Hedmark University College

Hamar Campus
Faculty of Education and Natural Sciences
Department of Natural Sciences and Technology

Ahmed Khaleel Mekhlif

Master thesis

Functional analysis of *Physcomitrella patens* DEK1-LG3 domain

Master's Degree in Applied and Commercial Biotechnology

November 2015

Contents

Acknowlegements:

This Master Thesis study was performed at the Department of Education and Natural

Sciences at Hedmark University College.

Firstly, I would like to express my sincere gratitude to my advisor Associate Prof. Wenche Johansen for the continuous support of my master study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis.

My sincere thanks also go to PhD student Ako Eugene Ako, thanks a lot my friend for all your support, patience, and encouragement. I am also grateful to Norwegian University of Life Sciences (NMBU) and Viktor Demko for their assistance during this project.

I take this opportunity to express gratitude to all of the biotechnology department members for their help and support during all these years.

Last but not the least; I would like to thank my family and my friends for supporting me spiritually throughout this thesis and my life in general.

Ahmed Khaleel Mekhlif

Hamar 2015

Abstract:

The *defective kernel 1* (*dek1*) is a highly conserved gene in land plants, encoding the single calpain in plants, which plays essential role in the development of both the epidermal cell layer in plant embryos and aleurone cell formation during seed development. The DEK1 protein comprises 21-23 transmembrane segments, a loop (DEK1-Loop) between transmembrane domains 9 and 10, a non-structured cytoplasmic arm (DEK1-Arm), and a highly conserved calpain domain composed of the CysPc-C2L domains. Recent unpublished bioinformatics study of the DEK1 protein shows that land plants DEK1-Arm sequences contain a domain homologous to the laminin globular 3 (LG3) domain. Our intention was to genetically investigate the function of the DEK1-LG3 domain in *P. patens* by using homologous recombination to precisely remove the genomic sequence corresponding to the LG3 domain. In order to achieve that, vector *pBHRF_PpArm_ΔLG3* plasmid was designed to harbour the *P. patens dek1 Arm-Δlg3* sequence. The plasmid was digested with restriction enzymes *RsrII* and *PacI*, and then transformed into the *P. pate*ns *dek1-ΔArm* mutant plant, thus creating the *P. patens dek1-Δlg3* mutant. The *P. patens dek1-Δlg3* mutant affected the gametophores development where the phyllids were small, missing marginal serration, have a blunt tip and short midrib. Deletion of the LG3 domain in *P. patens* DEK1 caused also failure in the opening of the apices, leaving the archegonia closed, and defect in the egg canal formation making the mutant plant sterile. The deletion of LG3 domain in the *P. patens* DEK1 may affects or regulates auxin biosynthesis, thus changing auxin concentration in the plant which leads to abnormality in both gametophore and gametangia development in the mutant plant.

1. Introduction

Reverse genetics is an approach to investigate the function of a gene by deleting or alternating the sequence of the gene, and then analyze the mutation effect on the phenotype. This process can be specifically achieved using methodology such as RNA silencing or homologous recombination (Tierney & Lamour, 2005). In *Physcomitrella patens* (*P. patens*), homologous recombination is usually used to create mutants (Schaefer & Zrÿd, 1997). Homologous recombination can also be used to manipulate the gene sequences by for example removing a specific region in the gene which encodes a protein domain in a multi-domain protein. In the present study, homologous recombination was used to precisely remove the genomic sequence corresponding to the laminin globular 3 (LG3) domain of the *P. patens* Defective Kernel 1 (DEK1) following characterization of the mutant phenotype.

1.1 Defective kernel 1

The *defective kernel 1* (*dek1*) is a highly conserved gene, encoding the single calpain of lands plants, which plays essential role in the development of both the epidermal cell layer in plant embryos and aleurone cell formation during seed development in *Zea mays*, *Arabidopsis thaliana* and *Nicotiana tabacum*. The *dek1* mutant was the first mutant to be phenotypically characterized and cloned in studies of aleurone cell fate specification in *Z. mays* endosperm development (Ahn, Kim, Lim, Kim, & Pai, 2004; Becraft & Asuncion-Crabb, 2000; Becraft, Li, Dey, & Asuncion-Crabb, 2002; Lid et al., 2002).

The DEK1 protein belongs to an ancient family of calpains harbouring a large transmembrane (TML) domain, the TML-calpain family. The broad phylogenetic distribution of TML-calpain suggests that the protein was established very early during eukaryote evolution (Zhao et al., 2012). TML-calpain has so far not been detected in chlorophyte green algae. Studies revealed a partial similarity in the DEK1 sequence between charophytes and streptophytes which support the hypothesis that a major shift in the DEK1 function occurred during the transition from single-celled charophytes to land plants (Demko et al., 2014). In streptophytes, DEK1 orthologous has been identified in all genomes sequenced today. Sequence analysis revealed that the DEK1 protein is highly conserved within land plants showing 70-98% sequence identity which may indicate an important function of DEK1 within land plants (Liang et al., 2013; Wang et al., 2003).

The DEK1 protein comprises 21 transmembrane segment, a loop (DEK1-Loop) between transmembrane domains 9 and 10, a non-structured cytoplasmic arm (DEK1-Arm), and a highly conserved calpain domain composed of the CysPc-C2L domains (Figure 1A). The calpain is the catalytic protease core domain and is essential for DEK1 activity (Kim Leonie Johnson, Faulkner, Jeffree, & Ingram, 2008; Lid et al., 2002; Zhao et al., 2012). Recent unpublished bioinformatics study of the DEK1 protein shows that land plants DEK1-Arm sequences contain a domain homologous to the laminin globular 3 (LG3) domain (W. Johansen, unpublished results). Laminins are large heterotrimeric glycoproteins (Sasaki, Fässler, & Hohenester, 2004). Laminin globular domains (LG domains) were identified in laminins, and found in many large and different extracellular proteins. LG domains consist of around 180 amino acid residues (Beckmann, Hanke, Bork, & Reich, 1998; Vuolteenaho, Chow, & Tryggvason, 1990). LG domains have different function including binding sites for heparin, sulphatides and the cell surface receptor dystroglycan (Tisi, Talts, Timpl, & Hohenester, 2000). Similar sequence of LG domain was found in Drosophila proteins and other extracellular proteins like the sex hormone-binding proteins, and Androgen-binding protein (ABP) (Joseph & Baker, 1992; Patthy, 1992; Sasaki, Costell, Mann, & Timpl, 1998).

Recently, it was suggested that DEK1 contain 23-transmembrane segments instead of 21 (Kumar, Venkateswaran, & Kundu, 2013), and that the transmembrane domain controls the activity of the calpain domain. In one study it was shown that DEK1 undergoes autolytic cleavage in the Arm domain and prior the start of the calpain domain, leading to release of an active calpain domain (Kim Leonie Johnson et al., 2008).

Studying the DEK1 protein in different plants, like *Arabidopsis*, *Z. mays*, and different higher plants, suggests that DEK1 has a conserved role in plant signal transduction. A comparison between *Z. mays* and *A. thaliana dek1* shows high conservation degree (70% identity) (Figure 1B), in spite of the two plants represent two main class of flowering plants, monocots and dicots, respectively. It has been suggested that DEK1 is the only calpain prototype detected in plants depending on broad searches in available database (Liang et al., 2013; Lid et al., 2002).

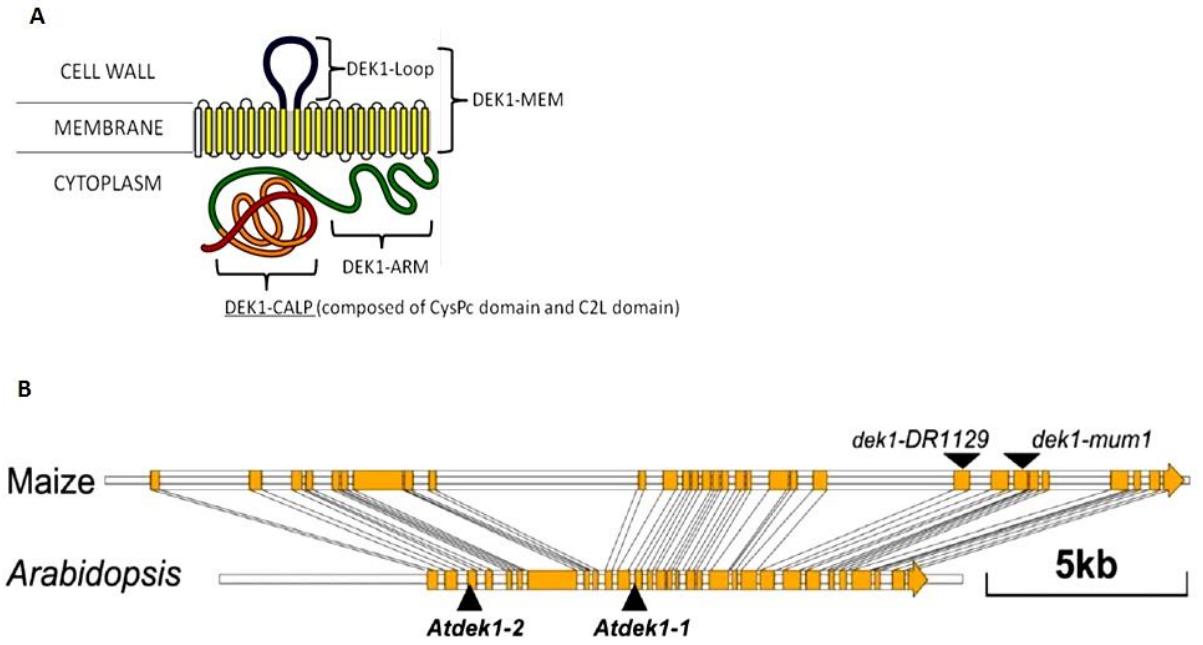


Figure 1: dek1 gene and protein. A) Predicted maize DEK1 structure showing 21 transmembrane domains, the loop region on the outside the cell, and the cysteine proteinase domain on the inside. The model is based on the TMHMM2.0program (Modified Figure). B) A comparison between Z. mays and A. thaliana dek1 showing intron-exon structure. Exons are shown in yellow (Modified Figure) (Lid et al., 2002).

1.2 DEK1 function:

As mentioned previously, the leaf epidermis layer in plants comprises particular cells that function to do important tasks for instance mechanical protection, restriction of transcription, and water and gas exchange and absorption. It has also been shown that epidermis cells have important role in directing the growth of the inner cells in the shoot. In the healthy *A. thaliana* embryo, the epidermal cells specify first as a single layer of protoderm cells. From this protoderm, the meristematic L1 layer will arise where it plays an essential acting in meristem function. In *Z. mays*, the epidermal cells of the endosperm develop a monolayer around external endosperm surface called aleurone layer. Studies have shown that DEK1 is essential for the aleurone layer formation in addition to CRINKLY4 (CR4) which is also involved in aleurone cell specification and epidermis differentiation (Tian et al., 2007).

Previous studies suggest the important function of the DEK1 protein, in all land plants including angiosperms, in specifying the correct cell division which was essential for early evolution to three-dimensional growth in land plants (Kim Leonie Johnson et al., 2008; Tian et al., 2007).

It has been shown that *dek1* mutant in *Z. mays* caused embryo-lethal phenotype. The mutation arrested the aleurone cell formation, and thus fail to support the grain matures (Lid et al., 2002). A relatively effect was observed on DEK1 mutation in *A. thaliana* (*At*DEK1). The loss of the *At*DEK1 lead to embryo arrest and caused early lethality by affecting the cell organization in the protoderm causing abnormal cell division (Kim L Johnson, Degnan, Ross Walker, & Ingram, 2005). On the other hand, over-expression of *At*DEK1 was performed under the 35S promoter control. It was observed some development phenotypes like global lack of trichomes, leaves showed incorrect dorsiventral symmetry and cell organization in the flower was abnormal (Lid et al., 2005).

DEK1 was also investigated in tobacco (*N. benthamiana*) using virus-induced gene silencing (VIGS) of *Nb*DEK1. The study shows that *Nb*DEK1 have different effect on different organ of the plant. The cell proliferation was prevented during vegetative growth of the plant, while the flower development was arrested during reproductive stage. This abnormality development of the plant led to formation of cylindrical cell mass instead of floral organs. The VIGS also suppressed the cell-fate determination of the leaves and many of palisade cells and mesophyll layers undergo premature cell death (Ahn et al., 2004).

Another study was performed on the *dek1* (called ADAXIALIZED LEAF1 gene, *adl1*,) in *Oriza sativa*. The ADL1 was found that it is a single copy of phytocalpains, consists of 2162 amino acids, and comprises 21 predicted transmembrane segments, an extracellular loop domain, an intracellular domain, and a calpain like cysteine domain. The *adh1* mutant plants showed semi-dwarf leaved and directed away from the axis. In the shoot apical meristem, L1 cells were larger than in the wild type (WT) plants. This suggests that *adl1* gene may play an important role in maintenance and specification of the leaf primordium position (Hibara et al., 2009).

In *P. patens,* the *Δdek1* mutant was not lethal as it was in *Z. mays* and *A. thaliana*. The *Δdek1* mutant shows absence of gametophores. It was found that there was no difference in the bud formation, at two-cell stage, between the WT and *Δdek1* mutant plant (Figure 2A, C). On the other hand, there was a distinctly difference in the first cell division plane of the apical cell WT and *Δdek1* mutant plant. In the WT the current cell wall divided the previous cell wall in a central position while in the *Δdek1* mutant plant it takes place at random planes (Figure 2B, D) (Perroud et al., 2014).

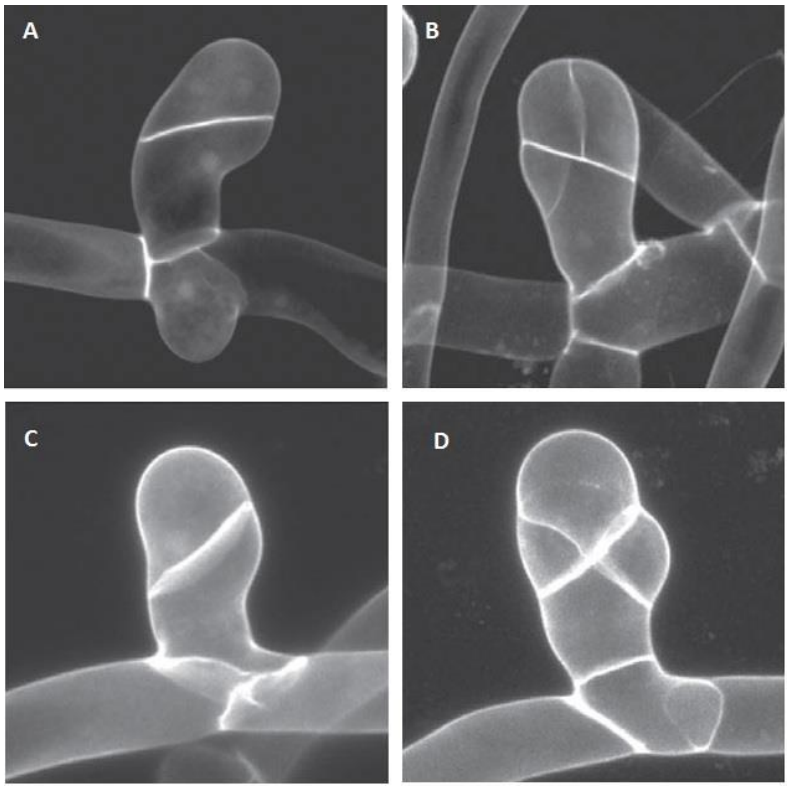


Figure 2: Bud development in P. patens WT and Δdek1 mutant plant using confocal microscopy. A) Two cell stage in WT. B) Cell division plane of the apical cell in WT. C) Two cell stage in Δdek1 mutant plant. D) Cell division plane of the apical cell in Δdek1 mutant plant (Perroud et al., 2014) .

As a result, *Δdek1* mutant plants lacked the gametophores comparing with WT plants as shown in Figure 3 (Perroud et al., 2014).

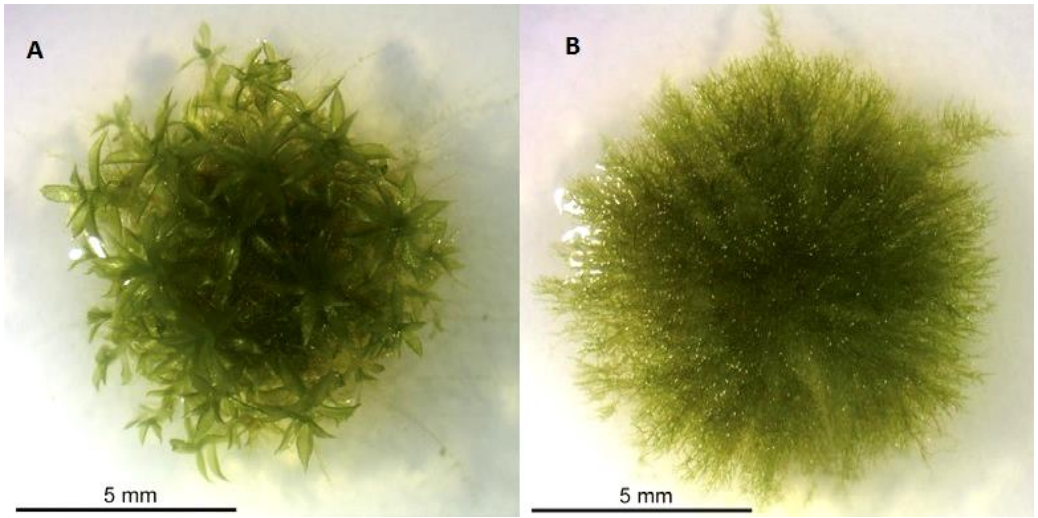


Figure3: P. patens WT & Δdek1 mutant. A) 3-weeks old WT plant showing gametophore; B) 3-weeks old delta Δdek1 mutant plant lucks the gametophore, Bar = 5 mm *(Perroud et al., 2014).*

As mentioned above, the loop is located in the transmembrane domains of DEK1 (Lid et al., 2002). A *P. pate*ns *dek1-Δloop* mutant was created to study the loop function. It was found that the *DEK1-Δloop* mutant plant could correctly position the division plane in the bud apical cell, contrary to the *Δdek1* mutant plant (Figure 4) (Demko et al., 2014).

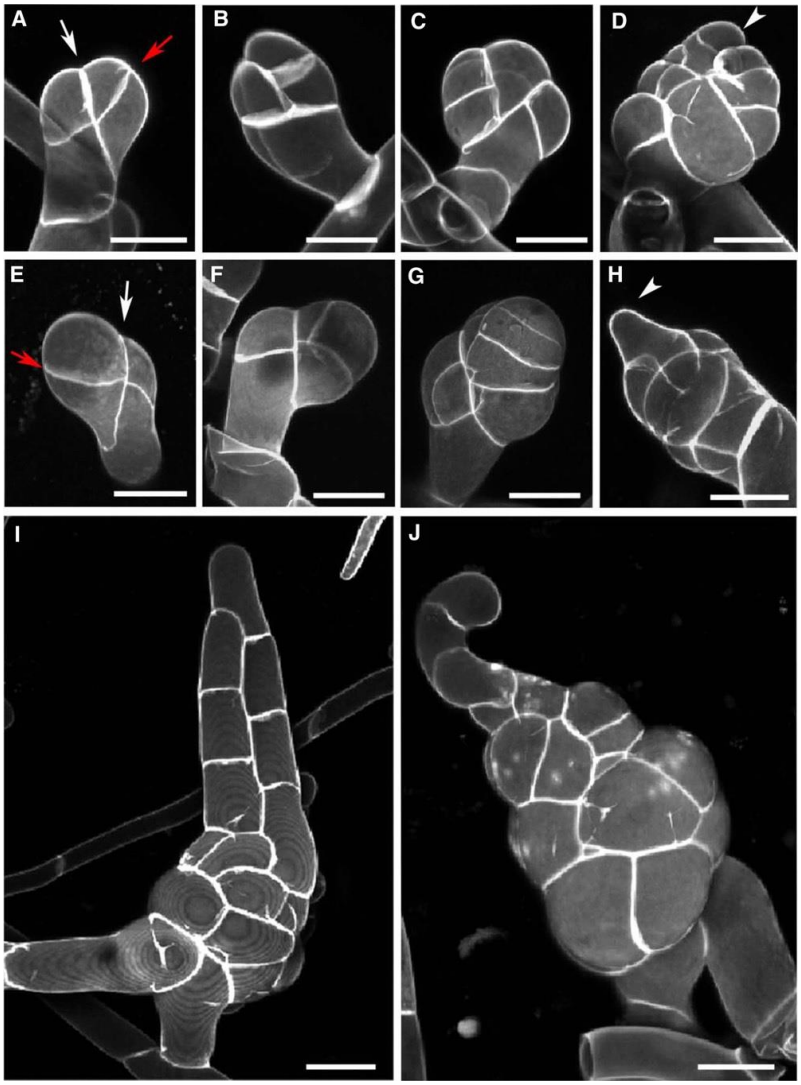


Figure 4: Bud development in P. patens WT & dek1-Δloop. From A to D, bud development in P. patens WT and from E to H bud development in DEK1-Δloop. I) Juvenile WT gametophore with emerging phyllid. J) Juvenile dek1-Δloop gametophore with filamentous protrusion outlined from the phyllid progenitor cell, Bar = 50 µm. *(Demko et al., 2014).*

However, the *P. patens dek1-Δloop* mutant failed to produce a completely WT phenotype. Instead, aberrant gametophores were observed with absence of developed phyllids. This was due to fail regulation of mitotic activity (Figure 5). The *P. patens dek1-Δloop* mutant were able to form gametophore apical stem cells but lack a phyllids stem (Demko et al., 2014).

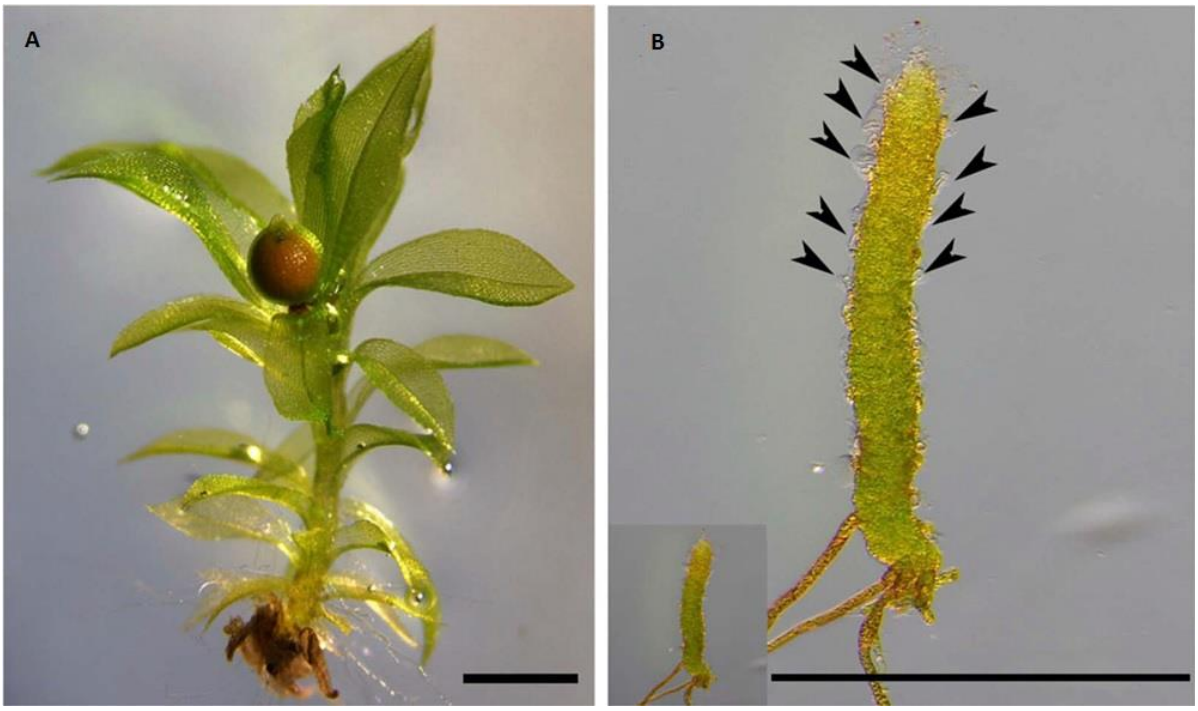


Figure 5: Gametophore development in P. patens WT and DEK1-Δloop. A) Mature WT gametophore with differentiated sporophyte. B) Mature P. patens DEK1-Δloop mutant showing filamentous protrusions formed on the DEK1-Δloop (pointed with arrowheads), Bar = 500 µm (Demko et al., 2014).

In addition to DEK1, CRINKLY4 (*CR4*), encoding a receptor-like protein kinase, is also known to have been associated in aleurone cell fate specification and epidermis differentiation. It has been shown that a mutation of *cr4* in *Z. mays* endosperm varied from lacking small patches of aleurone cells to large area of aleurone cells (Becraft, Stinard, & McCarty, 1996). Another protein has been identified to affect aleurone cell development which is SUPERNUMERARY ALEURONE LAYER1 (*SAL1*). The *sal1* encodes a predicted 204 amino acids protein which is a homolog of human CHMP1 and yeast DID2 (Howard, Stauffer, Degnin, & Hollenberg, 2001; Nickerson, West, & Odorizzi, 2006; Shen et al., 2003).

It was proposed a model for aleurone cell fate specification in *Z. mays* endosperm explaining the role that DEK1 plays (Figure 6). The model suggested that aleurone specification involves the function of DEK1, CR4, and SAL1, where the DEK1 is active just in the outer membrane on endosperm surface. The DEK1 is activated by the substrate which may be a cytoplasmic protein. The activation is under the control of DEK1 membrane where the loop region has an essential functional role. DEK1 protein moves laterally between the aleurone cells through the plasmodesmata with the help of CR4. Both DEK1 and CR4 are controlled by internalization and degradation at SAL1 positive endosperm (Tian et al., 2007).

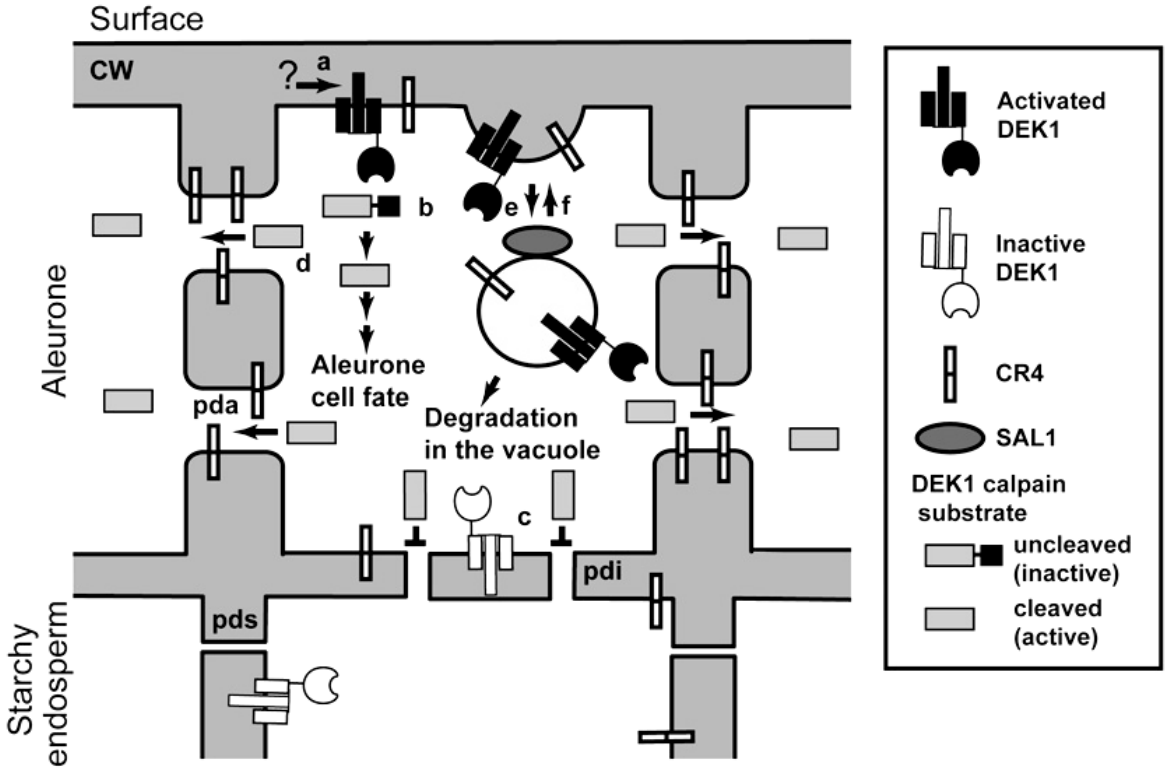


Figure 6: A Model for the Role of DEK1, CR4, and SAL1 in Aleurone Cell Specification. DEK1 at the surface of the endosperm is activated by an unknown mechanism (a), its calpain domain in the cytosol cleaving a postulated substrate (b) that leads to the specification of aleurone cell fate. DEK1 in all other positions is inactive (c). In cells with active DEK1 signaling, CR4 concentrates on plasmodesmata between aleurone cells (pda) and increases the plasmodesma exclusion limit, allowing the activated DEK1 substrate to move laterally between aleurone cells, thereby reinforcing the signal for aleurone cell fate specification (d). Plasmodesmata in cell walls between starchy endosperm cells are narrow (pds), whereas plasmodesmata in cell walls between aleurone cells and starchy endosperm cells are intermediary in width (pdi). DEK1 and CR4 are internalized by endocytosis (e) and traffic through endosomes. Whereas some DEK1 and CR4 molecules may be recycled back to the plasma membrane (f), others are sorted for degradation in the vacuole in a process that requires SAL1. Some endosomes are recycled back to the plasma membrane (f). Text and Figure are cited from (Tian et al., 2007).

1.3 Calpains:

Calpains are cysteine proteases family which their activity depends on the intracellular Ca^{2+} concentration. Calpains control signal transduction which influence cell differentiation, proliferation and cell death in animal system (Sato & Kawashima, 2001). Calpains perform restricted proteolysis of various substrates, thus they control numerous of intracellular processes. A typical animal calpain consist of four conserved domains: an N-terminal anchor helix (Nter), a catalytic protease core domain (CysPc) with tow subdomains PC1and PC2, a C2-like domain (C2L), and a penta-EF-hand domain (PEF). In the non-classical calpains, both Nter and the PEF domains are absence and they may contain extra domains in combination with CysPc (Zhao et al., 2012).

Calpain in the mammalians is found in two types, m-calpain and µ-calpain, depends on their *in vitro* requirements for the Ca^{2+}. Both types form a heterodimer containing a 80 kDa catalytic subunits and a 30 kDa regulatory sub-unit. The m-calpain half-maximal activity *in vitro* requires 300 µm Ca^{2+}, while the µ-calpain requires 50 µm of Ca^{2+}. Sequence alignment and three dimensional modeling indicate a relevant similarity between DEK1 and animal calpain, where DEK1 domain CysPc and C2L are 40-50% similarity and 30-40% identity to the same domains of the animal calpain (Wang et al., 2003).

Calpains in plants are known as phytocalpains. Phytocalpains belongs to another group of non-classical calpains (Ono & Sorimachi, 2012). Calpains in the land plants, usually contain one C2L domain at the C-terminus, and CysPc domain at the N-terminus of the calpain (CysPc-C2L domains) (Liang et al., 2013).

Johnson et al. studied DEK1 role in the *A. thaliana* and they showed that the active calpain domain is important for the DEK1 activity and it is sufficient alone to full complement the Δ*dek1* mutant phenotype. This complementation raised the questions about the role of the transmembrane domain of DEK1 activity (Kim Leonie Johnson et al., 2008).

1.4 *P. patens* as a model for studying development

*P. pate*ns are powerful plants to study gene functions using revers genetics tools. It has highly efficient gene targeting due to homologous recombination which allows specific mutation of specific genome sequence (Reski & Frank, 2005; Schaefer & Zrÿd, 2001). This process of homologous recombination (HR) leads to targeted gene replacement (TGR) between the construct and the targeted locus when transformation occurs (Kamisugi et al., 2006). Another benefits are the moss has both two and three dimensional growth and the whole genome sequence are known (Reski & Frank, 2005). *P. patens* is an easy plant to deal with where its requirement is not expensive and the plant does not take a large place (Schaefer & Zrÿd, 2001).

1.4.1 Homologous Recombination

Homologous recombination is a gene targeting method where a foreign DNA sequence can be cloned into a specific location due to the presence of isogenic genomic sequences on the introduced DNA (Figure 7). This similarity of isogenic genomic sequences will facilitate the integration of foreign DNA into the genomic DNA at the exact locus (Schaefer & Zrÿd, 1997).

HR is a normal metabolic process for repairing damages in the DNA like DNA gaps, DNA double-stranded breaks (DSBs), and DNA interstrand crosslinks (ICLs). HR plays an important role in preserving the genome in addition to supporting DNA replication and telomere maintenance (Li & Heyer, 2008).

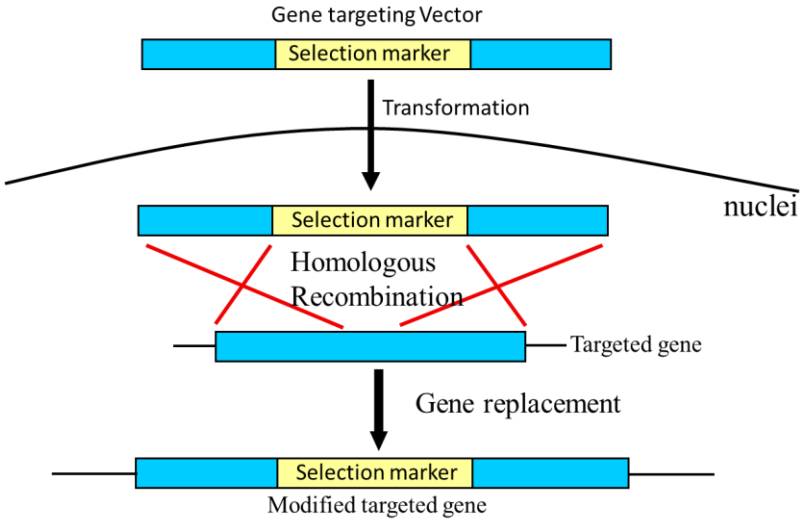


Figure 7: Homologous recombination mechanism. The Figure showing a schematic diagram of gene targeting principles (Kuwayama, 2012).

HR can be used to create deletion, insertion, or mutation in a specific DNA sequence. It helps also to detect the function of the plant gene and to improve the crops. By using HR it will be possible to create crops with higher values like disease resistance plants, enhanced nutritional qualities food, and modified oil or carbohydrate seed (Wright et al., 2005). The first efficient homologous recombination in *P. patens* was reported by Schaefer and Zryd (Schaefer & Zrÿd, 1997) making the *P. patens* a perfect plant model for analysis of gene function using reverse genetics.

The mechanism of HR is well known mostly in bacteria and yeast. Two proteins, Rad51 and Dmc1, was characterized in HR, where Rad51 appears to be take part in mitotic as well as meiotic recombination. Rad51 was analyzed in *P. pate*ns and it was found two related genes present in the *P. patens* genome. These genes are not interrupted by introns. The two intronless genes of Rad51 in *P. pate*ns suggests the reason where the HR is higher in this plant than in the flowering plants (Markmann-Mulisch et al., 2002).

1.5 *P. patens* life cycle:

Mosses are a very old group of plant originated about 500 million years ago. The life cycle of mosses contains variation between photoautotrophic haploid gametophyte and heterotrophic diploid sporophyte (Figure 8) (Heckman et al., 2001).

*P. pate*ns present a typical life cycle of a moss. The life cycle begins with the germination of the spores. After the germination, the first filaments that form are called chloronemata. The chloronemal cells are photosynthetically active and contain well-developed chloroplast. Later another type of filaments will form which is caulonemata. The caulonamal cells have fewer and smaller chloroplasts and it will grow radially, leading to plant branching. The bud will form from some caulonamal side branch where it will be developed to gametophore (N. Ashton, Grimsley, & Cove, 1979).

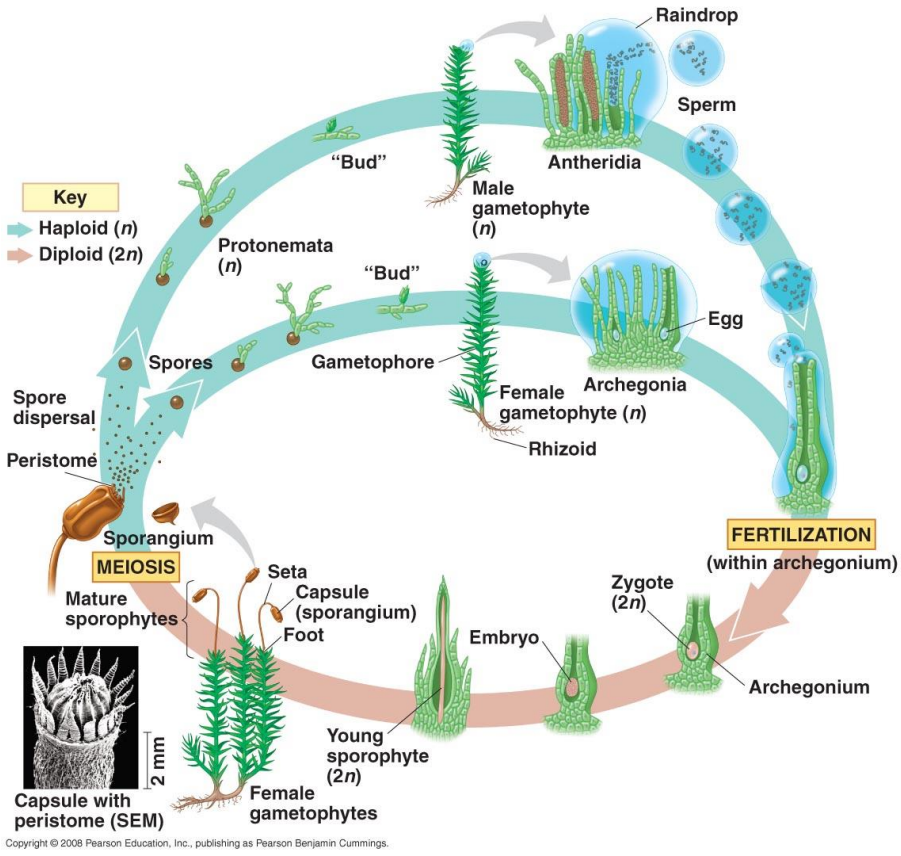


Figure 8: General life cycle of mosses. Mosses life cycle alternates between haploid gametophyte and diploid sporophyte(Campbell & Reece, 2009) .

*P. pate*ns is a monoecious plant. This indicates that both male and female gametes are produced in the same individual. The male gametes (Antherozoids) are found in the antherdia, while the female gametes (Oogonia) are in archegonia in the same gametophore. *P. pate*ns plants have the ability to regenerate the damage tissues. In both gametophytic and sporophytic stage, the protonemal tissue is generated first, followed by the gametophres (D. J. Cove & Knight, 1993).

Previous researches have studied bud development in the moss using cytokinin induced plants. In the mosses, bud formation starts at the initial cell. The apical region of the initial cell will change into dome shaped due to alternation of the cell expansion and elongation. The first division takes place in the initial cell is asymmetric resulting in to daughter cells of different developmental fates. The cell will elongate due to division, and then divided more frequently producing a larger and more complex bud. From this complex bud, a leafy gametophore will arise and it will be responsible to produce the sporophytes (Schumaker & Dietrich, 1998). The bud development stages are shown in Figure 9.

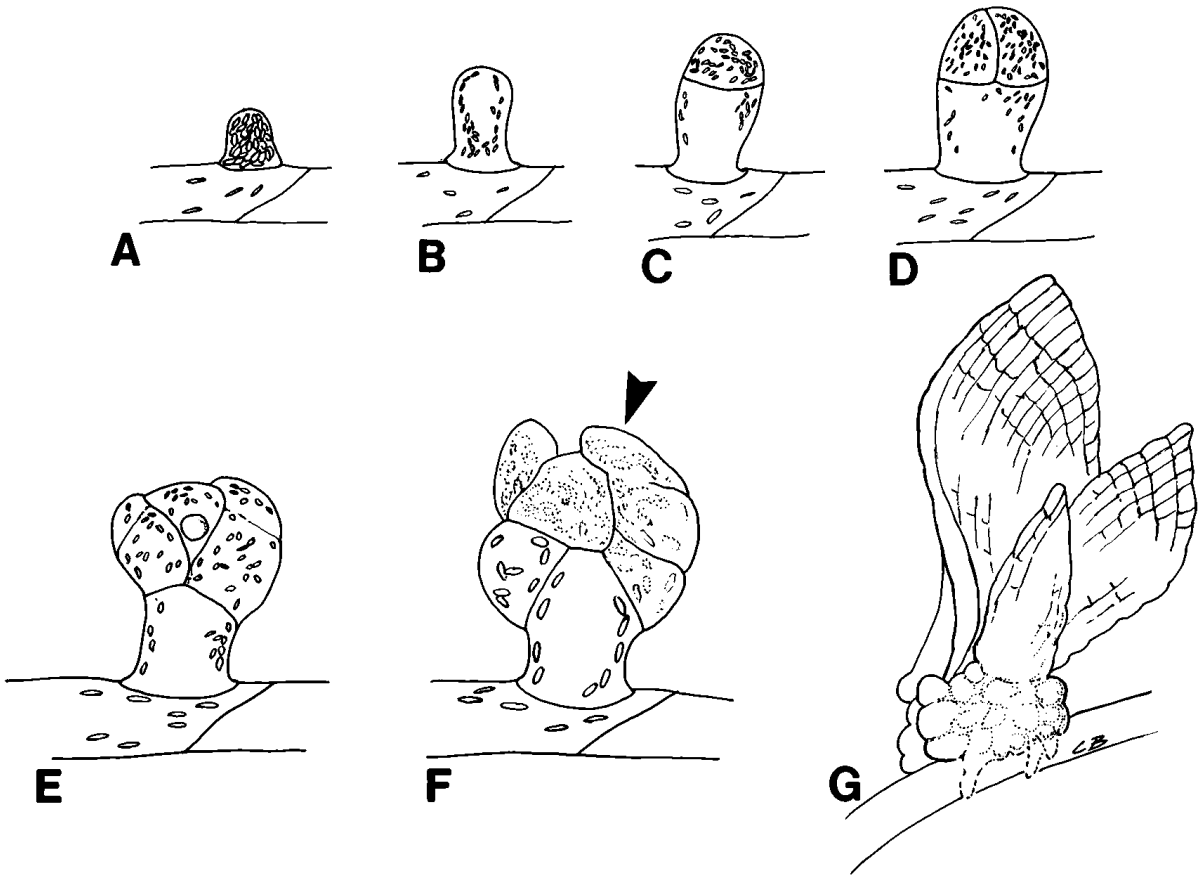


Figure 9: Bud formation in the moss. Developmental transition from filamentous to meristematic growth. Changes are apparent in the initial cell (A) within two to three hours of cytokinin addition. The first visible indication of bud assembly is a dramatic swelling of the initial cell (compare A and B). This is followed by an asymmetric division to produce a large, highly vacuolated stalk cell and a small, densely cytoplasmic apical cell (C). The apical cell divides longitudinally, resulting in two densely cytoplasmic cells (D). Subsequent unequal divisions give rise to a tetrahedral apical cell that continues to divide in three planes to form the relatively simple multicellular bud (E). The subapical cells of the bud divide more frequently than the apical cell to give rise to a larger, more complex bud (F). Subsequently, the leaf primordia (F, arrowhead), each of which will develop into a leaflet of the leafy shoot (G), arise as projections from the side of the bud. Text and Figure are cited from (Schumaker & Dietrich, 1998).

1.6 Aim of study

DEK1 belongs to an ancient family of calpains harboring large transmembrane (TML) domains. DEK1 plays essential roles in the development of both epidermal cell layer in plant embryo and for aleurone cell formation during seed development. The protein comprises 21-23 transmembrane segments, loop insertion between transmembrane segment 9 and 10, a non-structured cytoplasmic arm (DEK1-Arm), and a highly conserved calpain domain (CysPC-C2L). Recent unpublished study of DEK1 shows that land plants DEK1-Arm sequences contain a domain homologous to the LG3 domain (W. Johansen, unpublished results).

The aim of this study was to genetically investigate the function of the DEK1-LG3 domain in *P. patens*. This was achieved by deleting the genomic sequence corresponding to the LG3 domain thus creating the *P. patens dek1-Δlg3* mutant. The mutant plant (*P. patens dek1-*Δlg3) was then characterized both molecularly and phenotypically.

2. Materials and Methods

2.1 Bioinformatics analysis of the DEK1-Arm segment and LG3 domain

Orthologues land plant DEK1 protein sequences (Appendix 1) were obtained from the National Center for Biotechnology Information (NCBI) using reference protein sequence (Refseq) database. DEK1-Arm sequences, defined as the region between the transmembrane segment (TMs) 23 and the calpain domain, from various plant species were extracted using the Simple Modular Architecture Research (SMART) database. Extracted DEK1-Arm sequences were further submitted to the HMMER web server (http://www.ebi.ac.uk/Tools/hmmer/) to identify the LG3 domain sequences. Multiple sequence alignment, both using the full-length DEK1-Arm and DEK1-LG3 sequences were performed using the Clustal Omega algorithm (http://www.ebi.ac.uk/Tools/msa/clustalo/) using default parameters. The resulting multiple sequence alignments were analyzed using the CLC Main Workbench 6.9 to reveal the conserved regions.

2.2 Plasmid vector construction

Plasmid *pBHRF_ΔArm* (Figure 10, W. Johansen pers. comm.), containing the ampicillin and hygromycin resistance cassettes for selection in bacteria and plant, respectively, and 5′ and 3′ targeting sequences (TGS) corresponding to wild type (WT) *P. pate*ns genomic *dek1* sequence nucleotide 7249 to 8346 and 11545 to 12546 (relative to the ATG start site) respectively, was used to generate the LG3 mutant vector (*pBHRF_PpArm_ ΔLG3*) which was transformed into *P. patens dek1-ΔArm* mutant plant in this study.

Vector *pBHRF_PpArm_ΔLG3* was made in two steps; first by inserting the *P. patens dek1-Arm* sequence (nucleotide 8347 to 11544) into the linearized *pBHRF_ΔArm* plasmid creating plasmid *pBHRF_PpArmComp*, then by deleting the sequence corresponding to the *dek1-lg3* domain (nucleotide 10567 to 11239) from this vector thus creating the *pBHRF_PpArm_ΔLG3* plasmid (Figure 10).

To make the *pBHRF_PpArmComp* vector, the *P. patens dek1-Arm* sequence (called the insert) was PCR amplified from genomic *P. patens* DNA using primer pair

SP_Inf_PpArm_1 (Appendix 2A) and ASP_Inf_PpArm_1 (Appendix 2A). The PCR reaction (20 µl) contained 1x High-fidelity PCR buffer, 200 µM dNTPs, 0.5 µM sense and antisense primers, 0.02 U/ µl High-fidelity Hotstart Phusion DNA polymerase and 50 ng *P. patens* genomic DNA as template. PCR cycling was performed as follows: 98°C for 20 sec, 35 cycles of 98°C for 10 sec, 58°C for 20 sec, 72°C for 2 minutes, and a final elongation step of 72°C for 5 minutes.

In-verse PCR was used to linearize the *pBHRF_ΔArm* vector backbone, excluding the hygromycin resistance cassette, using primers SP_Inf_1 (Appendix 2A) and ASP_Inf_1 (Appendix 2A). The PCR reaction (50 µl) contained 1x High-fidelity PCR buffer, 200 µM dNTPs, 0.5µM sense and antisense primers, 0.02 U/ µl High-fidelity Hotstart Phusion DNA polymerase and 10 ng of plasmid *pBHRF_ΔArm*. PCR cycling was performed as follows: 98°C for 30 sec, 35 cycles of 98°C for 10 sec, 62°C for 30 sec, 72°C for 5 minutes, and final elongation step of 72°C for 7 minutes.

The PCR product (the linearized vector) was purified from 0.65% low-melting agarose gel using the QIAquick Gel Extraction kit according to the manufacture protocol.

In-Fusion cloning procedure was followed to ligate the insert with the linearized vector. 2 µl 5x In-Fusion HD Enzyme Premix was mixed with 2 µl of linearized vector (100 ng) and 2 µl of the insert PCR reaction, in a total reaction volume of 10µl. The reaction was incubated at 50°C for 15 minutes, and then 3 µl was transformed to *Stellar Competent Cells* (Section 2.3).

Vector *pBHRF_PpArm_ΔLG3* was constructed from the verified (see section 2.3) *pBHRF_PpArmComp* plasmid by PCR using primer pair SP_Inf_2 (Appendix 2A) and ASP_Inf_2 (Appendix 2A), to delete the sequence corresponding to the *dek1-lg3* segment (nucleotide 10567 to 11239). The components of the reaction and thermal amplification profile used were as follows: 1x Clone Amp HiFi PCR Premix, 0.2 µM sense and antisense primers, 10 ng *pBHRF_PpArmCom* plasmid, in a total volume of 25 µl. PCR cycling was performed as follows: 98°C for 10 sec, 35 cycles of 98°C for 10 sec, 55°C for 15 sec, 72°C for 45 sec, and final elongation step of 72°C for 1 minutes.

The linearized plasmid *pBHRF_PpArm_ΔLG3* was purified from 0.65% low melting agarose gel using QIAquick Gel Extraction kit according to the manufacture protocol. Finally, In-Fusion strategy was performed (Clontech Laboratory) using the In-Fusion® HD Cloning Kit, following manufacturer's instructions. Briefly, 2 µl of 5x In-Fusion HD Enzyme Premix was

mixed with 5 µl (15 ng) of the linearized *pBHRF_PpArm_ΔLG3* PCR product in a total reaction volume of 10 µl. The reaction was incubated at 50°C for 15 minutes, and then 3 µl was transformed into *Stellar Competent Cells* (Section 2.3). The different steps in plasmids constructions are shown schematically in Figure 10.

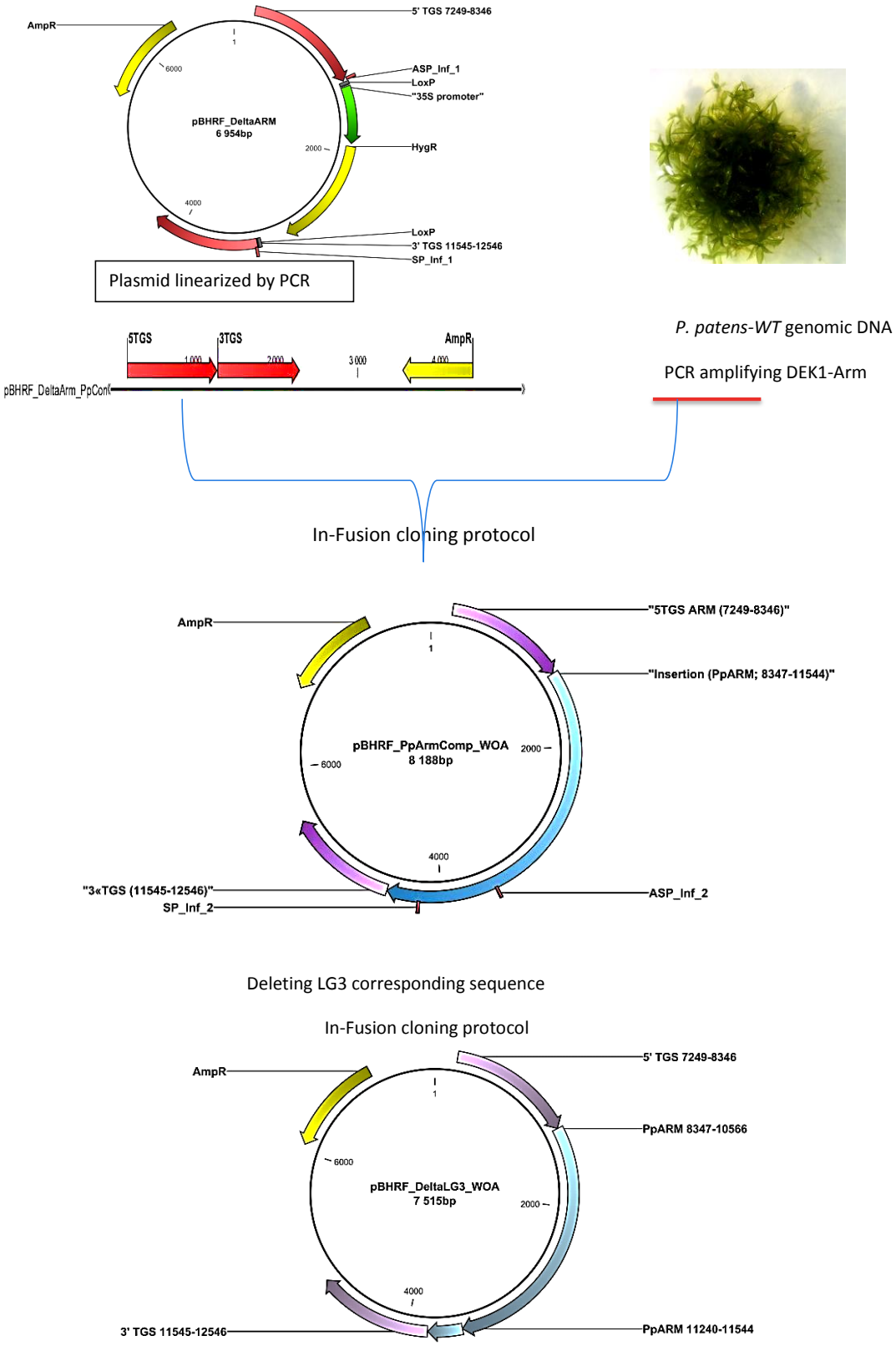


Figure10: The strategy of plasmids designing. The plasmid pBHRF_ ΔArm was linearized and the arm was amplified from the Physcomitrella patens genomic DNA followed by In-Fusion protocol to make the pBHRF_PpArmCom plasmid. Using specific primers, pBHRF_PpArmCom was linearized to delete the DEK1-LG3 sequence. In-Fusion protocol was performed on the linearized vector to make the final plasmid pBHRF_PpArm_ΔLG3.

2.3 Transformation of plasmids into *Stellar Competent Cells*, verification of constructs and plasmid preperation

Cloning reactions of *pBHRF_PpArmComp* and *pBHRF_PpArm_ΔLG3* were transformed into *Stellar Competent Cells* according to the manufacture protocol. In brief, 50 µl of cells were mixed with 3 µl of the cloning reaction, incubated on ice for 30 minutes and then the samples were heat shocked for exactly 45 sec at 42°C. 500 µl of pre-warmed (37°C) S.O.C medium was added to the cells and the samples were further incubated for 1 hour at 225 rpm in a shaking incubator. 20 µl of the cells were spread on Lysogeny broth (LB) plates (10 g/L peptone, 5 g/L yeast extract, 10 g/L NaCl and 16 g/L agar) containing ampicillin to a final concentration of 100 µg/ml. The plates were inverted and incubated at 37°C overnight.

To screen for putative positive clones harboring the *pBHRF_PpArmComp* or *pBHRF_PpArm_ΔLG3* plasmids, colony PCR was performed using the primer pair ΔArm IF (Appendix 2A) and ΔArmIR (Appendix 2A) or Arm_seq6 (Appendix 2A) and pBHRF_rev (Appendix 2A), respectively. For both PCR colony reactions the following components and thermal amplification profile were used: 1x AmpliTaq buffer, 0.2 µM sense and antisense primers, 1.5 mM $MgCl_2$, 200 µM dNTPs, and 0.01 µl AmpliTaq Gold DNA polymerase (1.25 Units/reaction) in a total reaction volume of 25 µl. PCR cycling was performed as follows: 95°C for 5 minutes, 35 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 1.5 minutes (for *pBHRF_PpArmComp*) and 2.5 minutes (for *pBHRF_PpArm_ΔLG3*), and final elongation step of 72°C for 3 minutes.

PCR positives colonies were selected and grown over-night in 4 ml LB medium supplemented with 100 µg/ml ampicillin. Plasmid was isolated using the Pure Yield™ Plasmid Miniprep System Kit (Promega) according to the manufactures instructions. Purified plasmids were analyzed by restriction digest analysis (RDA) using enzymes *XbaI and HincII* (BioLabs, New England) in two separate reactions. RDA positive plasmids were sequenced with specific primers spanning 500 bp domains (Appendix 2B) using the BigDye v.3.1 chemistry according to the stepped elongation time protocol (Platt, Woodhall, & George, 2007). DNA fragments were precipitated using sodium acetate:ethanol and finally sequenced by Capillary Electrophoresis using the 3130xL Genetic Analyzer (Life Technologies). CLC Genomic Workbench v 6.9 was used to analyze the sequences.

Plasmid *pBHRF_PpArm_ΔLG3* for *P. patens* transformation was prepared by inoculating 100 ml LB/ampicillin (100 µg/ml) with a single colony containing the sequence verified *pBHRF_PpArm_ΔLG3* plasmid. Plasmid was isolated using QIAGEN Plasmid Midi Kit according to manufacture protocol. The plasmid was digested with restriction enzymes *RsrII* and *PacI* for 6 hours at 37°C before DNA was precipitated with ethanol and concentrated to ~ 1 µg/µl.

2.4 Transformation and growth of *P. patens*

2.4.1 PEG-mediat protoplast transformation

The plasmid vector *pBHRF_PpArm_ΔLG3* was transformed into the *P. pate*ns *dek1-ΔArm* mutant to create the *P. patens dek1-Δlg3*. In the *P. patens dek1-Δarm* mutant, the entire *dek1-Arm* segment has been removed and replaced by the hygromycin resistant cassette using homologous recombination (Figure 11) (W. Johansen pers. comm.). *P. patens* was transformed using the PEG-mediate protoplast transformation method according to Cove et al (David J Cove et al., 2009) and as described below.

Three agar-plates containing 7 days old *P. patens dek1-Δarm* tissue were incubated with 20ml of driselase solution (15 ml mannitol (8.5%) + 5ml driselase (2%)) for 45minutes. The resulting protoplast solution was filtered twice, first through 100 µm then through 50 µm filters before the protoplasts were collected by centrifugation at 300 x g for 5 minutes at room temperature. The pellet was washed twice by re-suspending the pellet in the same volume of protoplast wash solution, PW, (Mannitol 8.5% + 1% $CaCl_2$ 1M) and re-centrifuged. The protoplast density was estimated using a hemocytometer then re-suspended with MMM solution (Appendix 3) to adjust the protoplast density to $1.6x10^6$ cells/ml. 300 µl protoplasts solution was mixed with 300 µl polyethylene glycol (PEG_{6000}) (Appendix 3) and 15 µg of *RsrII/PacI* digested *pBHRF_PpArm_ΔLG3* plasmid. The sample was then heat shocked in a water bath at 45°C for 5 minutes. The mixture was cooled to room temperature for 10 minutes, and then 10 ml PW solution was added gradually before the sample was centrifuged at 300 x g for 5 minutes at room temperature. The pellet was re-suspended with 5-6 ml PW and 7-8 ml PRMT solution (Appendix 3). The solution was spread on cellophane disc containing protoplast regeneration medium PRMB agar plates (Appendix 3) and incubated at 24°C for 6 days. After 6 days, the cellophane discs containing the tissues were

transferred to agar-plates containing BCDA media (Appendix 3) and incubated for 10 day using the same culturing conditions, before it was transferred to agar plates containing BCD medium (Appendix 3) and incubated for additional 10 days. The growth of both plant tissues and protoplasts were under long day conditions (16-h light [70–80 µmol m^{-2} s^{-1}]/8-h dark) at 25°C. In some cases, 50 µg/l of vancomycin was supplemented to the BCDA media to prevent microbial contamination.

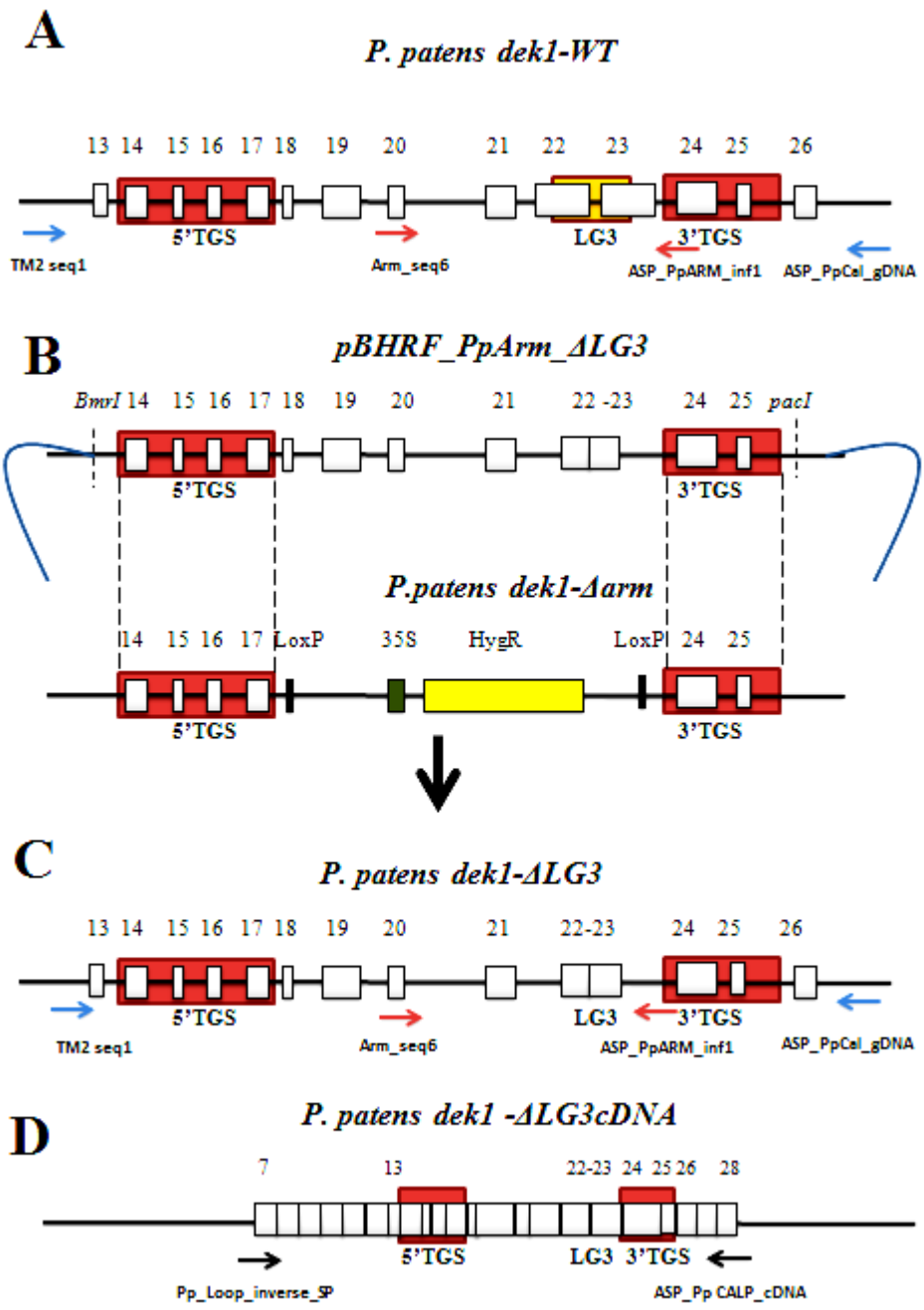


Figure11: Schematic presentation of P. patens dek1-WT, P. patens dek1-Δarm mutant plant, LG3 vector, genomic P. patens dek1-ΔLG3 mutant plant, and cDNA of P. patens dek-ΔLG3 mutant plant. A) P. patens dek1 locus, B) The pBHRF_PpArm_ΔLG3 plasmid was transformed into P. patens dek1-ΔArm mutant plant in order to make the P. patens dek1-ΔLG3 mutant plant. This procedure was performed depending on the homologous recombination due to presence if 5'TGS and 3'TGS, C) P. patens dek1-ΔLG3locus showing also the primers used for first and second genotyping, D) P. patens dek1-ΔLG3 cDNA locus showing the primers used for sequencing.

2.5 Selection and characterization of putative *P. patens dek1-Δlg3* mutant lines

2.5.1 Selection of putative *P. patens dek1-Δlg3* mutant lines

The line used for transformation in this study, the *P. patens dek1-ΔArm* mutant, has a *Δdek1* phenotype (W. Johansen, pers. comm.) characterized by developmental arrest and the absence of three-dimensional structures (gametophore). The strategy for selecting the putative transgenic lines generated in the present study was to select surviving plants which showed a different phenotype from the background line (*P. patens dek1-ΔArm* mutant). This characterization was performed by visual inspection of plants and by the use of a dissecting microscope (Nikon SMZ 1500)

2.5.2 Molecular characterization of putative *P. patens dek1-Δlg3* mutant lines

2.5.2.1 Genotyping PCR

Putative mutant plants were investigated by genotyping PCR using the Phire® Plant Direct PCR Kit (Thermo Scientific). The first genotyping was aimed at identifying mutant lines with loss of the lg3 segment using primers Arm_seq6 (Appendix 2A) and ASP_Inf_PpArm_1 (Appendix 2A) (Figure 11).

Second PCR genotyping was performed using primers TM2 seq1 (Appendix 2A) and ASP_PpCal_gDNA (Appendix 2A) (Figure 11) to confirm single-copy integration of the transgene. PCR reactions were analyzed by agarose gel electrophoresis (0.8% agar in 1x TAE buffer). For both PCR genotyping reactions, genomic DNA extracted from WT *P. patens* was used as positive control.

2.5.2.2 Southern Blot

Genomic DNA was isolated from 10 days old plants using DNA Extraction kit PHYTOPURE (GE Healthcare) according to the manufactures instructions. Approximately 1µg DNA was digested with the restriction enzyme *HindIII* at 37°C overnight. DNA fragments were separated by agarose/1x TAE gel electrophoresis for 18 hours at 37 v/h using 0.6% ultrapure agarose gel (Invitrogen).

The DNA was depurinated by soaking the gel in 250 mM HCl for 10 minutes at room temperature. After depurination, DNA was denaturated by adding denaturation buffer (0.5 M NaOH + 1.5 M NaCl) twice for 15 minutes each at room temperature before the DNA was neutralized with neutralization buffer (0.5 M Tris-HCl pH 7.0 + 1.5 M NaCl) two times for 15 minutes each at room temperature.

The gel was incubated in transfer buffer (87.65 g/l NaCl + 44.1 g/l trisodium citrate) for 5 minutes, and then the DNA was transferred to a positively charged nylon membrane (Roche) using tissue papers and 3mm filters and incubated for 5 hours. After the transfer, the membrane was air dried for 2 minutes before DNA was cross-linked two times at 120 mega joules with a UV crosslinker.

The membrane was pre-hybridized using hybridization buffer (Roche buffer) for 30 minutes at 42°C, then the membrane was incubated overnight with pre-heated DNA probe (25 ng/ml), diluted with Roche buffer in a 10 ml volume, at 42°C with constant rotation. The probe were in advance synthesized and labelled with digoxigenin using the DIG Probe PCR synthesis kit (Roche) following the manufacture manual. The primers used in probe synthesis were sense primer Armseq2 (Appendix 2A) and anti-sense primer ΔArm1R (Appendix 2A). PCR cycling was performed as follows: 95°C for 2 minutes, 30 cycles of 95°C for 30 sec, 62°C for 30 sec, 72°C for 80 sec, and final elongation step of 72°C for 5 minutes.

Washing step were performed first by washing the membrane two times for 5 minutes each with first washing buffer (35.06 g/l NaCl + 17.64 g/l trisodium citrate + 0.1% Sodium Dodecyl Sulfate) at room temperature, then two times for 30 minutes each at 65°C with pre-heated second washing solution (1.75 g/l NaCl + 0.88 g/l trisodium citrate + 0.1% Sodium Dodecyl Sulfate).

The membrane was blocked for 1 hour at room temperature with DIG1 solution (maleic acid 11.6 g/l, NaCl 8.76 g/l, NaOH pellets 2 g/l, pH 7.5). After blocking, the membrane was incubated with the antibody solution (Anti-DIG antibody + DIG1 blocking solution 1:25000) using rotation for 30 minutes at 25°C. Excess antibody was removed by washing the membrane twice in DIG1 buffer with 0.3% Tween20 for 45 minutes each on a shaker. The membrane was equilibrated for 5 minutes in equilibration buffer (100mM Tris-HCl pH9.6 +

100mM NaCl), before it was incubated with chemiluminescent substrates (CDP-Star®) and equilibration buffer (1:200) for 5 minutes at room temperature in the dark.

The last step was performed in the dark room were the membrane was wrapped in plastic cover and incubated with Amersham Hyperfilm (GE Healthcare) in a Hyper cassette (Amersham International plc) for 40 minutes. After incubation, the film was washed in a developer solution bath G150 (AGFA) for 5 minutes, followed by washing in water bath for few seconds before it washed with fixing solution bath (AGFA) for two minutes.

2.5.2.3 RT-PCR

Total RNA was isolated from *P. patens* tissue using the RNeasy® Plant Mini Kit (QIAGEN), following the manufacture instructions.

1 µg of total RNA was treated with 1 µl of 10x DNase buffer (Invitrogen) and 1 µl of DNase1 (Invitrogen) in a total reaction volume of 10 µl. The reaction was incubated at 37°C for 30 minutes. The DNase was inactivated by adding 1 µl of 25 mM EDTA to the reaction, and then incubated at 65°C for 10 minutes.

cDNA synthesis was performed by adding 1 µl of random hexamer primers (50 µM) (Amersham Pharmacia Biotech Inc) and 1 µl of 10 mM dNTPs to approximately 500 ng of *DNaseI*-treated RNA sample in a total reaction volume of 13 µl. The reaction was incubated at 65°C for 5 minutes. 7 µl of reverse transcription mixture, made by mixing 4 µl 5x First-Strand buffer, 1 µl 0.1 M DTT, 1 µl of 40 U/µl RNaseOUT™ and 1 µl of 200 U/µl *SuperScript™ III Reverse Transcriptase* (Invitrogen), was added to the solution. The mixture was incubated at 25°C for 5 minutes, then at 50 °C for 30 minutes and finally at 70 °C for 15 minutes to inactive the RT enzyme

A PCR reaction was performed to amplify part of the *dek1* cDNA using primers Pp_Loop_inverse_SP (Appendix 2A) and ASP_Pp CALP_cDNA (Appendix 2A) (Figure 11D). The PCR reaction (50 µl) contained 1x High-fidelity PCR buffer, 200 µM dNTPs, 0.5 µM sense and antisense primers, 0.02 U/µl High-fidelity Hotstart Phusion DNA polymerase and 1 µl of cDNA. PCR cycling was performed as follows: 98°C for 30 sec, 35 cycles of 98°C for 10 sec, 60°C for 30 sec, 72°C for 80 sec, and final elongation step of 72°C for 7 minutes.

The RT-PCR product was sequenced from exon 7 to exon 28 to confirm correct splicing at the mutant *dek1* locus. The PCR product was sequenced according to Platt et al (Platt et al., 2007) and as described below.

The reaction contained 2 µl of PCR product, 1x Big Dye Sequencing Buffer and 4 U *ExoI* in a total reaction volume of 10 µl. The samples were incubated at 37 °C for 60 min and 85 °C for 15 min. The *ExoI*-treated PCR product (5 µl) was sequenced by adding 0.32 µM overlapping specific primers using gene-specific primers (Appendix 2C), 2 µl of 5x BigDye Sequencing Buffer, and 0.5 µl of BigDye Terminator mix v3.1 in a total reaction of 10 µl. The cycling conditions: 96 °C for 1 min, 15 cycles: 96 °C for 10 sec, 50 °C for 5 sec, 60 °C for 75 sec, 5 cycles: 96 °C for 10 sec, 50 °C for 5 sec, 60 °C for 90 sec. and 5 cycles: 96 °C for 10 sec, 50 °C for 5 sec, 60 °C for 2 min. DNA fragments were precipitated using sodium acetate:ethanol and finally sequenced by Capillary Electrophoresis using the 3130xL Genetic Analyzer (Life Technologies). The Genomic Workbench Software was used to analyze the sequences.

2.5.2.4 Real time quantitative PCR (qPCR)

Total RNA was isolated from 6 and 16 days old *P. patens* WT and *dek1-Δlg3* mutant plants. RNA isolation and cDNA synthesis was performed as described previously (section 2.5.2.3). Real-Time PCR (qPCR) was performed in the 7500 Real–Time system (Applied Biosynthesis) using Eva Green Fire Pol qPCR reaction containing 5x Eva Green Fire Pol qPCR Mix (Solis Bio Dyne), 0.2 µM of each sense and anti-sense primers, and 1 µl of undiluted and 10-fold diluted cDNA in a total reaction volume of 15 µl. The reactions were performed in optical 96-Well Reaction plate (Applied Biosystems). The experimental primers used in the qPCR were CALPqF (Appendix 2A) and CALPqR (Appendix 2A), while the reference primers were SQSF (Appendix 2A) and SQSR (Appendix 2A). PCR cycling was performed as follows: 95°C for 15 minutes, 40 cycles of 95°C for 30 sec, 58°C for 20 sec, 72°C for 32 sec, and a final dissociation step.

The data from the qPCR reactions was analysed using LinRegPCR version 2012.2 (J. M. Ruijter, Amsterdam, the Netherlands). LinRegPCR is a program for analysing qPCR resulting from PCR reaction contains SYBR green or similar fluorescence dyes. The program performs a baseline fluorescence and baseline subtraction, determines a window-of-linearity, and then the PCR efficiencies per each sample are calculated. The mean PCR

efficiency per amplicon and the Ct value per sample are used to calculate a starting concentration per sample, expressed in arbitrary fluorescence units (Ramakers, Ruijter, Deprez, & Moorman, 2003; Ruijter et al., 2009).

2.5.3 Phenotypic characterization of the *P. patens dek1-Δlg3* mutant lines

Confocal microscopy (LEICA TCS SP5) was used to study early bud and gametophore development in both WT and mutant plants. To study phyllid morphology a dissection microscope was used. Gametangia were analysed from 3 months, old *P. patens* WT and *dek1-Δlg3* mutant plants cultivated on sterile soil block, to reveal if there is any difference between the two lines. The plant tissue was stained with Propidium iodide (PI) for 30 minutes, and then washed three times in sterile water before mounting in glass bottom dished (WillCo Wells B.V., Amsterdam, the Netherlands).

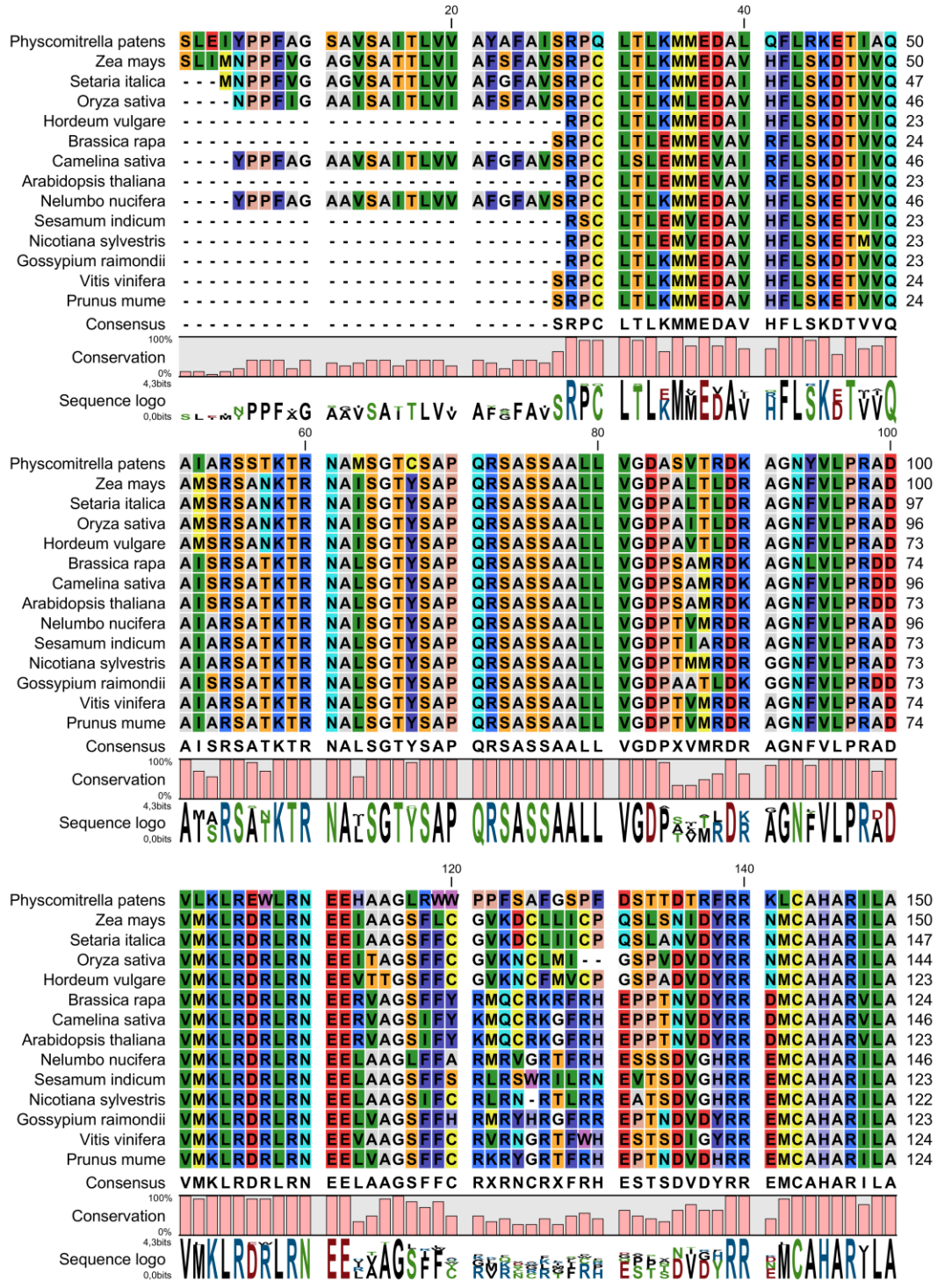
3. Results

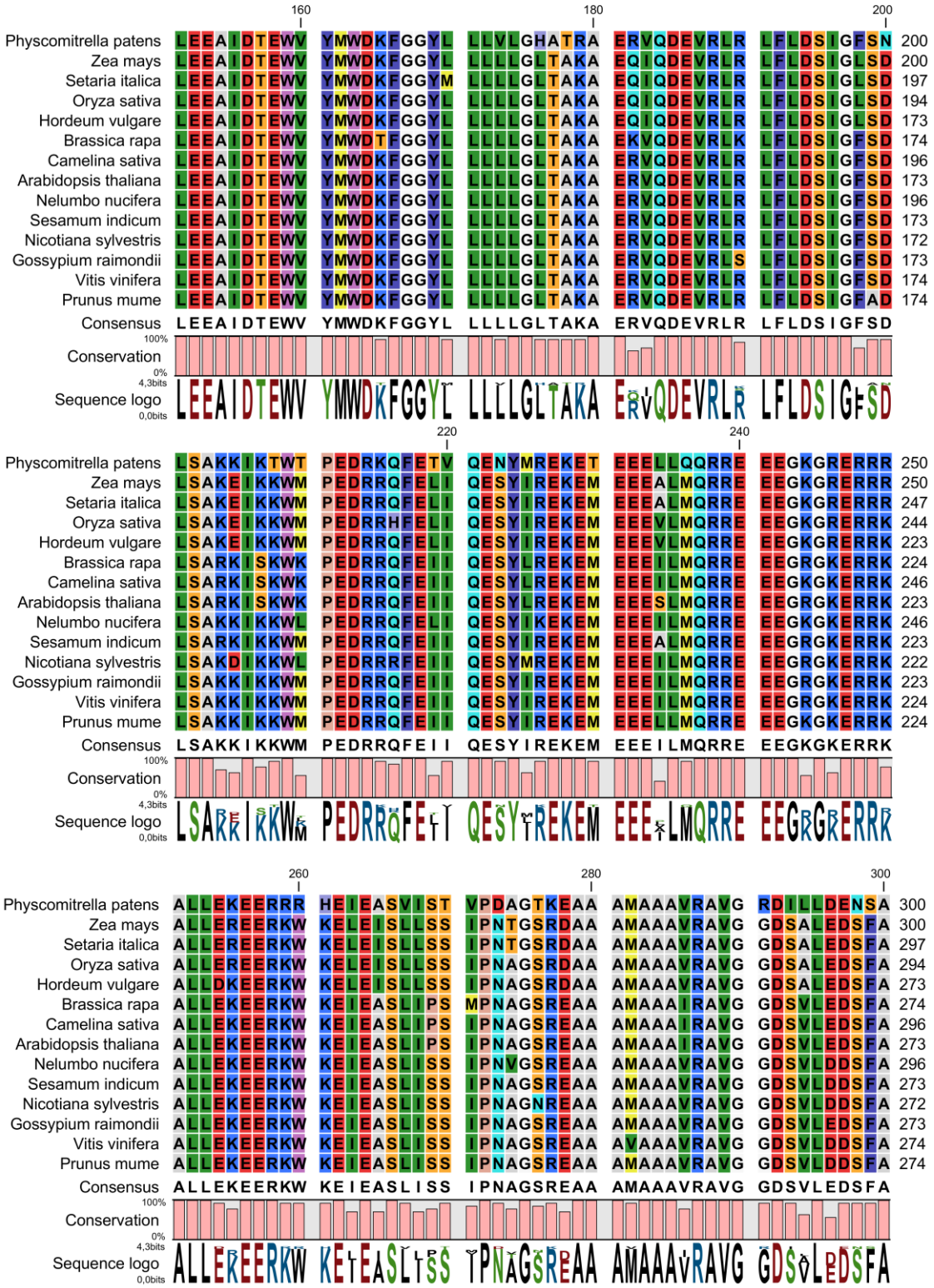
3.1 *In silico* analysis

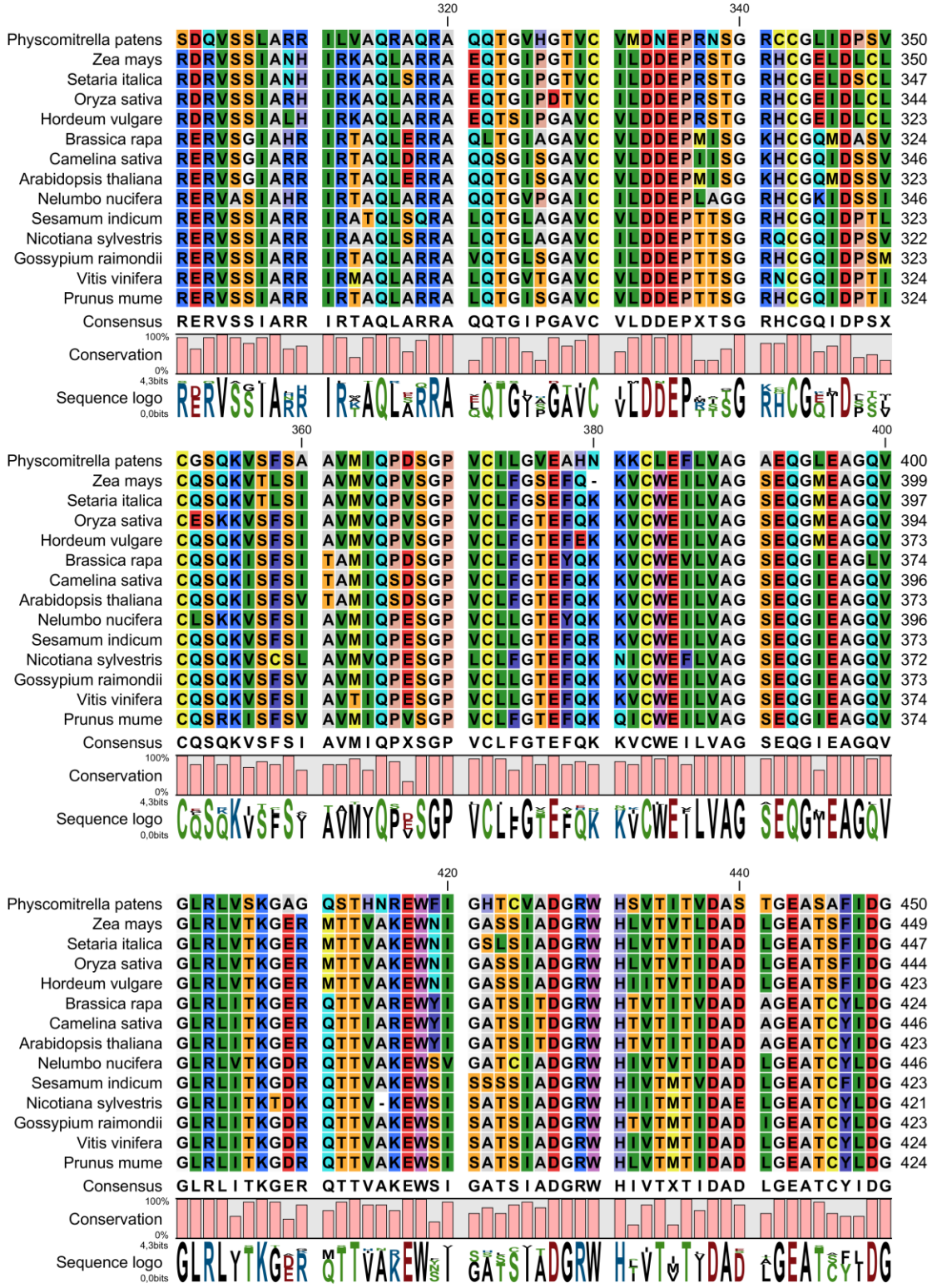
3.1.1 Multiple sequence alignment of the DEK1-Arm sequence

The DEK1 protein in plants harbour 21-23 transmembrane segments, a Loop insertion, a non-structured cytoplasmic Arm, and a highly conserved calpain domain (Kumar et al., 2013; Lid et al., 2002).

Not much work has been done on the DEK1-Arm until now. However, the *P. patens* DEK1-Arm are quite large containing approximately 613 amino acids according to SMART database. The amino acid sequences of the *P. patens* DEK1 protein and from other plants were acquired from the protein database at National Center for Biotechnology Information (NCBI). The DEK1-Arm sequences coordinates were determined using the SMART database based on pfam, and then multiple sequence alignment was performed using the Clustal Omega database at The European Bioinformatics Institute website in order to reveal the conserved region in the DEK1-Arm as shown in Figure 12.


20
40
Physcomitrella patens SLEIYPPFAG SAVSAITLVV AYAFAISRPQ LTLKMMEDAL QFLRKETIAQ 50
Zea mays SLIMNPPFVG AGVSATTLVI AFSFAVSRPC LTLKMMEDAV HFLSKDTVVQ 50
Setaria italica ---MNPPFVG AGVSATTLVV AFGFAVSRPC LTLKMMEDAI HFLSKDTVVQ 47
Oryza sativa ----NPPFIG AAISAITLVI AFSFAVSRPC LTLKMLEDAV HFLSKDTVVQ 46
Hordeum vulgare ---------- ---------- -------RPC LTLKMMEDAI HFLSKDTVIQ 23
Brassica rapa ---------- ---------- -------SRPC LTLEMMEVAV RFLSKDTVVQ 24
Camelina sativa ----YPPFAG AAVSAITLVV AFGFAVSRPC LSLEMMEVAI RFLSKDTIVQ 46
Arabidopsis thaliana ---------- ---------- -------RPC LTLEMMEVAV RFLSKDTIVQ 23
Nelumbo nucifera ----YPPFAG AAVSAITLVV AFGFAVSRPC LTLKMMEDAV HFLSKETVVQ 46
Sesamum indicum ---------- ---------- -------RSC LTLEMVEDAV HFLSKETVIQ 23
Nicotiana sylvestris ---------- ---------- -------RPC LTLEMVEDAV HFLSKETMVQ 23
Gossypium raimondii ---------- ---------- -------RPC LTLKMMEDAV HFLSKDTVVQ 23
Vitis vinifera ---------- ---------- -------SRPC LTLKMMEDAV HFLSKETVVQ 24
Prunus mume ---------- ---------- -------SRPC LTLKMMEDAV HFLSKETVVQ 24
Consensus ---------- ---------- -------SRPC LTLKMMEDAV HFLSKDTVVQ
100%
Conservation
0%
4,3bits
Sequence logo
0,0bits
60
80
100
Physcomitrella patens AIARSSTKTR NAMSGTCSAP QRSASSAALL VGDASVTRDK AGNYVLPRAD 100
Zea mays AMSRSANKTR NAISGTYSAP QRSASSAALL VGDPALTLDR AGNFVLPRAD 100
Setaria italica AMSRSANKTR NAISGTYSAP QRSASSAALL VGDPALTLDR AGNFVLPRAD 97
Oryza sativa AMSRSANKTR NAISGTYSAP QRSASSAALL VGDPAITLDR AGNFVLPRAD 96
Hordeum vulgare AMSRSANKTR NAISGTYSAP QRSASSAALL VGDPAVTLDR AGNFVLPRAD 73
Brassica rapa AISRSATKTR NALSGTYSAP QRSASSAALL VGDPSAMRDK AGNLVLPRDD 74
Camelina sativa AISRSATKTR NALSGTYSAP QRSASSAALL VGDPSAMRDK AGNFVLPRDD 96
Arabidopsis thaliana AISRSATKTR NALSGTYSAP QRSASSAALL VGDPSAMRDK AGNFVLPRDD 73
Nelumbo nucifera AIARSATKTR NAISGTYSAP QRSASSAALL VGDPTVMRDR AGNFVLPRAD 96
Sesamum indicum AIARSATKTR NALSGTYSAP QRSASSAALL VGDPTIARDR AGNFVLPRAD 73
Nicotiana sylvestris AIARSATKTR NALSGTYSAP QRSASSAALL VGDPTMMRDR GGNFVLPRAD 73
Gossypium raimondii AISRSATKTR NALSGTYSAP QRSASSAALL VGDPAATLDK GGNFVLPRDD 73
Vitis vinifera AIARSATKTR NALSGTYSAP QRSASSAALL VGDPTVMRDR AGNFVLPRAD 74
Prunus mume AIARSATKTR NALSGTYSAP QRSASSAALL VGDPTVMRDR AGNFVLPRAD 74
Consensus AISRSATKTR NALSGTYSAP QRSASSAALL VGDPXVMRDR AGNFVLPRAD
100%
Conservation
0%
4,3bits
Sequence logo
0,0bits
120
140
Physcomitrella patens VLKLREWLRN EEHAAGLRWW PPFSAFGSPF DSTTDTRFRR KLCAHARILA 150
Zea mays VMKLRDRLRN EEIAAGSFLC GVKDCLLICP QSLSNIDYRR NMCAHARILA 150
Setaria italica VMKLRDRLRN EEIAAGSFFC GVKDCLIICP QSLANVDYRR NMCAHARILA 147
Oryza sativa VMKLRDRLRN EEITAGSFFC GVKNCLMI-- GSPVDVDYRR NMCAHARILA 144
Hordeum vulgare VMKLRDRLRN EEVTTGSFFC GVKNCFMVCP GSPADVDYRR NMCAHARILA 123
Brassica rapa VMKLRDRLRN EERVAGSFFY RMQCRKRFRH EPPTNVDYRR DMCAHARVLA 124
Camelina sativa VMKLRDRLRN EERVAGSIFY KMQCRKGFRH EPPTNVDYRR DMCAHARVLA 146
Arabidopsis thaliana VMKLRDRLRN EERVAGSIFY KMQCRKGFRH EPPTNVDYRR DMCAHARVLA 123
Nelumbo nucifera VMKLRDRLRN EELAAGLFFA RMRVGRTFRH ESSSDVGHRR EMCAHARILA 146
Sesamum indicum VMKLRDRLRN EELAAGSFFS RLRSWRILRN EVTSDVGHRR EMCAHARILA 123
Nicotiana sylvestris VMKLRDRLRN EELAAGSIFC RLRN-RTLRR EATSDVGHRR EMCAHARILA 122
Gossypium raimondii VMKLRDRLRN EELVAGSFFH RMRYHRGFRR EPTNDVDYRR EMCAHARILA 123
Vitis vinifera VMKLRDRLRN EEVAAGSFFC RVRNGRTFWH ESTSDIGYRR EMCAHARILA 124
Prunus mume VMKLRDRLRN EELVAGSFFC RKRYGRTFRH EPTNDVDHRR EMCAHARILA 124
Consensus VMKLRDRLRN EELAAGSFFC RXRNCRXFRH ESTSDVDYRR EMCAHARILA
100%
Conservation
0%
4,3bits
Sequence logo
0,0bits


160
180
200
Physcomitrella patens LEEAIDTEWV YMWDKFGGYL LLVLGHATRA ERVQDEVRLR LFLDSIGFSN 200
Zea mays LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA EQIQDEVRLR LFLDSIGLSD 200
Setaria italica LEEAIDTEWV YMWDKFGGYM LLLLGLTAKA EQIQDEVRLR LFLDSIGLSD 197
Oryza sativa LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA EQIQDEVRLR LFLDSIGLSD 194
Hordeum vulgare LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA EQIQDEVRLR LFLDSIGLSD 173
Brassica rapa LEEAIDTEWV YMWDTFGGYL LLLLGLTAKA EKVQDEVRLK LFLDSIGFSD 174
Camelina sativa LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA ERVQDEVRLR LFLDSIGFSD 196
Arabidopsis thaliana LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA ERVQDEVRLR LFLDSIGFSD 173
Nelumbo nucifera LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA ERVQDEVRLR LFLDSIGFSD 196
Sesamum indicum LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA ERVQDEVRLR LFLDSIGFSD 173
Nicotiana sylvestris LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA ERVQDEVRLR LFLDSIGFSD 172
Gossypium raimondii LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA ERVQDEVRLS LFLDSIGFSD 173
Vitis vinifera LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA ERVQDEVRLR LFLDSIGFSD 174
Prunus mume LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA ERVQDEVRLR LFLDSIGFAD 174
Consensus LEEAIDTEWV YMWDKFGGYL LLLLGLTAKA ERVQDEVRLR LFLDSIGFSD
100%
Conservation
0%
4,3bits
Sequence logo
0,0bits
220
240
Physcomitrella patens LSAKKIKTWT PEDRKQFETV QENYMREKET EEELLQQRRE EEGKGRERRR 250
Zea mays LSAKEIKKWM PEDRRQFELI QESYIREKEM EEEALMQRRE EEGKGRERRR 250
Setaria italica LSAKEIKKWM PEDRRQFELI QESYIREKEM EEEALMQRRE EEGKGRERRR 247
Oryza sativa LSAKEIKKWM PEDRRHFELI QESYIREKEM EEEVLMQRRE EEGKGRERRK 244
Hordeum vulgare LSAKEIKKWM PEDRRQFELI QESYIREKEM EEEVLMQRRE EEGKGRERRK 223
Brassica rapa LSARKISKWK PEDRRQFEII QESYLREKEM EEEILMQRRE EEGRGKERRK 224
Camelina sativa LSARKISKWK PEDRRQFEII QESYLREKEM EEEILMQRRE EEGKGKERRK 246
Arabidopsis thaliana LSARKISKWK PEDRRQFEII QESYLREKEM EEESLMQRRE EEGRGKERRK 223
Nelumbo nucifera LSAKKIKKWL PEDRRQFELI QESYIKEKEM EEEILMQRRE EEGKGKERRK 246
Sesamum indicum LSARKIKKWM PEDRRQFEII QESYIREKEM EEEALMQRRE EEGRGKERRK 223
Nicotiana sylvestris LSAKDIKKWL PEDRRRFEII QESYMREKEM EEEILMQRRE EEGRGKERRK 222
Gossypium raimondii LSAKKIKKWM PEDRRQFEII QESYIREKEM EEEILMQRRE EEGRGKERRK 223
Vitis vinifera LSAKKIKKWM PEDRRQFEII QESYIREKEM EEEILMQRRE EEGRGKERRK 224
Prunus mume LSAKKIKKWM PEDRRQFEII QESYIREKEM EEELLMQRRE EEGKGKERRK 224
Consensus LSAKKIKKWM PEDRRQFEII QESYIREKEM EEEILMQRRE EEGKGKERRK
100%
Conservation
0%
4,3bits
Sequence logo
0,0bits
260
280
300
Physcomitrella patens ALLEKEERRR HEIEASVIST VPDAGTKEAA AMAAAVRAVG RDILLDENSA 300
Zea mays ALLEREERKW KELEISLLSS IPNTGSRDAA AMAAAVRAVG GDSALEDSFA 300
Setaria italica ALLEREERKW KELEISLLSS IPNTGSRDAA AMAAAVRAVG GDSALEDSFA 297
Oryza sativa ALLEREERKW KELEISLLSS IPNAGSRDAA AMAAAVRAVG GDSALEDSFA 294
Hordeum vulgare ALLDKEERKW KELEISLLSS IPNAGSRDAA AMAAAVRAVG GDSALEDSFA 273
Brassica rapa ALLEKEERKW KEIEASLIPS MPNAGSREAA AMAAAIRAVG GDSVLEDSFA 274
Camelina sativa ALLEKEERKW KEIEASLIPS IPNAGSREAA AMAAAIRAVG GDSVLEDSFA 296
Arabidopsis thaliana ALLEKEERKW KEIEASLIPS IPNAGSREAA AMAAAIRAVG GDSVLEDSFA 273
Nelumbo nucifera ALLEKEERKW KEIEASLISS IPNVGSREAA AMAAAVRAVG GDSVLDDSFA 296
Sesamum indicum ALLEKEERKW KEIEASLISS IPNAGSREAA AMAAAVRAVG GDSVLDDSFA 273
Nicotiana sylvestris ALLEKEERKW KEIEASLISS IPNAGNREAA AMAAAVRAVG GDSVLDDSFA 272
Gossypium raimondii ALLEKEERKW KEIEASLISS IPNAGSREAA AMAAAVRAVG GDSVLEDSFA 273
Vitis vinifera ALLEKEERKW KEIEASLISS IPNAGSREAA AVAAAVRAVG GDSVLDDSFA 274
Prunus mume ALLEKEERKW KEIEASLISS IPNAGSREAA AMAAAVRAVG GDSVLDDSFA 274
Consensus ALLEKEERKW KEIEASLISS IPNAGSREAA AMAAAVRAVG GDSVLEDSFA
100%
Conservation
0%
4,3bits
Sequence logo
0,0bits


320
340
Physcomitrella patens SDQVSSLARR ILVAQRAQRA QQTGVHGTVC VMDNEPRNSG RCCGLIDPSV 350
Zea mays RDRVSSIANH IRKAQLARRA EQTGIPGTIC ILDDEPRSTG RHCGELDLCL 350
Setaria italica RDRVSSIANH IRKAQLSRRA EQTGIPGTVC ILDDEPRSTG RHCGELDSCL 347
Oryza sativa RDRVSSIARH IRKAQLARRA EQTGIPDTVC ILDDEPRSTG RHCGEIDLCL 344
Hordeum vulgare RDRVSSIALH IRKAQLARRA EQTSIPGAVC VLDDEPRSTG RHCGEIDLCL 323
Brassica rapa RERVSGIAHR IRTAQLERRA QLTGIAGAVC VLDDEPMISG KHCGQMDASV 324
Camelina sativa RERVSGIARR IRTAQLDRRA QQSGISGAVC VLDDEPIISG KHCGQIDSSV 346
Arabidopsis thaliana RERVSGIARR IRTAQLERRA QQTGISGAVC VLDDEPMISG KHCGQMDSSV 323
Nelumbo nucifera RERVASIAHR IRTAQLARRA QQTGVPGAIC ILDDEPLAGG RHCGKIDSSI 346
Sesamum indicum RERVSSIARR IRATQLSQRA LQTGLAGAVC VLDDEPTTSG RHCGQIDPTL 323
Nicotiana sylvestris RERVSSIARR IRAAQLSRRA LQTGLAGAVC ILDDEPTTSG RQCGQIDPSV 322
Gossypium raimondii RERVSSIARR IRTAQLARRA VQTGLSGAVC ILDDEPTTSG RHCGQIDPSM 323
Vitis vinifera RERVSSIARR IRMAQLARRA LQTGVTGAVC VLDDEPTTSG RNCGQIDPTI 324
Prunus mume RERVSSIARR IRTAQLARRA LQTGISGAVC VLDDEPTTSG RHCGQIDPTI 324
Consensus RERVSSIARR IRTAQLARRA QQTGIPGAVC VLDDEPXTSG RHCGQIDPSX
100%
Conservation
0%
4,3bits
Sequence logo
0,0bits
360
380
400
Physcomitrella patens CGSQKVSFSA AVMIQPDSGP VCILGVEAHN KKCLEFLVAG AEQGLEAGQV 400
Zea mays CQSQKVTLSI AVMVQPVSGP VCLFGSEFQ- KVCWEILVAG SEQGMEAGQV 399
Setaria italica CQSQKVTLSI AVMVQPVSGP VCLFGSEFQK KVCWEILVAG SEQGMEAGQV 397
Oryza sativa CESKKVSFSI AVMVQPVSGP VCLFGTEFQK KVCWEILVAG SEQGMEAGQV 394
Hordeum vulgare CQSQKVSFSI AVMVQPVSGP VCLFGTEFEK KVCWEILVAG SEQGMEAGQV 373
Brassica rapa CQSQKISFSI TAMIQPDSGP VCLFGTEYQK KVCWEVLVAG SEQGIEAGLV 374
Camelina sativa CQSQKISFSI TAMIQSDSGP VCLFGTEFQK KVCWEILVAG SEQGIEAGQV 396
Arabidopsis thaliana CQSQKISFSV TAMIQSDSGP VCLFGTEFQK KVCWEILVAG SEQGIEAGQV 373
Nelumbo nucifera CLSKKVSFSI AVMIQPESGP VCLLGTEYQK KVCWEILVAG SEQGIEAGQV 396
Sesamum indicum CQSQKVSFSI AVMIQPESGP VCLLGTEFQR KVCWEILVAG SEQGIEAGQV 373
Nicotiana sylvestris CQSQKVSCSL AVMVQPESGP LCLFGTEFQK NICWEFLVAG SEQGIEAGQV 372
Gossypium raimondii CQSQKVSFSV AVMIQPESGP VCLLGTEFQK KVCWEILVAG SEQGIEAGQV 373
Vitis vinifera CQSQKVSFSI AVTIQPESGP VCLLGTEFQK KVCWEILVAG SEQGIEAGQV 374
Prunus mume CQSRKISFSV AVMIQPVSGP VCLFGTEFQK QICWEILVAG SEQGIEAGQV 374
Consensus CQSQKVSFSI AVMIQPXSGP VCLFGTEFQK KVCWEILVAG SEQGIEAGQV
100%
Conservation
0%
4,3bits
Sequence logo
0,0bits
420
440
Physcomitrella patens GLRLVSKGAG QSTHNREWFI GHTCVADGRW HSVTITVDAS TGEASAFIDG 450
Zea mays GLRLVTKGER MTTVAKEWNI GASSIADGRW HLVTVTLDAD LGEATSFIDG 449
Setaria italica GLRLVTKGER MTTVAKEWNI GSLSIADGRW HLVTVTIDAD LGEATSFIDG 447
Oryza sativa GLRLVTKGER MTTVAKEWNI GASSIADGRW HLVTVTIDAD LGEATSFIDG 444
Hordeum vulgare GLRLVTKGER MTTVAKEWNI GASSIADGRW HIITVTIDAD LGEATSFIDG 423
Brassica rapa GLRLITKGER QTTVAREWYI GATSITDGRW HTVTITVDAD AGEATCYLDG 424
Camelina sativa GLRLITKGER QTTIAREWYI GATSITDGRW HTVTITIDAD AGEATCYIDG 446
Arabidopsis thaliana GLRLITKGER QTTVAREWYI GATSITDGRW HTVTITIDAD AGEATCYIDG 423
Nelumbo nucifera GLRLVTKGDR QTTVAKEWSV GATCIADGRW HIVTVTIDAD LGEATCYLDG 446
Sesamum indicum GLRLITKGDR QTTVAKEWSI SSSSIADGRW HIVTMTVDAD LGEATCFIDG 423
Nicotiana sylvestris GLRLITKTDK QTTV-KEWSI SATSIADGRW HIITMTIDAE LGEATCYLDG 421
Gossypium raimondii GLRLITKGDR QTTVAKEWSI SATSIADGRW HTVTMTIDAD IGEATCYLDG 423
Vitis vinifera GLRLITKGDR QTTVAKEWSI SATSIADGRW HIVTMTIDAD LGEATCYLDG 424
Prunus mume GLRLITKGDR QTTVAKEWSI SATSIADGRW HLVTMTIDAD LGEATCYLDG 424
Consensus GLRLITKGER QTTVAKEWSI GATSIADGRW HIVTXTIDAD LGEATCYIDG
100%
Conservation
0%
4,3bits
Sequence logo
0,0bits

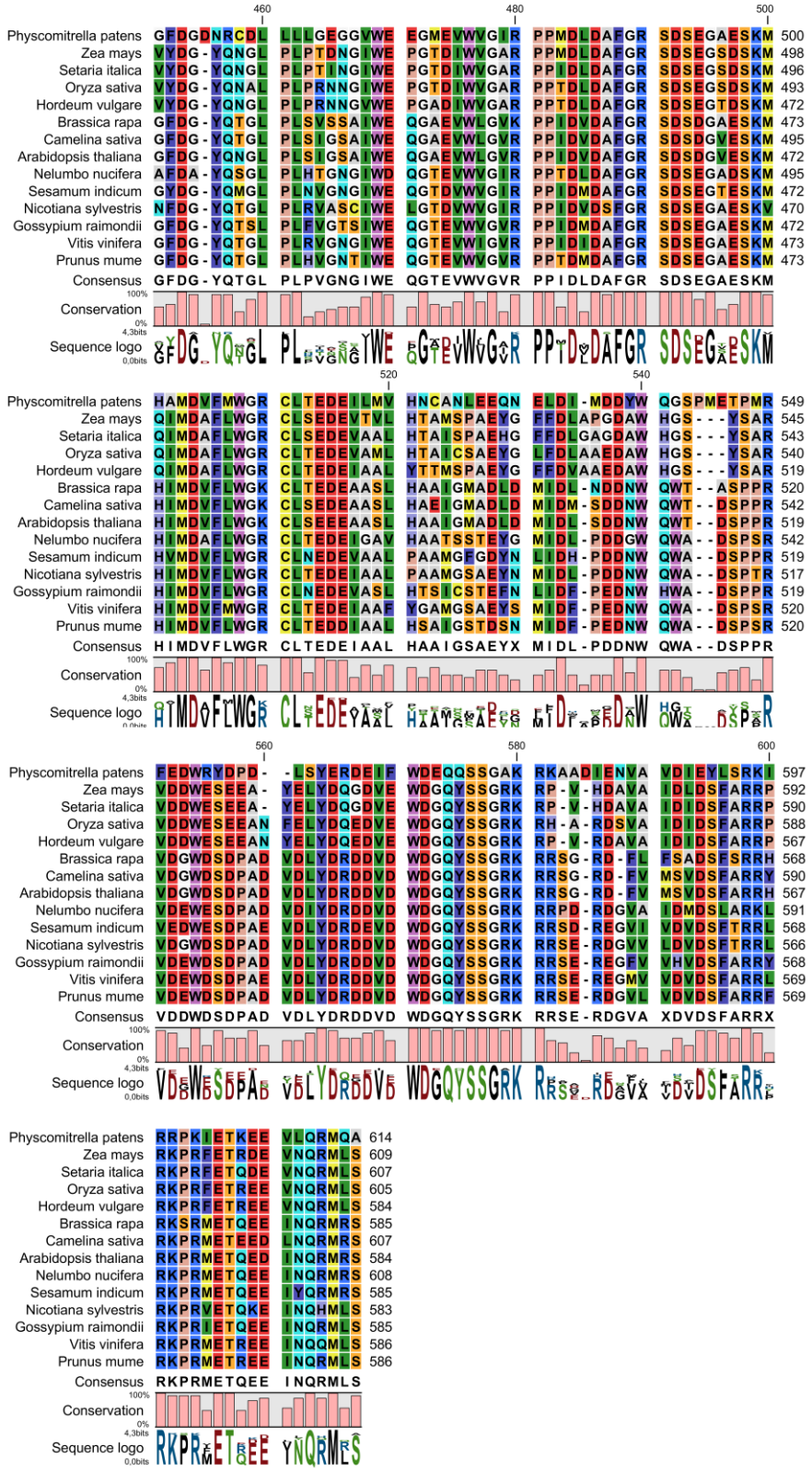


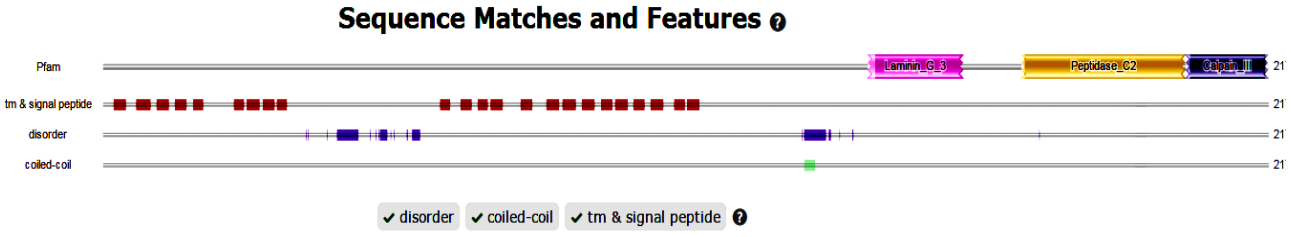
Figure 12: Multiple sequence alignment of the DEK1-Arm. Arm sequence was determined using SMART database based on pfam search, then multiple sequence alignment was performed in CLC Work Bench (using ClustalW)

The results from Figure 12 reveal high similarity between the *P. patens* DEK1-Arm and the Arm sequences from other plants. Highest percentage similarity of *P. patens* DEK1-Arm was observed in *Nelumbo nucifera* (64%), *Vitis vinifera* (61%) and *Sesamum indicum* (61%), while the lowest percentage similarity was observed in *Hordeum vulgare* (57%) and *Brassica rapa* (57%). The percentage similarity between all DEK1-Arm sequences used in this study is shown in Appendix 4A.

High sequence similarity suggests an important conserved function that the Arm provides to protein function.

3.1.2 Conserved domain in *P. patens* DEK1-Arm

Recent bioinformatics analysis of the DEK1-Arm shows that it contains a domain which is homologous to the LG3 domain (W. Johannsen, unpublished results). To verify this finding, the protein sequence of *P. patens* DEK1 was used as a query to perform the analysis using the hidden Markov models website server (HMMER) based on pfam database search. This analysis confirmed the presence of a domain with homology to the LG3 module (Figure 13).



am Matches — Advanced

Family		Clan	Description	Start	End	Domain E-values	
Id	Accession					Ind.	Cond.
Peptidase_C2	PF00648.16	CL0125 ✔	Calpain family cysteine protease	1715	2017	5.3e-76	1.4e-79
Calpain_III	PF01067.17	n/a	Calpain large subunit, domain III	2020	2170	1.6e-24	4.2e-28
Laminin_G_3 ▲	PF13385.1	CL0004 ✔	Concanavalin A-like lectin/glucanases superfamily	1429	1603	1.9e-05	5.0e-09
Laminin_G_2	PF02210.19	n/a	Laminin G domain	1474	1538	0.36	9.8e-05

ur search took:0.06 secs

Figure 13: P. patens DEK1-LG3 domain. The HMMER database was used to investigate the presence of LG3 domain in the P. patens DEK1 protein sequence.

In addition, the result from Figure 13 shows that the *P. patens* DEK1 protein contains 23 transmembrane segments, an LG3 domain (Accession number PF13385.1) with a length of 175 amino acids, the peptidase_C2 domain and the large calpain_III domain.

Another search, using the SMART database, confirmed the HMMER finding (Figure 14). This search was performed based on pfam and outlier homologous structure.

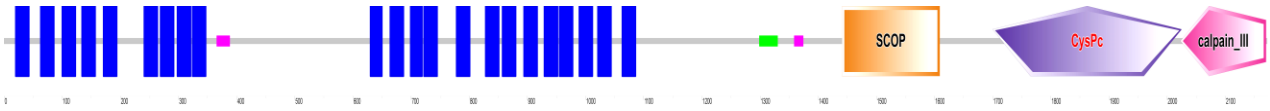


Figure 14: Domain analysis of the P. patens DEK1. The protein sequence of the DEK1 *was analyzed using SMART database, which present transmembrane domains (blue), coild coils structure (green), a SCOP domain, CysPc domain, and calpainIII domain.*

The SMART database search (Figure 14) shows a SCOP domain (d2sli_1) consisting of 168 amino acids and with an E-value of 5.00e-13. This domain was further investigated using the SCOP database and it showed that the domain belongs to the Concanavalin A-like lectins/glucanases domain. It is known that LG3 domains is a Concanavalin A-like lectins/glucanases family, thus both database searches identify a new conserved domain of the same family in DEK1.

The SCOPE and LG3 domain sequences were aligned together using the Clustal Omega database and the result is shown in Figure 15.

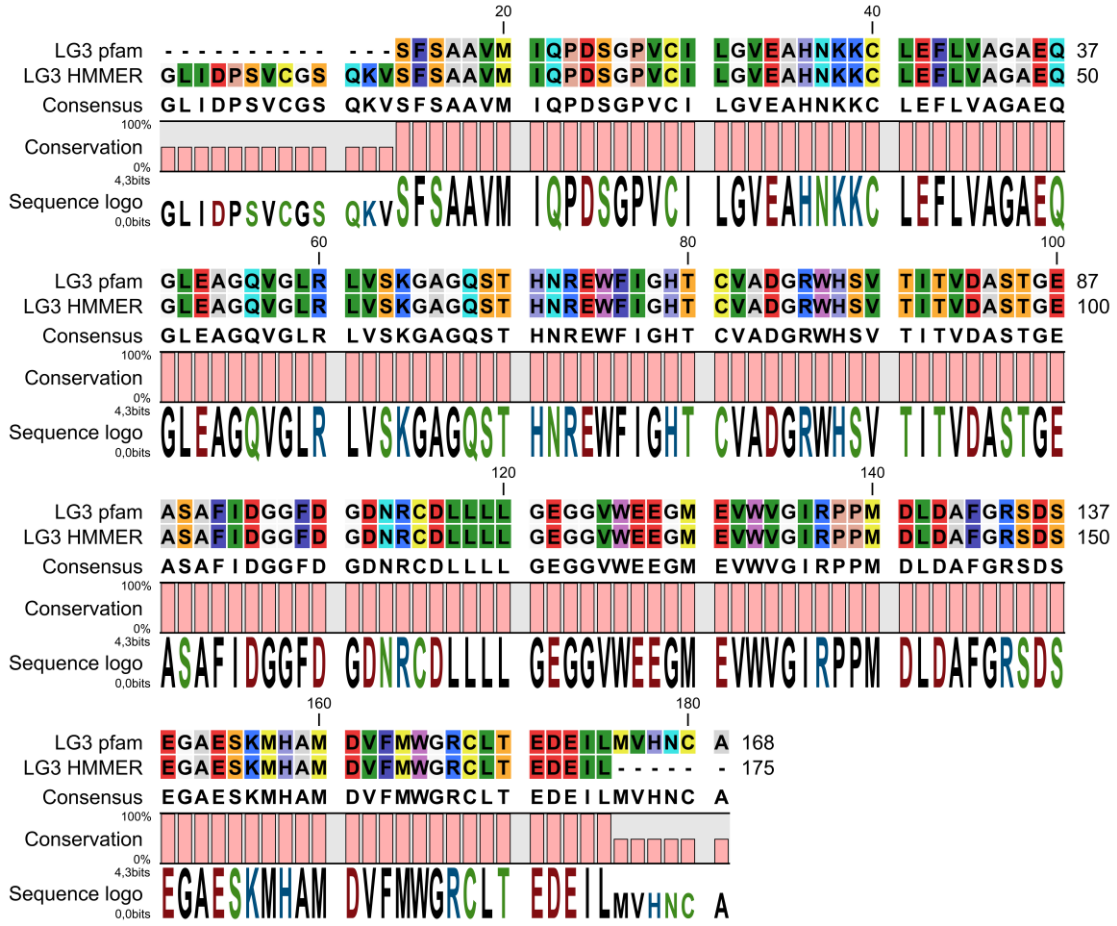


Figure 15: P. patens DEK1-LG3 domains. The two LG3 domains, from SMART and HMMER databases, were aligned together in CLC Work Bench (using ClustalW)

The alignment in Figure 15 shows that the LG3 domain protein sequence from HMMER is longer than the one in the SMART prediction server by 19 amino acid residues, otherwise they are identical. This difference can be due to variant statistical methods used in the two databases.

In order to reveal the conserved residues between the various plant DEK1-LG3 sequences, the DEK1-LG3 sequence from various plants were obtained from the SMART database and then a multiple sequence alignment (MSA) was performed in CLC Work Bench (using ClustalW) as shown in Figure 16.

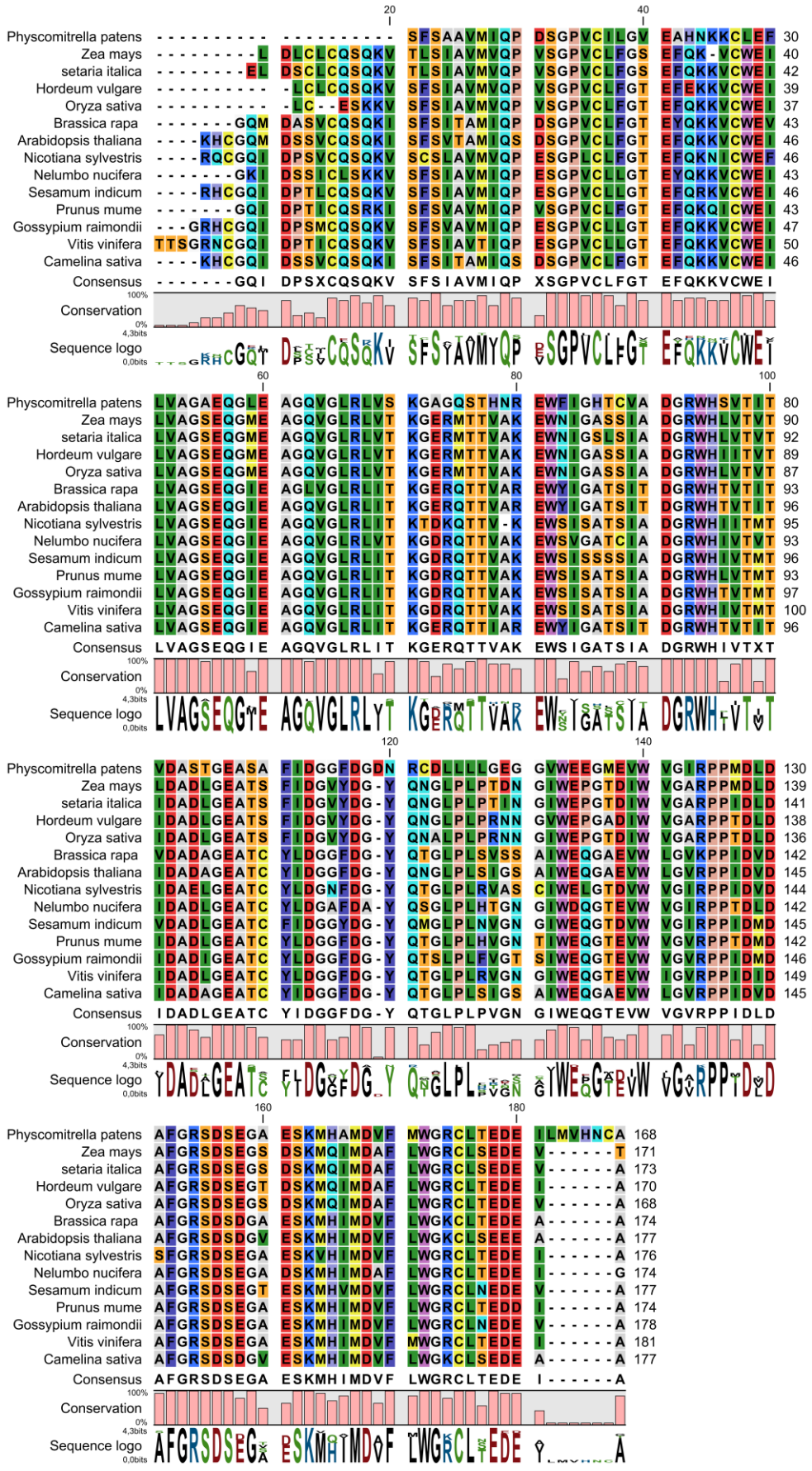


Figure 16: Multiple sequence alignment of DEK1- LG3 domain in CLC Work Bench (using ClustalW).

The result in Figure 16 reveals high similarity and conserved amino acids residues between *P. patens* DEK1-LG3 and the other lands plants. The percentage similarity between *P. patens* DEK1-ΔLG3 and other plants ranges from 60% (*Nelumbo nucifera*) to 54% (*Nicotiana sylvestris*). The percentage similarity between all DEK1-LG3 sequences used in this study is shown in Appendix 4B.

High conservation of the DEK1-LG3 domain suggests that it may perform an important function in the DEK1 protein.

3.2 Molecular characterization of potential LG3- mutant plant

3.2.1 First & second genotyping by PCR

A putative LG3 mutant plant displaying a phenotype different from the background line used for transformation was genotyped by PCR as outlined in section 2.5.2.1

First genotyping was performed using sense primer Arm_seq6 and anti-sense primer ASP_PpARM_inf1 aimed to amplify exons 21, 22 and 23 (Figure 17A), which is inside the targeting sequence, to confirm deletion of the LG3 domain (~670 bp). The expected size of both WT and LG3 mutant plants were 1327 bp and 645 bp, respectively, as shown in Figure 17A and B.

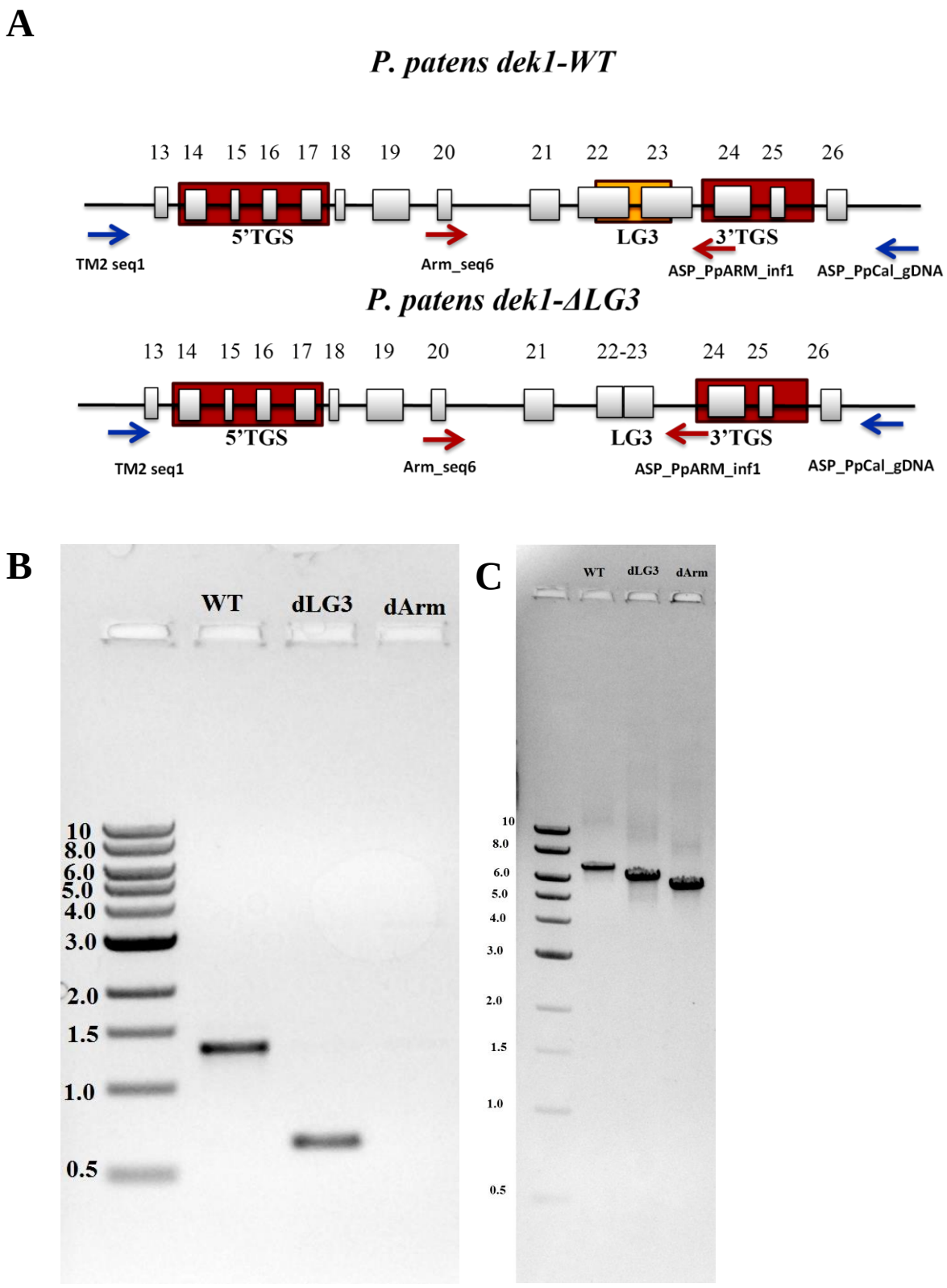


Figure 17: First & second genotyping of putative LG3 mutant plant. The genotyping was performed using Phire® Plant Direct PCR Kit where A) A sketch showing primers used for first (red arrows) and second (blue arrows) genotyping, B) First genotyping agarose gel, and C) second genotyping agarose gel.

The results from Figure 17B shows a PCR fragment between 1 kbp and 1.5 kbp for the WT plant as expected, as well as for the putative mutant plant which shows a PCR signal between 0.5 kbp and 1 Kb. The difference in size of the PCR products between the WT and the putative mutant are in agreement with deletion of the corresponding sequence of the LG3 domain from the putative mutant plant.

A second genotyping was performed in order to investigate in locus integration of both the 5` and 3` targeting sequences and to investigate the possibility of multi copy insertion of the DNA used for transformation. Second genotyping was performed using sense (TM2 seq1) and antisense (ASP_PpCal_gDNA) primers annealing outside the 5' and 3' targeting sequences, as shown in the Figure 17A. The expected PCR genotyping signal for WT, *dek1-Δlg3*, and *dek1-ΔArm* before cre (cre is a process whereby the antibiotic resistant marker is removed from the mutant plant) is approximately 6.8 kbp, 6.1 kbp, and 5.5 kbp, respectively.

The results in Figure 17C show a band between 6-8 kbp for the WT, a band at 6 kbp for the putative ΔLG3 mutant plant, and a band between 5-6 kbp for the ΔArm mutant plant, confirming at single copy integration at the *dek1* locus.

3.2.2 *dek1*-cDNA sequencing of putative ΔLG3 mutant plant

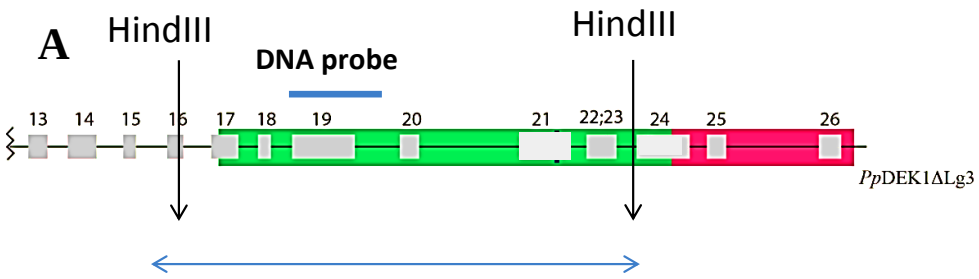
Sequencing of the *dek1* transcript produced by the putative ΔLG3 mutant plant was another approach in order to confirm correct splicing of the truncated *dek1* transcription. The procedure was performed as explained in section 2.5.2.3. The *dek1* cDNA that was sequenced was PCR amplified using primer pair Pp Loop_inverse and ASP_Pp CALP_cDNA, annealing to the target template outside the region used as targeting sequences (Figure 11D).

The sequencing result was analyzed and compared with wild type *P. patens dek1* cDNA, as a reference; using CLC Main Workbench v 6.9.

The sequencing confirmed deletion of the sequence corresponding to the LG3 domain and also revealed correct intron-exon splicing at the *dek1* locus (Data not shown).

3.2.3 Southern Blot:

Genomic DNA extracted from the mutant LG3 plant was analyzed with Southern blot to investigate the possibility of off-locus integration of the DNA during transformation. During this analysis, genomic DNA extracted from the P. patens WT was used as a positive control, while genomic DNA extracted from Δ*dek1* and *ΔArm* mutant plant was used as a negative control. The experiment was performed using the DNA probe shown in Figure 18.



Wt= 4465 bp

B *Δlg3* = 3793 bp

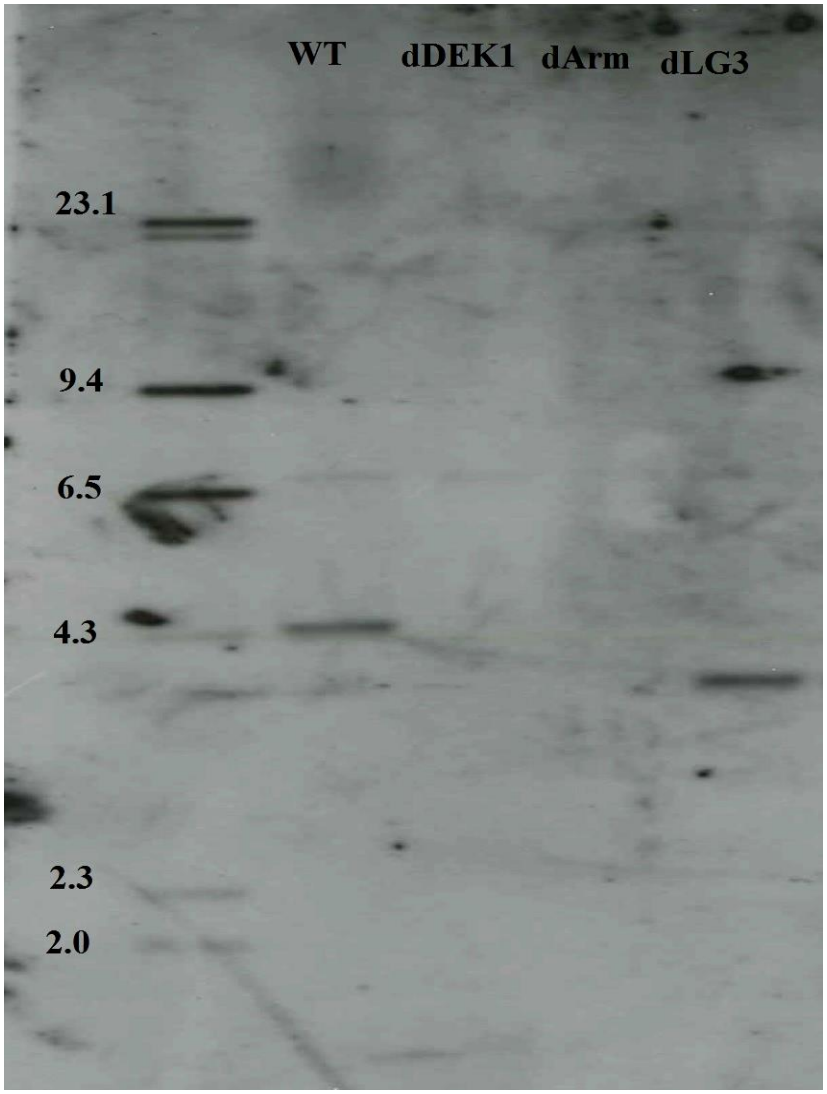


Figure 18: Southern blot analysis of WT, ΔDek1, ΔArm, and ΔLG3 plants. Genomic DNA were extracted from the plants and cut overnight with HindIII. A) A sketch showing the appropriate probe targeting exons 19 and 20, sites of the restriction enzymes, and the expected size of both WT and LG3 mutant plant. B) Agarose gel of southern blot.

The results from Figure 18 show a band near 4.3 kbp for the WT which is as expected (4465 bp), while the LG3 mutant plant show a band below 4.3 kbp where the expected band will be about 3.8 kbp. No other signal was detected confirming at locus single-integration of the DNA.

For the other mutant plants, *ΔArm* and *Δdek1*, the Figure 18 shows no band for both mutant plants which is as expected due to the missing arm sequence and the whole *dek1* gene respectively. However, a weak band can be seen at 6.5 kbp in all plants which could be due to unspecific binding of the probe.

3.2.4 Real time quantitative PCR (qPCR)

Quantitative PCR (qPCR) was performed to investigate whether there is any difference in the *dek1* transcription level between the WT and the mutant plant (*P. patens dek1-Δlg3* mutant).

Total RNA was isolated from 6 and 16 days old *P. patens* WT and *dek1-Δlg3* mutant plants. RNA isolation and cDNA synthesis was performed as described previously (section 2.5.2.3). The results are shown in Figure 19.

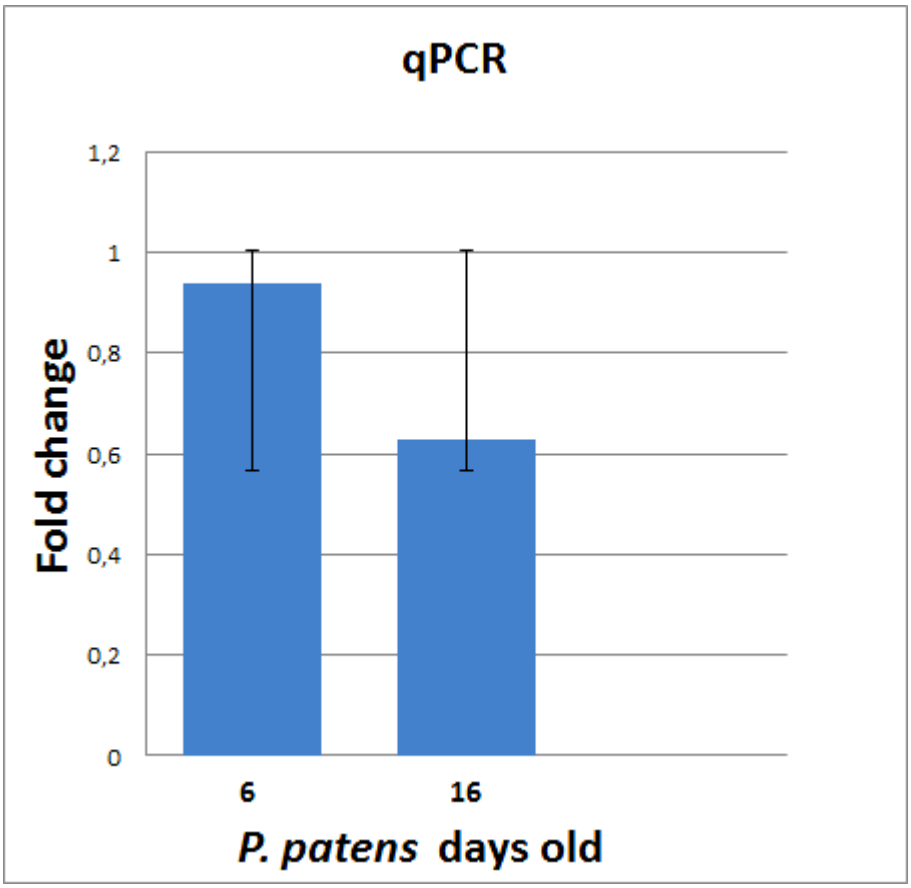


Figure 19: Quantitative PCR (qPCR) data for P. patens dek1-Δlg3 mutant plant. Total RNA was isolated from 6 and 16 days old P. patens - WT and dek1-Δlg3 mutant plants. RNA isolation and cDNA synthesis was performed and the data from the qPCR reaction was analysed using LinRegPCR version 2012.2 (J. M. Ruijter, Amsterdam, the Netherlands).

At 6 days, the fold change in the steady-state level of *dek1* transcript between WT and the *dek1-Δlg3* mutant were calculated to be 0.94, suggesting that there is no significant difference in the *dek1* expression level between the WT and the mutant plant.

On the other hand, at 16 days, the fold change in the steady-state level of *dek1* transcript between WT and the *dek1-Δlg3* mutant were calculated to be 0.63, suggesting that *dek1* transcript level is lower in the mutant than the WT plants.

3.3 Phenotype characterization:

The goal of this project was to investigate the function of the DEK1-LG3 domain by creating a *P. patens* mutant where the LG3 domain is deleted by homologous recombination. Studying the mutant plant phenotype is one technique used to achieve this goal.

Both *P. patens* WT and the mutant plant (*P. patens dek1-Δlg3*) were cultivated on BCD media and a dissecting microscope was used to characterize the difference in the phenotype between the WT and the mutant plants (Figure 20).

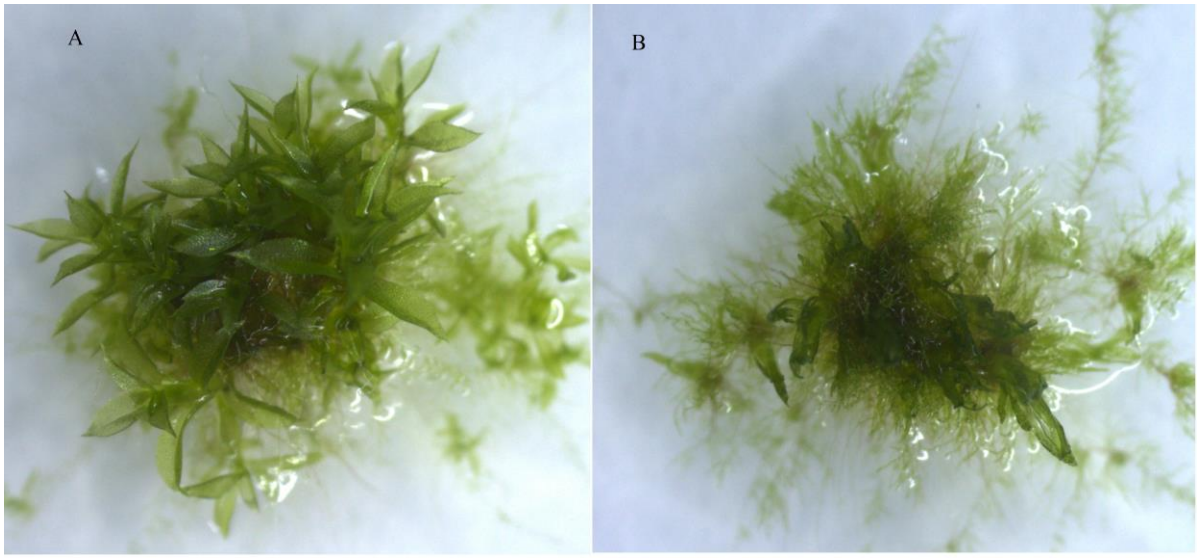


Figure 20: Phenotype of both WT (A) and ΔLG3 mutant plant (B). Both were cultivated on BCD media for three weeks.

From Figure 20, it can be observed that both WT and ΔLG3 mutant plant have gametophores. However, it is obvious that there are some differences between the two plants. The WT plant contains gametophores with large phyllids while the ΔLG3 mutant plant has gametophores with smaller phyllids.

The next step was to investigate the differences between the gametophores WT and ΔLG3 mutant plant. Both plants were cultivated in soil blocks for about one month before the gametophores were isolated carefully to reveal the differences as shown in Figure 21.

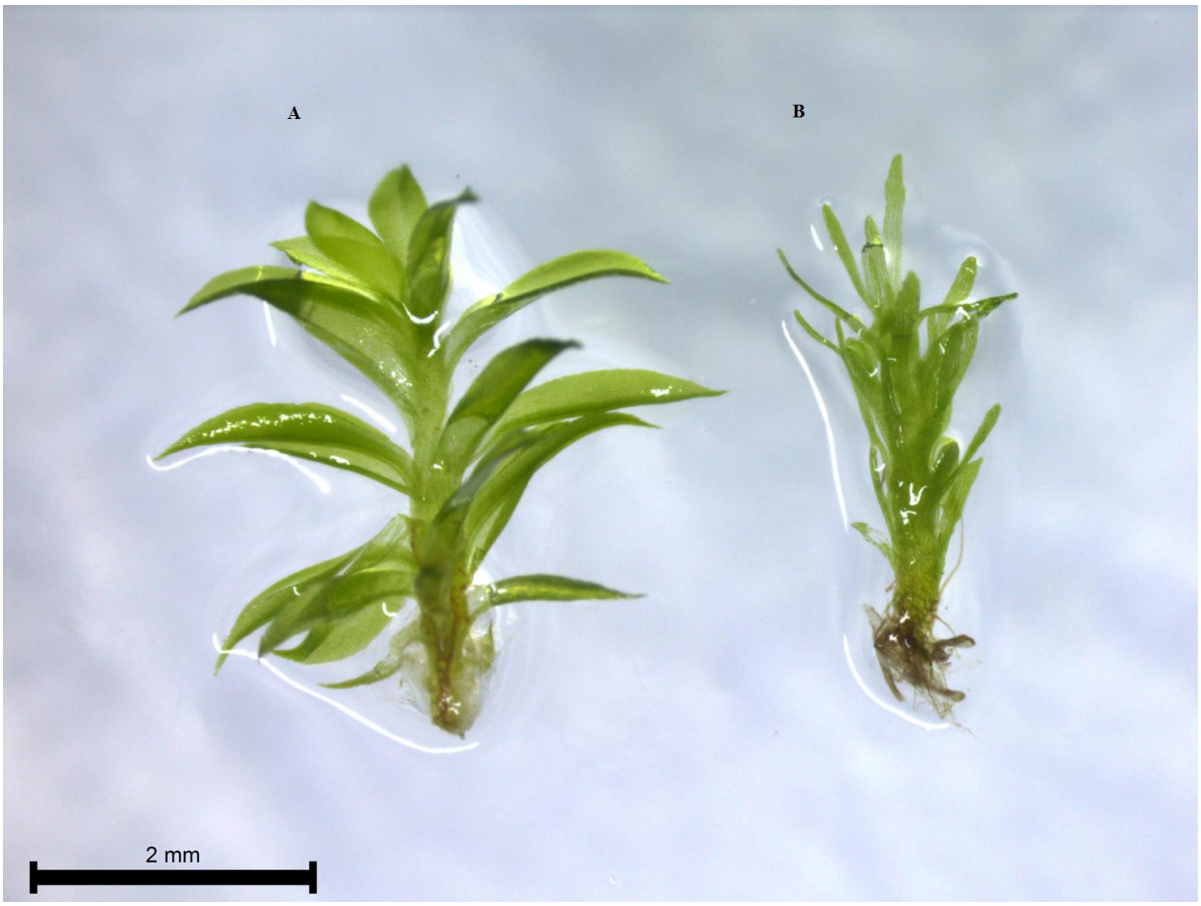


Figure 21: WT and ΔLG3 mutant plant gametophores. The gametophores were isolated form one month old WT plant (A) and ΔLG3 mutant plant (B) plants cultivated on soil blocks.

The WT gametophore shows larger and wider phyllids than ΔLG3 mutant plant, where the phyllids in the mutant plant look narrower and more compressed than the WT.

Further work on the phyllids was performed to explore the differences between the WT and ΔLG3 mutant plant. The phyllids were isolated very carefully from three different regions of the plants before the pictures were taken. The phyllids from both WT and ΔLG3 mutant plants are shown in Figure 22.

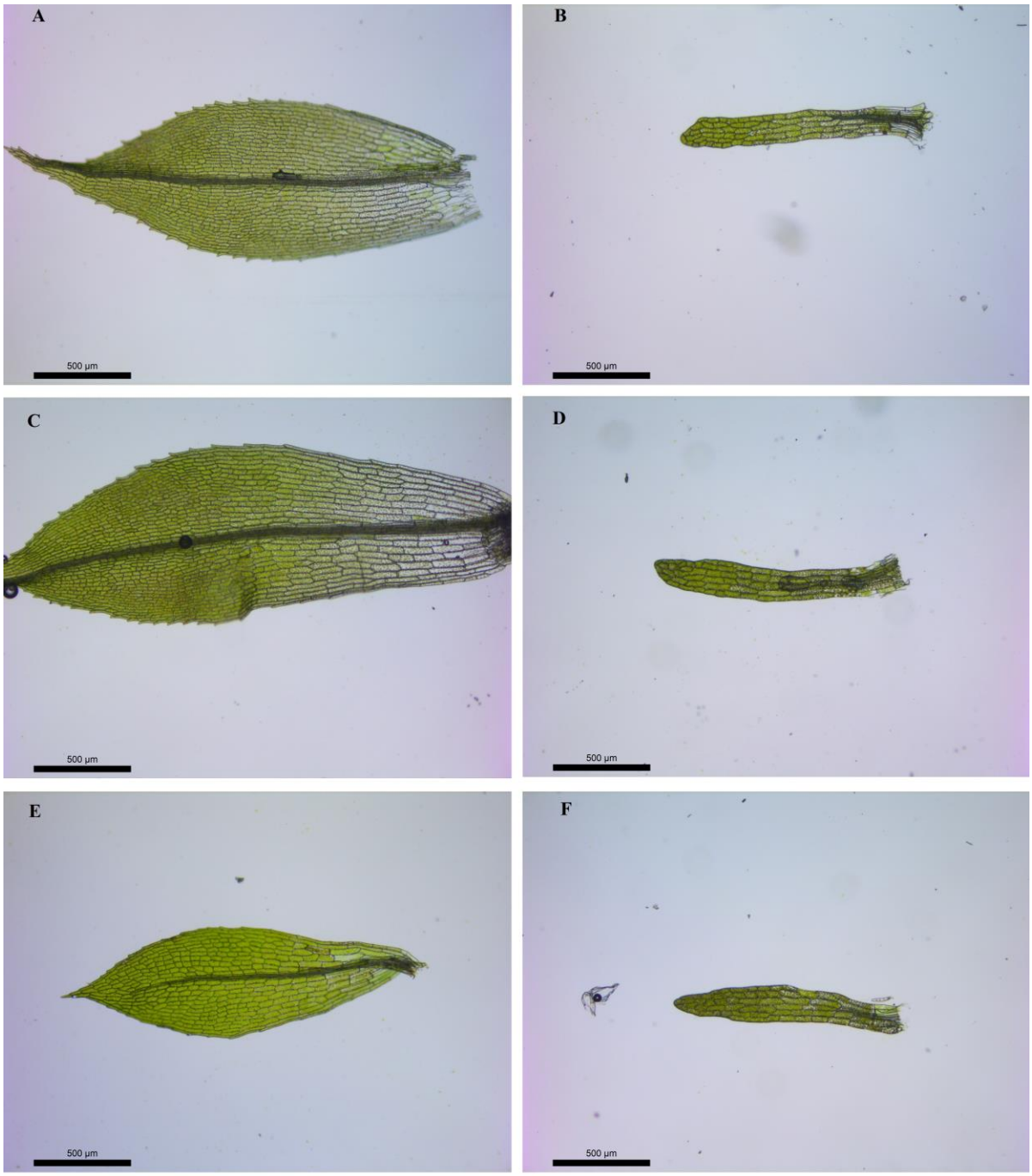


Figure 22: Phyllids from both WT and ΔLG3 mutant plant. The phyllids were isolated from both WT and ΔLG3 mutant plant from three different regions where A) WT phyllid from upper part, B) ΔLG3 mutant plant phyllid from upper part, C) WT phyllid from middle part, D) ΔLG3 mutant plant phyllid from middle part, E) WT phyllid from lower part, and F) ΔLG3 mutant plant phyllid from the lower part.

The results in Figure 22 revealed many differences between the WT and ΔLG3 mutant plant phyllids. The first clear difference is the size of the phyllid. The WT phyllids are larger and wider while the ΔLG3 mutant plant phyllids are narrower and smaller than the WT.

The cells of the phyllids show also difference between WT and ΔLG3 mutant plant. WT phyllids have smaller and more cells than ΔLG3 mutant plant phyllids which has large cells and fewer in number. Still it can be seen that WT phyllids has also large cells in the beginning of the phyllids especially the middle part phyllids (Figure 22C) but it becomes smaller and more during the phyllids growth. This may reveal a defect in the phyllid cell division in the LG3 mutant plant caused directly or indirectly by deletion of the LG3 domain.

It is obvious as well that the ΔLG3 mutant plant phyllids is missing the marginal serration, which is present in the WT and it has a blunt tip on the contrary to the WT which has sharp tip. The last apparent variance between WT and ΔLG3 mutant plant phyllids is the midrib. While the WT phyllids show long midrib in all phyllids from different parts, the ΔLG3 mutant plant phyllids show short midrib in all three phyllids (Figure 22B, D, and F).

Due to the difference between the gametophores in *P. Patens* WT and DEK1-ΔLG3 (Figure 22), the gametophores were further examined using confocal microscope to reveal any difference in the initiation steps between the WT and mutant plant (Figure 23), as described in section 2.5.3.

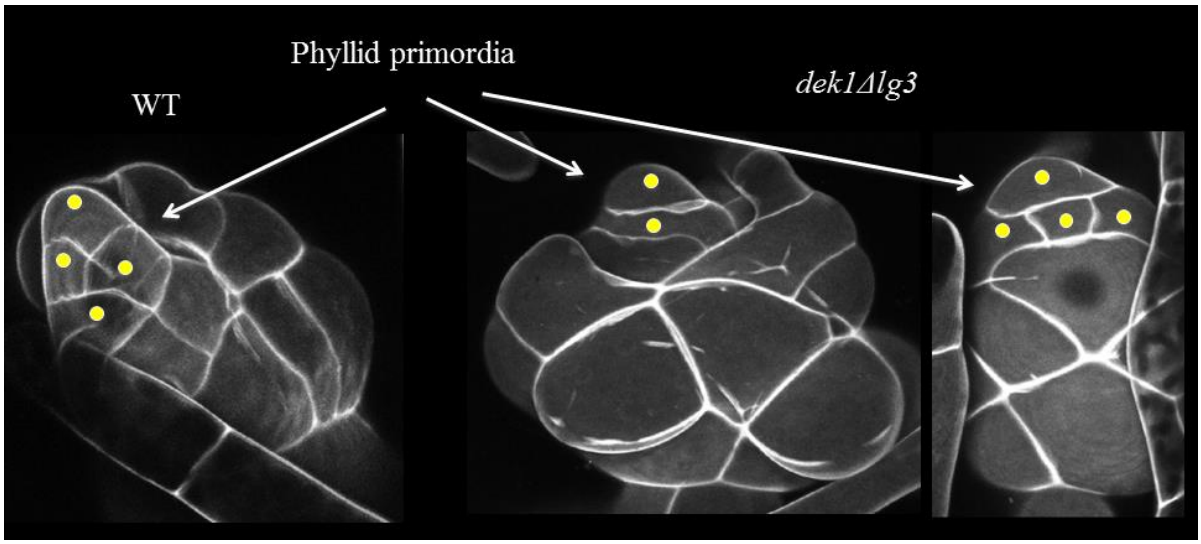


Figure 23: Bud development in P. patens WT and ΔLG3. Physcomitrella patens filaments containing later stages of bud development were stained with Propidium iodide and observed by confocal microscopy.

The earlier division stages of the *P. patens* DEK1-ΔLG3 bud occur as in WT plant (results not shown). Later stages in the gametophore formation in the mutant plant revealed differences with the WT. It can be seen, in Figure 23, that there are differences in in the cell division in the phyllid primordia between the mutant plant and WT. The mutant plant shows either less cells or different division pattern than in WT which is the reason that leads to different phenotype.

The *P. patens* DEK1-ΔLG3 transgenic line was cultivated on sterile soil blocks to investigate possible effect of the mutation (ΔLG3 domain) on the sporophytes (Figure 24).

Figure 24: P. patens sporophytes. A) WT sporophytes. B) P. patens DEK1-ΔLG3 sporophytes.

From figure 24, it can be seen that the sporophyte of *P. patens* DEK1-ΔLG3 has a growth defect and it is not fully mature as in the WT.

Further analysis of the *P. patens* DEK1-ΔLG3 gametangia revealed that the male gametangia (antheridia) did not show any phenotype deviation from the WT (data not shown). On the other hand, the female gametangia (archegonia) showed deviations in phenotype from the WT (Figure 25).

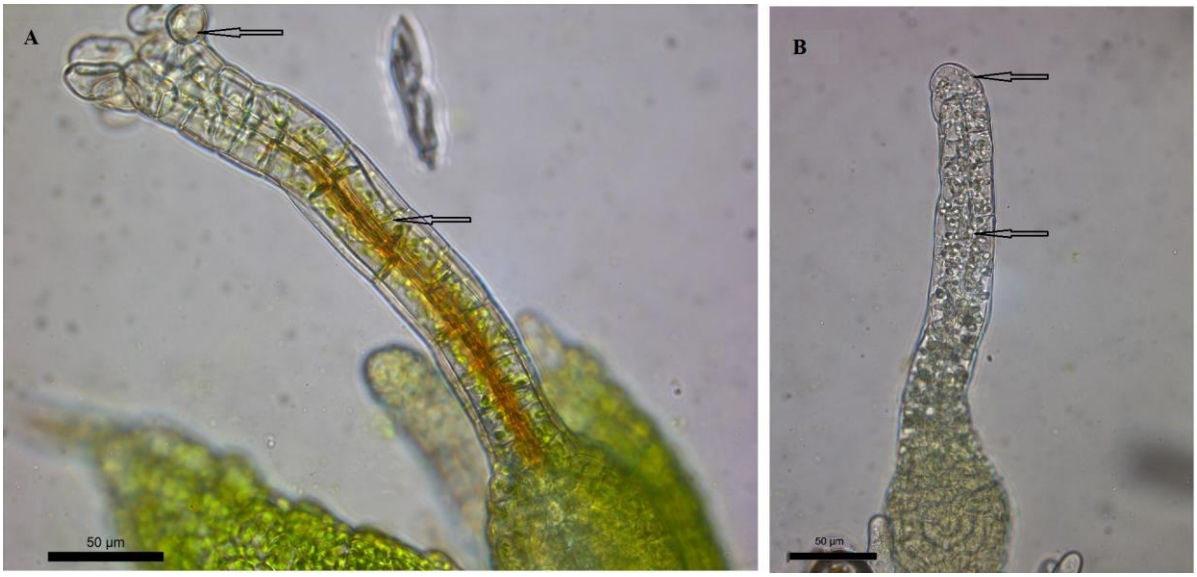


Figure 25: Female sporophytes (Archegonia) of P. patens. The figure shows the WT archegonia (A) and P. patens DEK1-ΔLG3 mutant archegonia (B).

Deletion of the LG3 domain in *P. patens* DEK1 caused failure in the opening of the apex, leaving the archegonia closed. Another effect can be seen in the egg canal, which seems to be absent in the LG3 mutant plant. The egg canal failed to form throughout the development due to division defect of the inner cells which is responsible for the formation of the canal in the *P. patens* archegonia. Thus as a consequence, the delta LG3 mutant is sterile and not able to produce the sporophyte as seen in Figure 25B.

4. Discussion:

4.1 Land plant DEK1-Arm segments are highly conserved and contain an domain with homology to the LG3 module:

The defective kernel 1 protein (DEK1), belongs to an ancient family of calpains harbouring large transmembrane (TML) domains, the TML-calpain family (Zhao et al., 2012). Previous studies have shown that DEK1 comprise 21-23 transmembrane segments, loop insertion between transmembrane segment 9 and 10, a non-structured cytoplasmic arm (DEK1-Arm, and a highly conserved calpain domain (CysPC-C2L) (Kim Leonie Johnson et al., 2008; Lid et al., 2002; Zhao et al., 2012). Sequence conservation analyses have revealed that DEK1 proteins are highly conserved within land plants, showing 70-98% amino acid sequence identity supporting the hypothesis that DEK1 has an important function in land plants (Liang et al., 2013; Wang et al., 2003).

In this study we show that the DEK1-Arm sequence is conserved in land plants as shown in Figure 12, the similarity varies from 57% to 64%. This conservation may indicate that the Arm segment harbours an important role in DEK1 function, thus it was important to preserve the DEK1-Arm sequence during plant evolution in order to maintain protein function. It has been reported that DEK1-Arm contains specific sites where DEK1 undergoes autolytic cleavage, leading to release of the active calpain domain (Kim Leonie Johnson et al., 2008). This suggests that the DEK1-Arm sequence is important to control the activation of DEK1.

An unpublished study has revealed that the DEK1-Arm segment of land plants contain a domain with homology to the laminin globular 3 domain (LG3) (W. Johansen, unpublished results). The bioinformatics search in this study confirms the presence a domain homology to the LG3 domain using both the HMMER and SMART conserved domain search servers. However, there was a difference in the length of the LG3 domain reported by these two servers despite the fact that both servers used pfam. For further investigation of the LG3 domain, we chose the sequence identified by HMMER because it has been reported that SMART may have problems to update the rapid expansion of the protein-sequence databases (Dickens & Ponting, 2003).

The results from the multiple sequence alignment of the DEK1-LG3 domain show that the domain is conserved within the land plants where the similarity ranges from 54% to 60%. LG3 domains in different proteins show only low sequence similarity, however these domains are structurally high conserved (Rudenko, Hohenester, & Muller, 2001). Whether the DEK1-LG3 structure is similar to other LG3 domains needs to be determined. However, preliminary results show that *P. patens* DEK1-LG3 domain, obtained by homology modelling, has a very similar structure to other LG3 domains (W. Johansen, unpublished results). This suggests that the DEK1-LG3 domain may have a function similar to that of other LG3 domains, which is to act as a reporter binding other proteins or ligand (Rudenko et al., 2001). Thus, it is tempting to speculate that DEK1-LG3 function as a binding site for DEK1 interactors of yet unknown identity.

4.2 Molecular characterization confirms the generation of the *P. patens dek1-Δlg3* mutant:

In order to confirm the deletion of the sequence corresponding to the LG3 domain in the putative *P. patens* mutant generated in this study, first and second genotyping were performed using genomic DNA extracted from the mutant as template. Both first and second genotyping showed PCR amplification signals of the expected size of the mutant (Figure 17), which confirm the deletion of the LG3 sequence from the locus. Sequencing of DEK1 cDNA also confirmed that the mutant correctly spliced the *dek1* transcript, generating a truncated calpain transcript without the *lg3* sequence. To verify that the mutant did not contain any off-locus DNA integrations, Southern Blot analysis was performed (Figure 18). In the *P.patens dek1-Δlg3* mutant, a single band of 3.8 kbp can be expected if the mutant only harbours DNA integration at the *dek1* locus, and this is what the Southern Blot analysis shows (Figure 18).

However, a signal above 6.5 kbp can also be observed, which should not be present. However, since this band is also observed in the *P. patens Δdek1* mutant sample, in which the entire *dek1* gene is deleted, we can conclude that this band is due to unspecific hybridization of the probe to the DNA bound on the membrane. There are many factors which can lead to unspecific binding during Southern Blot analysis.

Unspecific hybridization can arise despite the compatibility between the probe and the target. It is important that the specific probe-target is stable while other hybrids are unstable,

which will provide the stringency to the process. The major factors, which determine the stringency, are the buffer components and temperature. The buffer composition plays an important role in the specificity of the probe since the probe stability depends on the ionic strength and the presence of destabilizing agents. The temperature is also an important factor in the hybridization. For example if the probe is larger than 100 bp so it is preferable that the hybridization will perform at 68°C in high salt buffer. This will increase the stability of the probe-target interaction and remove any unspecific bindings (Brown, 2001).

In our case, the washing steps was performed at 65°C and later it was detected that the temperature was lower than 65°C due to inaccuracy of the incubator. According to the discussion above, we could maybe erase the unspecific signal by increasing the washing temperature to 68°C especially since our probe was 1.3 kbp in length. This could increase the specificity of our probe and remove the unspecific signal from the Southern Blot.

Real-time PCR was performed in order to investigate if there is any difference in the *dek1* transcript level between WT and the *dek1-Δlg3* mutant plants. As shown in Figure 19, there was no significant difference in the expression level of *dek1* transcript between the two lines grown for 6 days. On the other hand, there was a significant difference for 16 days old plant. Still the high standard deviation at 16 days old plant shows that the data is widely spread which make the result less reliable. Hence the results from 16 days old plants can be due to unprecise in the work since unpublished results show that there is no significant difference in the transcription level of *dek1* transcript between the *P. patens* WT and *P. patens dek1-Δlg3* mutant (W. Johansen, unpublished results).

However, the transcription level can not give us precise conclusion concerning the protein level in the plant since mRNA level may not indicate the correct protein level (Kendrick, 2014). In order to achieve this, it is important and necessary to perform western blotting on both WT and LG3 mutant plants to make final conclusion about DEK1 protein level. So far, every attempt to detect the *P. patens* DEK1 protein from native locus has been unsuccessful, probably because of low protein abundance and/or instability of the protein (W. Johansen, pers. comm.).

4.3 The *P. patens dek1-Δlg3* mutant is affected in phyllid and archegonia development

The mutant plants generated in this study, *P. patens dek1-Δlg3*, has a different phenotype than wild type plant as shown in Figures 20, 21, and 22.

Generally, the final shape and size of the leaf depends equally on cell division and cell expansion where timing, frequency, duration, and orientation must be regulated (Poethig & Sussex, 1985). In our study, deleting the LG3 domain from *P. patens* DEK1 led to the formation of gametophore with development defects (Figures 20, 21, 22, 24).

Deletion or partial deletion of *dek1* in *P. patens* give rise to mutant plants with different phenotypic characteristics (Demko et al., 2014; Perroud et al., 2014). Studies have reported that auxin plays important role in the development of *P. patens* and other plants (N. Ashton, Cove, & Featherstone, 1979; N. W. Ashton, 1998; D. Cove, 1984; De Smet & Jürgens, 2007; Geldner et al., 2003). Chopra et al. was the first who reported that bud formation in *P. patens* depends on the cytokinin/auxin ratio (Chopra & Rashid, 1969). It was also reported that auxin is involved in cell division, cell elongation, and cell differentiation (Macdonald, 1997). Auxin is used in long-range signaling in the communication between cells in plant and it has widely-known role in plant development (Bhalerao & Bennett, 2003). Auxin is also responsible for the transition from chloronema to caulonema differentiation during protonema development. Mosses treated with auxin have increased numbers of filaments that develop as caulonema and divers auxin mutant develop chloronema alone with little or no caulonema (Johri & DESAI, 1973; Prigge, Lavy, Ashton, & Estelle, 2010).

Interestingly, incubation of WT *P. patens* with 2.5µM of synthetic auxin (Beta-naphthalene acetic acid, beta-NAA) gives rise to gametophore with similar phenotypes (Barker & Ashton, 2013) to those observed in our mutant (Figure 21B, D, and F). In this study, the young leaves in WT *P. patens* which treated with auxin were long, narrow, composed of elongated cells and lacked marginal serration.

From all these facts, we can hypothesize that there might be a type of regulation between DEK1 and auxin in *P. patens*. This is also supported by the observation that DEK1 affects microtubule-associated protein 65 (MAP65), CLIP-associated protein (CLASP), including auxin-related genes which control cell division planes, cell wall orientation, and epidermal cell identity. This makes DEK1 the first initial positional sensor in cell division by regulating

the activity of other proteins (Dhonukshe et al., 2012; Liang, Brown, Fletcher, & Opsahl-Sorteberg, 2015).

Another link between DEK1 and auxin can also be identified in gametangia development. In *P. patens*, both male and female reproductive organs form from the tip of the gametophore shoot due to the effect of low temperature and short daylight (Hohe, Rensing, Mildner, Lang, & Reski, 2002). Auxin was also investigated if it contributes to male and female gametangia development in *P. patens* (Landberg et al., 2013). It was found that the SHORT INTERNODE/STYLISH (SHI/STY) family gene regulates auxin biosynthesis in both *P. patens* and *Arabidopsis* where the SHI/STY transcription factors in the *Arabidopsis* bind directly to the promoters of YUC auxin biosynthesis genes and activate their expression (Eklund et al., 2010b; Eklund et al., 2010a; Sohlberg et al., 2006; Ståldal, Sohlberg, Eklund, Ljung, & Sundberg, 2008).

In the Landberg et al. study, the effect of mutating the different *shi* genes (*shi1-1*, *shi2-1*, and *shi2-2*) on the reproductive system of *P. patens* was investigated. In their study they revealed that deletion of *Ppshi2-1* and *Ppshi2-2* genes prevented the opening of antheridia apex. However, they also showed that *shi1-1*, *shi2-1*, and *shi2-2* mutant also affect the archegonia development in *P. patens*. The archegonia failed to develop the egg canal and failed to form the apex opening. Briefly, both *Ppshi1* and *Ppshi2* affect the development of the *P. patens* archegonia, where the abnormality can be seen in the egg canal formation and apex opening (Landberg et al., 2013). This abnormality is very similar to the phenotype we observed in our *P. patens dek1-Δlg3* mutant (Figure 25B) suggesting that the *P. patens* DEK1 protein is involved in regulating *Ppshi1* expression or auxin biosynthesis, either directly or indirectly, and that the LG3 domain may have an important function in this process.

5. Conclusion:

In this study, the aim was to investigate the effect of deleting the LG3 domain from *P. patens* DEK1 protein in order to obtain functional information of this domain in the DEK1 protein. To achieve this, the plasmid *pBHRF_PpArm_ΔLG3*, was constructed, harbouring the *P. patens dek1 Arm-Δlg3* sequence. The plasmid was transferred to the *P. patens dek1-ΔArm* mutant plants using the PEG-mediate protoplast transformation method according to Cove et al (David J Cove et al., 2009).

The *P. patens dek1-Δlg3* mutant plants have gametophore development defects; phyllids were long, narrow, composed of elongated cells, lacked marginal serration, and show short midrib. An effect could also be observed in the gametangia, where the mutant plant showed defect in the archegonia development, specifically failure to form the egg canal formation and apex opening.

The conclusion in this study is that deletion of LG3 domain in the *P. patens* DEK1 may have an effect or regulates either the *Ppshi1* or auxin biosynthesis directly or indirectly. This failure to regulate the auxin concentration, led to abnormalities in both gametophore and gametangia organs in the mutant plant.

6. Further work:

In this study, qPCR analysis showed no significant difference in the *dek1* transcript level between WT and *P. Patens dek1-Δlg3* at 6 days, while at 16 days the mutant showed lower transcript level. However, the transcription level can not give us precise conclusion concerning the protein level in the plant. Thus, it important first to optimize a method to measure DEK1 protein concentration in both WT and LG3 mutant plants to investigate whether there is a difference in protein concentration or not between both plants.

The bioinformatics work showed high similarity and conserved amino acids residues between *P. patens* DEK1-LG3 and the other lands plants. In order to confirm the *dek1 lg3* homology, we can complement the *P. patens dek1-Δlg3* mutant from other plants lg3 domain like *A. thaliana* or *N. nucifera* and observe what if it possible to recover the *P. patens* WT phenotype.

Literature

Ahn, J. W., Kim, M., Lim, J. H., Kim, G. T., & Pai, H. S. (2004). Phytocalpain controls the proliferation and differentiation fates of cells in plant organ development. *The Plant Journal, 38*(6), 969-981.

Ashton, N., Cove, D., & Featherstone, D. (1979). The isolation and physiological analysis of mutants of the moss, Physcomitrella patens, which over-produce gametophores. *Planta, 144*(5), 437-442.

Ashton, N., Grimsley, N., & Cove, D. (1979). Analysis of gametophytic development in the moss, Physcomitrella patens, using auxin and cytokinin resistant mutants. *Planta, 144*(5), 427-435.

Ashton, N. W. (1998). Auxin structure-activity relationships in Physcomitrella patens. *Bates J. W., Ashton NW & Duckett JG (eds.), Bryology for the twenty-first century*, 175-188.

Barker, E. I., & Ashton, N. W. (2013). Heteroblasty in the moss, Aphanoregma patens (Physcomitrella patens), results from progressive modulation of a single fundamental leaf developmental programme. *Journal of Bryology, 35*(3), 185-196.

Beckmann, G., Hanke, J., Bork, P., & Reich, J. G. (1998). Merging extracellular domains: fold prediction for laminin G-like and amino-terminal thrombospondin-like modules based on homology to pentraxins. *Journal of molecular biology, 275*(5), 725-730.

Becraft, P. W., & Asuncion-Crabb, Y. (2000). Positional cues specify and maintain aleurone cell fate in maize endosperm development. *Development, 127*(18), 4039-4048.

Becraft, P. W., Li, K., Dey, N., & Asuncion-Crabb, Y. (2002). The maize dek1 gene functions in embryonic pattern formation and cell fate specification. *Development, 129*(22), 5217-5225.

Becraft, P. W., Stinard, P. S., & McCarty, D. R. (1996). CRINKLY4: a TNFR-like receptor kinase involved in maize epidermal differentiation. *Science, 273*(5280), 1406-1409.

Bhalerao, R. P., & Bennett, M. J. (2003). The case for morphogens in plants. *Nature cell biology, 5*(11), 939-943.

Brown, T. A. (2001). Southern Blotting and Related DNA Detection Techniques. *eLS*.

Campbell, N. A., & Reece, J. B. (2009). *Biology* (8th ed.). San Francisco: Pearson Benjamin Cummings.

Chopra, R., & Rashid, A. (1969). Auxin-cytokinin interaction in shoot-bud formation of a moss: Anoectangium thomsonii mitt. *Zeitschrift fur Pflanzenphysiologie*.

Cove, D. (1984). role of cytokinin and auxin in protonemal development in Physcomitrella patens and Physcomitrium sphaericum. *Journal of the Hattori Botanical Laboratory*.

Cove, D. J., & Knight, C. D. (1993). The Moss Physcomitrella patens, a Model System with Potential for the Study of Plant Reproduction. *Plant Cell, 5*(10), 1483-1488. doi: 10.1105/tpc.5.10.1483

Cove, D. J., Perroud, P.-F., Charron, A. J., McDaniel, S. F., Khandelwal, A., & Quatrano, R. S. (2009). The moss Physcomitrella patens: a novel model system for plant development and genomic studies. *Cold Spring Harbor Protocols, 2009*(2), pdb. emo115.

De Smet, I., & Jürgens, G. (2007). Patterning the axis in plants–auxin in control. *Current opinion in genetics & development, 17*(4), 337-343.

Demko, V., Perroud, P.-F., Johansen, W., Delwiche, C. F., Cooper, E. D., Remme, P., . . . Quatrano, R. (2014). Genetic Analysis of DEFECTIVE KERNEL1 Loop Function in

Three-Dimensional Body Patterning in Physcomitrella patens. *Plant Physiology, 166*(2), 903-919.

Dhonukshe, P., Weits, D. A., Cruz-Ramirez, A., Deinum, E. E., Tindemans, S. H., Kakar, K., . . . Sasabe, M. (2012). RETRACTED: A PLETHORA-Auxin Transcription Module Controls Cell Division Plane Rotation through MAP65 and CLASP. *Cell, 149*(2), 383-396.

Dickens, N. J., & Ponting, C. P. (2003). THoR: a tool for domain discovery and curation of multiple alignments. *Genome Biol, 4*(8), R52.

Eklund, D. M., Ståldal, V., Valsecchi, I., Cierlik, I., Eriksson, C., Hiratsu, K., . . . Ezcurra, I. (2010b). The Arabidopsis thaliana STYLISH1 protein acts as a transcriptional activator regulating auxin biosynthesis. *Plant Cell, 22*(2), 349-363.

Eklund, D. M., Thelander, M., Landberg, K., Ståldal, V., Nilsson, A., Johansson, M., . . . Ljung, K. (2010a). Homologues of the Arabidopsis thaliana SHI/STY/LRP1 genes control auxin biosynthesis and affect growth and development in the moss Physcomitrella patens. *Development, 137*(8), 1275-1284.

Geldner, N., Anders, N., Wolters, H., Keicher, J., Kornberger, W., Muller, P., . . . Jürgens, G. (2003). The Arabidopsis GNOM ARF-GEF mediates endosomal recycling, auxin transport, and auxin-dependent plant growth. *Cell, 112*(2), 219-230.

Heckman, D. S., Geiser, D. M., Eidell, B. R., Stauffer, R. L., Kardos, N. L., & Hedges, S. B. (2001). Molecular evidence for the early colonization of land by fungi and plants. *Science, 293*(5532), 1129-1133.

Hibara, K.-i., Obara, M., Hayashida, E., Abe, M., Ishimaru, T., Satoh, H., . . . Nagato, Y. (2009). The ADAXIALIZED LEAF1 gene functions in leaf and embryonic pattern formation in rice. *Developmental biology, 334*(2), 345-354.

Hohe, A., Rensing, S., Mildner, M., Lang, D., & Reski, R. (2002). Day Length and Temperature Strongly Influence Sexual Reproduction and Expression of a Novel MADS-Box Gene in the Moss Physcomitrella patens. *Plant Biology, 4*(5), 595-602.

Howard, T. L., Stauffer, D. R., Degnin, C. R., & Hollenberg, S. M. (2001). CHMP1 functions as a member of a newly defined family of vesicle trafficking proteins. *Journal of cell science, 114*(13), 2395-2404.

Johnson, K. L., Degnan, K. A., Ross Walker, J., & Ingram, G. C. (2005). AtDEK1 is essential for specification of embryonic epidermal cell fate. *The Plant Journal, 44*(1), 114-127.

Johnson, K. L., Faulkner, C., Jeffree, C. E., & Ingram, G. C. (2008). The phytocalpain defective kernel 1 is a novel Arabidopsis growth regulator whose activity is regulated by proteolytic processing. *The Plant Cell Online, 20*(10), 2619-2630.

Johri, M., & DESAI, S. (1973). Auxin regulation of caulonema formation in moss protonema. *Nature, 245*(146), 223-224.

Joseph, D., & Baker, M. (1992). Sex hormone-binding globulin, androgen-binding protein, and vitamin K-dependent protein S are homologous to laminin A, merosin, and Drosophila crumbs protein. *The FASEB journal, 6*(7), 2477-2481.

Kamisugi, Y., Schlink, K., Rensing, S. A., Schween, G., von Stackelberg, M., Cuming, A. C., . . . Cove, D. J. (2006). The mechanism of gene targeting in Physcomitrella patens: homologous recombination, concatenation and multiple integration. *Nucleic acids research, 34*(21), 6205-6214.

Kendrick, N. (2014). A gene's mRNA level does not usually predict its protein level. http://www.kendricklabs.com/WP1_mRNAvsProtein-New2014.pdf

Kumar, S. B., Venkateswaran, K., & Kundu, S. (2013). Alternative conformational model of a seed protein DeK1 for better understanding of structure-function relationship. *Journal of Proteins & Proteomics, 1*(2).

Kuwayama, H. (2012). Enhancement of Homologous Recombination Efficiency by Homologous Oligonucleotides.

Landberg, K., Pederson, E. R., Viaene, T., Bozorg, B., Friml, J., Jönsson, H., . . . Sundberg, E. (2013). The moss Physcomitrella patens reproductive organ development is highly organized, affected by the two SHI/STY genes and by the level of active auxin in the SHI/STY expression domain. *Plant Physiology, 162*(3), 1406-1419.

Li, X., & Heyer, W.-D. (2008). Homologous recombination in DNA repair and DNA damage tolerance. *Cell research, 18*(1), 99-113.

Liang, Z., Brown, R. C., Fletcher, J. C., & Opsahl-Sorteberg, H.-G. (2015). Calpain-mediated positional information directs cell wall orientation to sustain plant stem cell activity, growth and development. *Plant and Cell Physiology, 56*(9), 1855-1866.

Liang, Z., Demko, V., Wilson, R. C., Johnson, K. A., Ahmad, R., Perroud, P. F., . . . Otegui, M. S. (2013). The catalytic domain CysPc of the DEK1 calpain is functionally conserved in land plants. *The Plant Journal, 75*(5), 742-754.

Lid, S. E., Gruis, D., Jung, R., Lorentzen, J. A., Ananiev, E., Chamberlin, M., . . . Olsen, O.-A. (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. *Proceedings of the National Academy of Sciences, 99*(8), 5460-5465.

Lid, S. E., Olsen, L., Nestestog, R., Aukerman, M., Brown, R. C., Lemmon, B., . . . Olsen, O.-A. (2005). Mutation in the Arabidopisis thaliana DEK1 calpain gene perturbs endosperm and embryo development while over-expression affects organ development globally. *Planta, 221*(3), 339-351.

Macdonald, H. (1997). Auxin perception and signal transduction. *Physiologia Plantarum, 100*(3), 423-430.

Markmann-Mulisch, U., Hadi, M. Z., Koepchen, K., Alonso, J. C., Russo, V. E., Schell, J., & Reiss, B. (2002). The organization of Physcomitrella patens RAD51 genes is unique among eukaryotic organisms. *Proceedings of the National Academy of Sciences, 99*(5), 2959-2964.

Nickerson, D. P., West, M., & Odorizzi, G. (2006). Did2 coordinates Vps4-mediated dissociation of ESCRT-III from endosomes. *The Journal of cell biology, 175*(5), 715-720.

Ono, Y., & Sorimachi, H. (2012). Calpains—An elaborate proteolytic system. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1824*(1), 224-236.

Patthy, L. (1992). A family of laminin-related proteins controlling ectodermal differentiation in Drosophila. *FEBS letters, 298*(2), 182-184.

Perroud, P. F., Demko, V., Johansen, W., Wilson, R. C., Olsen, O. A., & Quatrano, R. S. (2014). Defective Kernel 1 (DEK1) is required for three-dimensional growth in Physcomitrella patens. *New Phytologist*.

Platt, A. R., Woodhall, R. W., & George, A. L. (2007). Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *BioTechniques, 43*(1), 58.

Poethig, R., & Sussex, I. (1985). The cellular parameters of leaf development in tobacco: a clonal analysis. *Planta, 165*(2), 170-184.

Prigge, M. J., Lavy, M., Ashton, N. W., & Estelle, M. (2010). Physcomitrella patens auxin-resistant mutants affect conserved elements of an auxin-signaling pathway. *Current Biology, 20*(21), 1907-1912.

Ramakers, C., Ruijter, J. M., Deprez, R. H. L., & Moorman, A. F. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience letters, 339*(1), 62-66.

Reski, R., & Frank, W. (2005). Moss (Physcomitrella patens) functional genomics—gene discovery and tool development, with implications for crop plants and human health. *Briefings in functional genomics & proteomics, 4*(1), 48-57.

Rudenko, G., Hohenester, E., & Muller, Y. A. (2001). LG/LNS domains: multiple functions–one business end? *Trends in biochemical sciences, 26*(6), 363-368.

Ruijter, J., Ramakers, C., Hoogaars, W., Karlen, Y., Bakker, O., Van den Hoff, M., & Moorman, A. (2009). Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic acids research, 37*(6), e45-e45.

Sasaki, T., Costell, M., Mann, K., & Timpl, R. (1998). Inhibition of glycosaminoglycan modification of perlecan domain I by site-directed mutagenesis changes protease sensitivity and laminin-1 binding activity. *FEBS letters, 435*(2), 169-172.

Sasaki, T., Fässler, R., & Hohenester, E. (2004). Laminin the crux of basement membrane assembly. *The Journal of cell biology, 164*(7), 959-963.

Sato, K., & Kawashima, S. (2001). Calpain function in the modulation of signal transduction molecules. *Biological chemistry, 382*(5), 743-751.

Schaefer, D. G., & Zrÿd, J.-P. (2001). The moss Physcomitrella patens, now and then. *Plant Physiology, 127*(4), 1430-1438.

Schaefer, D. G., & Zrÿd, J. P. (1997). Efficient gene targeting in the moss Physcomitrella patens. *The Plant Journal, 11*(6), 1195-1206.

Schumaker, K. S., & Dietrich, M. A. (1998). Hormone-induced signaling during moss development. *Annual review of plant biology, 49*(1), 501-523.

Shen, B., Li, C., Min, Z., Meeley, R. B., Tarczynski, M. C., & Olsen, O.-A. (2003). sal1 determines the number of aleurone cell layers in maize endosperm and encodes a class E vacuolar sorting protein. *Proceedings of the National Academy of Sciences, 100*(11), 6552-6557.

Sohlberg, J. J., Myrenås, M., Kuusk, S., Lagercrantz, U., Kowalczyk, M., Sandberg, G., & Sundberg, E. (2006). STY1 regulates auxin homeostasis and affects apical–basal patterning of the Arabidopsis gynoecium. *The Plant Journal, 47*(1), 112-123.

Ståldal, V., Sohlberg, J. J., Eklund, D. M., Ljung, K., & Sundberg, E. (2008). Auxin can act independently of CRC, LUG, SEU, SPT and STY1 in style development but not apical-basal patterning of the Arabidopsis gynoecium. *New Phytologist, 180*(4), 798-808.

Tian, Q., Olsen, L., Sun, B., Lid, S. E., Brown, R. C., Lemmon, B. E., . . . Otegui, M. S. (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. *The Plant Cell Online, 19*(10), 3127-3145.

Tierney, M., & Lamour, K. (2005). An introduction to reverse genetic tools for investigating gene function. *The Plant Health Instructor*.

Tisi, D., Talts, J. F., Timpl, R., & Hohenester, E. (2000). Structure of the C-terminal laminin G-like domain pair of the laminin α2 chain harbouring binding sites for α-dystroglycan and heparin. *The EMBO Journal, 19*(7), 1432-1440.

Vuolteenaho, R., Chow, L. T., & Tryggvason, K. (1990). Structure of the human laminin B1 chain gene. *Journal of Biological Chemistry, 265*(26), 15611-15616.

Wang, C., Barry, J. K., Min, Z., Tordsen, G., Rao, A. G., & Olsen, O.-A. (2003). The calpain domain of the maize DEK1 protein contains the conserved catalytic triad and functions as a cysteine proteinase. *Journal of Biological Chemistry, 278*(36), 34467-34474.

Wright, D. A., Townsend, J. A., Winfrey, R. J., Irwin, P. A., Rajagopal, J., Lonosky, P. M., . . . Voytas, D. F. (2005). High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *The Plant Journal, 44*(4), 693-705.

Zhao, S., Liang, Z., Demko, V., Wilson, R., Johansen, W., Olsen, O.-A., & Shalchian-Tabrizi, K. (2012). Massive expansion of the calpain gene family in unicellular eukaryotes. *BMC evolutionary biology*, *12*(1), 193.

Appendix:

Appendix1: DEK1 protein accession number for different land plants used in bioinformatics work in this study

Name	**Accession Number**
Zea mays	NP_001105528.1
Arabidopsis thaliana	NP_850967.1
Hordeum vulgare	ABW81402.1
Setaria italica	XP_004984908.1
Gossypium raimondii	XP_012471753.1
Sesamum indicum	XP_011089165.1
Vitis vinifera	XP_010651386.1
Camelina sativa	XP_010501149.1
Nelumbo nucifera	XP_010257671.1
Nicotiana sylvestris	XP_009766184.1
Brassica rapa	XP_009147506.1
Prunus mume	XP_008222910.1
Oryza sativa	NP_001047890.1

Appendix 2:

A: Table of different primer sets used in this study containing information of the purpose of the primers, names, and the primer sequences

Purpose	**Primers name**	**Sequence 5'-3'**
Amplify *P. patens DEK1-Arm* sequence	SP_Inf_PpArm_1	TGACATTATTACTATTGGTTTACCTCAC
	ASP_Inf_PpArm_1	ATAGATACACGACCGGCAG
Linearize the plasmid *pBHRF_ ΔArm*	SP_Inf_1	CGGTCGTGTATCTATCTTGTTCTC
	ASP-Inf_1	ATAGTAATAATGTCATATGCGTACAC
Construct vector *pBHRF_PpArm_ΔLG3*	SP_Inf_2	GAGCAGAATGAGCTGGATATTATGG
	ASP_Inf_2	CAGCTCATTCTGCTCACTATTACGAGGCTCATTGTCCATC
pBHRF_PpArmComp colony PCR	ΔArm IF	CTTTGACTCTACAACGGATA
	ΔArmIR	CAGAGTTCTCATCGAGTAAA
pBHRF_PpArm_ΔLG3 colony PCR	Arm_seq6	TGCAGGTACCAAAGAAGCAGC
	pBHRF_rev	AGGAAACAGCTATGACCATGA
P. patens 1st genotype	Arm_seq6	TGCAGGTACCAAAGAAGCAGC
	ASP_Inf_PpArm_1	ATAGATACACGACCGGCAG
P. patens 2nd genotype	TM2 seq1	TACCTAGGGTGGGCAATTGC
	ASP_PpCal_gDNA	TCAATCTCCTCTCCAGCACCT
ΔLG3 probe	Armseq2	GGTTCTTGGTCATGCTACACGA
	ΔArm1R	CAGAGTTCTCATCGAGTAAA
Amplify DEK1-LG3 cDNA for sequencing	Pp_Loop_inverse_SP	TGGGTCTTCTTCAGTGTGATC
	CALP_cDNA	CGACCTCTCGTACCTGTAAAAGAG
experimental primers used in the qPCR	CALPqF	TGGGCTAATGAAGTTGAATGG
	CALPqR	AAATCTTGCCATGACATCCAG
reference primers used in the qPCR	SQSF	AGGTTTACACTGTCTGAACGA
	SQSR	CAGAATCGAAGATTTGGTTGGT

B: Primers used for sequencing the *pBHRF_PpArmComp* and *pBHRF_PpArm_ΔLG3*

***pBHRF_PpArmComp* sequencing primers:**

Primer name	Primer sequence 5' to 3'
pBHRF_fw	GCCTCTTCGCTATTACGCCA
ARM 5' Fw	GAGGTTCTTTTTGTCTGTCT
ARM 5TS seq	CCCGCGGTGGTTGTATACTC
TM2 seq2	GCCTTCTTGGTTCTTTATGGGA
ARM 5' rev	TAATGTCATATGCGTACACC
PpARM_inf1 _SP	TGACATTATTACTATTGGTTTACCTCAC
ARMseq1	TGCAAGTTCAGCAGCTCTGC
ARMseq2	GGTTCTTGGTCATGCTACACGA
ARMseq3	TGTTTTAGCACGGCTATTCTTTTC
ARM seq 10 R	CTGAGATTCGAGAAGCCAATGC
ARMseq4	TGATCTTCAGTTTTGGGCATAGA
ARMseq5	TGCATCGGAACAAGAATCTAGTGTA
ARMseq6	TGCAGGTACCAAAGAAGCAGC
ARMseq7	GCATATTGGGCGTTGAAGCT
ARMseq8	GATGGAAGTATGGGTTGGCATC
ARMseq9	GCAAAGAGGAAGGCAGCAGA
PpARM_inf1 _ASP	ATAGATACACGACCGGCAG
CALP seq1	AAAGAGGAGGTCTTGCAGCG
ARM 3TS seq	CCGCCATCAGATCAGTCGCT
PpCALP_ gDNA	TTCATGAACACCATTTGAGCG
pBHRF_Rev	AGGAAACAGCTATGACCATGA

***pBHRF_PpArm_ΔLG3* sequencing primers:**

Primer name	Primer sequence 5' to 3'
pBHRF_fw	GCCTCTTCGCTATTACGCCA
ARM 5' Fw	GAGGTTCTTTTTGTCTGTCT
ARM 5TS seq	CCCGCGGTGGTTGTATACTC
TM2 seq2	GCCTTCTTGGTTCTTTATGGGA
ARM 5' rev	TAATGTCATATGCGTACACC
PpARM_inf1 _SP	TGACATTATTACTATTGGTTTACCTCAC
ARMseq1	TGCAAGTTCAGCAGCTCTGC
ARMseq2	GGTTCTTGGTCATGCTACACGA
ARMseq3	TGTTTTAGCACGGCTATTCTTTTC
ARM seq 10 R	CTGAGATTCGAGAAGCCAATGC
ARMseq4	TGATCTTCAGTTTTGGGCATAGA
ARMseq5	TGCATCGGAACAAGAATCTAGTGTA
ARMseq6	TGCAGGTACCAAAGAAGCAGC
ARMseq9	GCAAAGAGGAAGGCAGCAGA
PpARM_inf1 _ASP	ATAGATACACGACCGGCAG

CALP seq1	AAAGAGGAGGTCTTGCAGCG
ARM 3TS seq	CCGCCATCAGATCAGTCGCT
PpCALP_ gDNA	TTCATGAACACCATTTGAGCG
pBHRF_Rev	AGGAAACAGCTATGACCATGA

C: *P. patens dek1-Δlg3* cDNA sequencing primers:

Primer name	Primer sequence 5' to 3'
TM2 seq1	TACCTAGGGTGGGCAATTGC
ARM 5' Fw	GAGGTTCTTTTTGTCTGTCT
TM2 seq2	GCCTTCTTGGTTCTTTATGGGA
ARMseq1	TGCAAGTTCAGCAGCTCTGC
ARMseq2	GGTTCTTGGTCATGCTACACGA
ARMseq6	TGCAGGTACCAAAGAAGCAGC
ARMseq7	GCATATTGGGCGTTGAAGCT
ARMseq8	GATGGAAGTATGGGTTGGCATC
ARMseq9	GCAAAGAGGAAGGCAGCAGA
ARM 3TS seq	CCGCCATCAGATCAGTCGCT

Appendix 3: Different media and solutions used in this study

Solution B:		**Final concentration:**
$MgS0_4.7\ H_2O$	25 g	0.1 M
Distilled H_2O	to 1 l	
Solution C:		
KH_2PO_4	25 g	184 mM
Distilled H_2O	to 1 l	
Adjust pH to 6.5 with minimal volume of 4 M KOH		
Solution D:		
KNO_3	101 g	1 M

Distilled H_2O	to 1 l	
BCD medium:		
Stock solution B	10 ml	1 mM $MgS0_4$
Stock solution C	10 ml	1.84 mM KH_2PO_4
Stock solution D	10 ml	10 mM KNO_3
Trace element solution	1 ml	
$CaCl_2$	111 mg	1 mM
$FeS0_4.7H_2O$	12.5 mg	45 µM
Agar	7 g	0.7% (w/v)
Distilled H_2O	to 1 l	
BCDA medium:		
Stock solution B	10 ml	1 mM $MgS0_4$
Stock solution C	10 ml	1.84 mM KH_2PO_4
Stock solution D	10 ml	10 mM KNO_3
Trace element solution	1 ml	
$CaCl_2$	111 mg	1 mM
$FeS0_4.7H_2O$	12.5 mg	45 µM
di-ammonium (+) tartrate	920 mg	5 mM
Agar	7 g	0.7% (w/v)
Distilled H_2O	to 1 l	

PRMB (protoplast regeneration medium – bottom layer)		
Stock solution B	10 ml	1 mM $MgS0_4$
Stock solution C	10 ml	1.84 mM KH_2PO_4
Stock solution D	10 ml	10 mM KNO_3
Trace element solution	1 ml	
$CaCl_2$	1.1 g	1 mM
$FeS0_4.7H_2O$	12.5 mg	45 µM
Agar	7 g	0.7% (w/v)
di-ammonium (+) tartrate	920 mg	5 mM
D-mannitol	60 g	6% (w/v)
Distilled H_2O	to 1 l	
PRMT (protoplast regeneration medium – top layer)		
Stock solution B	10 ml	1 mM $MgS0_4$
Stock solution C	10 ml	1.84 mM KH_2PO_4
Stock solution D	10 ml	10 mM KNO_3
Trace element solution	1 ml	
$CaCl_2$	1.1 g	1 mM
$FeS0_4.7H_2O$	12.5 mg	45 µM
Agar	4 g	0.4% (w/v)
di-ammonium (+) tartrate	920 mg	5 mM
D-mannitol	80 g	8% (w/v)

Distilled H_2O	to 1 l	
Trace element solution		
$Al_2K(S0_4)_3.K_2SO_4.24H_2O$	110 mg	0.006% (w/v)
$CoCl_2.6H_2O$	55 mg	0.006% (w/v)
$CuSO_4.5$ H2O	55 mg	0.006% (w/v)
H_3BO_3	614 mg	0.061% (w/v)
KBr	28 mg	0.003% (w/v)
KI	28 mg	0.003% (w/v)
LiCl	28 mg	0.003% (w/v)
$MnCl_2.4H_2O$	389 mg	0.039% (w/v)
$SnCl_2.2H_2O$	28 mg	0.003% (w/v)
$ZnS0_4.7H_2O$	55 mg	0.006% (w/v)
Distilled H_2O	to 1 l	
MMM solution		
D-mannitol	910 mg	9.1%
2-[*N*-morpholino] ethanesulfonic acid (MES) (1% w/v, pH 5.6)	1 ml	10%
$MgCl_2$	150 µl	15 mM
Distilled H_2O	8.85 ml	

Polyethylene glycol solution (PEG)		
$CaCl_2.H_2O$	109 mg	
Distilled H_2O	10 ml	
Polyethylene glycol (PEG) (MW 6000)	5 g	

Appendix 4:

A: DEK1-Arm sequences similarity

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Physcomitrella patens	1		58,83	58,67	59,55	60,68	64,24	56,96	61,33	60,68	57,28	57,44	61,00	58,74	60,03
Zea mays	2	58,83		95,41	90,51	69,66	75,37	66,39	71,62	71,45	86,58	67,54	70,47	68,68	71,13
Setaria italica	3	58,67	95,41		90,62	71,64	75,90	67,54	71,48	71,97	87,34	69,02	71,48	69,34	72,13
Oryza sativa	4	59,55	90,51	90,62		71,76	76,85	68,14	73,40	74,22	89,46	69,29	72,09	70,44	74,06
Camelina sativa	5	60,68	69,66	71,64	71,76		78,45	88,96	77,63	79,44	68,97	93,41	75,49	75,33	80,26
Nelumbo nucifera	6	64,24	75,37	75,90	76,85	78,45		75,00	83,55	83,55	73,40	75,16	79,93	79,61	80,10
Brassica rapa	7	56,96	66,39	67,54	68,14	88,96	75,00		80,03	81,74	70,19	93,68	78,16	78,16	82,25
Vitis vinifera	8	61,33	71,62	71,48	73,40	77,63	83,55	80,03		91,81	76,66	80,72	87,71	87,54	88,57
Prunus mume	9	60,68	71,45	71,97	74,22	79,44	83,55	81,74	91,81		76,15	82,76	86,69	86,69	90,10
Hordeum vulgare	10	57,28	86,58	87,34	89,46	68,97	73,40	70,19	76,66	76,15		71,50	74,91	72,87	75,26
Arabidopsis thaliana	11	57,44	67,54	69,02	69,29	93,41	75,16	93,68	80,72	82,76	71,50		79,32	79,15	83,93
Sesamum indicum	12	61,00	70,47	71,48	72,09	75,49	79,93	78,16	87,71	86,69	74,91	79,32		86,32	85,64
Nicotiana sylvestris	13	58,74	68,68	69,34	70,44	75,33	79,61	78,16	87,54	86,69	72,87	79,15	86,32		84,96
Gossypium raimondii	14	60,03	71,13	72,13	74,06	80,26	80,10	82,25	88,57	90,10	75,26	83,93	85,64	84,96	

B: DEK1-ΔLG3 sequences similarity

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Physcomitrella patens	1		54,75	55,00	58,19	58,29	57,46	55,43	54,35	59,67	58,70	58,56	57,84	57,98	55,98
Zea mays	2	54,75		94,80	88,95	90,12	66,67	67,80	67,23	74,14	75,14	72,41	70,22	69,61	67,80
setaria italica	3	55,00	94,80		88,44	89,60	67,82	70,06	68,36	75,29	76,84	72,99	71,91	71,27	70,06
Hordeum vulgare	4	58,19	88,95	88,44		92,94	68,39	68,93	69,49	75,86	75,71	73,56	70,22	71,82	68,93
Oryza sativa	5	58,29	90,12	89,60	92,94		67,24	67,80	67,23	76,44	74,58	74,14	71,35	70,72	67,80
Brassica rapa	6	57,46	66,67	67,82	68,39	67,24		89,83	73,45	74,14	74,01	79,31	79,21	77,35	90,40
Arabidopsis thaliana	7	55,43	67,80	70,06	68,93	67,80	89,83		72,32	72,32	75,71	77,97	80,90	76,80	97,18
Nicotiana sylvestris	8	54,35	67,23	68,36	69,49	67,23	73,45	72,32		74,58	79,10	81,36	82,02	81,77	73,45
Nelumbo nucifera	9	59,67	74,14	75,29	75,86	76,44	74,14	72,32	74,58		78,53	82,76	80,34	81,77	73,45
Sesamum indicum	10	58,70	75,14	76,84	75,71	74,58	74,01	75,71	79,10	78,53		83,05	87,64	86,19	76,84
Prunus mume	11	58,56	72,41	72,99	73,56	74,14	79,31	77,97	81,36	82,76	83,05		87,64	87,29	78,53
Gossypium raimondii	12	57,84	70,22	71,91	70,22	71,35	79,21	80,90	82,02	80,34	87,64	87,64		89,50	81,46
Vitis vinifera	13	57,98	69,61	71,27	71,82	70,72	77,35	76,80	81,77	81,77	86,19	87,29	89,50		78,45
Camelina sativa	14	55,98	67,80	70,06	68,93	67,80	90,40	97,18	73,45	73,45	76,84	78,53	81,46	78,45	