δloop* gametophore with filamentous protrusion formed from the phyllid progenitor cell. White arrows point to the first asymmetric division of the bud initial. Red arrows point to the first division of the bud apical cell. Arrowheads point to the phyllid primordials. Barr: 50 µm.

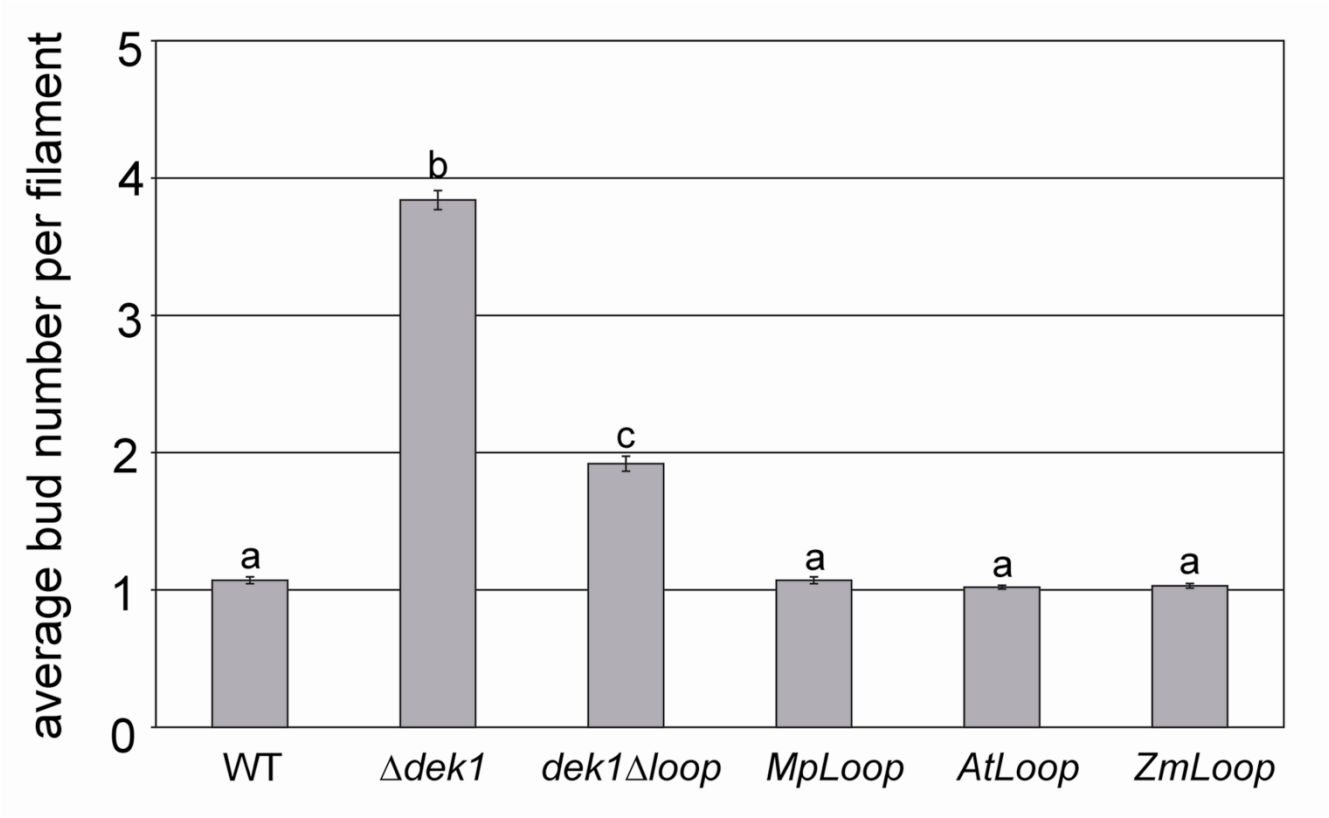

Figure 4. Bud initiation in WT, *Δdek1*, *dek1Δloop*, *MpLoop*, *AtLoop* and *ZmLoop* lines. Graph showing average number of buds formed per 15 filament cells in WT, *Δdek1*, *dek1Δloop*, *MpLoop*, *AtLoop* and *ZmLoop* lines. Values shown are means ± s.e. Different letters denote significant differences at P = 0.05 (ANOVA, LSD-test).

Figure 5. Gametophore development in WT and *dek1Δloop* complementation lines, six weeks old. A, WT gametophore. B, *dek1Δloop* gametophore. C, *PpLoop - dek1Δloop* complemented with the Loop coding sequence from *Physcomitrella patens* showing the WT phenotype. D, *MpLoop - dek1Δloop* complemented with the Loop coding sequence from *Marchantia polymorpha*. E, *AtLoop - dek1Δloop* complemented with the Loop coding sequence from *Arabidopsis thaliana*. F, *ZmLoop - dek1Δloop* complemented with the Loop coding sequence from *Zea mays*. G-J, isolated phyllids from apical, sub-apical, middle and basal part of the gametophores. G, WT; H, *MpLoop*; I, *AtLoop*; J, *ZmLoop*. Bar: 1 mm (A-F), 500 µm (G-K).

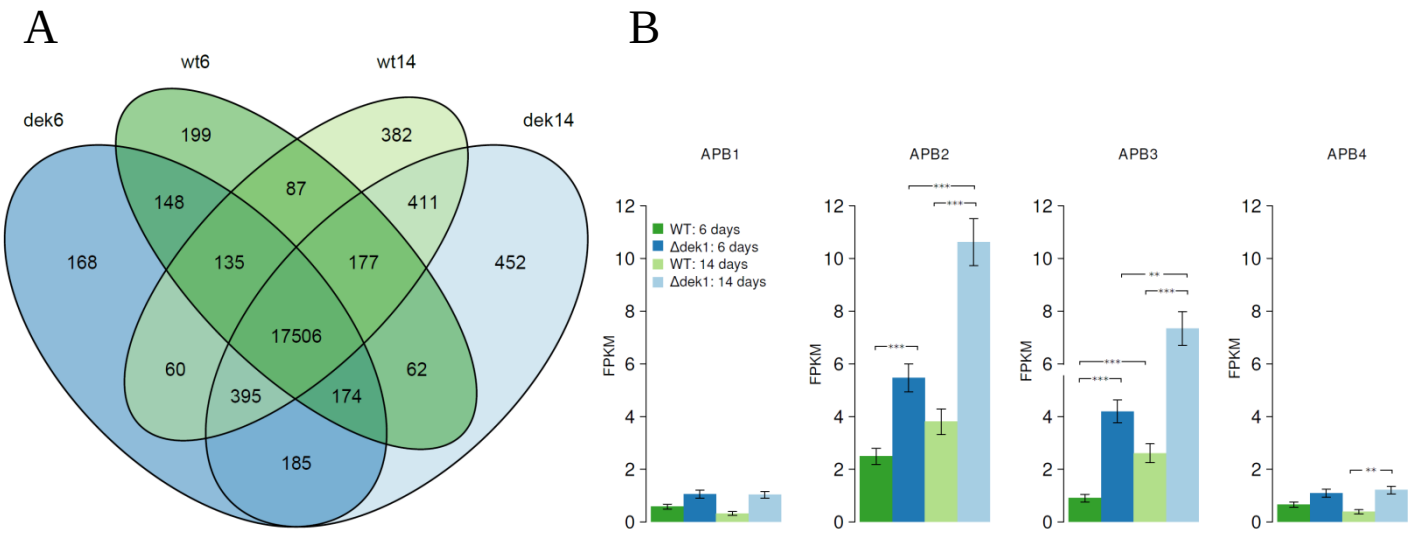

Figure 6. Transcriptome comparison between WT and *Δdek1*. A, Genes present in different conditions. As expected we find the majority of genes to be expressed (FPKM > 1) in all four conditions. However, there is also a substantial number of genes that are only expressed under certain conditions. B, Expression of AP2-type transcription factors. Expression of the AP2-type transcription factors, APB1, APB2, APB3, and APB4. The height of the bars corresponds to the reported FPKM, and the error bars represent the standard error (n=3). Asterisks indicate significance based on the adjusted p-value (**: p-value < 0.01; ***: p-value < 0.001).

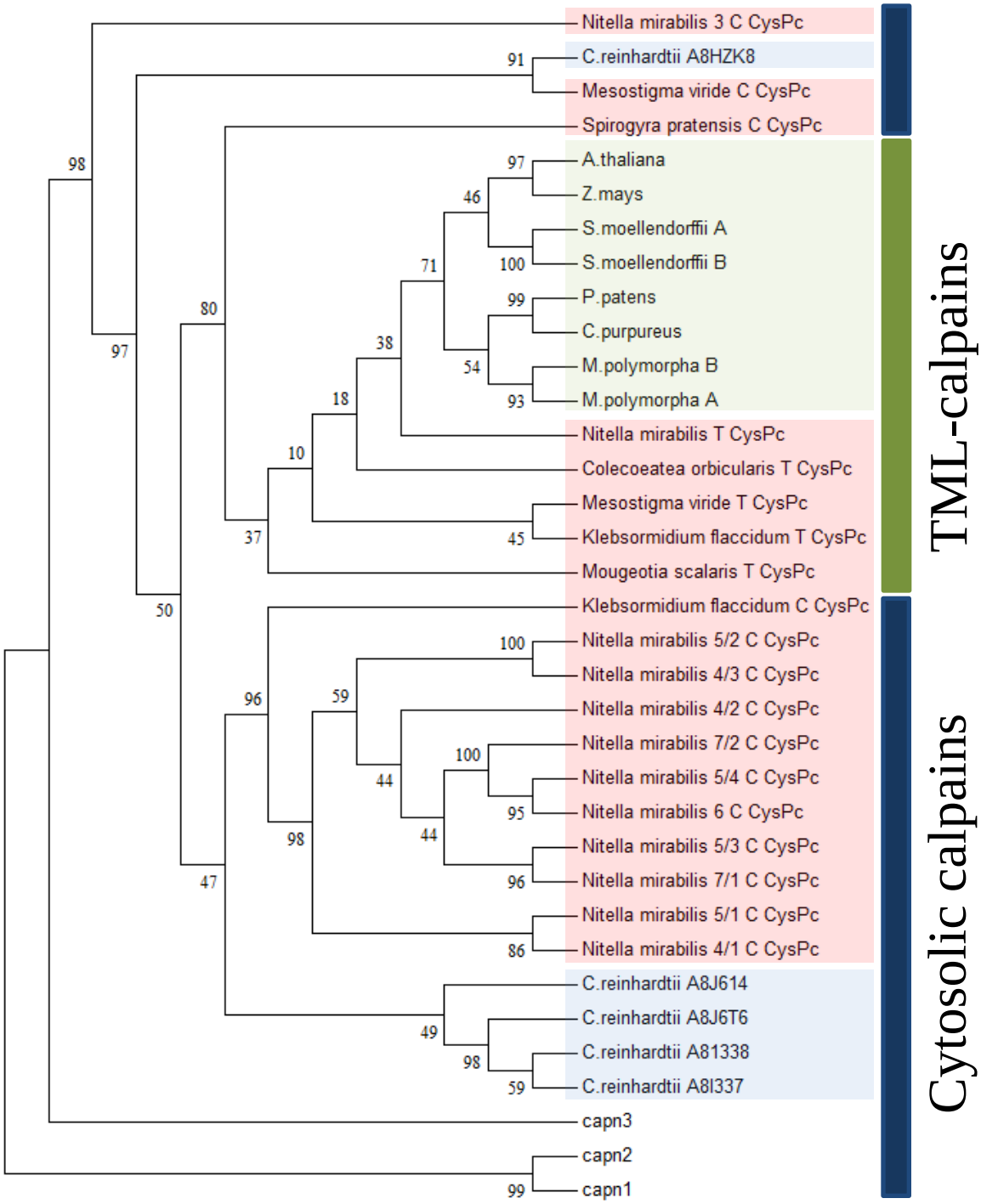

Figure 7. Phylogenetic analysis of streptophyte and representative chlorophyte CysPc sequences. The CysPc sequences were aligned with MAFFT v.7.058b. The tree was constructed using RAxML with 1000 bootstrap replicates using the WAG model with GAMMA distributed rate heterogeneity. TML-calpains cluster in a separate group from the cytosolic calpains. Land plant sequences are highlighted in green, charophyte and chlorophyte algae sequences in red and blue, respectively. capn1-3 represent cytosolic rat calpains.