Hedmark University
of Applied Sciences

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

Joseph Innocent Massawe

Master thesis

Genetic analyses of functionally conserved residues in the *Defective-kernel1* laminin G3-like domain of the moss *Physcomitrella patens*

Master's Degree in Applied and Commercial Biotechnology

June 2016

Consent to lending by University College Library	YES ☐ X	NO☐
Consent to accessibility in digital archive Brage	YES ☐ X	NO☐

Acknowledgement

Firstly, I would like to thank my supervisors Associate Prof. Wenche Johansen and Dr. Viktor Demko for giving me opportunity to work on this project. I sincerely appreciate Associate Prof. Wenche Johansen for her support, guidance, encouragement and motivation. Thank you for every piece of your input, you make me someone else.

My gratitude also goes to PhD. student Ako Eugene Ako, for all his support and guidance especially during the lab work.

My sincere thanks also to Lånekassen for financing my master's studies in Norway.

I would also like to express my gratitude to the biotech staff members for their support during my study. Thank you all.

To all my family members and friends, thank you for all your support and encouragement during the whole duration of my master's studies.

Abbreviations

3D	Three dimensional
cDNA	Complimentary deoxyribonucleic acid
ddNTP	Dideoxynucleotide triphosphates
DDT	1, 4 dithiothretiol
DEK1	Defective Kernel1
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
HF	High fidelity
HR	Homologous recombination
LG3	Laminin-like globular domain 3
MSA	Multiple sequence alignment
NaOAc	Sodium Acetate
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
RNA	Ribonucleic acid
RPM	Revolution per minute
RT	Room temperature
RT-PCR	Reverse transcriptase Polymerase chain reaction
TGS	Targeting sequence
WT	Wild type

Table of Contents

Abstract

DEFECTIVE KERNEL1 (DEK1), the single calpain of land plants, is highly conserved over the past 450 million years and it has been shown that the protein is essential for determining cell wall orientation in 3D growth. A recent study reveals that the DEK1-Arm segment harbour a conserved LG3 domain belonging to the Laminin-like globular domain family (LG3). In the current study, assessment of conserved residues in DEK1-LG3 for DEK1 function in *Physcomitrella patens* was investigated. This was achieved by substituting residues predicted to be functionally important in the DEK1-LG3 domain with alanine (Ala). Five (5) plasmid vectors for targeted insertion of mutagenized version of DEK1-LG3 were constructed and then re-targeted to the *DEK1*Δ*LG3* locus in the *dek1*Δ*lg3* mutant background. The resulting mutant plants were subsequently characterized both molecularly and phenotypically to assess the importance of the conserved amino acids for DEK1 function. Results suggest that the amino acids E1477, Q1478, E1481 and S1497 have no important function for protonemata growth or gametophore development as the mutant plants displayed the WT phenotype. However, the mutations introduced into the *dek1lg3:R*^{1514} mutant affected post-transcriptional processing of the *DEK1* transcript, resulting in a Δ*dek1* mutant phenotype. Targeting of the pBHRF_JI_LG3:G^{1574}R^{1575}S^{1576}D^{1577}S^{1578}E^{1579} vector into the *DEK1*Δ*LG3* locus in the *dek1*Δ*lg3* mutant did not result in any re-generating plants, suggesting that the introduced mutations are lethal to *P. patens.*

1. Introduction

1.1 Calpains

The calpain family (EC 3.4.22.17, Clan CA, family C02) is a group of complex intracellular enzymes that share a Ca^{2+}-dependent cysteine protease activity (Margis and Margis-Pinheiro 2003; Campbell and Davies 2012; Ono and Sorimachi 2012). Calpains show a large variation in domain architecture, however, common to all calpains is the catalytic core domain, CysPc. Calpain was first reported in the 1960s when calcium-dependent proteolytic activity in soluble extracts of rat brain was observed (Guroff 1964). Members of the calpain family are broadly present in different organisms ranging from mammals to plants, and some are constitutively expressed, while others show temporal and spatial expression pattern (Branca 2004). While the human genome contains 15 genes encoding calpains (Ono and Sorimachi 2012), plants encode only a single calpain variant, the DEFECTIVE KERNEL1 (DEK1) protein (Lid et al. 2002; Wang et al. 2003). Based on phylogenetic inferences, four calpain architectures named CysPc, CysPc-C2L, MIT-CysPc-C2L and MEM-CysPc-C2L are believed to have formed in the early evolution of eukaryotes through assembly of ancient domains (Zhao et al. 2012). Calpains are grouped into the classical and non-classical calpain variants (Figure 1). Classical calpains are composed of two subunits, the large catalytic subunit and the small regulatory subunit (Figure 1). The catalytic subunit of classical calpains harbour four (4) conserved domains; an N-terminal anchor helix domain (Nter), a catalytic protease core domain (CysPc) composed of the sub-domains PC1and PC2, a C2-like domain (C2L) and a penta-EF-hand domain (PEF). The regulatory subunit contains an N-terminal Gly-rich domain (GR) and penta-EF-hand domains (PEF) (Figure 1). Non-classical calpain variants, which lack the regulatory subunit, are composed of the CysPc domain but lacks the Nter and PEF domains, and may in addition harbour additional domains (Ono and Sorimachi 2012). The catalytic protease core domain (CysPc) contains the active sites catalytic triad formed by a cysteine residue (located in sub-domain PC1) and the histidine and asparagine residues (located in sub-domain PC2). Classical calpain are dependent on Ca^{2+} for activation; upon binding of two Ca^{2+} atoms, coordinated by several amino acids in the CysPc domain, the catalytic cleft of the CysPc domain is re-aligned into an active conformation (Moldoveanu et al. 2002). Calpains are mainly located in the cytosol of the cell as an inactive enzyme, their activation is highly

regulated, and activity is often associated with membranes (Suzuki et al. 2004). As opposed to conventional proteases, calpains are regulatory enzymes; they do not degrade their target substrate, instead they modulate the activity of a restricted set of protein substrates by cleavage at one or a few specific positions using complex substrate-recognition mechanisms. Proteins known to be substrates for calpains includes membrane-bound or membrane-associated proteins, such as calcium-ATPase, the epidermal growth factor (EGF) receptor, the ryanodine receptor, the calcium receptor, the N-methyl-D-aspartate (NMDA) receptor (a glutamic acid receptor), β-integrins and transcription factors (Croall and Ersfeld 2007). Calpains have been described to play important roles in various cellular processes in animals including cell proliferation, exocytosis, apoptosis, differentiation, signal transduction and endocytosis (Sato et al. 1995; Tompa et al. 2001). Dis-regulation of calpain activity is implicated in various human diseases (Huang and Wang 2001). For example, over-activation of calpain 1 has long been tied to acute neurological disorders like stroke and traumatic brain injury and in addition to Alzheimer's disease (Grynspan et al. 1997; Huang and Wang 2001).

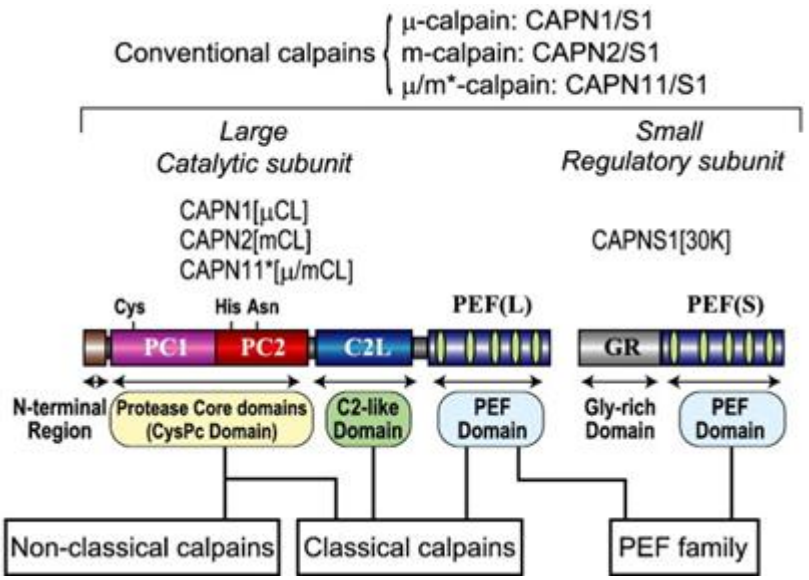


Figure 1. Structures of the calpain superfamily members. Calpains are composed of a large catalytic and a small regulatory subunit. Classical calpain contains the N-terminal anchor helix domain (Nter), a catalytic protease core domain (CysPc) composed of the sub-domain PC1and PC2, a C2-like domain (C2L) and a penta-EF-hand domain (PEF). Non-classical calpains contain the catalytic protease core domain (CysPc) and the C2-like domain (C2L) (Ono and Sorimachi 2012).

1.2 DEK1, the single calpain of land plants

Land plants evolved from a small group of freshwater green algae, the Charophyceae (Becker and Marin 2009; Kenrick et al. 2012). The evolution of land plants was one of the most important events in evolution, in which the origin and diversification of plants is placed at a minimum of 450 million years ago (Waters 2003; Rensing et al. 2008), with the mosses and seed plants sharing their last common ancestor at least 400 million years ago (Theißen et al. 2001). The ability to determine cell wall placement orientation in predictable planes was a novel feature that evolved in the transition from charophyte algae to the first land (Graham et al. 2000). Functional analysis in the moss *Physcomitrella patens* has shown that the *DEK1* gene, encoding a transmembrane (TML) calpain proteases, is necessary for determining cell wall placement and that the gene is required for three dimensional growth in this organism (Perroud et al. 2014) thus supporting a central role for DEK1 in land plant evolution. All land plants examined today, for which full genome sequence are available, harbour a highly conserved *DEK1* gene (Zhao et al. 2012; Liang et al. 2013). The common ancestor of chlorophyte and charophyta algae possessed both TML-calpains and cytosolic calpains, however TML calpains were subsequently lost from the chlorophyte alga, while both cytosolic and TML-calpains were retained in the charophyte alga (Demko et al. 2014). Cytosolic calpains were subsequently lost in land plants leaving DEK1 as the only calpain of land plants (Demko et al. 2014). Genetic complementation studies in the *Arabidopsis thaliana dek1-3* mutant has also shown that the *P. patens* DEK1 catalytic domain, CysPc–C2L is functional in *A. thaliana,* complementing the *dek1-3* mutant phenotype, providing evidence for functional conservation of DEK1 for at least 450 million years (Liang et al. 2013). Examination of DEK1 sequence conservation in land plants reveal a high degree of conservation of up to 80% and 60% amino acid identity for the CysPc and C2L domains, respectively (Liang et al. 2013). Thus the high degree of DEK1 sequence conservation between angiosperm and moss over the period of evolution implies an important conserved function for the *DEK1* gene (Tian et al. 2007).

1.3 DEK1 structure

The identification of the first plant calpain historically named “phytocalpain” was shown by (Lid et al. 2002) when the *DEK1* gene was cloned from maize. Now, analysis of plant genomes reveals that all land plants, from mosses to angiosperms harbour only one calpain variant, DEK1 (Tian et al. 2007; Zhao et al. 2012; Liang et al. 2013). DEK1 encodes a large protein of approximately 240 kDa (2,159 amino acid residues in maize) (Lid et al. 2002). The DEK1 protein has been predicted to contain a transmembrane domain (MEM) composed of 23 transmembrane segments (TMs) interrupted by a Loop (300 amino acids long) located between the ninth and tenth TMs, speculated to be involved in either perception and/or transmission of positional signals (Tian et al. 2007). DEK1-MEM is further connected to the C-terminal calpain protease domain composed of the catalytic CysPc and C2L regulatory domains by the DEK1-Arm segment of approximately 620 amino acids (Figure 2). Recent searches in conserved domain databases have reveal that the C-terminal region of the DEK1-Arm segment harbour a domain belonging to the Laminin-like globular domain family (LG3) (Johansen W, - manuscript in preparation). The LG3 domain belongs to the Concanavalin A-like lectin/glucanases superfamily and was initial identified as one of five modules (LG1-LG5) building the large globular domain of the α chain C-terminus of the heterotrimeric glycoprotein laminin (Beck et al. 1990). LG3 modules share low to moderate sequence similarity (Timpl et al. 2000) and are also found to share approximately 20-25% sequence similarities with domains found in neurexin (Ushkaryov et al. 1992) and the sex hormone-binding globulin (SHBG) (Joseph and Baker 1992). These domains are therefore often referred to as LNS domain (for LG, neurexin and SHBG). Based on available information in protein domain database, LG/LNS domains, which contain 180-200 residues, are found as singletons or in tandem arrays in more than 1200 different protein architectures in a diverse range of protein families in both prokaryotic and eukaryotic organisms. The 3D structure of DEK1 is unknown. However, the CysPc-C2L structure has been predicted using homology modelling showing that the DEK1-calpain module has similar structure to classical animal calpain Cys-CL2 domains (Wang et al. 2003; Liang et al. 2013). Specifically, the active-site residues cysteine (Cys-71) is located on the α-helix on sub-domain PC1 (IIa), and histidine (His-229) and asparagine (Asn-249) is located on sub-domain PC2 (IIb) in the maize DEK1 calpain (Figure 3), thus showing a similar configuration of these amino acids as in animal calpains (Wang et al. 2003). In addition, the amino acids shown to be involved in Ca^{2+} binding in animal calpains (Moldoveanu et al.

2002) are conserved in land plant DEK1 proteins (Liang et al. 2013). Further, modelling of the structure show that the corresponding amino acids in the *A. thaliana* DEK1 CysPc domain has the potential to bind two Ca^{2+} in the corresponding positions to the animal calpains (Liang et al. 2013).

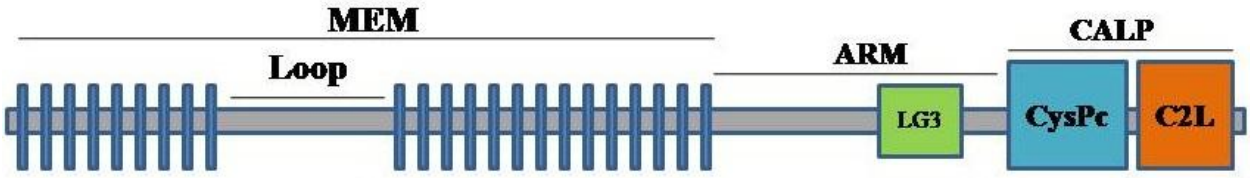


Figure 2. Schematics of the predicted DEK1 structure. The MEM segment is composed of 23 TMs interrupted by the loop segment located between the ninth and tenth TMs. The calpain module (CALP) is composed of the CysPC and C2L domains and is connected to the MEM via the ARM segment. The LG3 domain is located towards the C-terminal end of the ARM.

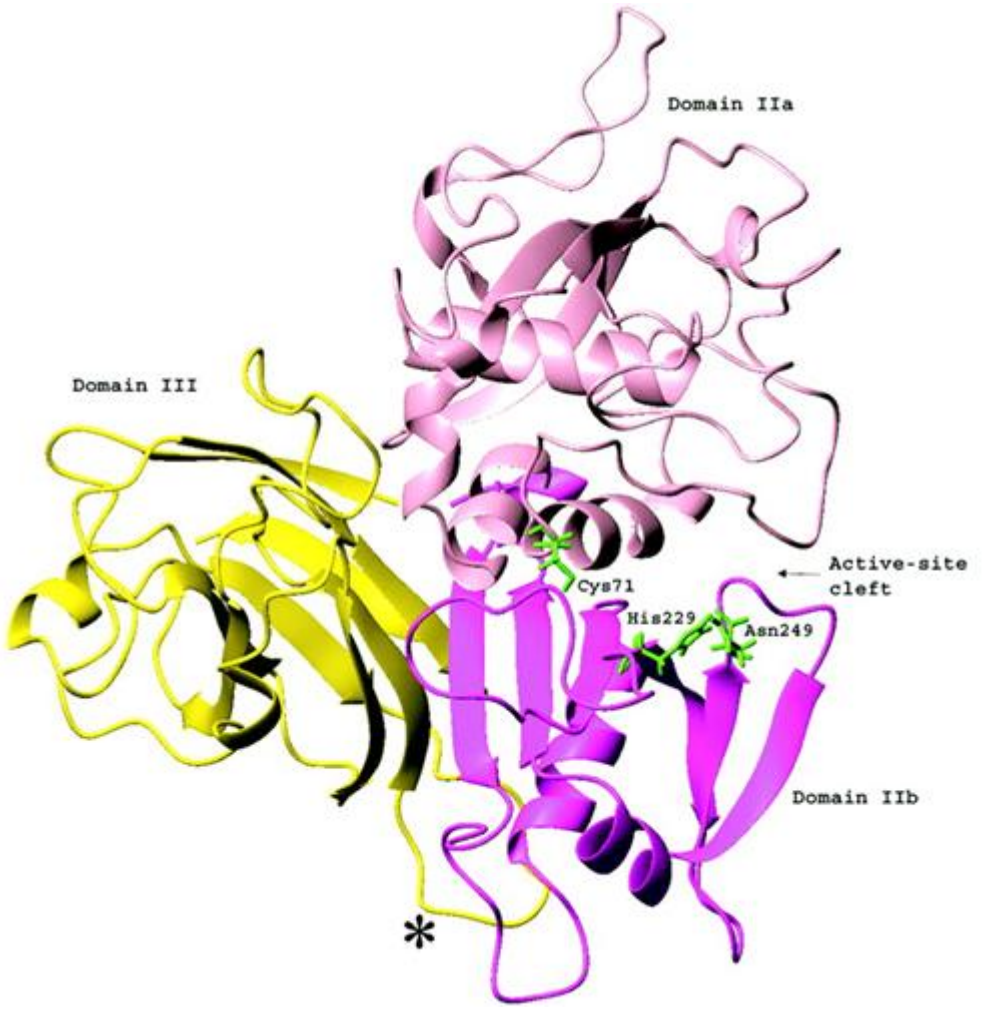


Figure 3. Predicted three-dimensional structure of maize DEK1 calpain. The catalytic triad active site residues Cys-71, His-229 and Asn-249 are shown in the structure. Cys-71 is located in Domain IIa (PC1) while His-229 and Asn-249 are located in the Domain IIb (PC2). The figure is modified from (Wang et al. 2003).

1.4 DEK1 function

The high degree of DEK1 sequence conservation from moss to angiosperms indicates that the DEK1 protein has an important function in land plants.

The moss *P. patens* Δ*dek1* mutant is characterized by the lack of gametophore (Figure 4) (Perroud et al. 2014). Closer examination of the mutant plant reveals that the first division of the bud apical cell fail to position in correct orientation and instead occurs at random planes compared to WT where the new cell wall orient with the previous cell wall in a clear median and perpendicular position (Figure 5A III and B III). Thus as a result of misorientation in the

division of the bud apical cell, bud development is arrested which subsequently leads to block of gametophore development.

Deletion of the DEK1-LOOP (*dek1*Δ*loop*) has been shown to affect expansion of phyllids (Demko et al. 2014). Examination of the *dek1*Δ*loop* mutant shows irregular cell divisions after the first asymetrical division and the first division of the bud apical cell. As a result of irregular patterns in cell division, subsequent developments were arrested and phyllids failed to expand (Figure 5C IV) as compared to wild-type plant.

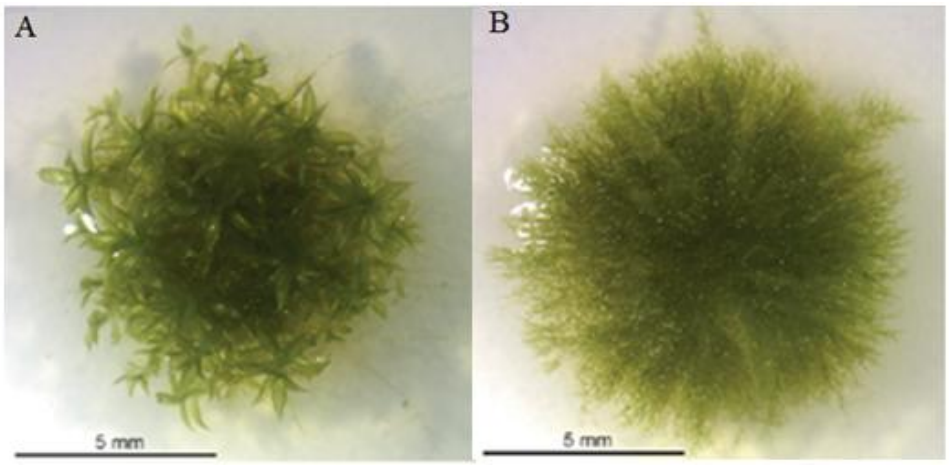


Figure 4. WT and Δ*dek1* mutant grown on BCD medium. (A) *P. patens* 3-weeks-old wild-type (WT) plant showing well developed gametophores, (B) 3-weeks-old Δ*dek1*mutant plant showing filamentous growth without gametophores. Modified figure from (Perroud et al. 2014).

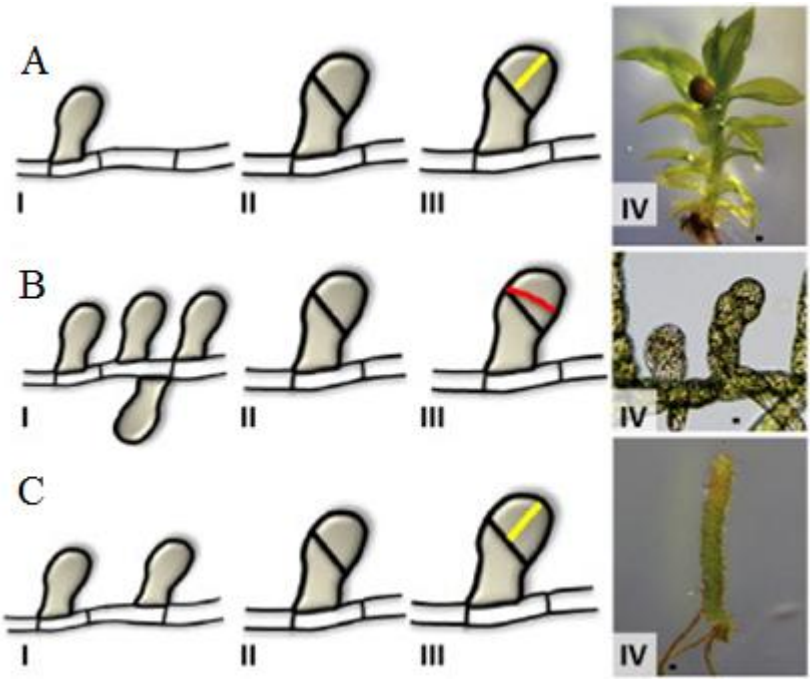


Figure 5. Effects of DEK1 mutation in moss *P. patens.* (A) WT development with fully developed gametophore. (I) Protonema cell with one bud, (II) Bud development after first asymmetrical division, (III) New cell wall development traverse the previous cell wall (yellow), (IV) Gametophore with leaf-like phyllids. (B) Development of *P. patens* DEK1 deletion mutant. (I) Protonema with four buds, (II) Normal bud development after first asymmetrical division, (III) Misoriented second division as compared to WT (red) and (IV) Arrested gametophore development due to misorientation of cell division plane. (C) Development of *dek1* Δ*loop* mutant. (I) Development of protonema cells with two buds, (II) Asymeric first division, (III) Perpendicular second cell division (yellow) and (IV) Development of *dek1*Δ*loop* mutant stem lacking phyllids. Figure retrieved from (Demko et al. 2014; Olsen et al. 2015).

Recent searches in conserved domain databases also reveal that the C-terminal region of the DEK1-Arm segment harbour a domain belonging to the Laminin-like globular domain family (LG3) (Johansen W, - manuscript in preparation). A deletion mutant of DEK1-LG3 (*dek1*Δ*lg3)* has shown significant defects in phyllid development. In the mutant plant the phyllids are narrow, smaller, with fewer and large cells as compared to wild-type phyllids. Moreover the phyllids in the mutant plant are lacking marginal serration, having blunt tip and short midrib as compared wild-type phyllids (Johansen W, - manuscript in preparation). Examination of phyllid morphology in the *dek1*Δ*lg3* mutant suggests that both phyllid apical cell activity and sectorial cell divisions within the phyllid are impaired in the mutant.

The *dek1*Δ*lg3* mutant also shows defects in gametangia development. In *dek1*Δ*lg3* mutant, the archegonia apex failed completely to open but also the egg canal is absent which result in sterile plants (Johansen W, - manuscript in preparation).

In maize, Δ*dek1* mutant has been shown to have effect on endosperm development (Lid et al. 2002). The mutant plant shows altered development with defective aleurone layer of the endosperm (Figure 6). Thus, the *DEK1* gene was shown to be involved in maintaining aleurone cell fate at an early developmental stage and also by restricting the aleurone cell fate to the surface layer of the maize endosperm (Lid et al. 2002).

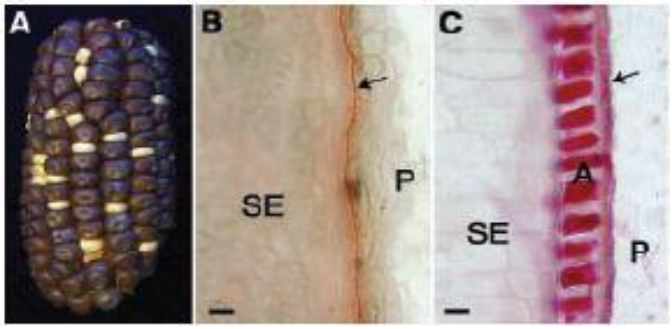


Figure 6. The phenotype of *dek1-mum1* homozygous maize kernels (A) The ear segregates 3:1 for wild-type (dark) and DEK1 grains, representing that the two mutations are both in the *DEK1* gene. (B) Hand section of *dek1-mum1* kernel starchy endosperm cells (SE) is located in the periphery of the endosperm close to the remnants of the nucleus (arrow); P representing the maternal pericarp. (C) Section of wild-type kernel with peripheral aleurone cells. Figure modified from (Lid et al. 2002).

In *A. thaliana*, DEK1 has been shown to be essential in early embryonic development and epidermal activity (Johnson et al. 2005). The assessment of *A. thaliana dek1-3* mutant embryo reveal abnormal divisions in embryo proper and in the apical portion of the suspensor where subsequent periclinal cell divisions in the protoderm leads to abnormal embryo surface which results to embryo lethal (Figure 7B III, IV and V) as compared to WT with normal embryo development throughout (Figure 7A). In addition, down-regulation of the DEK1 in *A. thaliana* using RNA interference (RNAi) leads to apparent loss of epidermal identity where palisade cells were not aligned in correct positions (Johnson et al. 2005).

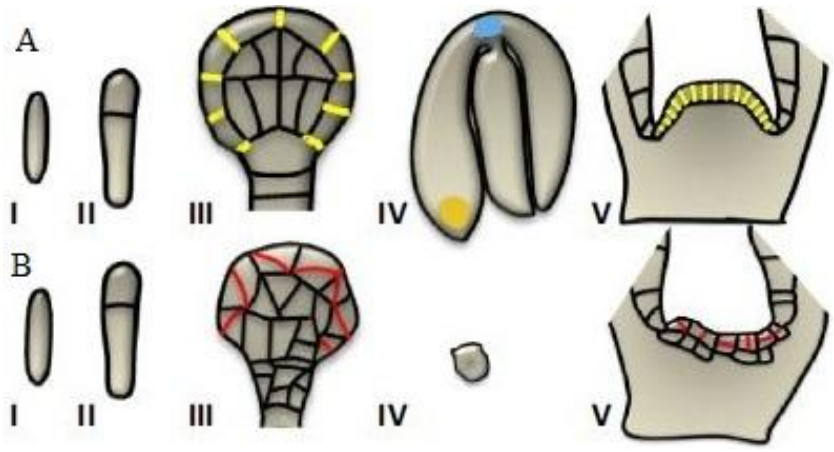


Figure 7. Effect of *dek1* mutation in *A. thaliana. (A)* Early cell division in *A. thaliana* wild-type plant. (I) fertilized zygote, (II) First asymmetrical division of the zygote, (III) Globular embryo development with protoderm initials, (IV) Mature embryo with apical and root meristems, (V) Apical meristem with L1 layer formed by division. (B) Development of *A. thaliana dek1* mutant embryo. (I and II) normal first asymmetric division, (III) Failure of subsequent division (cell wall marked in red) which results in lack of protoderm, (IV) embryo lethal after failure to orient properly and (V) mutant plant lacking epidermal identity as a result of down-regulation of DEK1 activity. Modified figure from (Olsen et al. 2015).

To summarize, DEK1 has a common function in the orientation of cell division plane in different plant species as described in this section. In *P. patens,* the DEK1 deletion mutant shows misorientation in the first division of the bud apical cell which fail to orient its wall perpendicular to the previous wall, which result in arrest of gametophore development. In *A. thaliana,* early mitotic division failed to orient correctly leading to failure in the establishment of a protoderm subsequently leading to arrest in embryo development.

1.5 *Physcomitrella patens* as a model organism

The moss *P. patens* is the main model system for basal plants. Due to its relatively simple body plan, well-physiological reactions and its high ratio of homologous recombination compared to other land plants, it has been widely used as the model plant (Schaefer and Zrÿd 1997; Hohe and Reski 2003). As a bryophyte, *P. patens* is suited to study the function of DEK1 because bryophytes are the earliest diverging group of land plant thus they harbour all ancent evolutionary features that enable them to live teristial environment (Rensing et al. 2008). Also the easiness of culture, growth under defined and controlled

environment (Cove et al. 2006) enable *P. patens* to be used as a model organism to study the function of DEK1.

1.5.1 Life cycle of *Physcomitrella patens*

Like other vascular plants, *P. patens* shows alternation of generations between the haploid gametophyte and diploid sporophyte in which the haploid gametophyte is the dominant phase (Schaefer and Zrÿd 1997). The life cycle of *P. patens* (Figure 8) begins by germination of a haploid spore into filamentous structure called protonema. Protonema produce two cell types, the chloronema cells with large chloroplast and caulonema cells which grow very fast. As the plant continue to develop a transition in the side branch of chloronemal filaments occur which then develops into gametophore, a leaf-like structure with male (antheridia) and female (archegonia) sex organs. Under moist conditions, flagellate sperm, which are produced in antheridia, swim towards the egg in archegonia and fertilization occur, finally producing a diploid zygote. The resulting diploid zygote begins the sporophyte phase that produce thousands of spores through meiosis (Prigge and Bezanilla 2010).

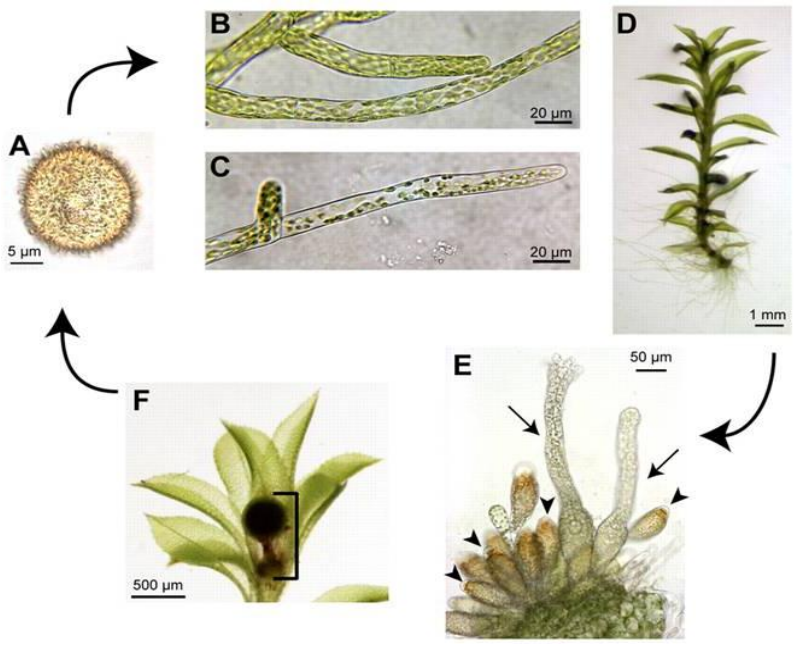


Figure 8. Life cycle of *P. patens*. (A) A haploid spore germinates into (B) chloronemal cells, which differentiate into (C) caulonemal cells. (D) Gametophore, emerge from protonemal filaments. (E) Both female, archegonia (arrows), and male, antheridia (arrow heads), organs form at the top of gametophore. A motile flagellate sperm fertilizes the egg and the (F) sporophyte (marked with a bracket) develops at the apex of the gametophore. Modified figure from (Prigge and Bezanilla 2010).

The body plan of multicellular organism is determined by the number of planes during cell division. For example, in *P. patens,* division in one plane leads to the formation of unbranched filaments while division in two planes leads to branched filaments and division in three planes leads to rise of gametophores (Niklas 2000). The changes in the body plan in *P. patens,* for example growth from two-dimensional to three-dimensional is determined at the single cell level (Harrison et al. 2009). During development of *P. patens*, eight (8) types of self-differentiate cells (stem cells) are formed (Kofuji and Hasebe 2014). During development, spores released from the sporophyte undergo division to form chloronema apical stem cell that can expand through tip growth to produce filamentous body (Cove and Knight 1993; Menand et al. 2007). The expansion of chloronema apical stem cell produce caulonema apical stem cell that develops into caulonema cells by tip growth (Cove and Knight 1993; Menand et al. 2007). The developed caulonema cells form side branch initial cells in which less than 5% of the cell develop into bud apical stem cells while the rest develops into caulonemal filaments and non-dividing cells (Cove and Knight 1993). The bud apical stem cells produce cells that leads to the formation of leaf apical stem. The epidermal cells of the leafy shoot cells (gametophores) develops filamentous rhizoids with rhizoid apical stem cells (Sakakibara et al. 2003). Under low temperature (15 °C), low light intensity (20 µmol/m^2/s) and short day conditions (8 hours), the tip of gametophore shoot produce male (antheridia) and female (archegonia) reproductive organs from stem cells (Hohe et al. 2002). After the formation of male and female reproductive organs, antheridia swimm towards archegonia for fertilization and produce a zygote. The resulting zygote divide asymmetrically in its first cell divison to produce cytoplasm-rich apical cell (sporophyte apical stem cell) and a basal cell (Tanahashi et al. 2005).

1.5.2 Leaf development

Development of body plan of plants and correct patterning of all organs merely depends on orientation of asymmetric cell division; while misorientation of cell divison pattern leads to abnormal morphogenesis (Smith et al. 1996; Heidstra 2007). In flowering plants, stems cells are formed in diploid generation and maintained in root meristem and shoot meristem, specifically in peripheral zone and as a result of repeated cell division leads to initiation of all lateral organs (Yruela 2015). In moss plants, a gametophore bud generated from caulonema stem cell play a role as meristematic cell and generates bushy leafy shoot

(gametophores) which contains phyllids (Kofuji and Hasebe 2014). In *P. patens*, after a series of asymmetric division of the bud initial cell, a tetrahedral initial cell is established on its apex (Harrison et al. 2009). The establieshed tetrahedral initial cell start to divide and their daughter cell bulge out and commence growth as leaves. The phytohormone auxin has been shown to be important throughout plant growth and development, from embryo to post-embronic development (Finet and Jaillais 2012). The final shape and complexity of the leaf depends on the balance between different members of proteins family, KNOX-PIN-CUC (Hepworth and Pautot 2015).

1.5.3 Homologous recombination

Development of reverse genetic approaches has facilitated rapid progress in the study of gene function. Using different approaches, the functions of different genes in plants are studied by altering the gene of interest and then analyze the mutant phenotype. In *P. patens*, the most popular method to study gene function is to utilize the organism's capability to insert, by homologous recombination (HR), foreign DNA into the genome (Strepp et al. 1998; Puchta 2002). In genomes, two different methods has been identified to alter covalent linkage in DNA, homologous recombination and non-homologous end joining (Puchta 2002). Gene targeting approach, via HR, involves replacement of a desired gene region with a selection cassette by the use of flanking sequences which is homologous to the targeted region (Figure 9) (Frank et al. 2005; Kuwayama 2012). This method is commonly employed to produce knock-out mutants by removing entire genes, but can also be used to replace or remove specific regions of the desired gene, there by producing mutants harbouring site-directed mutations in the gene of interest.

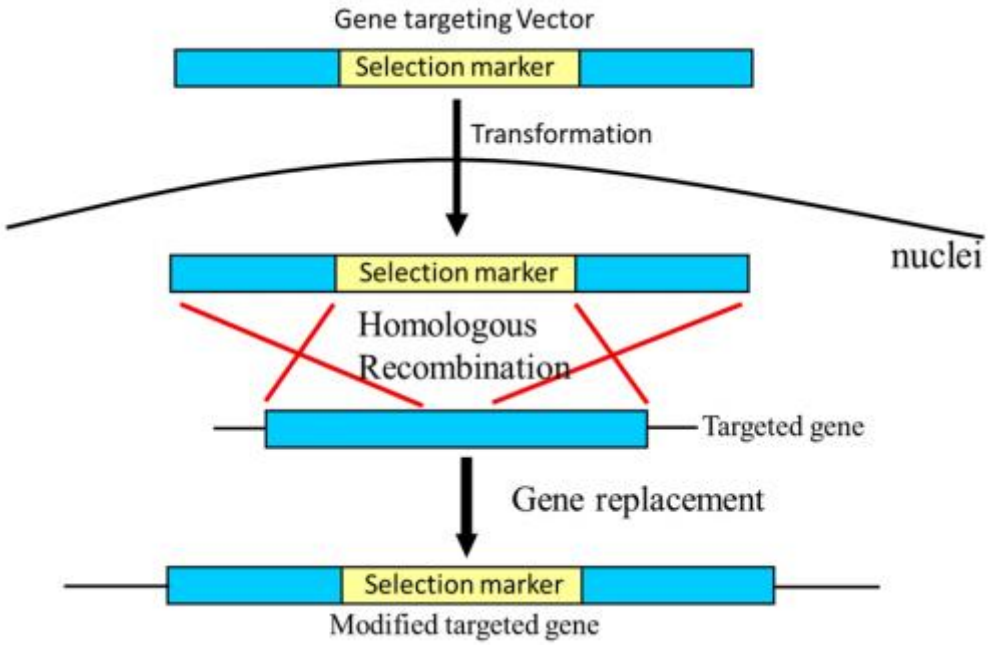


Figure 9. The principle of gene targeting.The gene targeting vector harbouring a selection marker with the flanking sequences is transformed into the nuclei of the cell and recombine with targeting gene via HR to replace targeted gene generating modified targeted gene. Figure adopted from (Kuwayama 2012).

In the moss *P. patens*, homologous recombination has shown to be the dominant way to transform exogenous DNA to the genome (Schaefer 2001), with a transformation efficiency of up to 100 % being reported (Kamisugi et al. 2005), compared to 10^{-4} in flowering plants (Britt and May 2003). The transformation efficiency by HR in the moss has been postulated to be due to predominance of the gametophytic haplophase of this organism (Schaefer and Zrÿd 1997). Also the presence of non-intronic RAD51 genes in *P. patens* compared to RAD51 gene in other eukaryotes has been associated with the high frequency of transformation by HR in the moss (Markmann-Mulisch et al. 2002).

1.6 Aim of the study

The relationship between gene structure and function has been widely deduced through analysis of the mutant lines of the desired gene. DEK1 of land plants is a member of one of four ancestral calpain variants with high conservation for at least 450 million years. Multiple sequence alignment (MSA) of DEK1- LG3 sequences in land plants and the ConSurf server were used to predict several conserved residues in *P. patens* DEK1- LG3 domain. The aim of this study is to assess the importance of conserved residues in DEK1-LG3 domain for DEK1 function in *P. patens*. This was done by substituting highly conserved residues of DEK1-LG3 domain predicted to be functionally important to alanine (Ala) and retarget the mutegenized version of DEK1-LG3 into *DEK1*Δ*LG3* locus. The resulting mutants were characterized by PCR-based genotyping, Southern blot hybridization, RT-PCR, sequencing, bright-field and dissecting microscopy techniques to assess the importance of these amino acids for *P. patens* development.

2. Material and Methods

2.1 Bioinformatic analyses of the DEK1-LG3 domain

The DEK1 sequences used in the present study are listed in Appendix 1. The sequences corresponding to DEK1-LG3 domain were extracted using the Simple Modular Architecture Research Tool (SMART) database (http://smart.embl-heidelberg.de). Multiple sequence alignment (MSA) of the DEK1-LG3 sequences was constructed using MAFFT v7 with default parameters (http://www.ebi.ac.uk/Tools/msa/mafft/). The resulting MSA was subsequently analyzed using CLC Main Workbench v6 (www.clcbio.com). The DEK1-LG3 sequence logo, a graphical representation of sequence conservation of the amino acids, was generated using the WebLogo tool (http://weblogo.berkeley.edu/logo.cgi). Conserved residues in the *P. patens* DEK1-LG3 domain were predicted using the ConSurf server, which were run in the "ConSeq" mode (http://consurf.tau.ac.il/index_proteins.php). For predicting the three-dimensional structure of the DEK1-LG3 domain the amino acids sequence corresponding to the LG3 domain (residues 1442-1609; XP 001774206.1) was submitted to Phyre2 fold recognition server. Structural alignment was performed using the PDBeFold server (http://www.ebi.ac.uk/msd-srv/ssm/cgi-bin/ssmserver) and superimposed structures were viewed using PyMOL software (v 0.99).

2.2 Construction of vectors for gene targeting

In this study, five different plasmid construct for targeted insertion of mutagenized version of the DEK1-LG3 were constructed, in addition to one vector that contained the WT sequence (Table 1). Plasmids were constructed using In-Fusion cloning technology as outlined in Figure 10.

Table 1. Plasmid constructs made in the present study. The name of each plasmid, WT sequence and mutated residues are given.

Plasmid name (pBHRF_JI_LG3mut)	WT sequence	Mutated residues
$E^{1477}Q^{1478}$	^{1472}LVAGAEQGLEAG	LVAGAEQGLEAGQV
E1481	^{1476}AEQGLEAGQVG	AEQGLEAGQVG
S1497	^{1492}KGAGQSTHNRE	KGAGQSTHNRE
R1514	^{1509}CVADGRWHSVT	CVADGRWHSVT
$G^{1574}R^{1575}S^{1576}D^{1577}S^{1578}E^{1579}$	^{1569}DLDAFGRSDSEGAESK	DLDAFGRSDSEGAESK
pBHRF_JI_LG3		

2.2.1 Generation of 5'_TGS, cloning and Mutagenesis reaction

First, the 5' targeting sequence (5'_TGS) was PCR amplified from genomic DNA extracted from WT *P. patens* tissue using Phusion® High-Fidelity PCR kit (Thermo SCIENTIFIC, Cat ≠HF-549L). In a 50 µl reaction volume, 50 ng genomic DNA was PCR amplified with 1X HF buffer, 200 µM of dNTPs, 0.02 units/µl of Phusion DNA polymerase and 0.5 µM each of gene specific primers JI_5TGS_SP_Info and JI_5TGS_ASP_Info (Appendix 3A). PCR amplification was initiated with an initial denaturation step at 98 °C for 30 sec., followed by 40 cycles of 98 °C for 10 sec., 64 °C for 30 sec., and 72 °C for 20 sec. and a final extension step at 72 °C for 5 min.

The PCR amplified 5'_TGS was cloned to zero Blunt pCR vector using the Zero Blunt® PCR Cloning Kit (Invitrogen, Cat ≠44-0302) generating vector pCR_5TGS. In a 10 µl reaction volume, 0.5 µl of PCR amplified 5'_TGS were ligated with 1 µl of Zero Blunt PCR plasmid vector (10 ng) using 1 µl ExpressLink™ T4 DNA Ligase (5 U) and 2 µl of 5X ExpressLink™ T4 DNA Ligase Buffer. The reaction mixture was incubated at RT for 30 min. and then transformed into 50 µl of One Shot® TOP10 cells (Invitrogen Cat ≠C404003) following manufacturer's instructions. The mixture of competent cells and ligation reaction was incubated on ice for 30 min. and then subjected to heat shock at 42 °C for 45 sec. Transformed cells were cooled on ice for 2 min. and then mixed with 250 µl of RT SOC

medium (Appendix 8H). The cell culture was incubated at 37 °C with shaking (225rpm) for 1 hr and then overnight on LB-agar medium supplemented with 50 µg/ml Kanamycin.

Methylation and mutagenesis reactions were used to introduce mutations into the vector pCR_5TGS, using GeneArt® Site-Directed Mutagenesis PLUS kit (Invitrogen, Cat ≠A14551) generating the vectors collectively named pCR_5TGSmut. In a 20 µl reaction volume, 50 ng pCR-5TGS were PCR amplified with 0.5 µM each of gene specific primers, 1X of AccuPrime™ *Pfx* reaction buffer, 1X of Enhancer, 4.8 U of DNA methylase, 1X of SAM and 1.5 Units of AccuPrime™ *Pfx*. The gene specific primers used were F8_9 SP and F8_9 ASP; F10E_SP and F10E_ASP; F14_SP and F14_ASP; F16_SP and F16_ASP, and F23_28 SP and F23_28 ASP (Appendix 3C). PCR cycling was performed with an initial methylation step at 37 °C for 20 min., and then a two steps PCR with initial denaturation step at 95 °C for 2 min. followed by 35 cycles of 95 °C for 15 sec., 68 °C for 6 min. and a final extension at 72 °C for 5 min. The mutagenesis reactions were transformed into One Shot® MAX Efficiency® DH5α ™-T1R Competent Cells (Invitrogen Cat ≠44-0097) as described above. Inserts 5'_TGSmut were then PCR amplified from pCR_5TGSmut using the primers JI_5TGS_SP_Inf and JI_5TGS_ASP_Inf (Appendix 3D) as described above.

2.2.2 Generation of 3'_TGS and cloning

The 3' targeting sequence (3'_TGS) was PCR amplified from genomic DNA extracted from WT *P. patens* as described in section 2.2.1 except using the primers JI_3TGS_SP_Info and JI_3TGS_ASP_Info (Appendix 3A) and an annealing temperature of 64 °C for 20 sec.

Vector pBHRF was linearized by "Inverse PCR" using Phusion® High-Fidelity PCR kit as described in section 2.2.1 except using the primers pBHRF_1_SP and pBHRF_1_ASP (Appendix 3B) and PCR cycling with an initial denaturation at 98°C for 30 sec. followed by 40 cycles of 98 °C for 10 sec., 62 °C for 20 sec. and 72 °C for 2 min. and a final extension step at 72 °C for 7 min. The linearized vector pBHRF was then gel purified using QIAquick gel extraction kit (QIAGEN, Cat ≠28704) following manufacturer's instructions.

PCR amplified 3'_TGS was then cloned into the linear vector pBHRF generating vector pBHRF_JI_3TGS. In brief, the PCR amplified 3'_TGS was first purified using cloning enhancer, where 5 µl of the PCR amplified 3'_TGS were treated with 2 µl of the cloning

enhancer and incubated at 37 °C for 20 min., and then at 80 °C for 20 min. In-Fusion cloning reaction was then performed in a total volume of 10 µl with 2 µl of purified PCR product 3'_TGS with 0.5 µl of linearized vector pBHRF using 2 µl of 5X In-Fusion HD Enzyme Premix. The mixture was incubated at 50 °C for 15 min. and then transformed as described in section 2.2.1 but using the Stellar™ Competent cells (Clontech, Cat ≠636763) and incubated overnight on LB-agar medium supplemented with 100 µg/ml Amplicilin.

The resulting vector pBHRF_JI_3TGS was linearized using 2 µl of XhoI with 5 µl of plasmid DNA and 2 µl of CutSmart® buffer in a total reaction volume of 20 µl and then incubated at 37 °C for 4 hrs.

Inserts 5'_TGSmut and 5'_TGSWT were cloned into XhoI digested pBHRF_JI_3TGS and then transformed into Stellar™ Competent cells as described in section 2.2.1; resulting in the final vectors pBHRF_JI_LG3mut and pBHRF_JI_LG3, respectively.

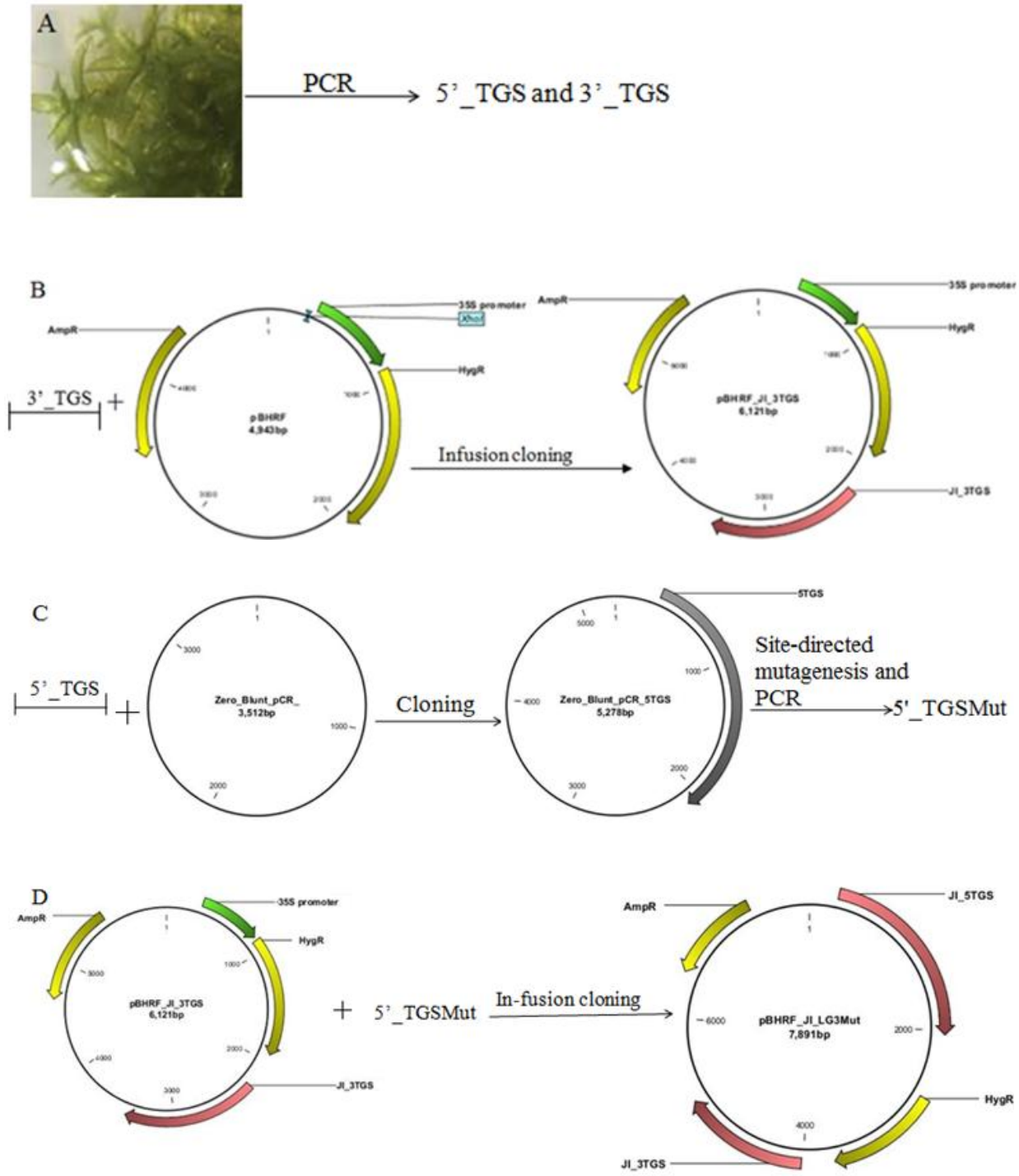


Figure 10. Cloning strategy. (A) PCR amplification of 5'_TGS and 3'_TGS from *P. patens* WT, (B) Infusion cloning of 3'_TGS into vector pBHRF to generate vector pBHRF_JI_3TGS, (C) Cloning of 5'_TGS into pCR vector to generate vector pCR-5TGS, Site directed mutagenesis in vector pCR-5TGS and PCR amplification of inserts 5'_TGSMut (D) Infusion cloning of vector pBHRF_JI_3TGS with 5'_TGSMut to generate final vector pBHRF_JI_LG3Mut.

2.3 Plasmid isolation and Construct verification

2.3.1 Colony PCR

In order to screen for putative *E. coli* colonies harbouring plasmid with cloned insert, "colony PCR" was performed. In a 25 µl reaction volume, a small amount of the *E. coli* colony was PCR amplified with 0.2 µM each of primers Armseq7, 35s_Rev1 (to screen for putative pBHRF_JI_LG3Mut), JI_3TGS_SP_Inf and JI_3TGS_ASP_Inf (to screen for putative pBHRF_JI_3TGS) (Appendix 3E), 200 µM of dNTP's, 2.5 mM of $MgCl_2$, 0.05 units/µl of HOT FIREPol DNA polymerase, and 1X of Buffer B1. Amplification was initiated at 95 °C for 15 min. to activate the enzyme then followed by 30 cycles of 95 °C for 45 sec., 58 °C for 45 sec., and 72 °C for 80 sec. and a termination step of 72 °C for 5 min.

2.3.2 Plasmid isolation

PCR-positive colonies were cultured in 4 ml LB medium containing 50 µg/ml Kanamycin (for pCR-5TGS constructs) and 100 µg/ml Amplicilin (for pBHRF_JI_3TGS construct and pBHRF_JI_LG3mut) at 37 °C with shaking (225 rpm) overnight. Plasmid was isolated from the overnight culture using PureYield™ Plasmid Minprep System kit (Promega Cat ≠A1222) following the manufacturer's instructions.

2.3.3 Restriction Digestion Analysis

All generated constructs were verified by restriction digestion analysis. In a 10 µl reaction volume, 500 ng of plasmid DNA were digested with 1 µl of EcoRV enzyme in 1X reaction buffer (3.1 (NEB)) and incubated for 2 hrs at 37 °C. The restriction digestion reactions were analysed by electrophoresis using 0.5 % agarose gel.

2.3.4 Cycle sequencing and capillary electrophoresis

All generated constructs were verified by cycle sequencing and capillary electrophoresis using the BigDye® terminator v3.1 cycle sequencing kit (Applied Biosystems, Cat

≠4336915). In a 10 µl reaction volume, 250 ng of plasmid DNA were cycle sequenced with 3.2 ρmol each of gene specific sequencing primers TER_F, pBHRF_R, ASP-PpARM-Info, ARM 3'_F, CALP Seq1, EX25-F, ΔARM 3'_ R, T7, M13R, ArmSeq6, CSMW_5R, pBHRF_F, 35s_Rev1, and Armseq7 (Appendix 3F), 2 µl of 5X BigDye sequencing buffer and 0.5 µl of BigDye® terminator v3.1. Cycle sequencing steps were performed according to the Stepped elongation Time Protocol (Platt et al. 2007).

Cycle sequencing steps were followed by purification of the extension products to remove excess primers, dNTP, and ddNTP. To the sequencing reactions, 10 µl of dH_2O, 2 µl of 125 mM EDTA, 2 µl of 3M NaOAc (pH 5.2) and 52 µl of 96 % EtOH were added. The reaction mixture was incubated at RT for 15 min. and DNA was precipitated at maximum speed (14800 rpm) in a Thermo Scientific Heraeus Fresco 21 Refrigerated microcentrifuge (Thermo Scientific, EW-17703-30) at 4 °C for 30 min. The DNA pellet was washed with 70 µl of 70% EtOH and centrifuged (Thermo SCIENTIFIC, HERAEUS PICO21) at maximum speed (14800 rpm) for 10 min. The DNA was air dried for 30 min. to remove excess EtOH and resuspended in 10 µl of deionized formamide and finally sequenced by Capillary Electrophoresis using the 3130xL Genetic Analyzer (Life Technologies). CLC Main Workbench v6 (www.clcbio.com) was used to analyze the sequences.

2.4 PEG-mediated *Physcomitrella patens* transformation

P. patens protoplast isolation and transformation was performed as described by Cove et al. (2009) using 15 µg of linearized plasmid DNA per transformation. Prior to transformation, the plasmids (Table 1) were completely digested with restriction enzymes PacI and BmrI, and then the digested DNA was precipitated by standard ethanol precipitation. The plasmids were then transformed into the *P. patens DEK1ΔLG3* locus (Johansen W, - manuscript in preparation). Protoplasts were grown under long day conditions [16 hrs light (70-80 µmol/m^{-2}/S^{-1}) and 8 hrs dark]. Regeneration and selection of transformed plants was performed by transferring the tissue cultures to different media in the following sequence; 7 days of protoplast regeneration in protoplast regeneration medium (PRM) (Appendix 7B), 7 days of selection on BCDA medium (Appendix 7A) supplemented with 20 µg/ml of hygromycin, 14 days of growth on BCDA medium and 7 days on BCDA supplemented with 20 µg/ml of hygromycin. Hygromycin resistance plants were then transferred to fresh BCDA medium.

Cre recombinase procedure was performed to remove the hygromycin resistance cassette from the transformants. Protoplast was isolated and transformation was carried out using 20 µg of pAct-Cre plasmid (Trouiller et al. 2006). Protoplast were diluted four times (25,000 counted protoplasts per petri dish) to avoid mixing during picking and plated on Petri dishes containing cellophane. Protoplast regeneration and selection of putative positive Cre lines were performed as follow; 7 days of protoplast regeneration on protoplast regeneration medium, 7 days of growth on BCDA medium, picking of individual plant colony on BCDA medium and grow for 7 days, and finally, replica picking of individual plant colony on BCDA medium supplemented with 20 µg/ml of hygromycin. Lines showing loss of resistance cassette were grown on BCDA medium, and analyzed.

2.5 Molecular characterization of mutants

2.5.1 Genotyping of transformants

Putative transformed and hygromycin resistance plants were PCR genotyped using the Phire Plant Direct PCR Kit (Thermo scientific, Cat ≠F-130WH) following the manufacturer's instructions. Three rounds of PCR genotyping were performed. First PCR genotyping was performed to screen for insertion of the DEK1-LG3 sequence using primers LG3_Fw and LG3_Rev. Second PCR genotyping was performed to select for line harbouring single-copy insertion at the locus using primers ArmSeq3 and EX30_R. In order to confirm proper 5' and 3' targeting a third PCR genotyping was performed using primers ARMSeq3 and 35sRev1, and Term_Fw and EX30_R, respectively. Genomic DNA was isolated from hygromycin resistant plants by crushing a small amount of plant tissue in 20 µl of Dilution Buffer, which was used as a template for PCR. In a 20 µl reaction volume, 0.5 µl of DNA was PCR amplified with 0.5 µM each of gene specific primers (Appendix 3G), 0.4 µl of Phire Hot Start II DNA polymerase and 2X Phire Plant PCR Buffer. PCR amplification were as follow: initial denaturation at 98°C for 5 min. then followed by 30 cycles of 98 °C for 5 sec., 58 °C for 5 sec. and 72 °C for 1 min. (first PCR genotyping), 72 °C for 3.30 min. (second PCR genotyping) and 72 °C for 2.30 min. (third PCR genotyping). The final extension step was carried out at 72 °C for 5 min.

2.5.2 Southern Blot

Southern blot analysis was performed to investigate the possibility of off target integration into the *DEK1* locus. Genomic DNA for Southern-blot analysis was extracted using the PhytoPure Genomic DNA Extraction Kit (GE HealthCare, Cat ≠RPN-8511) according to the manufacturer's instructions. Approximately 1 µg of genomic DNA was digested using *Bgl*II for 6 hrs. Fragmented DNA was separated by agarose gel electrophoresis using 0.6 % ultrapure agarose at 37° C for 18 hrs. The DNA was transferred to positively charged nylon membrane (Roche), and then pre-hybridized using DIG Easy Hyb (Roche) at 42 °C for 30 min. and finally probed using 5' and 3' target probes at 42 °C overnight. The probes were synthesized using PCR DIG probe synthesis kit (Roche, Cat≠11636090910) and primers pair Armseq5/CSMW_5R and 3TGS_SP/3TGS_ASP (Appendix 3I) for the 5' and 3' target probes, respectively. Development of the Southern blot was performed using X-Ray films.

2.5.3 RT-PCR

RT-PCR and subsequent sequencing of the product were performed to analyse the *DEK1* transcript to verify the introduced mutations and to investigate if the transcript were properly spliced. Total RNA was isolated from *P. patens* mutant tissue using the RNAqueous®-Micro kit (Life technologies, Cat≠AM 1931) according to the manufacturer's instructions. 8 µl of RNA was treated with 1 µl of 100 units DNase I-Amplification Grade (Invitrogen, Cat ≠18068-015) and 1 µl of 10X DNase I buffer. The reaction was incubated at RT for 15 min., then 1 µl of 25 mM EDTA was added to inactivate DNase I and incubated at 65 °C for 15 min. 8 µl of DNase I-treated total RNA was primed with 1 µl of 50 µM Random hexamer primers and 1 µl of 10 mM dNTP in a total volume of 13 µl. The reaction was incubated at 65 °C for 5 min. and placed on ice for 1 min. The reaction mixture was reverse transcribed using 1 µl of 200 units of Superscript III® Reverse Transcriptase (Invitrogen, Cat ≠18080-044), 1 µl of 40 units of RNaseOUT™ (Invitrogen, Cat ≠10777-019), 4 µl of 5X First-strand buffer and 1 µl of 0.1 M DTT. The reaction was incubated at RT for 5 min. and then at 50 °C for 1 hr. Finally, the Reverse transcriptase enzyme was inactivated by incubation at 70 °C for 15 min. Phusion® High-Fidelity DNA polymerase was used to amplify the target sequence spanning a region from exon 7 to exon 30 (transcript nucleotides 1888 to 6444) as follow: 1µl undiluted cDNA template was PCR amplified with primers PpLoop_Inverse SP and

Ex30_R (Appendix 3H) using the following cycling conditions: 98 °C for 10 sec., 40 cycles of 98 °C for 10 sec., 62 °C for 20 sec. and 72 °C for 1.5 min. and a final elongation step at 72 °C for 5 min. The PCR products was treated with Exonuclease I (Fermentas) to remove excess primers, and then cycle sequencing reactions were performed using primers LG3_R, ArmSeq7, ArmSeq8 and CSMW_5R and the ABI BigDye® terminator v.3.1 chemistry according to the StePP method (Platt et al. 2007). DNA fragments were precipitated using sodium-acetate/ethanol and finally sequenced by Capillary Electrophoresis using the 3130xL Genetic Analyzer. The CLC Main Workbench v6 software was used to analyse the sequences.

2.6 Microscopy

The phenotype of the mutants was characterized using dissecting (Nikon SMZ 1500) and fluorescence (Nikon ECLIPSE Ti) microscopy. The mutants generated in this study were grown side by side on BDC medium for 3 weeks, phyllids were dissected from the middle part of the gametophores, mounted in a glass slide and observed under dissection microscope.

To determine the growth and spreading of protonema tissue, generated mutants were grown side by side for 2 weeks and colonies were photographed directly by dissecting microscope while growing on BCD medium.

To observe buds initial developments, mutant tissue was stained in Fluorescent Brightener 28 for 30 min., washed two times in sterile water, then mounted in a glass slide and observed under fluorescence microscope.

3. Results

3.1 *In silico* analyses of DEK1-LG3

3.1.1 The N-terminal end and middle part of the DEK1-LG3 domain is conserved in land plants

In order to analyze the degree of sequence conservation, 94 DEK1-LG3 sequences of land plants were aligned using MAFFT (Appendix 5). As shown by the sequence logo representation (Figure 11), the N-terminal end and the middle part of the DEK1-LG3 domain are more highly conserved than the more divergent C-terminal end. Investigation of the multiple sequence alignment showed that the amino acid identity decreases with evolutionary distance. For example, the sequence identity between the angiosperm *A. thaliana* and mosses *P. patens* is 55%.

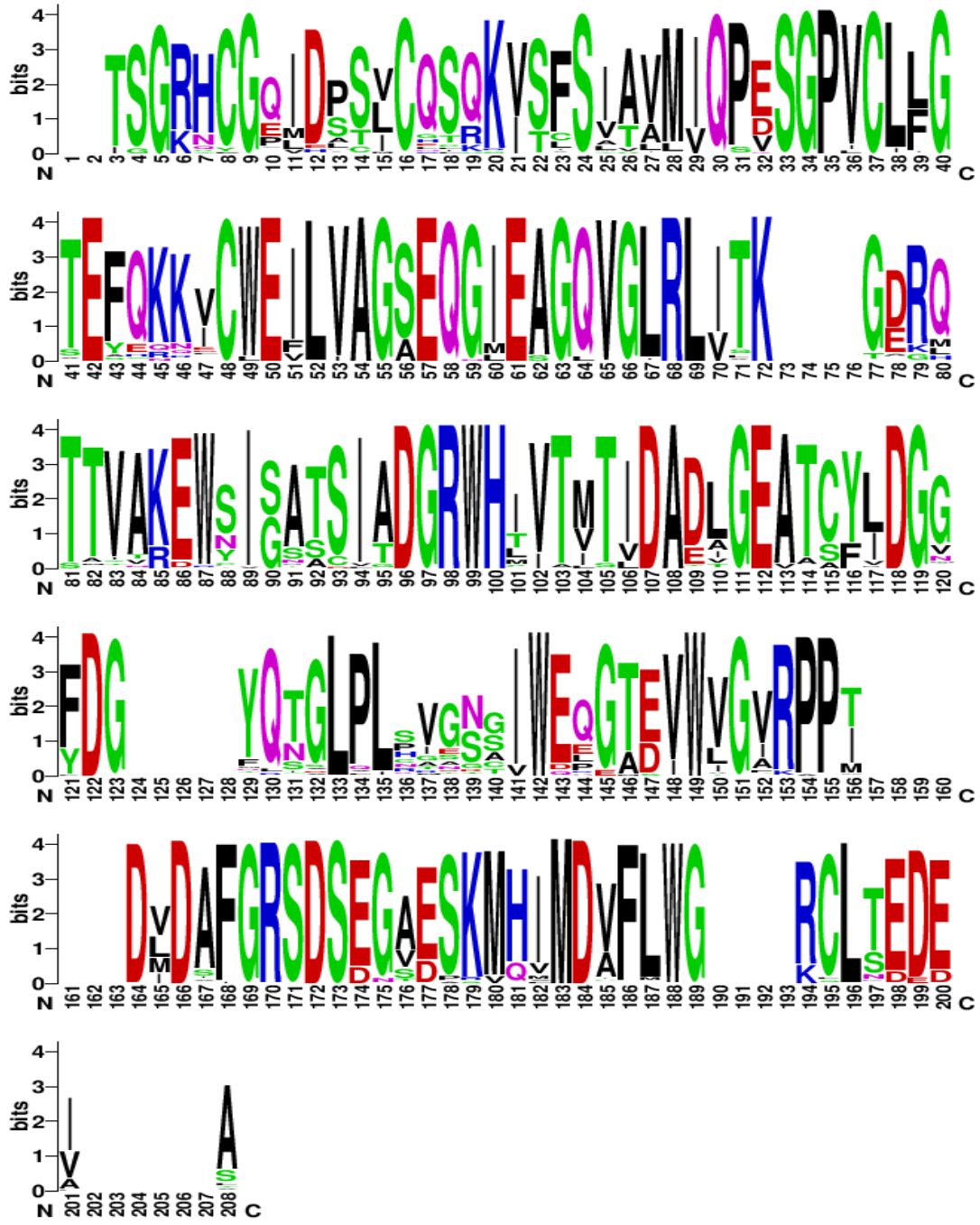


Figure 11. Graphical representation of land plant DEK1-LG3 sequence conservation (using WebLogo v3). The N-terminal end and the middle part are more highly conserved while the C-terminal end is more divergent. The colour scheme of the residues represent their chemical properties; Green = Hydrophilic, blue = basic, red = acidic and black = hydrophobic amino acids.

3.1.2 Predicted functional and structural residues in *P. patens* DEK1-LG3 sequence

The functionally and structurally important residues in the DEK1-LG3 domain were predicted by submitting the multiple sequence alignment to the ConSurf server, which estimates evolutionary conservation of residues based on the phylogenetic relationship between homologous sequences (Armon et al. 2001). Figure 12 shows the output result of the prediction; functional residues (marked ‘‘f”) are highly conserved and exposed while predicted structural residues (marked “s”) are highly conserved and buried inside the protein. The analysis shows that important functionally and structurally residues are evenly distributed throughout the sequence.

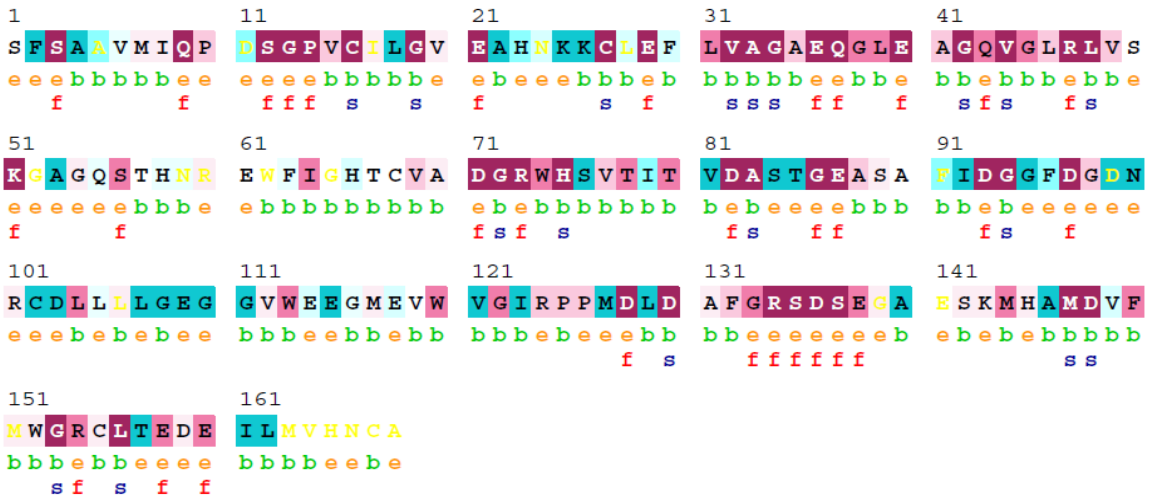


Figure 12. Predicted functional and structural residues in DEK1-LG3 domain using ConSurf server. The conservation of functionally and structurally predicted residues is evenly spread throughout the sequence. Conserved functional and structural residues are marked “f” and “s” respectively, while “b” and “e” represent buried and exposed residues, respectively.

3.1.3 The predicted *P. patens* DEK1-LG3 3D structure adopts a fold similar to the lectin fold

The 3D structure of the DEK1-LG3 domain was predicted by submitting the *P. patens* DEK1-LG3 sequence (10618-11230; XP 001774206.1) to the Phyre2 fold recognition server. The predicted model is based on C3flpJ, the SAP-like pentraxin from *Limulus polyphemus*. DEK1-LG3 is predicted to contain ten (10) anti-parallel β-sheets (yellow) which are arranged in concave and convex shape, two (2) α-helix (red) and loop regions connecting β-sheets on both edges (Figure 13A). The predicted structure form a structural

fold which is similar to the β-sandwich or lectin fold, formed by two anti-parallel β-sheets (Rudenko et al. 2001). The surface view of the model was modified using PyMOL structure viewer (v0.99) software to visualize the location of the predicted functionally important residues (Figure 13B) on the surface of the structure. The surface positions marked blue show the positions of the residues that were mutated in the current study. The mutated residues are also given.

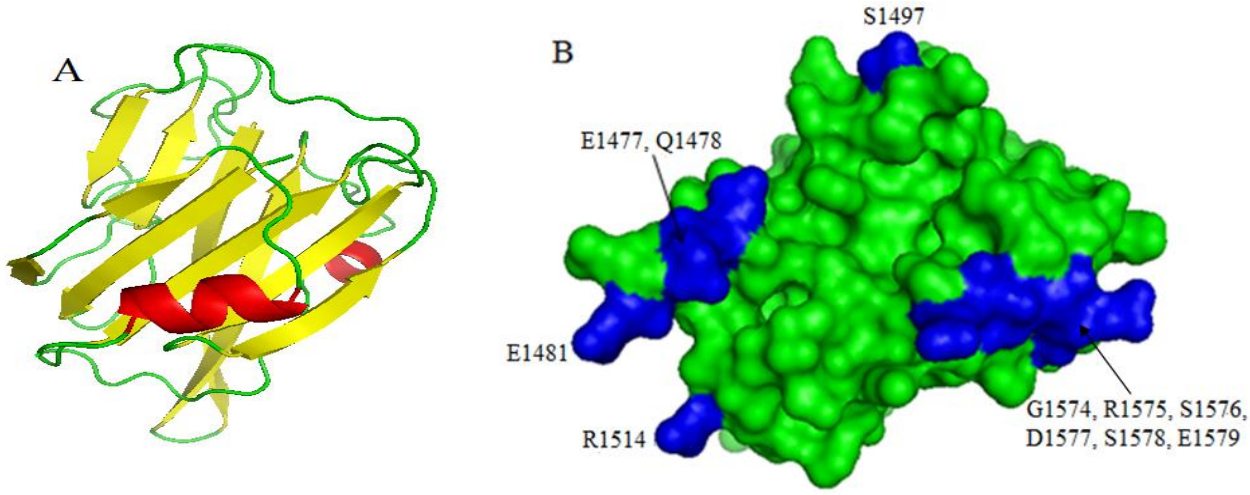


Figure 13. The predicted 3D structure of *P. patens* DEK1-LG3 domain using Phyre2 server (A) Cartoon representation of the 3D structure with 10 anti-parallel β-sheets (yellow), 2 α-helix (red) and loop regions connecting the β-sheets on both edges. (B) Surface view of the 3D structure indicating the location of mutated residues. The positions marked blue show the location of the residues that were mutated in the current study, namely $E^{1477}Q^{1478}$, E^{1481}, S^{1497}, R^{1514} and $G^{1574}R^{1575}S^{1576}D^{1577}S^{1578}E^{1579}$. The surface view was modified using PyMOL (v0.99) structure viewer.

The obtained DEK1-LG3 3D structure was pairwise superimposed to the database model 3FLP using the PDBeFold server, which shows that the two structures display a similar fold (Figure 14). The structural homology search identified a root mean square deviation (RMSD) of 0.718 Å between the structures, which also shows that the structures of the two proteins are similar. The structure of heptameric SAP-like pentraxin has been reported to have a fold resembling the Concanavalin A superfamily of proteins. Proteins in this family have a highly conserved fold despite low sequence similarity (Emsley et al. 1994; Shrive et al. 2009).

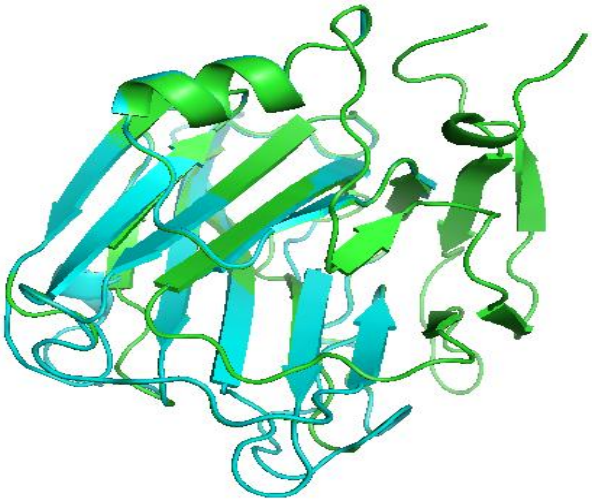


Figure 14. The superimposed structural alignment between native heptameric SAP-like pentraxin from *Limulus polyphemus* (PDB, 3FLP) (green) and the predicted structure of DEK1-LG3 domain using PDBeFold (light blue).

3.2 Molecular characterization of *P. patens* mutants

3.2.1 Analysis of *P. patens* transformants show targeted on-locus events

To assess the dependence of the residues predicted to be functionally important in the LG3 domain (Figure 13B), constructs carrying mutagenized versions of the LG3 sequences (Table 1) were re-targeted to the *DEK1ΔLG3* locus. Figure 15 shows the schematic of recombination strategy employed in PEG-mediated *P. patens dek1Δlg3* and Cre recombinase transformations. Table 2 shows the number of stable transformants (hygromycin resistant), obtained in the experiments. Transformation of construct pBHRF_JI_LG3G^{1574}R^{1575}S^{1576}D^{1577}S^{1578}E^{1579} did not give any stable transformant even after several rounds of transformation. The hygromycin resistant plants were PCR-genotyped to identify lines with proper targeting as shown schematically in Figure 16 and the result of this genotyping is provided in Table 2. As an example, the PCR genotyping result for 5' and 3' targeting is shown in Figure 17. This result confirms that DEK1-LG3 sequences were successfully re-targeted to the *DEK1ΔLG3* locus.

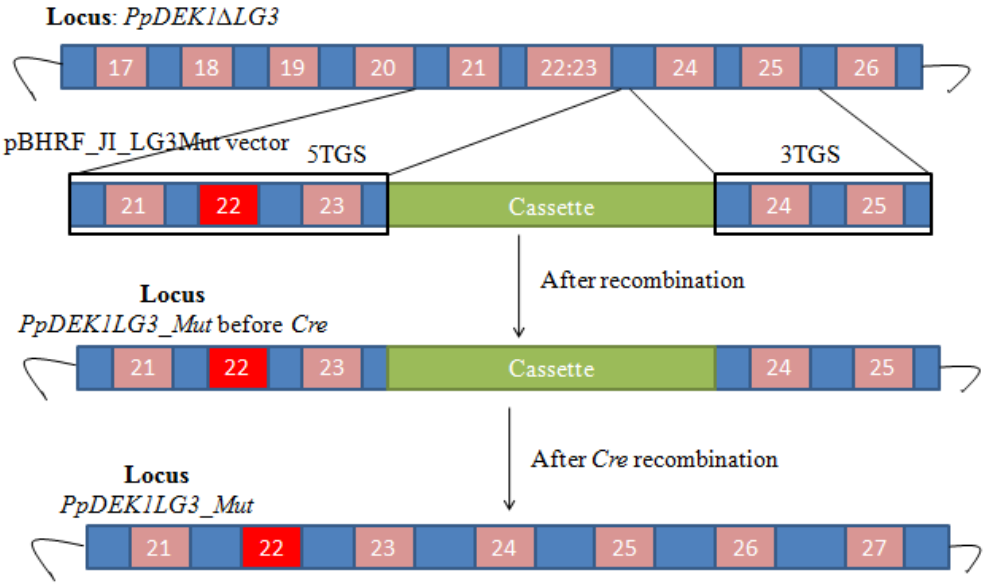


Figure 15. Homologous recombination strategy after PEG-mediated transformation of *Physcomitrella patens*. The mutant locus *PpDEK1ΔLG3* was transformed with vectors pBHRF_JI_LG3Mut creating mutant loci collectively named *PpDEK1LG3_Mut* before Cre. After transformation to remove the resistance cassette, *PpDEK1LG3_Mut* loci were generated. The number in the boxes corresponds to the *P. patens* DEK1 exons. Exon 22 (red) contains the mutations.

Table 2. Number of stable transformants, transformants re-targeted with DEK1-LG3, transformants with single copy insertion, and double targeting transformants obtained after PEG-mediated transformation of each construct.

Mutants	Number of stable transformants (HygR)	Number of transformants with DEK1-LG3 insertion	Number of transformants with single-copy insertion	Number of double targeting
dek1lg3:E^{1477Q1478}	9	4	0	3
dek1lg3:E^{1481}	12	8	2	5
dek1lg3:S^{1497}	44	16	2	8
dek1lg3:R^{1514}	19	8	1	3
dek1lg3: $G^{1574}R^{1575}S^{1576}D^{1577}S^{1578}E^{1579}$	0	0	0	0
dek1-wtlg3	6	5	0	1

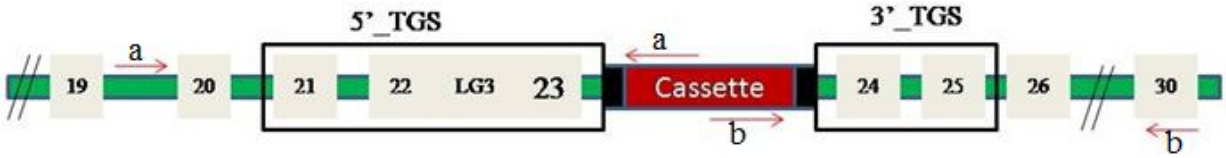


Figure 16 Schematics of the PCR genotyping strategy. Red arrows show primer annealing sites; primers marked "a" and primers marked "b" was used for 5' and 3' targeting genotyping PCR, respectively.

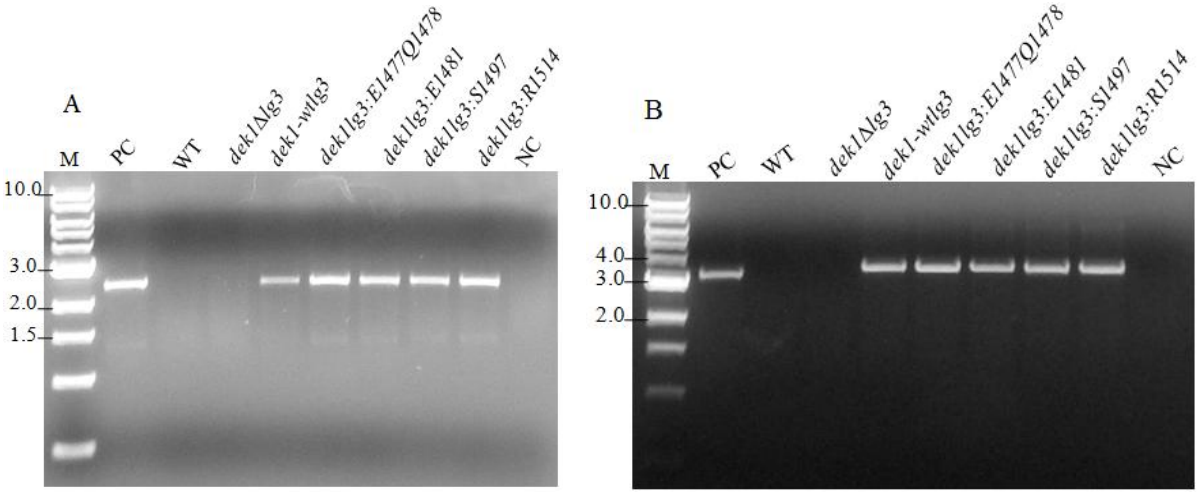


Figure 17. Genotyping of *Physcomitrella patens dek1lg3_mut*. Gel photos showing amplicons generated from PCR genotyping of 5' (A) and 3' (B) targeting by electrophoresis on 0.5% agarose. Positive control sample denoted as PC represents DNA extracted from the *P. patens dek1Δlg3* mutant complemented with *Marchantia polymorpha* DEK1-LG3. The negative control (NC) contains no template; WT and *dek1Δlg3* were also used as negative control. The expected fragment size for *dek1-wtlg3*, *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481}, *dek1lg3:S*^{1497} and *dek1lg3:R*^{1514} mutant lines are approximately 2.4 kbp and 3.4 kbp, respectively.

PCR-based genotyping suggest that *dek1lg3:E*^{1481}, *dek1lg3:S*^{1497} and *dek1lg3:R*^{1514} mutants harbour a single copy insertion of DNA at the targeted locus and that *dek1-wtlg3* and *dek1lg3:E*^{1477}*Q*^{1478} mutants harbour multicopy integration of DNA either at the targeted locus or off target locus (data not shown).

PCR positive double targeted lines, *dek1-wtlg3 #1*, *dek1lg3:E*^{1477}*Q*^{1478} *#1*, *dek1lg3:E*^{1481} *#1*, *dek1lg3:S*^{1497} *#1* and *dek1lg3:R*^{1514} *#1* were subjected to Cre mediated transformation to remove the resistance cassette and further analyzed by Southern blot to investigate the possibility of off target integration.

3.2.2 Southern blot analyses

Southern blot analysis was performed for PCR positive genotyped lines to investigate the possibility of off-locus integration of DNA during transformation. Figure 18 represents schematics of the Southern-blot in which the restriction fragments were generated using *Bgl*II.

The Southern blot analysis before Cre recombinase show mutant plants *dek1lg3:E*^{1481}, *dek1lg3:S*^{1497} and *dek1lg3:R*^{1514} have the two expected restriction fragments of size 15.8 kbp and 5.0 kbp (Figure 19A) suggesting that these lines harbour a single copy insertion of DNA at the targeted locus. Mutant plants *dek1-wtlg3* and *dek1lg3:E*^{1477}*Q*^{1478} display several restriction fragments which suggest that these lines harbour multicopy integration of DNA and also suggests the possibility of off locus targeting of the vector (Figure 19A). Southern blot analysis for mutant plants after Cre recombinase removal of the resistance cassette show that the mutant plants *dek1-wtlg3* and *dek1lg3:E*^{1481} have a single restriction fragment of size 18.8 kbp, the same size as the WT plant suggesting that these lines harbour a single copy integration of DNA at the locus (Figure 19B). However, several restriction fragments were detected in the mutant plants *dek1lg3:E*^{1477}*Q*^{1478} and *dek1lg3:S*^{1497} suggesting that these lines harbour off locus DNA integration (Figure 19B).

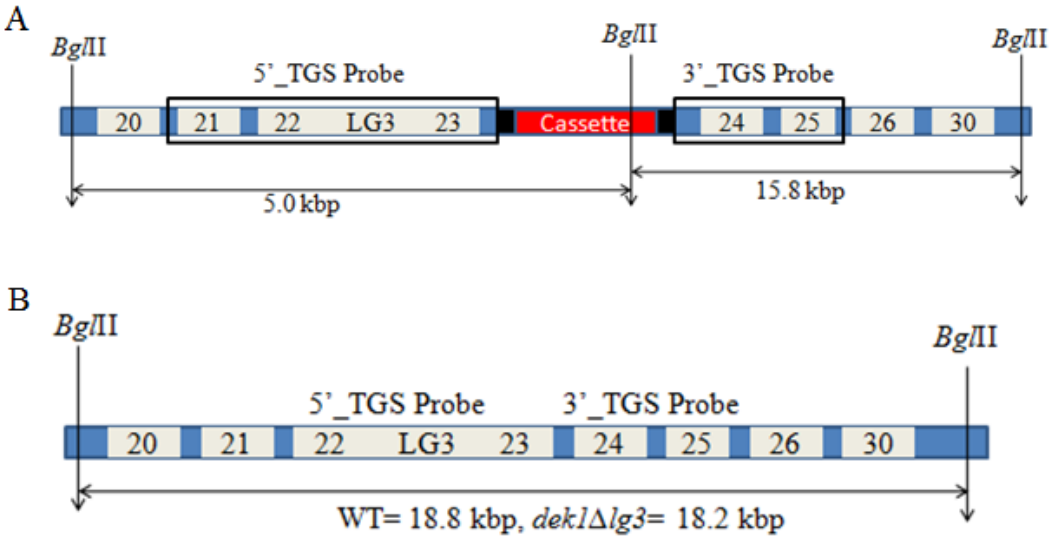


Figure 18. Schematics of southern blot strategy showing the expected restriction fragments for generated mutants (A) and WT and *dek1*Δ*lg3* mutant (B) using *Bgl*II. The expected restriction fragments for generated mutants using 5' and 3' targeting probes are 5.0 kbp and 15.8 kbp, respectively. The expected restriction fragments for WT plant and *dek1*Δ*lg3* mutant are 18.8 kbp and 18.2 kbp, respectively.

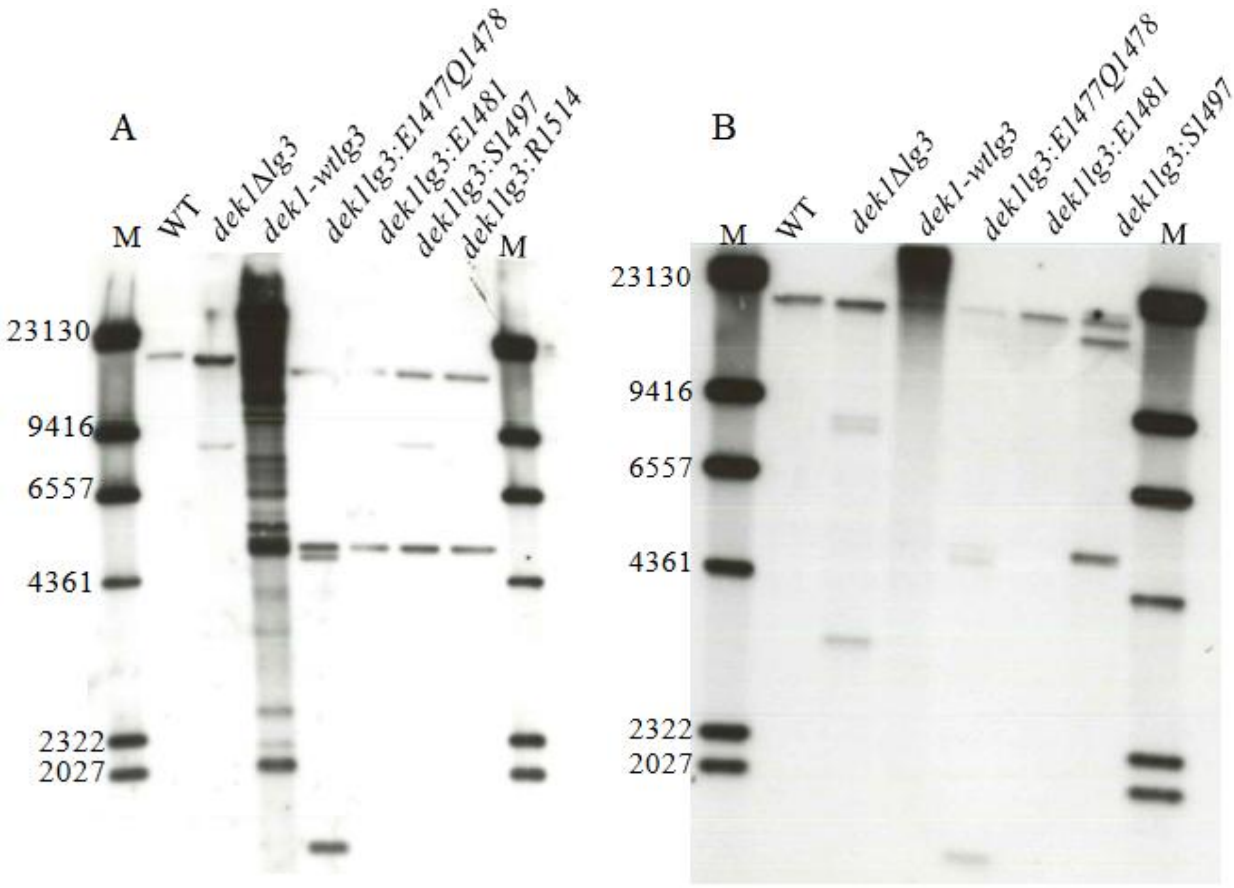


Figure 19. Southern blot analysis for generated mutants before (A) and after Cre recombinase removal of the resistance cassette (B). *Bgl*II was used to create restriction fragments. (A) The two expected restriction fragments of size 5.0 kbp and 15.8 kpb in the *dek1lg3:E*^{1481}, *dek1lg3:S*^{1497} *and dek1lg3:R*^{1514} suggesting these mutants harbour a single integration of DNA at the locus, and several restriction fragments were detected in the mutant plants *dek1-wtlg3* and *dek1lg3:E*^{1477}*Q*^{1478} suggesting these lines harbour multicopy DNA integration and possible off locus targeting of the vector. (B) A single restriction fragment of size 18.8 kpb in the *dek1-wtlg3, and dek1lg3:E*^{1481} same as the WT suggesting these mutants harbour a single copy integration of DNA at the locus and several restriction fragments were detected in the mutant plants *dek1lg3:E*^{1477}*Q*^{1478}, *and dek1lg3:R*^{1514} suggesting these lines harbour off locus integration. Mutant *dek1*Δ*lg3* was used as negative control.

3.2.3 RT-PCR and cDNA sequencing

The *DEK1* transcript from the mutant plants were analyzed by RT-PCR using primers PpLoop_Inverse SP_ and Ex30_R, which bind to DEK1 cDNA in positions flanking the targeting sequences, to verify the mutant transcripts. Figure 20 shows the gel picture of the cDNA products amplified from the various mutant plants (*dek1-wtlg3*, *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481}, *dek1lg3:S*^{1497}) in addition to WT with the expected size of approximately 4.6 kbp. This result shows that the mutant transcript was produced and the expected amplicon sizes.

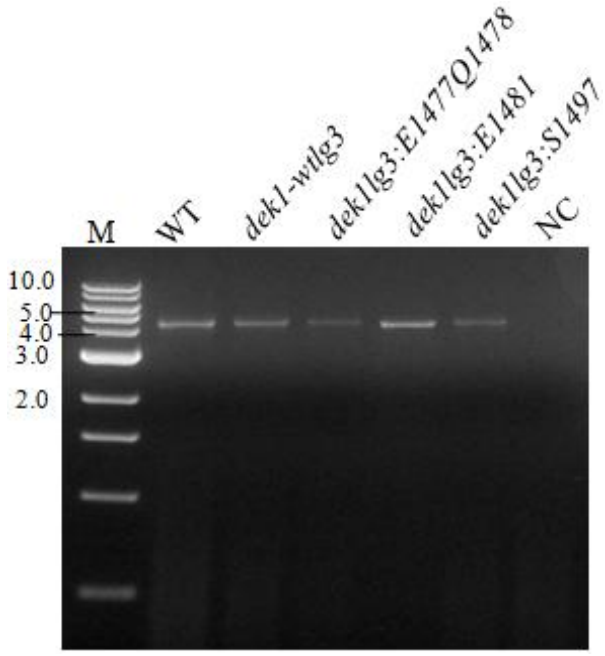


Figure 20. Reverse transcriptase (RT) PCR analysis of WT plant, *dek1-wtlg3*, *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481}, *dek1lg3:S*^{1497} mutant plants. The expected amplicon size for WT is 4.6 kbp. The mutant plants displayed an amplicon of similar size. Negative control (NC) sample contains no template.

Subsequently, *DEK1* cDNA from all generated mutants were sequenced to verify production of mutated transcripts. Figure 21 show part of the *DEK1* cDNA sequencing result with location of the various mutations marked “Conflict”. The sequencing results show that the transcripts display the introduced mutations, in addition to proper splicing of the transcript in the *dek1-wtlg3*, *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481}, *dek1lg3:S*^{1497} mutants (Data not shown). However, DEK1 sequencing of cDNA in the mutant *dek1lg3:R*^{1514} as shown in Figure 22 revealed that intron 22 was not spliced out rather it was retained in the transcript.

Examination of the sequenced data also shows that the remained intron 22 in the transcript introduced a frame shift leading to a stop codon in the transcript.

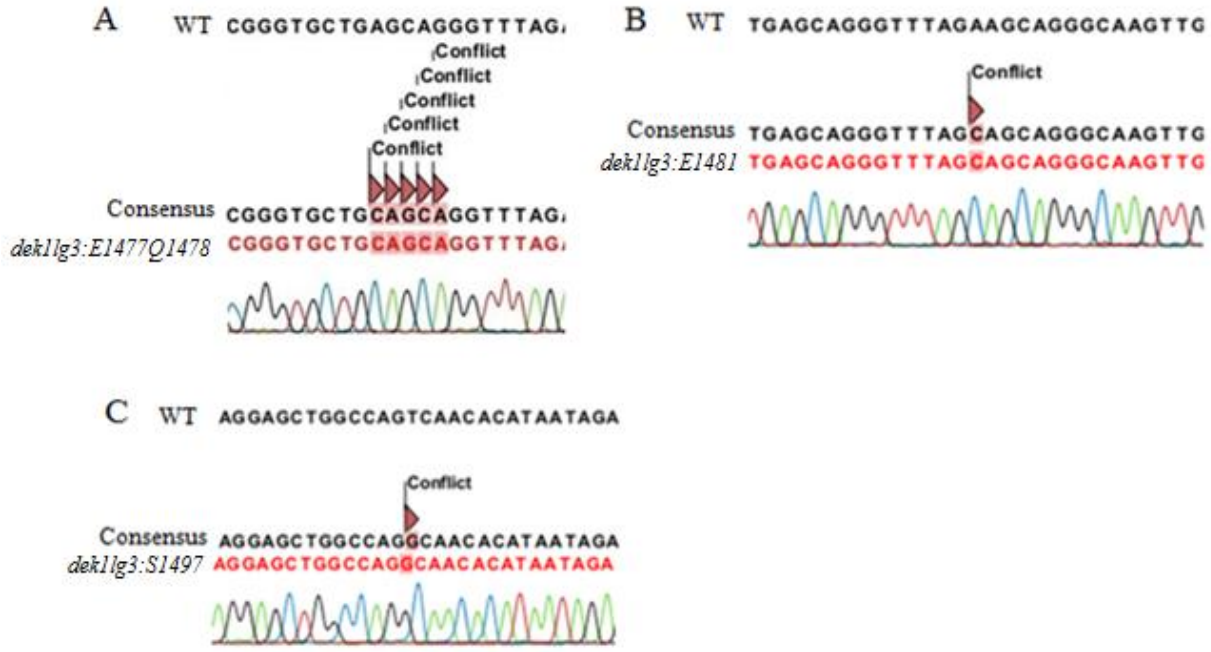


Figure 21. Part of the *DEK1* transcript sequencing confirming the presence of introduced mutations in the transcripts. (A) Part of *dek1lg3:E*^{1477}*Q*^{1478} cDNA sequencing showing substituted nucleotides (CAGCA) marked "Conflict". (B) Part of *dek1lg3:E*^{1481} cDNA sequencing showing substituted nucleotide (C) marked "Conflict". (C) Part of *dek1lg3:S*^{1497} cDNA sequencing showing substituted nucleotide (G) marked "Conflict".

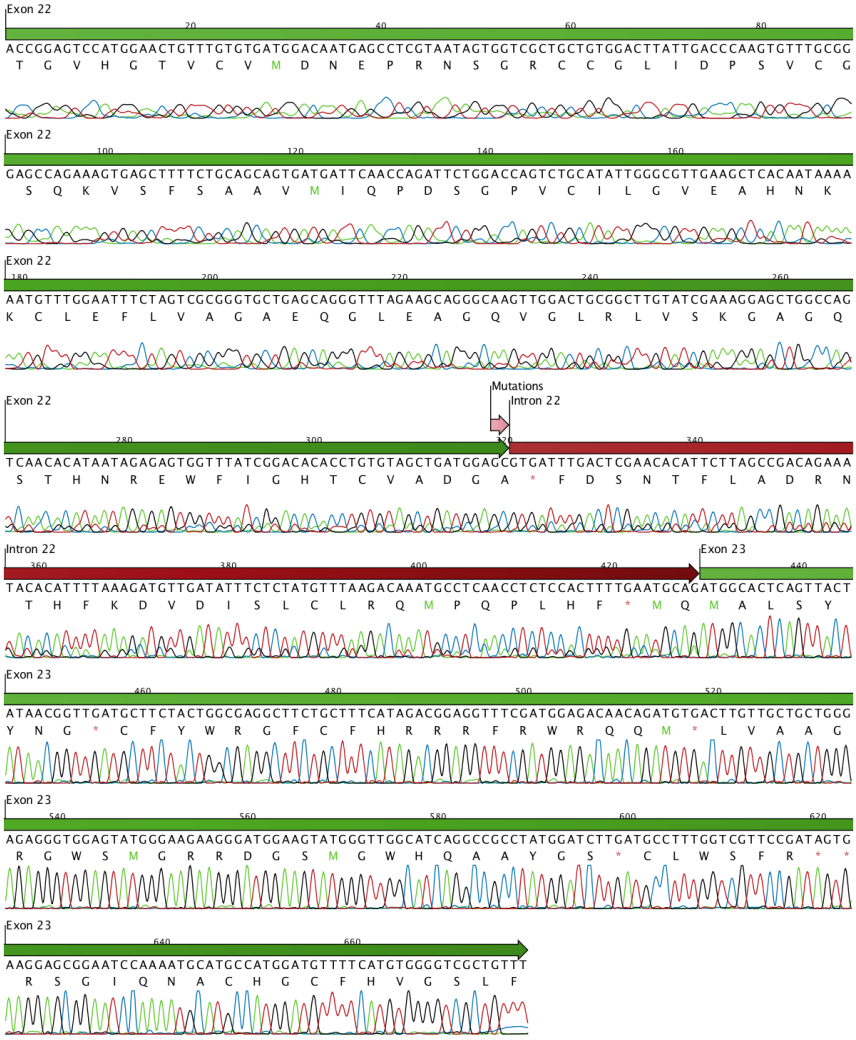


Figure 22. DEK1 sequence of cDNA in the mutant *dek1lg3:R*^{1514}. The introduced mutations (AG to GC) at the end of exon 22 interfere with splicing of the transcript, leaving intron 22 which introduces a frame shift leading to a stop codon and subsequently result in absence of active DEK1 protein.

In summary, the molecular characterization techniques used in this study, namely PCR-based genotyping, Southern blotting, RT-PCR and sequencing confirmed generation of *dek1-wtlg3* harbouring WT sequence, and the *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481}, *dek1lg3:S*^{1497} mutant plants harbouring the respective substituted nucleotides. Also, molecular characterization suggests that *dek1-wtlg3* and *dek1lg3:E*^{1481} mutant plants harbour a single integration of DNA at the locus while *dek1lg3:E*^{1477}*Q*^{1478} and *dek1lg3:S*^{1497} mutant plants harbours off locus integration of DNA. Also *DEK1* cDNA sequencing confirmed that intron 22 was not spliced out in the *dek1lg3:R*^{1514} mutant plant.

3.3 Phenotypic characterization

All mutants for phenotypic characterization were grown side by side.

To investigate the functionality of the complementation assay, the WT DEK1-LG3 sequence was re-targeted to the *DEK1ΔLG3* locus. The resulting mutant, *dek1-wtlg3* completely reverted to the WT phenotype with well developed gametophore and expanded phyllids (Figure 23 C and 24 C). This result confirms that the assay and the background used for transformation are working properly.

3.3.1 Mutant plants *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481}, and *dek1lg3:*S^{1497} display WT gametophore development and phyllid morphology

In the *dek1lg3:E*^{1477}*Q*^{1478} mutant two conserved amino acids residues glutamic acid (E1477) and glutamine (Q1478) were mutated to alanine (Ala) while in the *dek1lg3:E*^{1481} and *dek1lg3:S*^{1497} mutant plants a single amino acid, glutamic acid (E1481) and serine (S1497) were mutated to alanine (Ala), respectively.

Examination of gametophores morphology in the *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481} and *dek1lg3:S*^{1497} mutants (Figure 23) show that the WT phenotype was restored in these mutants. Gametophore morphology in mutants appears the same as WT gametophore with well developed and expanded phyllids compared to completely different gametophore morphology in the *dek1Δlg3* mutant, which harbour small and narrow phyllids. This result suggests that the amino acids E1477, Q1478, E1481 and S1497 are not important for *P. patens* gametophores development.

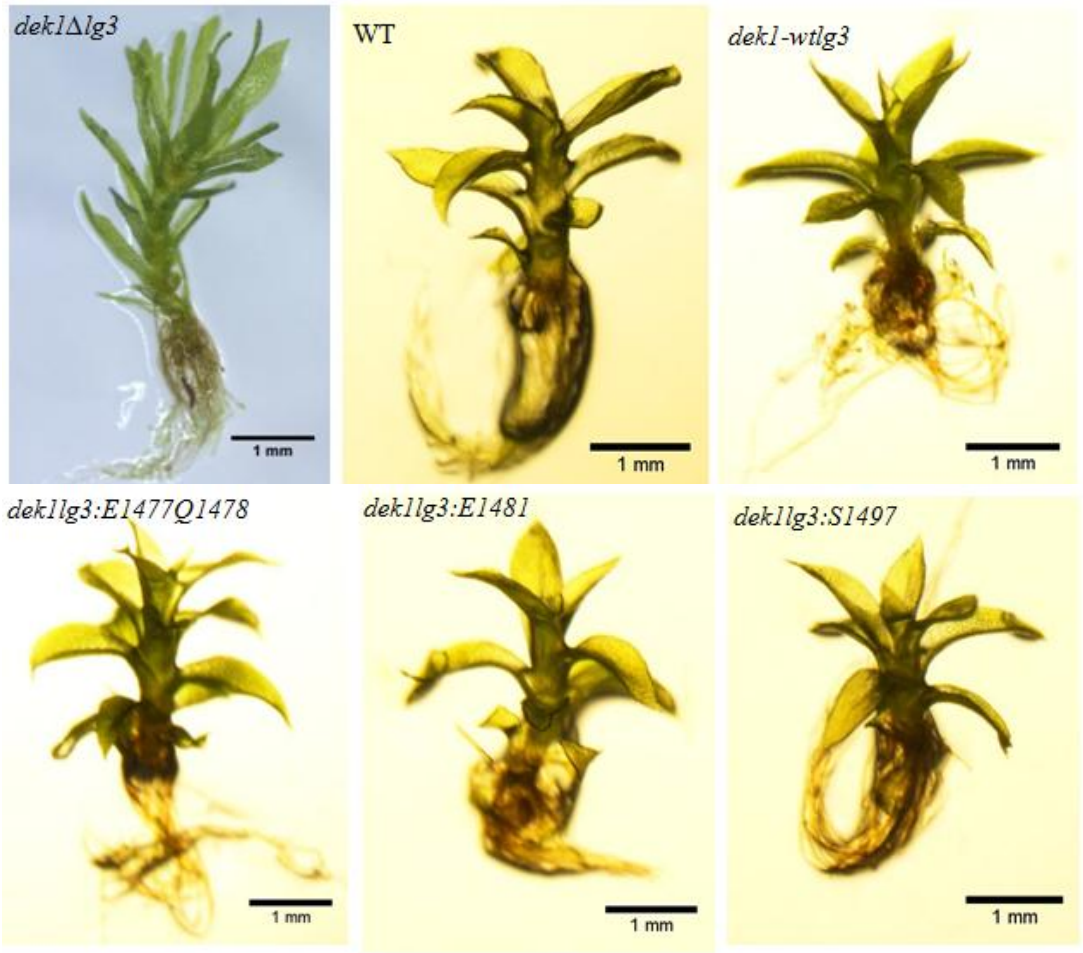


Figure 23. Gametophore morphology in *dek1*Δ*lg3*, WT and generated mutants *dek1-wtlg3, dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481}, and *dek1lg3:S*^{1497}. The *dek1*Δ*lg3* mutant plant gametophore with narrow and compressed phyllids, WT plant gametophore with well developed and expanded phyllids, *dek1-wtlg3* retargeted with WT DEK1-LG3, *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481} and *dek1lg3:S*^{1497} mutant plants gametophore display WT phenotype with well developed and expanded phyllids after re-targeting mutagenized versions of DEK1-LG3. All presented gametophore are 3 weeks old except *dek1*Δ*lg3* mutant gametophore obtained from (Johansen W.).

To investigate phyllid morphology, phyllids were dissected from the middle part of the gametophore and analyzed by dissecting microscopy. Analysis of phyllid morphology (Figure 24) shows that the phyllids from *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481} and *dek1lg3:S*^{1497} mutants appeared as the WT phyllid with expanded phyllids having clear marginal serration, small cells, long midrib and sharp tip compared to the *dek1*Δ*lg3* mutant phyllid which is small, narrow, lacking marginal serration, having short midrib, blunt tip and large cells. This result suggests that the mutated residues are not important for shaping the phyllids.

Figure 24. Phyllids morphology in *dek1*Δ*lg3* mutant, WT and generated mutants. The *dek1*Δ*lg3* mutant phyllid which is small and narrow, WT phyllid showing normal morphology, *dek1-wtlg3* phyllid re-targeted with WT DEK1-LG3 revert to WT, *dek1lg3:*$E^{1477}Q^{1478}$, *dek1lg3:*E^{1481} and *dek1lg3:*S^{1497} mutant phyllids display WT phenotype by re-targeting mutagenized version of DEK1-LG3. The *dek1*Δ*lg3* mutant phyllid obtained from (Johansen W.). All presented phyllids are 3 weeks old except *dek1*Δ*lg3* mutant phyllid.

3.3.2 Protonemata growth and spreading is not affected in mutants *dek1lg3:E^{1477}Q^{1478}*, *dek1lg3:E^{1481}* and *dek1lg3:S^{1497}*

Development of protonema tissue was analyzed by dissecting microscopy to investigate the growth and spreading of protonema in mutant plants. Analysis of protonema growth in all generated mutants, namely *dek1-wtlg3, dek1lg3:E^{1477}Q^{1478}*, *dek1lg3:E^{1481}* and *dek1lg3:S^{1497}* show that protonema is spreading similar to WT (Figure 25) suggesting that the mutated residues are not important for the growth and spreading of protonemata tissue. However, there is a difference in the spreading of protonemata tissue between generated mutants and the *dek1*Δ*lg3* mutant plant (Figure 25). In generated mutants the protonema is spreading much more compared to *dek1*Δ*lg3* mutant where the protonema is spreading less.

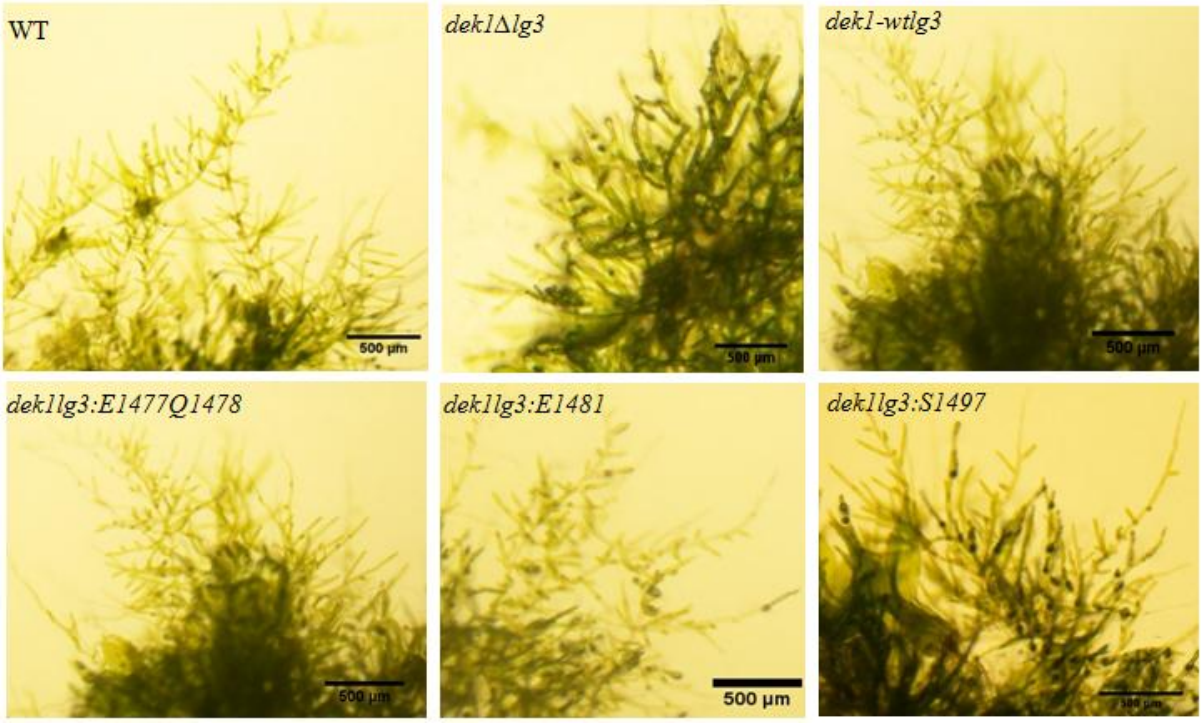


Figure 25. Growth and spreading of protonema tissue in WT, *dek1*Δ*lg3* and generated mutants. WT tissue with well spreading protonema, *dek1*Δ*lg3* tissue showing less spreading protonema, *dek1-wtlg3*, *dek1lg3:E1477Q1478*, *dek1lg3:E^{1481}* and *dek1lg3:*S^{1497} protonema tissue growing and spreading in the same pattern as in WT. All presented colony are 2 weeks old.

3.3.3 Mutant *dek1lg3:R*^{1514} has arrested bud development

Investigation of the mutant *dek1lg3:R*^{1514} show that the mutant display the Δ*dek1* mutant phenotype characterized by the lack of gametophore (Perroud et al. 2014). Closer examination of the bud initial development in the *dek1lg3:R*^{1514} mutant shows that there is no difference in the first asymmetrical division of the bud initial cell between the WT (Figure 26A, arrow) and *dek1lg3:R*^{1514} mutant (Figure 26D, arrow). However, the first cell division plane of the bud apical cell is clearly different between WT (Figure 26B, arrow) and *dek1lg3:R*^{1514} (Figure 26E, arrow) where in the WT the first cell division of the bud apical cell occurs perpendicular to the first asymmetrical division, new cell wall plane in the *dek1lg3:R*^{1514} mutant occurs at random positions. Thus, due to misorientation of cell wall in the first division of the bud apical cell, mutant bud fails to expand as in WT (Figure 26C), which subsequently leads to the arrest in bud development (Figure 26F). This result suggests that the mutant *dek1lg3:R*^{1514} is not producing a functional DEK1 protein.

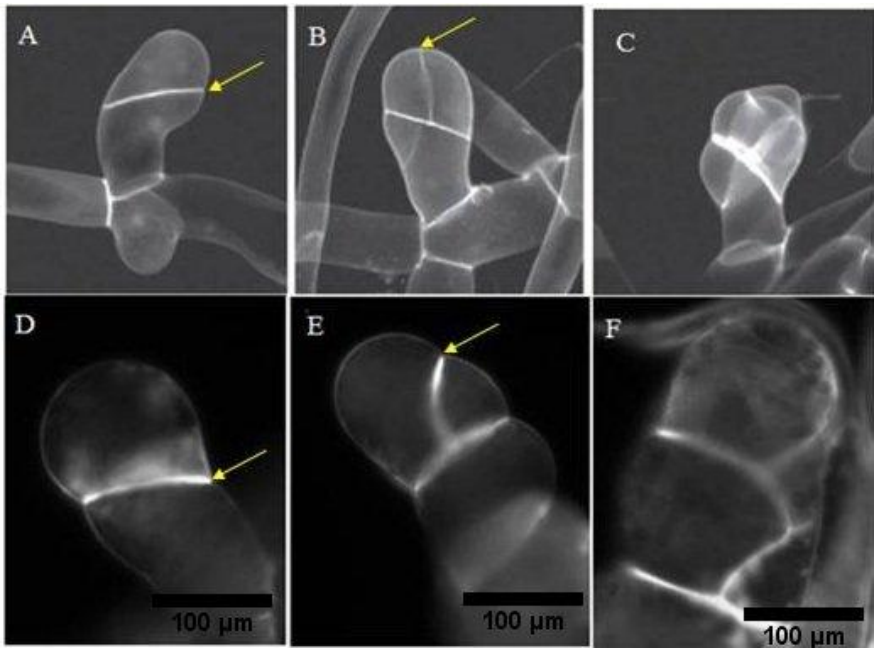


Figure 26 Bud development in *dek1lg3:R*^{1514} mutant. (A) Wild-type bud development at two-cell stage, the apical and basal cells, (B) First division of the bud apical cell (arrow) which occur in perpendicular to the prevision cell wall, (C) Globular WT bud showing seven cells in a three-dimension organization, (D) *dek1lg3:R*^{1514} mutant bud development at two-cell stage, the apical and basal cells, (E) First division of the bud apical cell (arrow) in *dek1lg3:R*^{1514} mutant which occurs in random position rather than perpendicular to the first division, (F) *dek1lg3:R*^{1514} arrested bud development. Figures A, B and C; Scale bar 50 µm and are modified from (Perroud et al. 2014).

3.3.4 Sporophyte development

All generated mutants were subjected to sporophyte growth conditions as described by (Perroud et al. 2011), to investigate the effects of mutated residues on gametangia and sporophyte development. However, due to time frame since induction until observation of sporophytes (about 2 month), sporophytes development data were not available to be included in this thesis.

4. Discussion

The aim of this study was to assess the importance of conserved DEK1-LG3 residues for DEK1 function in *Physcomitrella patens*. First, conserved and functionally important residues were predicted by bioinformatic analyses. Secondly, a series of *P. patens* mutants carrying site-specific mutations of the predicted functional important DEK1-LG3 residues were created. The resulting mutants were phenotypically characterized to assess the importance of these amino acids for DEK1 function. The result of the study was the generation of four different *dek1lg3* mutant lines: three mutants each carrying a single amino acid substitution to the amino acid alanine (Ala), namely *dek1lg3:E*^{1481}, *dek1lg3:S*^{1497}, *dek1lg3:R*^{1514} and one mutant carrying double amino acid substitutions to Ala, namely *dek1lg3:E*^{1477}*Q*^{1478} .

Acidic and polar residues are predicted to be functionally conserved in land plant DEK1-LG3 domain

Bioinformatics analysis predicted acidic residues E1477, E1481, D1577 and E1579, polar residues Q1478, S1497, S1576 and S1578, basic residues R1514, R1575 and non-polar G1574 as functionally important in DEK1-LG3 domain. These residues were predicted by ConSurf server which estimates the evolutionary conservation of residue according to its phylogenetic relation with other homologous sequence by employing empirical Bayesian method or a Maximum likelihood method (Armon et al. 2001; Landau et al. 2005). These predicted amino acids are highly conserved in land plant lineage which suggests that they may be involved in an important function in DEK1.

Phyllids morphology is not affected in *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481} and *dek1lg3:S*^{1497}mutants

The functionality of the complementation assay was confirmed by re-targeting the WT DEK1-LG3 sequence into *DEK1*Δ*LG3* locus and the mutant was complemented and completely reverted to the WT phenotype with well developed gametophores and expanded phyllids. All generated mutant plants before Cre recombinase showed Δ*dek1* phenotype

characterized by the lack of gametophore. The observed Δ*dek1* mutant phenotype was a result of absence of active calpain protein due to the insertion of the resistance cassette in an intron which probably disturbs splicing of the primary transcript and subsequently leads to the observed null phenotype. After Cre mediated excision removal of the resistance cassette, all mutants; *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481} and *dek1lg3:S*^{1497}except mutant *dek1lg3:R*^{1514} showed expanded phyllids indistinguishable from WT. Also the study of protonemata growth suggests there is no observable difference in the growth of protonemata between WT and generated mutants. These results suggest that these amino acids do not have an important function for protonemata growth or gametophores development. However, this study cannot rule out the possibility of importance of these amino acids in the sporophyte generation as the sporophyte development data were not available.

The conservation of amino acids throughout the period of evolution implies their importance, because slowly evolving amino acids in proteins are often very important either for structural or functional role (Celniker et al. 2013). In case where conserved residues suggested being not important can sometimes explained by the nature of amino acids substitution, whereby non-conservative substitution have high possibility of displaying the effect than conservative substitution. In this study, alanine substitution of conserved residues was carried out to study function of conserved residues, but study found that the mutated residues are not important for gametophores development and phyllids morphology. A study by (Baumberger and Baulcombe 2005) shows that a completely conserved residue can also turn out not being important. Baumberger study showed that a completely conserved residue (G758) in AGO1, an RNA-binding protein involved in RNA silencing, was mutated to serine and transiently expressed in *Nicotiana benthamiana*; however, an AGO1 protein was still able to mediate *in vitro* cleavage of *PHAVOLUTA* RNA as in WT plant. However, it is interestingly that sometimes even a single amino acid substitution can cause a change in phenotype as shown in the study by (Roeder et al. 2012), when *A. thaliana dek1-4* mutant created by a single base substitution leads to the phenotype where sepals lack giant cells.

In this study, Consurf server was used to predict functionally important residues. The Consurf server predicts evolutionary conservation of residues based on the phylogenetic relations between homologous sequences (Armon et al. 2001). The accuracy of the server in prediction of functionally and structurally conserved residues might also have contributed to

the observed results. However its robustness of using statistical data in the predictions of evolutionary rate has shown to give the server accuracy and confidence in estimation of evolutionary rate (Celniker et al. 2013). Moreover the efficiency of this server in prediction of functional regions has been shown to be successful previously in revealing two evolutionarily conserved regions at the ends of cytoplasmic membrane protein TatC that were proposed to mediate interaction of TatC with some other proteins (Rollauer et al. 2012). In addition, the accuracy of this server was also shown to be successful in mapping the patches of conserved residues onto the surface of well known domains, the Src homology 2 (SH2), a phosphotyrosine binding module involved in signal transduction and the phosphotyrosine binding domain (PTB) involved in peptide binding and recognition (Armon et al. 2001). However besides all the successes shown by the server, studying of the functionally important residues in the core or on the surface of a protein can only be accurately identified when its 3D structure is known (Montelione and Anderson 1999).

The introduced mutations in *dek1lg3:R*^{1514} disturbs *DEK1* transcript processing

In the mutant *dek1lg3:R*^{1514} two mutations were introduced at the end of exon 22 by substituting nucleotides AG to GC. After Cre mediated excision, this mutant still displayed the Δ*dek1* mutant phenotype, with lack of gametophore (Perroud et al. 2014). The mutant was phenotypically investigated and found to resemble the Δ*dek1* mutant phenotype; the first division of the bud apical cell was not orienting perpendicular to the previous cell division. This phenotype suggests that the mutant do not produce an active DEK1 protein. Sequencing of the DEK1 cDNA in this mutant revealed that the mutant did not correctly process the *DEK1* transcript because intron 22 was not spliced out. Failure to splice out intron 22 is suggested to be the result of the substituted nucleotides (AG to GC) which subsequently cause disturbance of splicing signals as the nucleotides are located at the exon-intron junction. Splicing errors caused by the introduced mutation is in agreement with studies by (Zorio and Blumenthal 1999; Blencowe 2000; Black 2003) when they reported that nucleotides AG at the exon-intron junction is important to bind splicing factor U2AF (Auxilliary factor), a subunit of U2 (small nuclear ribonucleoprotein (snRNP)) which is important to coordinate splicing process. The rate of intron retention has shown to be high in plants (30%) compared to (10%) reports in humans (Ner-Gaon et al. 2004). In order for an

intron to be excised from pre-mRNA special splicing signals that define exon-intron boundaries are required (Cartegni et al. 2002), these signals include the 5' and 3' splice sites, the branching point and the polypyrimidine tract which is located upstream of 3' splicing site. Therefore, the substituted nucleotides AG are suggested to be important splicing signals special for coordinating proper splicing of pre-mRNA.

The introduced mutations in the vector pBHRF-JI-LG3 $G^{1574}R^{1575}S^{1576}D^{1577}S^{1578}E^{1579}$ is suggested to be lethal in *P. patens*

To make construct pBHRF_JI_LG3$G^{1574}R^{1575}S^{1576}D^{1577}S^{1578}E^{1579}$, thirteen (13) mutations were introduced to exon 23. Generated construct was re-targeted into the *DEK1*Δ*LG3* locus in parallel with all other constructs. However, after several rounds of transformation no stable transformants were obtained. This observation suggests that the introduced mutation is toxic to the protoplasts allowing no growth and death of the protoplasts. The same observation was reported by (Perroud et al. 2014) when expressing *A. thaliana Calpain* cDNA and *Zea mays Calpain* cDNA to complement the *P. patens* Δ*dek1* mutant phenotype. Perroud et al reported the failure of complementation of either gametophore formation or overbudding phenotype after several rounds of transformation.

Southern blot analyses suggests off locus integration

In this study, southern blot analysis of the mutants before and after Cre recombinase shows conflicting results. Southern blot before Cre recombinase shows two expected restriction fragments in the mutant *dek1lg3:S*^{1497} suggesting that the mutant harbour a single integration of DNA at the targeted locus while Southern blot after Cre for the same mutant showed several restriction fragments suggesting that the mutant harbours off locus integration of DNA. Based on Southern blot for this mutant, there is a conflict between Southern blot before and after Cre recombinase. However, this is difficult to confirm if there is off locus integration or not, but if there is off locus integration, it is suggested that the DEK1 activity is not affected because the mutant displays WT phenotype.

Predicted DEK1-LG3 domain 3D structure resembles lectin and pentraxin crystal structures fold

To predict DEK1-LG3 domain 3D structure, the DEK1-LG3 sequence (10618-11230; XP 001774206.1) was submitted to the Phyre2 fold recognition server. The predicted structure of DEK1-LG3 domain displays a β-sandwich folds which is formed by two ant-parallel 10 β-sheets, 2 α-helix and loop region. Despite low sequence similarity (10-15 %) to legume lectins and pentraxins, LG has shown similar fold of β-sandwich to these classes of protein (Rudenko et al. 2001). The sub-family legume lectins and pentraxin proteins together with Laminin G-like module are all found in the super family Concanavalin A-like domain (ConA) which is characterized by sandwich fold of 12-14 β-strands in two sheets. Among the studied group of lectin protein is ConM from the seeds of leguminous plant *Canavalia maritima* which has been shown to have affinity to carbohydrates including mannose and glucose (Ramos et al. 1996). The study of crystal structure of this group of lectin (ConM) has revealed interaction of this group of protein with some active and inactive form of Indole-3- acetic acid (IAA) (Delatorre et al. 2013). Furthermore Delatorre et al reveal that this complex interaction between ConM-IAA might be important in restricting the availability of this phytohormone to the cell to protect its inhibitory activity during the early stages of germination. However, ConM residues shown to be involved in interaction with IAA, Ser108 and Asn131 (Delatorre et al. 2013) are not conserved in DEK1-LG3 domain.

On the other hand, the *dek1*Δ*lg3* mutant phenotype shows phyllids morphology that resembles the phenotype of *P. patens* auxin mutants. The *P. patens* pinA pinB disruptant mutant has shown phenotype with irregular shaped leaves that are longer and thinner than WT plant (Bennett et al. 2014) which is similar to the phenotype observed in *dek1*Δ*lg3* mutant plant. Moreover, the knockout mutant of SHORT INTERNODE/STYLISH (*SHI/STY*) family genes in *P. Patens,* the *Ppshi*1 and *Ppshi*2 genes has been shown to have effects on reproductive organs development similar to the defects observed in *dek1*Δ*lg3* mutant (Landberg et al. 2013). Landberg et al shows that the *Ppshi*1 and *Ppshi*2 mutants have defects in the formation of the archegonia canal, the arrested development of egg cell and unopened apex throughout the development. Interestingly it was also suggested that phytohormone auxin interact with the *Ppshi*1 gene, in which active form of auxin is required by *Ppshi1* gene for proper development of reproductive organs.

Taking into account the ConM-IAA interaction, with the fact that Laminin G-like module belongs to the same super family of Concanavalin A-like domain with β-sandwich fold and the similarity in *dek1*Δ*lg3* mutant phenotype to that of auxin mutants and the *SHI/STY* mutants it can also be speculated that DEK1-LG3 is involved in direct or indirect interaction with IAA.

Moreover, some other known crystal structures of LG3 domain have been shown to be involved in various functions including the LG3 of endorepellin, a calcium-dependent domain which is implicated in endothelia cell actin stress fibre and focal adhesion (Bix et al. 2004). However, the structure-function relationship prediction based on the jellyroll fold of the other available LG domains has been shown to be difficult due to differences in ligand interaction in each LG domain (Le et al. 2011).

5. Conclusion

The purpose of this study was to assess the importance of conserved DEK1-LG3 residues for DEK1 function in *Physcomitrella patens*. In the course of this study, four plasmid vector constructs harbouring mutations of the predicted functionally residues were successfully generated and re-targeted into *DEK1ΔLG3* locus. Phenotypic characterization of all generated mutant plants, namely *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481} and *dek1lg3:S*^{1497} suggest that these amino acids have no important function for DEK1 function during protonemata growth or gametophore development.

6. Further work

In this study, mutant plants *dek1lg3:E*^{1477}*Q*^{1478}, *dek1lg3:E*^{1481} and *dek1lg3:S*^{1497} were generated and this study suggests that these amino acids have no important function for DEK1 function during protonemata growth or gametophore development. Future study should subject all the mutants into the sporulation growth conditions to determine whether these mutations have effect in gametangia and sporophyte development.

In the present study, amino acids E1477, Q1478, E1481 and S1497 were mutated in DEK1-LG3 domain but this study suggest these residues have no important function for DEK1 function during protonemata growth or gametophore development. Further study needs to try to determine effects of combined mutation of these amino acids.

In this study, the Consurf server predicted several functionally/structural important residues in DEK1-LG3 domain, however only five (5) functionally predicted residues; E1477, Q1478, E1481, S1497 and R1514 have been addressed in this study. Future study needs also to focus on studying the functions of other predicted functionally/structurally conserved residues.

7. References

Armon A, Graur D, Ben-Tal N. 2001. ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *Journal of molecular biology* **307**(1): 447-463.

Baumberger N, Baulcombe D. 2005. Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proceedings of the National Academy of Sciences of the United States of America* **102**(33): 11928-11933.

Beck K, Hunter I, Engel J. 1990. Structure and function of laminin: anatomy of a multidomain glycoprotein. *The FASEB journal* **4**(2): 148-160.

Becker B, Marin B. 2009. Streptophyte algae and the origin of embryophytes. *Annals of Botany* **103**(7): 999-1004.

Bennett TA, Liu MM, Aoyama T, Bierfreund NM, Braun M, Coudert Y, Dennis RJ, O'Connor D, Wang XY, White CD. 2014. Plasma membrane-targeted PIN proteins drive shoot development in a moss. *Current Biology* **24**(23): 2776-2785.

Bix G, Fu J, Gonzalez EM, Macro L, Barker A, Campbell S, Zutter MM, Santoro SA, Kim JK, Höök M. 2004. Endorepellin causes endothelial cell disassembly of actin cytoskeleton and focal adhesions through α2β1 integrin. *The Journal of cell biology* **166**(1): 97-109.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annual review of biochemistry* **72**(1): 291-336.

Blencowe BJ. 2000. Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends in biochemical sciences* **25**(3): 106-110.

Branca D. 2004. Calpain-related diseases. *Biochemical and biophysical research communications* **322**(4): 1098-1104.

Britt AB, May GD. 2003. Re-engineering plant gene targeting. *Trends in plant science* **8**(2): 90-95.

Campbell R, Davies P. 2012. Structurefunction relationships in calpains1. *Biochemical Journal* **447**(3): 335.

Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Reviews Genetics* **3**(4): 285-298.

Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N. 2013. ConSurf: using evolutionary data to raise testable hypotheses about protein function. *Israel Journal of Chemistry* **53**(3-4): 199-206.

Cove D, Bezanilla M, Harries P, Quatrano R. 2006. Mosses as model systems for the study of metabolism and development. *Annu Rev Plant Biol* **57**: 497-520.

Cove DJ, Knight CD. 1993. The moss Physcomitrella patens, a model system with potential for the study of plant reproduction. *The Plant Cell* **5**(10): 1483.

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

Croall DE, Ersfeld K. 2007. The calpains: modular designs and functional diversity. *Genome Biol* **8**(6): 218.

Delatorre P, Silva-Filho JC, Rocha BAM, Santi-Gadelha T, da Nóbrega RB, Gadelha CAA, do Nascimento KS, Nagano CS, Sampaio AH, Cavada BS. 2013. Interactions between indole-3-acetic acid (IAA) with a lectin from Canavalia maritima seeds reveal a new function for lectins in plant physiology. *Biochimie* **95**(9): 1697-1703.

Demko V, Perroud P-F, Johansen W, Delwiche CF, Cooper ED, Remme P, Ako AE, Kugler KG, Mayer KF, Quatrano R. 2014. Genetic analysis of DEFECTIVE KERNEL1 loop function in three-dimensional body patterning in Physcomitrella patens. *Plant physiology* **166**(2): 903-919.

Emsley J, White HE, O'Hara BP, Oliva G, Srinivasan N, Tickle IJ, Blundell TL, Pepys MB, Wood SP. 1994. Structure of pentameric human serum amyloid P component. *Nature* **367**(6461): 338-345.

Finet C, Jaillais Y. 2012. Auxology: when auxin meets plant evo-devo. *Developmental biology* **369**(1): 19-31.

Frank W, Decker E, Reski R. 2005. Molecular tools to study Physcomitrella patens. *Plant Biology* **7**(3): 220-227.

Graham LE, Cook ME, Busse JS. 2000. The origin of plants: body plan changes contributing to a major evolutionary radiation. *Proceedings of the National Academy of Sciences* **97**(9): 4535-4540.

Grynspan F, Griffin W, Cataldo A, Katayama S, Nixon R. 1997. Active site-directed antibodies identify calpain II as an early-appearing and pervasive component of neurofibrillary pathology in Alzheimer's disease. *Brain research* **763**(2): 145-158.

Guroff G. 1964. A neutral, calcium-activated proteinase from the soluble fraction of rat brain. *The Journal of biological chemistry* **239**(1): 149-155.

Harrison CJ, Roeder AH, Meyerowitz EM, Langdale JA. 2009. Local cues and asymmetric cell divisions underpin body plan transitions in the moss Physcomitrella patens. *Current Biology* **19**(6): 461-471.

Heidstra R. 2007. Asymmetric cell division in plant development. In *Asymmetric Cell Division*, pp. 1-37. Springer.

Hepworth SR, Pautot VA. 2015. Beyond the divide: boundaries for patterning and stem cell regulation in plants. *Frontiers in plant science* **6**.

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.

Hohe A, Reski R. 2003. A tool for understanding homologous recombination in plants. *Plant Cell Reports* **21**(12): 1135-1142.

Huang Y, Wang KKW. 2001. The calpain family and human disease. *Trends in Molecular Medicine* **7**(8): 355-362.

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

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, Cuming AC, Cove DJ. 2005. Parameters determining the efficiency of gene targeting in the moss Physcomitrella patens. *Nucleic acids research* **33**(19): e173-e173.

Kenrick P, Wellman CH, Schneider H, Edgecombe GD. 2012. A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **367**(1588): 519-536.

Kofuji R, Hasebe M. 2014. Eight types of stem cells in the life cycle of the moss Physcomitrella patens. *Current opinion in plant biology* **17**: 13-21.

Kuwayama H. 2012. *Enhancement of Homologous Recombination Efficiency by Homologous Oligonucleotides*. INTECH Open Access Publisher.

Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. 2005. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic acids research* **33**(suppl 2): W299-W302.

Landberg K, Pederson ER, Viaene T, Bozorg B, Friml J, Jönsson H, Thelander M, 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.

Le BV, Kim H, Choi J, Kim J-H, Hahn M-J, Lee C, Kim KK, Hwang H-Y. 2011. Crystal structure of the LG3 domain of endorepellin, an angiogenesis inhibitor. *Journal of molecular biology* **414**(2): 231-242.

Liang Z, Demko V, Wilson RC, Johnson KA, Ahmad R, Perroud PF, Quatrano R, Zhao S, Shalchian-Tabrizi K, Otegui MS. 2013. The catalytic domain CysPc of the DEK1 calpain is functionally conserved in land plants. *The Plant Journal* **75**(5): 742-754.

Lid SE, Gruis D, Jung R, Lorentzen JA, Ananiev E, Chamberlin M, Niu X, Meeley R, Nichols S, 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.

Margis R, Margis-Pinheiro M. 2003. Phytocalpains: orthologous calcium-dependent cysteine proteinases. *Trends in plant science* **8**(2): 58-62.

Markmann-Mulisch U, Hadi MZ, Koepchen K, Alonso JC, Russo VE, 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.

Menand B, Calder G, Dolan L. 2007. Both chloronemal and caulonemal cells expand by tip growth in the moss Physcomitrella patens. *Journal of experimental botany* **58**(7): 1843-1849.

Moldoveanu T, Hosfield CM, Lim D, Elce JS, Jia Z, Davies PL. 2002. A Ca 2+ switch aligns the active site of calpain. *Cell* **108**(5): 649-660.

Montelione GT, Anderson S. 1999. Structural genomics: keystone for a Human Proteome Project. *Nat Struct Biol* **6**(1): 11-12.

Ner-Gaon H, Halachmi R, Savaldi-Goldstein S, Rubin E, Ophir R, Fluhr R. 2004. Intron retention is a major phenomenon in alternative splicing in Arabidopsis. *The Plant Journal* **39**(6): 877-885.

Niklas KJ. 2000. The evolution of plant body plans—a biomechanical perspective. *Annals of Botany* **85**(4): 411-438.

Olsen O-A, Perroud P-F, Johansen W, Demko V. 2015. DEK1; missing piece in puzzle of plant development. *Trends in plant science* **20**(2): 70-71.

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

Perroud PF, Cove DJ, Quatrano RS, McDaniel SF. 2011. An experimental method to facilitate the identification of hybrid sporophytes in the moss Physcomitrella patens using fluorescent tagged lines. *New Phytologist* **191**(1): 301-306.

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

Platt AR, Woodhall RW, George AL. 2007. Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *BioTechniques* **43**(1): 58.

Prigge MJ, Bezanilla M. 2010. Evolutionary crossroads in developmental biology: Physcomitrella patens. *Development* **137**(21): 3535-3543.

Puchta H. 2002. Gene replacement by homologous recombination in plants. In *Functional Genomics*, pp. 173-182. Springer.

Ramos JA, Zenser N, Leyser O, Callis J. 2001. Rapid degradation of auxin/indoleacetic acid proteins requires conserved amino acids of domain II and is proteasome dependent. *The Plant Cell* **13**(10): 2349-2360.

Ramos MV, Moreira RdA, Oliveira JTA, Cavada BS, Rougé P. 1996. The carbohydrate-binding specificity and molecular modelling of Canavalia maritima and Dioclea grandiflora lectins. *Memórias do Instituto Oswaldo Cruz* **91**(6): 761-766.

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist EA, Kamisugi Y. 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**(5859): 64-69.

Roeder AH, Cunha A, Ohno CK, Meyerowitz EM. 2012. Cell cycle regulates cell type in the Arabidopsis sepal. *Development* **139**(23): 4416-4427.

Rollauer SE, Tarry MJ, Graham JE, Jääskeläinen M, Jäger F, Johnson S, Krehenbrink M, Liu S-M, Lukey MJ, Marcoux J. 2012. Structure of the TatC core of the twin-arginine protein transport system. *Nature* **492**(7428): 210-214.

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

Sakakibara K, Nishiyama T, Sumikawa N, Kofuji R, Murata T, Hasebe M. 2003. Involvement of auxin and a homeodomain-leucine zipper I gene in rhizoid development of the moss Physcomitrella patens. *Development* **130**(20): 4835-4846.

Sato K, Saito Y, Kawashima S. 1995. Identification and Characterization of Membrane-Bound Calpains in Clathrin-Coated Vesicles from Bovine Brain. *European Journal of Biochemistry* **230**(1): 25-31.

Schaefer DG. 2001. Gene targeting in Physcomitrella patens. *Current opinion in plant biology* **4**(2): 143-150.

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

Shrive AK, Burns I, Chou H-T, Stahlberg H, Armstrong PB, Greenhough TJ. 2009. Crystal structures of Limulus SAP-like pentraxin reveal two molecular aggregations. *Journal of molecular biology* **386**(5): 1240-1254.

Smith LG, Hake S, Sylvester AW. 1996. The tangled-1 mutation alters cell division orientations throughout maize leaf development without altering leaf shape. *Development* **122**(2): 481-489.

Strepp R, Scholz S, Kruse S, Speth V, Reski R. 1998. Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein FtsZ, an ancestral tubulin. *Proceedings of the National Academy of Sciences* **95**(8): 4368-4373.

Suzuki K, Hata S, Kawabata Y, Sorimachi H. 2004. Structure, activation, and biology of calpain. *Diabetes* **53**(suppl 1): S12-S18.

Tanahashi T, Sumikawa N, Kato M, Hasebe M. 2005. Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss Physcomitrella patens. *Development* **132**(7): 1727-1736.

Theißen G, Münster T, Henschel K. 2001. Why don't mosses flower? *New Phytologist* **150**(1): 1-5.

Tian Q, Olsen L, Sun B, Lid SE, Brown RC, Lemmon BE, Fosnes K, Gruis DF, Opsahl-Sorteberg H-G, Otegui MS. 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* **19**(10): 3127-3145.

Timpl R, Tisi D, Talts JF, Andac Z, Sasaki T, Hohenester E. 2000. Structure and function of laminin LG modules. *Matrix biology* **19**(4): 309-317.

Tompa P, Emori Y, Sorimachi H, Suzuki K, Friedrich P. 2001. Domain III of calpain is a Ca 2+-regulated phospholipid-binding domain. *Biochemical and biophysical research communications* **280**(5): 1333-1339.

Trouiller B, Schaefer DG, Charlot F, Nogue F. 2006. MSH2 is essential for the preservation of genome integrity and prevents homeologous recombination in the moss Physcomitrella patens. *Nucleic acids research* **34**(1): 232-242.

Ushkaryov YA, Petrenko AG, Geppert M, Siidhof TC. 1992. Neurexins: Synaptic Cell Surface Proteins Related to the. *Science* **257**: 3JULY.

Wang C, Barry JK, Min Z, Tordsen G, Rao AG, 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.

Waters ER. 2003. Molecular adaptation and the origin of land plants. *Molecular phylogenetics and evolution* **29**(3): 456-463.

Yruela I. 2015. Plant development regulation: Overview and perspectives. *Journal of plant physiology* **182**: 62-78.

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.

Zorio DA, Blumenthal T. 1999. Both subunits of U2AF recognize the 3′ splice site in Caenorhabditis elegans. *Nature* **402**(6763): 835-838.

Appendices

Appendix 1: Land plant DEK1 sequences and their accession number used in this study

Species	Accession
Aegilops tauschii	EMT33050.1[A]
Amborella trichopoda	XP_006856301.1[A]
Aquilegia coerulea	Aquca_009_00510.1[B]
Arabidopsis lyrata	XP_002894501.1[A]
Arabidopsis thaliana	NP_175932.2[A]
Beta vulgaris	XP_010673464.1[A]
Brachypodium distachyon	XP_003570209.1[A]
Brassica napus_A	CDY33052.1[A]
Brassica napus_B	XP_013706153.1[A]
Brassica rapa	XP_009147506.1[A]
Camelina sativa_B	XP_010501149.1[A]
Camelina sativa_C	XP_010501148.1[A]
Camptotheca acuminata	GACF01058706.1[A]
Cannabis sativ	JP475882.1[A]
Capsella rubella	XP_006303131.1[A]
Capsicum annuum	JW063188.1[A]
Carica papaya	evm.TU.supercontig_119.40[B]
Ceratodon purpureus	SRS140252[C]
Chorispora bungeana	KA022282.1[A]
Cicer arietinum	XP_004504206.1[A]
Citrus clementina	XP_006445587.1[A]

Citrus sinensis	XP_006488938.1[A]
Coffea canephora	CDP18596.1[A]
Costus pictus	JW231520.1[A]
Cucumis melo	XP_008451014.1[A]
Cucumis sativus	Cucsa.142290.1[B]
Curcuma longa	JW811525.1[A]
Elaeis guineensis_A	XP_010936786.1[A]
Elaeis guineensis_B	XP_010940022.1[A]
Erythranthe guttata_A	EYU25999.1[A]
Erythranthe guttata_B	EYU39270.1[A]
Eucalyptus grandis	XP_010032857.1[A]
Eutrema salsugineum	XP_006392645.1[A]
Fragaria vesca	XP_004294954.1[A]
Genlisea aurea	EPS66151.1[A]
Glycine max_A	XP_003532791.1[A]
Glycine max_B	Glyma05g30080[B]
Glycine soja_A	KHN06483.1[A]
Glycine soja_B	KHN48081.1[A]
Gossypium arboreum	KHG02979.1[A]
Gossypium raimondii	Gorai.003G153800.1[B]
Hevea brasiliensis	JT914256.1[A]
Hordeum vulgare	ABW81402.1[B]
Jatropha curcas	KDP30593.1[A]
Lactuca serriola	JO020465.1[A]
Linum usitatissimum_A	Lus10010313[B]
Linum usitatissimum_B	Lus10013411[B]

Malus domestica_A	XP_008354991.1[A]
Malus domestica_B	MDP0000245785[B]
Malus domestica_C	MDP0000094595[B]
Manihot esculenta	cassava4.1_000045m[B]
Marchantia polymorpha_A	D
Marchantia polymorpha_B	D
Medicago truncatula	XP_003629937.1[A]
Mimulus guttatus_A	mgv1a023650m.g [B]
Mimulus guttatus_B	mgv1a000044m.g[B]
Morus notabilis	XP_010112666.1[A]
Musa acuminata	XP_009403924.1[A]
Nelumbo nucifera	XP_010257664.1[A]
Nicotiana benthamiana	AAQ55288.2[A]
Nicotiana tomentosiformis	XP_009619217.1[A]
Oryza brachyantha	XP_006647780.1[A]
Oryza sativa	AAL38190.1[A]
Panicum virgatum	Pavirv00022988m[B]
Phaseolus vulgaris	XP_007159560.1[A]
Phoenix dactylifera_A	XP_008787933.1[A]
Phoenix dactylifera_B	XP_008799338.1[A]
Physcomitrella patens	XP_001774206.1[A]
Populus euphratica_A	XP_011025140.1[A]
Populus euphratica_B	XP_011048187.1[A]
Populus euphratica_C	XP_011048187.1[A]
Populus trichocarpa_A	XP_002299263.2[A]
Populus trichocarpa_B	Potri.001G003900.1[B]

Prunus mum	XP_008222910.1[A]
Prunus persica	XP_007208412.1[A]
Pyrus x bretschneideri_A	XP_009339183.1[A]
Pyrus x bretschneideri_B	XP_009375947.1
Ricinus communis	XP_002523419.1[A]
Selaginella moellendorffii_B	236021 (fgenesh1_pm.C_scaffold_108000004)[B]
Selaginella moellendorffii_A	XP_002990425.1[A]
Sesamum indicum	XP_011089164.1[A]
Setaria_italica	XP_004984907.1[A]
Solanum lycopersicum	Solyc12g100360.1.1[B]
Solanum tuberosum	XP_006367593.1[A]
Sorghum bicolor	XP_002468005.1[A]
Tarenaya hassleriana_A	XP_010534866.1[A]
Tarenaya hassleriana_B	XP_010552284.1[A]
Thellungiella halophila	Thhalv10011175m[B]
Theobroma cacao	Thecc1EG038725t2[B]
Thlaspi arvense	GAKE01002389.1[A]
Utricularia gibba	Scf00134.g10074.t1
Vitis vinifera	XP_002285732.1[A]
Zea mays	NP_001105528.1[A]

[A]NCBIGenbank; [B]Phytozome; [C]Sequence retrieved from the SRS140252 library deposit at NCBI GenBank and [D]Sequences provided by Katsuyuki T. Yamato and Takayuki Kohchi (Liang et al. 2013).

Appendix 2: Mutated DEK1-LG3 nucleotide sequences (Indicated by small letters)

$E^{1477}Q^{1478}$

AGCTTTTCTGCAGCAGTGATGATTCAACCAGATTCTGGACCAGTCTGCATATTGGGCGTT
GAAGCTCACAATAAAAAATGTTTGGAATTTCTAGTCGCGGGTGCTGcagcaGGTTTAGAAG
CAGGGCAAGTTGGACTGCGGCTTGTATCGAAAGGAGCTGGCCAGTCAACACATAATAG
AGAGTGGTTTATCGGACACACCTGTGTAGCTGATGGAAGGTGATTTGACTCGAACACAT
TCTTAGCCGACAGAAATACACATTTTAAAGATGTTGATATTTCTCTATGTTTAAGACAAA
TGCCTCAACCTCTCCACTTTTGAATGCAGATGGCACTCAGTTACTATAACGGTTGATGCT
TCTACTGGCGAGGCTTCTGCTTTCATAGACGGAGGTTTCGATGGAGACAACAGATGTGA
CTTGTTGCTGCTGGGAGAGGGTGGAGTATGGGAAGAAGGGATGGAAGTATGGGTTGGC
ATCAGGCCGCCTATGGATCTTGATGCCTTTGGTCGTTCCGATAGTGAAGGAGCGGAATC
CAAAATGCATGCCATGGATGTTTTCATGTGGGGTCGCTGTTTAACCGAAGATGAAATTTT
AATGGTGCATAATTGTGCT

E^{1481}

AGCTTTTCTGCAGCAGTGATGATTCAACCAGATTCTGGACCAGTCTGCATATTGGGCGTT
GAAGCTCACAATAAAAAATGTTTGGAATTTCTAGTCGCGGGTGCTGAGCAGGGTTTAGc
AGCAGGGCAAGTTGGACTGCGGCTTGTATCGAAAGGAGCTGGCCAGTCAACACATAAT
AGAGAGTGGTTTATCGGACACACCTGTGTAGCTGATGGAAGGTGATTTGACTCGAACAC
ATTCTTAGCCGACAGAAATACACATTTTAAAGATGTTGATATTTCTCTATGTTTAAGACA
AATGCCTCAACCTCTCCACTTTTGAATGCAGATGGCACTCAGTTACTATAACGGTTGATG
CTTCTACTGGCGAGGCTTCTGCTTTCATAGACGGAGGTTTCGATGGAGACAACAGATGT
GACTTGTTGCTGCTGGGAGAGGGTGGAGTATGGGAAGAAGGGATGGAAGTATGGGTTG
GCATCAGGCCGCCTATGGATCTTGATGCCTTTGGTCGTTCCGATAGTGAAGGAGCGGAA
TCCAAAATGCATGCCATGGATGTTTTCATGTGGGGTCGCTGTTTAACCGAAGATGAAATT
TTAATGGTGCATAATTGTGCT

S^{1497}

AGCTTTTCTGCAGCAGTGATGATTCAACCAGATTCTGGACCAGTCTGCATATTGGGCGTT
GAAGCTCACAATAAAAAATGTTTGGAATTTCTAGTCGCGGGTGCTGAGCAGGGTTTAGA

AGCAGGGCAAGTTGGACTGCGGCTTGTATCGAAAGGAGCTGGCCAGgCAACACATAATA
GAGAGTGGTTTATCGGACACACCTGTGTAGCTGATGGAAGGTGATTTGACTCGAACACA
TTCTTAGCCGACAGAAATACACATTTTAAAGATGTTGATATTTCTCTATGTTTAAGACAA
ATGCCTCAACCTCTCCACTTTTGAATGCAGATGGCACTCAGTTACTATAACGGTTGATGC
TTCTACTGGCGAGGCTTCTGCTTTCATAGACGGAGGTTTCGATGGAGACAACAGATGTG
ACTTGTTGCTGCTGGGAGAGGGTGGAGTATGGGAAGAAGGGATGGAAGTATGGGTTGG
CATCAGGCCGCCTATGGATCTTGATGCCTTTGGTCGTTCCGATAGTGAAGGAGCGGAAT
CCAAAATGCATGCCATGGATGTTTTCATGTGGGGTCGCTGTTTAACCGAAGATGAAATTT
TAATGGTGCATAATTGTGCT

R^{1514}

AGCTTTTCTGCAGCAGTGATGATTCAACCAGATTCTGGACCAGTCTGCATATTGGGCGTT
GAAGCTCACAATAAAAAATGTTTGGAATTTCTAGTCGCGGGTGCTGAGCAGGGTTTAGA
AGCAGGGCAAGTTGGACTGCGGCTTGTATCGAAAGGAGCTGGCCAGTCAACACATAAT
AGAGAGTGGTTTATCGGACACACCTGTGTAGCTGATGGAgcGTGATTTGACTCGAACACA
TTCTTAGCCGACAGAAATACACATTTTAAAGATGTTGATATTTCTCTATGTTTAAGACAA
ATGCCTCAACCTCTCCACTTTTGAATGCAGATGGCACTCAGTTACTATAACGGTTGATGC
TTCTACTGGCGAGGCTTCTGCTTTCATAGACGGAGGTTTCGATGGAGACAACAGATGTG
ACTTGTTGCTGCTGGGAGAGGGTGGAGTATGGGAAGAAGGGATGGAAGTATGGGTTGG
CATCAGGCCGCCTATGGATCTTGATGCCTTTGGTCGTTCCGATAGTGAAGGAGCGGAAT
CCAAAATGCATGCCATGGATGTTTTCATGTGGGGTCGCTGTTTAACCGAAGATGAAATTT
TAATGGTGCATAATTGTGCT

G^{1574}R^{1575}S^{1576}D^{1577}S^{1578}E^{1579}

AGCTTTTCTGCAGCAGTGATGATTCAACCAGATTCTGGACCAGTCTGCATATTGGGCGTT
GAAGCTCACAATAAAAAATGTTTGGAATTTCTAGTCGCGGGTGCTGAGCAGGGTTTAGA
AGCAGGGCAAGTTGGACTGCGGCTTGTATCGAAAGGAGCTGGCCAGTCAACACATAAT
AGAGAGTGGTTTATCGGACACACCTGTGTAGCTGATGGAAGGTGATTTGACTCGAACAC
ATTCTTAGCCGACAGAAATACACATTTTAAAGATGTTGATATTTCTCTATGTTTAAGACA
AATGCCTCAACCTCTCCACTTTTGAATGCAGATGGCACTCAGTTACTATAACGGTTGATG
CTTCTACTGGCGAGGCTTCTGCTTTCATAGACGGAGGTTTCGATGGAGACAACAGATGT
GACTTGTTGCTGCTGGGAGAGGGTGGAGTATGGGAAGAAGGGATGGAAGTATGGGTTG
GCATCAGGCCGCCTATGGATCTTGATGCCTTTGcagcagCaGcagcaGcAGGAGCGGAATCCA

AAATGCATGCCATGGATGTTTTCATGTGGGGTCGCTGTTTAACCGAAGATGAAATTTTAA
TGGTGCATAATTGTGCT

Appendix 3: Primer sequences

A. Primer sets used for Amplification of In-Fusion inserts form *P. patens* WT

F= Forward Primer, R= Reverse Primer

Insert	Primer Name	Primer sequence (5'→ 3')
5'_TGS	JI_5TGS_SP_Inf	F=TACGTCGCGACTCGATGCATGATTAATATGATCTTCA
	JI_5TGS_ASP_Inf	R=ACGAAGTTATCTCGACATGTGCTTCGTGATATGC
3'_TGS	JI_3TGS_SP_Inf	F=CGCCACGCGTGATATGTTTAATTGAGTCAGTAATTAG
	JI_3TGS_ASP_Inf	R= ATGTTAACATGCATGACTGATACGTAAACGAAGATA

B. Primers for inverse PCR of the vector, F= Forward Primer, R= Reverse Primer

Vector	Primer name	Primer sequence (5'→ 3')
pBHRF	pBHRF_1_SP	F=CATGCATGTTAACATCGATCCATGG
	pBHRF_1_ASP	R= ATATCACGCGTGGCGCCACTAG

C. Primers for Methylation and Mutagenesis reaction, F= Forward Primer, R= Reverse Primer

Construct	Primer Name	Primer sequence (5'→ 3')
pCR_5TGSF8_9	F8_9 SP	F=TAGTCGCGGGTGCTGcagcaGGTTTAGAAGCAGG
	F8_9 ASP	R= CCCTGCTTCTAAACCtgctgCAGCACCCGCGACTA
pCR_5TGSF10	F10E_SP	F= CTGAGCAGGGTTTAGcAGCAGGGCAAGTTGG
	F10E_ASP	R= CCAACTTGCCCTGCTgCTAAACCCTGCTCAG

pCR_5TGS14	F14_SP	F= AAAGGAGCTGGCCAGgCAACACATAATAGAG
	F14_ASP	R= CTCTATTATGTGTTGcCTGGCCAGCTCCTTT
pCR_5TGS16	F16_SP	F= TGTGTAGCTGATGGAgcGTGATTTGACTCGAA
	F16_ASP	R= TCGAGTCAAATCACgcTCCATCAGCTACACA
pCR_5TGS23_28	F23_28 SP	F=CCGCCTATGGATCTTGATGCCTTTGcagcagCaGcag caGcAGGAGCGGAATCCAAAATGCATGCC
	F23_28 ASP	R=GGCATGCATTTTGGATTCCGCTCCTgCtgctgCtGctgctgC AAAGGCATCAAGATCCATAGGCGG

D. Primers for amplification of insert (5'_TGSmut) from the vector

Primer name	Primer sequence (5'→ 3')
JI_5TGS_SP_Inf	TACGTCGCGACTCGATGCATGATTAATATGATCTTCA
JI_5TGS_ASP_Inf	ACGAAGTTATCTCGACATGTGCTTCGTGATATGC

E. Primers for colony PCR

Primer name	Primer sequence (5'→ 3')
Armseq7	GCATATTGGGCGTTGAAGCT
35s_R	TAAAGTGACAGATAGCTGGG
JI_3TGS_SP_Inf	CGCCACGCGTGATATGTTTAATTGAGTCAGTAATTAG
JI_3TGS_ASP_Inf	ATGTTAACATGCATGACTGATACGTAAACGAAGATA

F. Sequencing primers

Primer name	Primer sequence (5'→3')
TER_F	AGGGTTCTTATAGGGTTTCGCTCATG
pBHRF_R	AGGAAACAGCTATGACCATGA

ASP-PpARM-Info	CTGCCGGTCGTGTATCTAT
ARM 3'_F	CCGCCATCAGATCAGTCGCT
CALP Seq1	AAAGAGGAGGTCTTGCAGCG
EX25-F	AACAAGGGCAAGATTCTCGG
ΔARM 3'_ R	AATGGACTACAAACTGATACG
T7	TAATACGACTCACTATAGGG
M13R	CAGGAAACAGCTATGAC
ArmSeq6	TGCAGGTACCAAAGAAGCAGC
CSMW_5R	GCGGCTTGTATCGAAAGGAG
pBHRF_F	GCCTCTTCGCTATTACGCCA
35s_Rev1	TAAAGTGACAGATAGCTGGG
Armseq7	GCATATTGGGCGTTGAAGCT
ArmSeq8	GATGGAAGTATGGGTTGGCATC

G. Primers for genotyping of transformed lines

Primer name	Primer sequence (5'→ 3')
ArmSeq1	TGCAAGTTCAGCAGCTCTGC
ArmSeq3	TGTTTTAGCACGGCTATTCTTTTC
35sRev1	TAAAGTGACAGATAGCTGGG
Term_Fw	AGGGTTCTTATAGGGTTTCGCTCATG
EX30_R	GTTACACGATTGTTCCAACCACA
LG3_Fw	TTGGCACATTTCAGACCGGA
LG3_Rev	GGCAGGGATCTCCAATGGAA

H. Primers for amplifying DEK1 cDNA

Primer name	Primer sequence (5'→ 3')
PpLoop_Inverse_SP	TGGGTCTTCTTCAGTGTGATC
Ex30_R	GTTACACGATTGTTCCAACCACA

I Primers for probe synthesis

Probe	Primer name	Primer sequence (5'→ 3')
5TGS	ArmSeq5	TGCATCGGAACAAGAATCTAGTGTA
	CSMW_5R	GCGGCTTGTATCGAAAGGAG
3TGS	JI_3TGS_SP_Inf	CGCCACGCGTGATATGTTTAATTGAGTCAGTAATTAG
	JI_3TGS_ASP_Inf	ATGTTAACATGCATGACTGATACGTAAACGAAGATA

Appendix 4: Sequences

A. *P. patens* genomic sequence

ATGGGGGCGTCCGTGGACGGTGCACCTGTGCACGCTGTGGTCAAATCTTGTGCGCTGTTT
GGGTCACTCTTCGTGGTTCTTTGTGCACTCTCCGTCGTTATTCTAGTTGCTGTAAATTGGC
GCCCCTGGCGGATCTACAGGTACTTATCTCAGCTGCAGCTTGTCTGTATATTGGTTCTTG
GTTCTTTGGATAGTTCTTTCACTCTCTGATTCACACTTGGATATTGGGTTGCGGGTGATGC
CGATGAACCTAGGAATCAATACCTAGGAGTTCTTATTGCCATCCGCTTTTGCTTGGATTG
GGCAGTTCTCAGCGGTTGAGTAGTTGGTCCTATGCAATCTTTCGTCCGATAACCTATGTT
TCATGAATAGATTAAGATTTTACTGGCGATTACATTTCAGAGGTATATCTTAACACAGTG
TATGGATTTCTCGTTTTGGCACATACTGAAGGATCCAATTGAGCACGGCACTCTTCTCTA
CTGATCACAGATGGAGTAGAAAGTCGAGTGTAATGTCGAGTAGTGGCTGCAATCCTGTC
ATCCTTACTCTTAGCTTATATAGGGCACGTAATGCAGAGACGTGCAAGCTCAGCCTATTG
ACAATAGTGTTACCTGTCATGGAAGACCGTTGCAACTAGCCTTATATTCTTGAAATGAA
AGAAAACAATGAATTACTTTGTAAAGTAGTGGTCTTGTTTATTAACATGAGTTTGCAATC
GTTTGACATCGTGTATCGATGTCAACGCTTGTAATTTATAAAGTTAGGGATGAAACCCA
GCCACTGGTTACATAGGATCCTGTCCTTTCATTAGAGTTATGCCGATTGTCTGCATAACG
TCGTAGGTGTTGTCGTGTGAACGACTGTATTTGATGCGCATTGTTTGATGCTTTCGTAGG

CAGCAGTCATTTCCCGACTTATGACGATTTATACATCTATTGTTTAAATACGCGACTAGA
TACCCATTATTTGAGACAAAAGCGGTCAACAAGAGGTCTTAGTCAGCCTACGGCAATTG
ATGATGGCTTGCGTGCTACAGAATGTGAATGAAACCTCCGTGCTTGTTTTACAGCTAGA
AAGTTTCTGAAACATACTTAGCATGCCTCCTCACTGGCCAAGTTATAAAAGTGAGAGTA
GTTATCGTTTGTTTTTCCATAGCAGCCATTCCTGCTGTTTCTGATATAATTGTGCCATGAA
ATAGATATTTTATTGCTTAGTTTAGTGTGTTTGTACGTCGTCTTCGTATTTAATGTGGATT
TCCCACAGCTTCTCTCATCTCATGTTGTGGAGTGTGATTGCTCCTTGTTCTCCTTGTTGTA
ACATCTCTCCCTGATTATTGCTGTCGTTGTTGCAGCTGGATATTTGCCCGCAGATGGAGA
TCGTTTATGCAAGGCCATCGGCTCAGCATAATTACAGCAGTACTTGCCGCAGCTGCATG
GACCATAGTCCTGTCCCCAATCGCGGTTCTTATCTTGTGGGGAGCTAGACTCATAATTTT
ACTGAATCATGACACAATCGGCCTTGCAGTGATCCTGGCAGGCACAGCTCTACTGTTAG
CATTTTATGCTATCATGCTTTGGTGGAGGACGCAATGGCAAAGCTCACGTATGTATTCCA
TTAGATCTCATTAGGCTACATTTCTGATCGGCTACATTGTAAGTTCATTTCTACATTTGGT
AGTACCCTGTTCTGGATTGTTCTGAAGGTTTTATTTATATGTTTTCTTGTCCTGCCCATCC
CAGGTGCTGTTGCAATCCTGCTTCTTCTAGCTGTATCGTTATTATGCGCCTACGAGCTTTC
AGCAGTCTACGTTACAGCAGGACGAAGTGCTTCAAATCAGTTTTCTCCATCAGCTTTCTT
TTTCGGGGTTTCAGCGATCGCAATGGGCATTAACATGCTCTTCATATGTAAAATGGTTTT
CAATGGTGAGTAATGGAAACTTGATGCCATTGTCGGCAGAAGTACAATTCTTATAAATG
GAAAGTGTCTTTGGTAATGATCGTCTGCACTTCACAGCTCTTGATTTAGATTTTTCCTGTT
TCGGGTAAACAAATCTTTTGAGTGTTTACCTAGCTTTCTGTGGCTAACAGTTCGAAAGTT
CTGTTTACAGGGGCAGGGCTTGATGTGGATGAGTATGTGAGAAGGTCCTACAAGTTTGC
GCATGCAGAAACGTTAGAAGTTGGTCCGATTGCATGCTTGCCAGAGCCTCCTGAGCCTC
GCGACGCAAGCATTCAGAGGAAAAGCAGGTGGTTCTTCTCTTGATGTAATTGTATTGTG
GTTGTGATATTTATGAACTACTTCTTGATTCCTGTCTTTCATACGGTGAAATTCCATGTAA
TCCTGCATGTTTATTTCTTCTGTACCCATGTTTTGACAAGACATCAGTTTTGAAGTGCATC
CAAGTGGTCTCCAAGTGTTAGATTCTTAGATATACTGCAAATGTTCAAGTTGGTGAATTC
TTGGGCTTGTTGTTTTTGAGCGAACCCGTGATAGCGACGTTGCCTAGTGCTTTGCTTTTCC
AATTATTAGTGGCTCTTGATTTTCTTAGCTAGCGTGAGTGAAAGCCTTAGCTATAATCAG
CAACTTACAGTTACGTCGCGATCTCAGAATTTTAAGTTTTACAGGGGAGGAATCCCGAG
GTCGTTTAACACATTCTTGTAGAACCTAGTCTCATCTCTTATTGCTTGCAGTACATGGCG
GCTGGCCCAGTTATATCAAGCATCTGTTGTGGTTCTGGCGGCATATTCAGTGCTGTATGG
ACTGACAGCAAGGGAAGCACGATGGCTTGGAGGAGTTACTTCGGCCGCTGTGATCATTT
TGGGTATGTTTTTTCTCAATCAATGACGAATGCGCACCAAAATTCTTTGATGGAGTCTTT
GTTCTTTATACTCAGTTGTGTGCTCAGCTACTAGTTGTCTCGTTTACTTTCTTTAGAAGCA
GTACTATTTCAGTACTTGCTTAAGTAATTTCTGTTTTGTATGAGGTGGTTTTCTTAACGCT
GCGTAGTATTTTTTACAGTTCTTAACCATGTATCATTTACTGTCAATAATTTGTACCTGGT
GTTGTGGCCTCGATCATTTATGAGGCAGATCACGTACTCACAACGCATTGGTTTCTCGTT

GAATACAGATATGAACATAGGAGCTTGCTTGTTTGGTTTCAAACTCCTCAAGAGCCGCA
TTGCAGCTCTCCTGGTAGCTGGGTCCTACAGAATGATATTGATCTGCTTTGGTGTACATT
TTTGGTAGGTCATGCTTGTTTCGGTTTGTCTTTGTTATCTCCAGATTATTAGGCATGTTTT
CAGTTGGAATTTACATAGATGTAGGTGATTAATTTGCTATTAACTCAAAGATATTCTGTG
ATACTAGAAACACTATACCAGGTCATTTGGGAATCATGGCAGGTCAATTTGAACTTGAG
GGCTTATGGAATATTTCAGGATTTTATTGTATTAGTTAAGAACACCTCTTGATCTATGTT
ACACTAGATTTGGTAGAGCAGTGTCAGATTGACATTAAAATTGGATGTCAACTTCTATG
AAGCAGGATAAATCGTGTTGTGTCTGTATCTGCATAAATAATTTTGGGTGTTTGCAGGTA
TTTGGGCCACTGTGTAGGGTACTCCGTGGTCGCTTCTGTGTTATTGGGAGCTGTGGTAGT
TCGTCATGTGTCTGTTGTTAATCCAGAAGCAGCTCGTCGAGCTGCCTTAAAGAACACTGT
AATTCGACTGCGGGAGGGTTTCGGCAGGCGAGGTCCAGGAAGCTCGTCGAGTGTGTCGG
AAGGCCGAACTTCTAGTGTTGTCCATAGTAGCGTTGGAGCTGAGCAAATTGGAGCTGCG
ATAGAGCTGATTAGTAGGGCAAATGCAAGGCCTCAAGGTGAAACGTTAGGTGCTGGATT
TGTTGGACACAATGGAGGAACTGGGCTTACTGACGCTTTTACAGCAGGTTTCGGCCTTC
AAAGTGTGGATTCTCATGTGTCAACTGAAAGAGTTGGTAGCTCTCAGCAGCCCTCTTCAC
CAGATCATGACACTACAGACTCCAACCGCATCTCTTGCGCGGGATCTGTCGCTGCAATT
GAGCCATTGAACGCTGTGGAACGGAGCAGCACATTTGCATCATTGACCGATCAACAGAC
GCTTGATTTGAACCTAGCTTTGATGTTTCAAGAGAGAATGAATGACCCTAGAATTACCTC
AATTCTTCGTCAAAATCCGGAGTTCGCCCGGGGTGGAGAACACGAGCTGATGACATTGC
TTCAGGATAAGGGACTTGATCCAAACTTTGCAACCATGTTGAAAGAGAAAGGTTTAGAT
CCGACAATCTTAGCCTTGCTACAACGGAGCAGCATTGATGCTGGCAGAGATCCAGGGGG
AAACGAAAACTCAACTGCAGGCTTGAAACAGTCGCAATCACTCGCTCCAGATGATGCAG
TTGGCTGGACAGGGGAATCCCATAAACATGCTTGTGGGAATTGGTTCAAGAGTGTTGAA
GATGTTGTACAGTTCTTTGTGGGGACACCAGAGCGAGCATGGGTCTTCTTCAGTGTGATC
TTTGTTGTGGAATGTGTTATCGTCGCTGTATTCCGTCCCACAACAGTGACTGTCATCAAC
GGAAGACACGAGCAGGTTAGCAATCAACAGCTAACTATCTTGTATACGGGTTATCAAGA
GAGGGAGGATATATATAATACTGGTAAAGTGCACAAAGCTGGCTTATTCACCGAATCAT
ATAGAGACCTTCAACTATAGATTGTGTCAGCATTTTGAGGTAGCTAGTGAGTTAGAAAA
TGTGCAGTTTCACATTTCCACACGCTGTCCTACATATGAATCCTTTCGTCTTCTTTTAAAT
TTGGGGATTCATTGTGATTTTTGTTTGTTTGTTTGTCTTGTTTTGTTTTGTAGAAGTCGATT
GTGCGCTCTACGTTTTTTCGCCGCTATCTGTCATTCCAGCTTGCGTTACTTGTTTGTATAT
TGAGAACCCGGCTTTATTTTGCAGTTTGAATTTGGCTTTTCGGCACTTCTGCTGTCTCCTG
TCACATGCTCATTACTAGCTTTCCTACGGTCTCTGCAAGCAGAGAATATGGCCCTTACGA
AAAAGGTCCGAAAGGTATGTTTCGATTCCTGGGAAGTGAGGAATCATTGTTACCCACCA
TTCTAATTCTTGAACTTGCCAGTTTTTTCGGTCTTAAGCATTAGCGCTTCTATTTGTATTT
GTATGTATAGATGGCTCTTATTGTGATGGCACGAACACATGATAACCCTTTGAGACTTAA
TTTTTCTATTCTCTGACACTGATTATTGTAACTTTCAGTTTGGATTGATTGCATGGCTTTT

GAGTACTGGAGTGGGACTTCTCTTGGCTTTTCTCAGGTTTGATCCTTGCTTTAAGATATTC
AGTCCTGCAGTTAGAATACTATTATGGGTGCTAAAGAGATCTTGTGGGAGTGCATATCG
GTTGTGCCTTTGTTATTGTGCGTGCTGGTGTTATTGTTTTATTTTCACTAATTTATCTCTCC
GACAATTGTTCTAAGCATATGCTTTGTTTTCTCTGCTCTCTGACATCGAGCAACATCTTTT
TTGTGGCTGTTTTTTTTTTTTTATTGATGCAACAGTAAGTCATCGATCATCTTAGGGCTCG
CTGTGACTGTCCCGTTAATGTTGGCTGCCCTGTCTGTTGCACTGCCTATTTGGGTGCACA
ATGGGTATAATTTCTATCAAGCTCCTGTCTTGGAGTATCATGCTACTCGAGCTTGGGAAC
AACAGGCAAGGCGTGGCAGAACGAAAGAGGTGAACTTCGCTTATTTATTTGTTTTGTTTT
TCCTCAACCAATTTCTCGTGCATTTCATTATGTTGCTTAACTGAAAATTTTGCTCAGAAA
GGGTTTCCTTAGTGTTTAAATATCTTCAAGTAATGTAATTGAACTGTGTGTCGTAGTGTA
GTATCAGCTTTTTCACAGGTAGCAACCAAGTGCATTTGCTATAGTTTGTTGTTGATTGAG
ACACATATGGTGTGGTGTGACATCGTTGTAACAGATCTGTTATAGATGCGCAGCGTCAC
ATTGTGATGTTTGGTTTCGGTTATTGAAATGCCGTTCTGAAAAGTTGAAATTTCACTCTCT
TTTAGTGTTACAAGTTCATTGTTTACTGGCTCATTGGTGCGTTAGGTCTAGCATGATGAC
GCAAAATTGCTGCTTTATTTCTGCAGGATAATGTAATGACACTTTGCATCTTGGTCACTG
TTATCTGCATAATTGCTCTGGGGGTTATCATCTCATTCAGCCCTCTCGATGATATTAAAT
ATCATAGCTGGAGCAGTGTGACAAGATATTCCACCTCTCCTTACACGTCACCTTTATACC
TAGGGTGGGCAATTGCCTCAGCATTTGCCCTTGTGATAACTGGAGTATTACCAGCTATCT
CTTGGTTTGCGACTTATCGTTTTTCTCTTTCATCAGCTGTGTGTGTTGCCATCTTCACCGG
TACTTGTTATACTGTAACCTTAACTTGACTTTGGCACGCTTTGCCTGTTCGGCTTGCTTAG
CAACTTAATATTTGTGTTGTTTTGCTGTATCCCTTTAAACACTATAGAAGTATAACTTACG
ACTCCATTCGCTGGTGTCACACTAGAGGAATGTGCACATTTGAATGGTCTCATTAGAGGT
GTATGAAAAGTATAATGACTTATGTAATAATATTTTATTTTCTTATCAAATGGCAACATT
TGCAGTGGTATTGATGACATTCTGTGGAGGATCTTACGTGGGCATTGTTCGGAGCCGTAT
CGGTAAAACACCGACAGAAGCCGATTTTCTAGCTGCGTTGCTGCCGTTGGTCTGCATAC
CAGCTGTTTTCTCTCTTGGTTGTGGCCTTTACAAGTGGTGAGTAGTTGTGTCATTGATCAC
GAGCGAAAGCTTTCGACTTTATTGGTTGTTTGAAGATTTGTTTGAAAATAATATCGGAAT
TTCCGCTGGATACAATGTCGTGACCTAATCTTCTTCTGATAGTTTTTTCACGTCTCTTTTG
TAATCATTAGGAGAGACGAAGGTTGGCGGTTCTCGAAGGGTGCGTACGTTTTTGTTGGT
CTTGGGCTCACGCTTCTGTTGGGTGCAATCTCAGCGGTCATTGCTACCATTAATCCATGG
ATGGTAAGGAAAATTGGCTATTGGTCGTGAAAGACTCACACCGTTATTTCCTTCTGGATC
TGTCTTTTAAAAAAAAGGTATATATTTATTGATTTATTTTCATCTTGTGCTAATCTAGCTT
TTGTGATGACAGGTTGGCGCAGCTTTCCTGTTGGTTTTGGTGCTTGTCGTTTTGACGATA
GCTGTTATACATCATTGGGCGTCGCATAACTTTTACTTGACTCGCATGGAGGTTCTTTTT
GTCTGTCTGGCGGCGTTGGTACTGGCTCTAGCTGCTTTCCTTATTGGACTCCTAGCGGGT
GAGGACCGTCTTTGGTGATGTTTTCAGCTGTTTTGTGTTATGATTGCTAGATATTTGTTTG
CAATATTCTAGAATTGCACTATACACTCCAGCAAACATATATTGTGGATCGCTTTGAGCT

ACACTGGTGTTCTACCTGTAGTTATTTCTGTTACTTTTTTGTTTGTTGCAGATGAACCATT
CGTGGGAGCATCAGTGGGATATTTCGCTTTCCTAGGAGTTCTGGCTGGAAGGTCTTTCAC
GGTAATGTTTGCGTAGAATGCTCCTAACTATCACTTGGTTCTCGTCAGGTTCGAGAACAT
CAATATGTCGCCATAATTTCTTGTTTGGTTTTTTGGTGGGGTGAGTATATTTTGCCGAATT
GACTGACTTGACAAGGTCAGTACAATGATTCTATTTGTCATCTATATTAACATCGGCCCC
ATGTTTATTCTATGTAACGGACAGGTTCTCTTGTCTCCCGCGGTGGTTGTATACTCTCCTA
GAGTTCTCCCCGTCTACGTCTACGATGCCCATGCAGACTCTGCTAAGAATGTCAGGTAA
GTTTCGATGATTTATTGTTCTATTCATGTTCCATTGGAGCAACCTCATTTCTGAGATCTTG
AAGAGCAATTGTGTACATTTTTAGCTGTACGTAAAGAACTCTGCAGTGCATAAGAGGAT
CTGTTTTACATATGTTTTTCTGGAAATTTTCTCATGGAAACTTGTATCACACCAGTGGCG
CCTTCTTGGTTCTTTATGGGATAGCCCTGGCCACAGCGGGATGGGGAGTGGTTGCGAGTT
TGGAGATCTATCCTCCCTTTGCAGGATCAGCGGTTTCAGCCATCACCTTGGTCGTGGCAT
ATGCTTTTGCTATTTCCCGGCCACAGCTTACATTAAAGGTGTACGCATATGACATTATTA
CTATTGGTTTACCTCACATTCATATTTGATTTGATAATTGCAGCCGTGTTGAGTACAGTG
AAATGATTCTTTTACTCAGTTTTCTTGCGCTTTTTTTGTTCTTGTAGATGATGGAAGATGC
TCTTCAGTTTTTGCGAAAGGAGACTATTGCTCAAGCTATTGCTCGCTCCTCAACTAAGGT
AACTTATGGGTGATGCATTATTTCATAGTATTTCTATCCTGCCATATTTTGGCTTCAGCTA
TTTCGTTAGATTAGGCGAGTTCAGCAAACGAGCAACTTTGATCACTCATGGCCTTAAACT
CACATTGTTGCTGTTTTCCAGACTAGGAATGCCATGTCGGGAACATGTTCAGCTCCTCAA
CGATCTGCAAGTTCAGCAGCTCTGCTTGTAGGTGATGCTTCTGTTACCCGTGATAAAGCT
GGAAATTATGTTCTTCCTCGTGCTGATGTATTGAAGCTGCGGGAATGGCTTCGCAACGA
GGAACATGCCGCGGGGCTTCGATGGTGGCCTCCGTTCTCTGCCTTCGGCTCTCCCTTTGA
CTCTACAACGGATACTAGGTTCAGAAGGAAGTTGTGCGCTCATGCGCGAATATTGGCAC
TTGAGGAGGCTATAGACACTGAGTGGGTGTACATGTGGGATAAGTTTGGTGGCTATTTG
CTTCTGGTTCTTGGTCATGCTACACGAGCCGAAAGAGTACAGGTACGACCAAGTCCTGA
AGCACTGATTTCCAGAAATTCTTTGGCATGTATTCCACCTTATAGAGGAGATTTTTTTCTT
CTTGATGATTGATGTGCGGATATTTTTTAAGGGTGGGATTCTTTTGAAATTTAGTCTTCTT
CAACTTTAGCAAACTGGATCTATTCTCCATTGAGTTCTAGAATAAATTTTGAGTTCTTTTT
TTTGTTCTCATATTCCTGAGAATAGTTCGCTGTTTTAGCACGGCTATTCTTTTCTATTTAA
ACAGTCTTTTTATTTATTTTCCTCCTTACAGGATGAAGTGAGGCTACGGTTATTTCTGGAC
AGCATTGGCTTCTCGAATCTCAGTGCAAAGAAAATCAAGACATGGACACCTGAAGATAG
GAAGCAGTTCGAGACAGTGCAAGAGAAGTAAGCGCGTGCTTTCCTTCTTTCTCTCTCTCT
CTCTCTTTTTTTACTTTCTTCTTCTTCTTCTTCTTCTTCATAATCATCATTATCATC
ATCATCATCATCTTTTCCTCTGCATGATTAATATGATCTTCAGTTTTGGGCATAGATGAA
GTATACCCTTTCCAAGTTAGTAGATCATAGAGTTATCATTCGGTGGTATTATTCCGAACT
AGAGCACATATAACCGATTCCTCTGAAGTGTTTGATATCGTAAATTTAGTACATTAAGTC
GTAATCTTAAGTCCACTATTCCTGATTTATGGTGTCTAAAACTTTTATTTAAAATTAATTA

TAAGTGTTGGTGGGTTTCTCTGCACTTGCGTCTCCAAGTGTTTGGTCCTAAAGTTCTTTAT
AAATGACGTGTCTTGCATCGGAACAAGAATCTAGTGTACAGTCATCGATGTTGGACAAC
ATTTCATATAGTACACAATGTTTTGATGCCGAATTTAGTTGCATACTGTGTTGCATTCCTA
CTGTACTAAGTAGATGATGGTGGGTGCAATAGTCTCGTTTCAAGGTGCTTCAAAAGTGG
AATTTCTGAATATCTGATTCAAGAATGTGGCACATGACTACGTATGTTATTCAGTTATAT
GCGGGAGAAGGAAACAGAAGAGGAGCTTTTGCAGCAAAGGCGTGAGGAGGAAGGAAA
AGGTCGTGAGAGGCGCAGGGCTTTGCTCGAAAAGGAAGAACGACGACGTCACGAAATT
GAGGCGTCAGTTATTTCAACTGTTCCTGATGCAGGTACCAAAGAAGCAGCAGCGATGGC
AGCTGCTGTTCGTGCAGTTGGAAGAGATATTTTACTCGATGAGAACTCTGCCAGTGATC
AGGTGTCGAGTTTGGCCCGGCGAATATTAGTTGCTCAAAGAGCCCAACGTGCGCAGCAG
GTCTGAGAACGAAGGGCTTATTTTCCGTGCGATGCAGGACTGCTTTTTAAAATTGTTTAA
ATTTATGAATAGTTCTGCGTTAAAGCTGGAAATTGTTTTACAAACCTTTTTAATATATTG
ATTGGCACATTTCAGACCGGAGTCCATGGAACTGTTTGTGTGATGGACAATGAGCCTCG
TAATAGTGGTCGCTGCTGTGGACTTATTGACCCAAGTGTTTGCGGGAGCCAGAAAGTGA
GCTTTTCTGCAGCAGTGATGATTCAACCAGATTCTGGACCAGTCTGCATATTGGGCGTTG
AAGCTCACAATAAAAAATGTTTGGAATTTCTAGTCGCGGGTGCTGAGCAGGGTTTAGAA
GCAGGGCAAGTTGGACTGCGGCTTGTATCGAAAGGAGCTGGCCAGTCAACACATAATA
GAGAGTGGTTTATCGGACACACCTGTGTAGCTGATGGAAGGTGATTTGACTCGAACACA
TTCTTAGCCGACAGAAATACACATTTTAAAGATGTTGATATTTCTCTATGTTTAAGACAA
ATGCCTCAACCTCTCCACTTTTGAATGCAGATGGCACTCAGTTACTATAACGGTTGATGC
TTCTACTGGCGAGGCTTCTGCTTTCATAGACGGAGGTTTCGATGGAGACAACAGATGTG
ACTTGTTGCTGCTGGGAGAGGGTGGAGTATGGGAAGAAGGGATGGAAGTATGGGTTGG
CATCAGGCCGCCTATGGATCTTGATGCCTTTGGTCGTTCCGATAGTGAAGGAGCGGAAT
CCAAAATGCATGCCATGGATGTTTTCATGTGGGGTCGCTGTTTAACCGAAGATGAAATTT
TAATGGTGCATAATTGTGCTAATCTGGAGGAGCAGAATGAGCTGGATATTATGGATGAC
TATTGGCAGGGATCTCCAATGGAAACGCCAATGAGAGTAAGTTGGGCATATCACGAAGC
ACATGGTTTAATTGAGTCAGTAATTAGATGTGGGGAATTGTAAGCTTGGAGTCTGCATGT
TCTTTAGATTGCATGTCAAAACAACCCTCATGCTTCTAAGTTACTTATGTTTGATGTATTC
ATCCGAAGTGGTCTGACCATTTGGTTTTGTTGAAGGTGCTTAAGTTTTTATAACTCATAC
TGTGATTCTCATGTGTGTGTGTTTTGATGTCCTGCCGGTCGTGTATCTATCTTGTTCTCTA
GTTTGAGGATTGGAGATATGATCCGGATCTCAGCTACGAGCGAGATGAAATTTTTTGGG
ATGAACAACAATCCTCAGGAGCAAAGAGGAAGGCAGCAGACATCGAGAATGTGGCGGT
AGATATTGAATATTTATCCAGGAAGATCCGACGCCCCAAAATTGAAACGAAAGAGGAG
GTCTTGCAGCGCATGCAGGCTGTTGAATTGGCTATTAAGGAGGCTTTGGTAGCAAGAGG
CGAGAAGAGATTTACCGACCAAGAGTTTCCGCCATCAGATCAGTCGCTTTTTGTTGATAC
TGATCACCCTGCTGCTAAGCTTCAGGTAAGCTAATATCCTTAACTCGTCTTTTAGTGTCC
AGGAATATTTAGATGTAATACATACAGGCTATTTTGGTGTGGTGACTGCCAATTAAAAG

TGCATTTTTATGTTTGCCAGGTTGTTAACAAGTGGTCGCGCCCAGAAGAGGTTTTGCGCG
AACAAGGGCAAGATTCTCGGCCGTGCCTGTTTGCAGGCACTGCCAACCCTTCAGACGTC
TGCCAGGTTCGAAGTGAACTTTATTTGTCTCCTTCAAGTTTTACTCAACCCGTTCATTGAT
TCATTCATGAACACCATTTGAGCGCTTTCGTGTTAAGCTGCATGTAGATAGAATACTTAA
ACGAGCACCATTTTCTGAGTCCTTCACCAGGGCTACATACGCGGTCATTGAAGCAACTT
GGAAACATTTTCGGTCATATCAAGCTAACAGCTTTTTAATAGTAATAGTAGTAACACTTT
CGCCAAATTCTTAGGGGCATAGCACGTTATAGGGTAAGAATTATGTCTAATCGCAGACA
CAGAGAATCTTTACCCACATCTCATGGAAACAAAGCATATTAAAAATTGCAGTAAAGGA
AGGTTGAGCACCTTCGAATCAGCTCCTGTTTATCTTCGTTTACGTATCAGTTTGTAGTCC
ATTAGTTTTCCTTTTTACGAACATGCCCTTCTAGAAAGAAGTAAAAAGTTGAACTTAACG
GTTCTAATTAAGCTCTAATAGTATATATCGGATTTATATATGGAAGCCCATGTCAGTTTC
TGGGAAGCTCAATGTCTAACATTGTTCTGTAACCTATACTTGTTATGCAGGGCCACCTAG
GAGATTGCTGGTTTTTGAGTGCAGTTGCTGTGTTAACCGAGGCATCACGGATATCTGATG
TCATGATTACCCCCGAATTCAACGAAGAGGGGATCTATACAGTTAGGTTTTGTATTCAG
GTATGAGGATGAAGATCATAATGATGTTTCCGTACTTGTTGTATTTTCATATTTGATATG
GTTACTTTTGAGTACTCTTTCATTCTACTCTTTTCAGTTTCCTCCAAATATGTTTTCAACC
ATTGTATTTAATGCATGCAGGGTGAGTGGGTGCCTGTGGTCGTTGACGACTGGATACCTT
GCGAGGCCAGGGGAAAACCGGCCTTCGCCACTAGCAGAAAAGGAAATGAACTTTGGGT
GTCTATTTTGGAGAAAGCGTATGCGAAGTTACACGGGTCTTACGAGGCCTTGGAAGGTG
GTCAGGTCCACGATGCGCTTGTCGACCTGACGGGAGGTGCTGGAGAGGAGATTGATCTA
ACAAATGAAGTTGCGCAGTTGGATCTTGCTAGTGGCCATCTATGGTCCCAGTTGCAACG
CTTTAAGCAAGAAGGATTTCTTCTCGGAGCAGGCAGCCCATCGGGTTCTGATGTTCATGT
TTCATCTAGTGGAATTGTACAAAATCATGCATATTCTCTTTTACAGGTGAGCACGATCTG
TCCTTTCCAATTCCAAACTACTACCTTGGTTATTTTTGTCAGGTTTTATTCACTTCAAATT
TTTGTAAATACAGTTATTATTACAGCATGCTTCAACATTTCATATTACGATTGAAGTACG
GCAATGTTTCTAAACGATCGATCATATAAATGTATCAATGTATCAGGTACGAGAGGTCG
ACGGTCACAAGCTCGTGCAGATCCGCAACCCATGGGCTAATGAAGTTGAATGGAATGGC
CCGTGGTCGGATTTTTCCCCAGAGTGGACAGATAGAATGAAACACAAGCTCAAGTATTC
CCCTCAGGTATTTGGAATTTAGTTCTGTGTTTTACAGTTCTGATCCCTTTCCATTGGTAGG
ATATGGCTCTTTCGTTGTGCAACTGATGGATAGTGTCCCACTTTGTAGCCTTTCTTTTGCT
GCAAAGGGGTCTCAGCGTGTGCCACTCGCGTGTTGCCTTAAGCACCTTTCATATTTTCCA
TCGAAGGATCTCAAGTATCTCAATGTGTTTCTAGGCTGCAAATGGAGTATTCTGGATGTC
ATGGCAAGATTTCCAGCTTCATTTTCGTTCTCTATACGTGTGTCGAATTTATCCACCTGA
GATGAAGTATTCAGTTCGGGGTCAGTGGCGTGGACCAACTGCTGGTGGATGCCAGGATT
ACGAAACTTGGCATTTAAATCCTCAATTTCATTTGAAAGCCGTGGGTAGTGATGCCCGA
GAACCCATTCATGTCTTCGCAACACTTACTCAGGTTTGCAATGTTCACTTCTTCAATAGT
GGGTGGCCTCAACAAGTAGATGAGATGCATTTGTTAATTGATTAGTTTTATTTATTTATT

TATTTATTTTTATCATACCCTGGAGTTTAGTTGTGTGCCTCTCATGAAGATATGATTTGCT
GAACAGGGTGTGCAGTCCACATCTCGATCTACAGCCAGCTTTGGGAATTACCAATTAGT
TGGAGATGCACCTCGATTTTATATCGGCATGCGTGTCATAAAAACCGGTGGTAGGCGAT
CTGGAAAGAATATCTTTATGCATGAAGCAGTGAATGGAACAGATTATGTGAATGCAAGA
GAAATATCTTGCGAAATGGTACTTGATCCTGATCCTAAAGGTTACACGATTGTTCCAACC
ACACATGCGCCGGGTGAAGAATGCCAGTTCTTGCTCTCAGTTTTCACGAAGGCTTCTATC
ATCTTGGAACCGCTTTAG

B. *P. patens* cDNA sequence

ATGGGGGCGTCCGTGGACGGTGCACCTGTGCACGCTGTGGTCAAATCTTGTGCGCTGTTT
GGGTCACTCTTCGTGGTTCTTTGTGCACTCTCCGTCGTTATTCTAGTTGCTGTAAATTGGC
GCCCCTGGCGGATCTACAGCTGGATATTTGCCCGCAGATGGAGATCGTTTATGCAAGGC
CATCGGCTCAGCATAATTACAGCAGTACTTGCCGCAGCTGCATGGACCATAGTCCTGTC
CCCAATCGCGGTTCTTATCTTGTGGGGAGCTAGACTCATAATTTTACTGAATCATGACAC
AATCGGCCTTGCAGTGATCCTGGCAGGCACAGCTCTACTGTTAGCATTTTATGCTATCAT
GCTTTGGTGGAGGACGCAATGGCAAAGCTCACGTGCTGTTGCAATCCTGCTTCTTCTAGC
TGTATCGTTATTATGCGCCTACGAGCTTTCAGCAGTCTACGTTACAGCAGGACGAAGTGC
TTCAAATCAGTTTTCTCCATCAGCTTTCTTTTTCGGGGTTTCAGCGATCGCAATGGGCATT
AACATGCTCTTCATATGTAAAATGGTTTTCAATGGGGCAGGGCTTGATGTGGATGAGTAT
GTGAGAAGGTCCTACAAGTTTGCGCATGCAGAAACGTTAGAAGTTGGTCCGATTGCATG
CTTGCCAGAGCCTCCTGAGCCTCGCGACGCAAGCATTCAGAGGAAAAGCAGTACATGGC
GGCTGGCCCAGTTATATCAAGCATCTGTTGTGGTTCTGGCGGCATATTCAGTGCTGTATG
GACTGACAGCAAGGGAAGCACGATGGCTTGGAGGAGTTACTTCGGCCGCTGTGATCATT
TTGGATATGAACATAGGAGCTTGCTTGTTTGGTTTCAAACTCCTCAAGAGCCGCATTGCA
GCTCTCCTGGTAGCTGGGTCCTACAGAATGATATTGATCTGCTTTGGTGTACATTTTTGG
TATTTGGGCCACTGTGTAGGGTACTCCGTGGTCGCTTCTGTGTTATTGGGAGCTGTGGTA
GTTCGTCATGTGTCTGTTGTTAATCCAGAAGCAGCTCGTCGAGCTGCCTTAAAGAACACT
GTAATTCGACTGCGGGAGGGTTTCGGCAGGCGAGGTCCAGGAAGCTCGTCGAGTGTGTC
GGAAGGCCGAACTTCTAGTGTTGTCCATAGTAGCGTTGGAGCTGAGCAAATTGGAGCTG
CGATAGAGCTGATTAGTAGGGCAAATGCAAGGCCTCAAGGTGAAACGTTAGGTGCTGG
ATTTGTTGGACACAATGGAGGAACTGGGCTTACTGACGCTTTTACAGCAGGTTTCGGCCT
TCAAAGTGTGGATTCTCATGTGTCAACTGAAAGAGTTGGTAGCTCTCAGCAGCCCTCTTC
ACCAGATCATGACACTACAGACTCCAACCGCATCTCTTGCGCGGGATCTGTCGCTGCAA
TTGAGCCATTGAACGCTGTGGAACGGAGCAGCACATTTGCATCATTGACCGATCAACAG
ACGCTTGATTTGAACCTAGCTTTGATGTTTCAAGAGAGAATGAATGACCCTAGAATTAC

CTCAATTCTTCGTCAAAATCCGGAGTTCGCCCGGGGTGGAGAACACGAGCTGATGACAT
TGCTTCAGGATAAGGGACTTGATCCAAACTTTGCAACCATGTTGAAAGAGAAAGGTTTA
GATCCGACAATCTTAGCCTTGCTACAACGGAGCAGCATTGATGCTGGCAGAGATCCAGG
GGGAAACGAAAACTCAACTGCAGGCTTGAAACAGTCGCAATCACTCGCTCCAGATGATG
CAGTTGGCTGGACAGGGGAATCCCATAAACATGCTTGTGGGAATTGGTTCAAGAGTGTT
GAAGATGTTGTACAGTTCTTTGTGGGGACACCAGAGCGAGCATGGGTCTTCTTCAGTGT
GATCTTTGTTGTGGAATGTGTTATCGTCGCTGTATTCCGTCCCACAACAGTGACTGTCAT
CAACGGAAGACACGAGCAGTTTGAATTTGGCTTTTCGGCACTTCTGCTGTCTCCTGTCAC
ATGCTCATTACTAGCTTTCCTACGGTCTCTGCAAGCAGAGAATATGGCCCTTACGAAAA
AGGTCCGAAAGTTTGGATTGATTGCATGGCTTTTGAGTACTGGAGTGGGACTTCTCTTGG
CTTTTCTCAGTAAGTCATCGATCATCTTAGGGCTCGCTGTGACTGTCCCGTTAATGTTGG
CTGCCCTGTCTGTTGCACTGCCTATTTGGGTGCACAATGGGTATAATTTCTATCAAGCTC
CTGTCTTGGAGTATCATGCTACTCGAGCTTGGGAACAACAGGCAAGGCGTGGCAGAACG
AAAGAGGATAATGTAATGACACTTTGCATCTTGGTCACTGTTATCTGCATAATTGCTCTG
GGGGTTATCATCTCATTCAGCCCTCTCGATGATATTAAATATCATAGCTGGAGCAGTGTG
ACAAGATATTCCACCTCTCCTTACACGTCACCTTTATACCTAGGGTGGGCAATTGCCTCA
GCATTTGCCCTTGTGATAACTGGAGTATTACCAGCTATCTCTTGGTTTGCGACTTATCGTT
TTTCTCTTTCATCAGCTGTGTGTGTTGCCATCTTCACCGTGGTATTGATGACATTCTGTGG
AGGATCTTACGTGGGCATTGTTCGGAGCCGTATCGGTAAAACACCGACAGAAGCCGATT
TTCTAGCTGCGTTGCTGCCGTTGGTCTGCATACCAGCTGTTTTCTCTCTTGGTTGTGGCCT
TTACAAGTGGAGAGACGAAGGTTGGCGGTTCTCGAAGGGTGCGTACGTTTTTGTTGGTC
TTGGGCTCACGCTTCTGTTGGGTGCAATCTCAGCGGTCATTGCTACCATTAATCCATGGA
TGGTTGGCGCAGCTTTCCTGTTGGTTTTGGTGCTTGTCGTTTTGACGATAGCTGTTATACA
TCATTGGGCGTCGCATAACTTTTACTTGACTCGCATGGAGGTTCTTTTTGTCTGTCTGGCG
GCGTTGGTACTGGCTCTAGCTGCTTTCCTTATTGGACTCCTAGCGGATGAACCATTCGTG
GGAGCATCAGTGGGATATTTCGCTTTCCTAGGAGTTCTGGCTGGAAGGTCTTTCACGGTT
CTCTTGTCTCCCGCGGTGGTTGTATACTCTCCTAGAGTTCTCCCCGTCTACGTCTACGATG
CCCATGCAGACTCTGCTAAGAATGTCAGTGGCGCCTTCTTGGTTCTTTATGGGATAGCCC
TGGCCACAGCGGGATGGGGAGTGGTTGCGAGTTTGGAGATCTATCCTCCCTTTGCAGGA
TCAGCGGTTTCAGCCATCACCTTGGTCGTGGCATATGCTTTTGCTATTTCCCGGCCACAG
CTTACATTAAAGATGATGGAAGATGCTCTTCAGTTTTTGCGAAAGGAGACTATTGCTCA
AGCTATTGCTCGCTCCTCAACTAAGACTAGGAATGCCATGTCGGGAACATGTTCAGCTC
CTCAACGATCTGCAAGTTCAGCAGCTCTGCTTGTAGGTGATGCTTCTGTTACCCGTGATA
AAGCTGGAAATTATGTTCTTCCTCGTGCTGATGTATTGAAGCTGCGGGAATGGCTTCGCA
ACGAGGAACATGCCGCGGGGCTTCGATGGTGGCCTCCGTTCTCTGCCTTCGGCTCTCCCT
TTGACTCTACAACGGATACTAGGTTCAGAAGGAAGTTGTGCGCTCATGCGCGAATATTG
GCACTTGAGGAGGCTATAGACACTGAGTGGGTGTACATGTGGGATAAGTTTGGTGGCTA

TTTGCTTCTGGTTCTTGGTCATGCTACACGAGCCGAAAGAGTACAGGATGAAGTGAGGC
TACGGTTATTTCTGGACAGCATTGGCTTCTCGAATCTCAGTGCAAAGAAAATCAAGACA
TGGACACCTGAAGATAGGAAGCAGTTCGAGACAGTGCAAGAGAATTATATGCGGGAGA
AGGAAACAGAAGAGGAGCTTTTGCAGCAAAGGCGTGAGGAGGAAGGAAAAGGTCGTG
AGAGGCGCAGGGCTTTGCTCGAAAAGGAAGAACGACGACGTCACGAAATTGAGGCGTC
AGTTATTTCAACTGTTCCTGATGCAGGTACCAAAGAAGCAGCAGCGATGGCAGCTGCTG
TTCGTGCAGTTGGAAGAGATATTTTACTCGATGAGAACTCTGCCAGTGATCAGGTGTCG
AGTTTGGCCCGGCGAATATTAGTTGCTCAAAGAGCCCAACGTGCGCAGCAGACCGGAGT
CCATGGAACTGTTTGTGTGATGGACAATGAGCCTCGTAATAGTGGTCGCTGCTGTGGAC
TTATTGACCCAAGTGTTTGCGGGAGCCAGAAAGTGAGCTTTTCTGCAGCAGTGATGATT
CAACCAGATTCTGGACCAGTCTGCATATTGGGCGTTGAAGCTCACAATAAAAAATGTTT
GGAATTTCTAGTCGCGGGTGCTGAGCAGGGTTTAGAAGCAGGGCAAGTTGGACTGCGGC
TTGTATCGAAAGGAGCTGGCCAGTCAACACATAATAGAGAGTGGTTTATCGGACACACC
TGTGTAGCTGATGGAAGATGGCACTCAGTTACTATAACGGTTGATGCTTCTACTGGCGA
GGCTTCTGCTTTCATAGACGGAGGTTTCGATGGAGACAACAGATGTGACTTGTTGCTGCT
GGGAGAGGGTGGAGTATGGGAAGAAGGGATGGAAGTATGGGTTGGCATCAGGCCGCCT
ATGGATCTTGATGCCTTTGGTCGTTCCGATAGTGAAGGAGCGGAATCCAAAATGCATGC
CATGGATGTTTTCATGTGGGGTCGCTGTTTAACCGAAGATGAAATTTTAATGGTGCATAA
TTGTGCTAATCTGGAGGAGCAGAATGAGCTGGATATTATGGATGACTATTGGCAGGGAT
CTCCAATGGAAACGCCAATGAGATTTGAGGATTGGAGATATGATCCGGATCTCAGCTAC
GAGCGAGATGAAATTTTTTGGGATGAACAACAATCCTCAGGAGCAAAGAGGAAGGCAG
CAGACATCGAGAATGTGGCGGTAGATATTGAATATTTATCCAGGAAGATCCGACGCCCC
AAAATTGAAACGAAAGAGGAGGTCTTGCAGCGCATGCAGGCTGTTGAATTGGCTATTAA
GGAGGCTTTGGTAGCAAGAGGCGAGAAGAGATTTACCGACCAAGAGTTTCCGCCATCA
GATCAGTCGCTTTTTGTTGATACTGATCACCCTGCTGCTAAGCTTCAGGTTGTTAACAAG
TGGTCGCGCCCAGAAGAGGTTTTGCGCGAACAAGGGCAAGATTCTCGGCCGTGCCTGTT
TGCAGGCACTGCCAACCCTTCAGACGTCTGCCAGGGCCACCTAGGAGATTGCTGGTTTTT
GAGTGCAGTTGCTGTGTTAACCGAGGCATCACGGATATCTGATGTCATGATTACCCCCG
AATTCAACGAAGAGGGGATCTATACAGTTAGGTTTTGTATTCAGGGTGAGTGGGTGCCT
GTGGTCGTTGACGACTGGATACCTTGCGAGGCCAGGGGAAAACCGGCCTTCGCCACTAG
CAGAAAAGGAAATGAACTTTGGGTGTCTATTTTGGAGAAAGCGTATGCGAAGTTACACG
GGTCTTACGAGGCCTTGGAAGGTGGTCAGGTCCACGATGCGCTTGTCGACCTGACGGGA
GGTGCTGGAGAGGAGATTGATCTAACAAATGAAGTTGCGCAGTTGGATCTTGCTAGTGG
CCATCTATGGTCCCAGTTGCAACGCTTTAAGCAAGAAGGATTTCTTCTCGGAGCAGGCA
GCCCATCGGGTTCTGATGTTCATGTTTCATCTAGTGGAATTGTACAAAATCATGCATATT
CTCTTTTACAGGTACGAGAGGTCGACGGTCACAAGCTCGTGCAGATCCGCAACCCATGG
GCTAATGAAGTTGAATGGAATGGCCCGTGGTCGGATTTTTCCCCAGAGTGGACAGATAG

AATGAAACACAAGCTCAAGTATTCCCCTCAGGCTGCAAATGGAGTATTCTGGATGTCAT
GGCAAGATTTCCAGCTTCATTTTCGTTCTCTATACGTGTGTCGAATTTATCCACCTGAGA
TGAAGTATTCAGTTCGGGGTCAGTGGCGTGGACCAACTGCTGGTGGATGCCAGGATTAC
GAAACTTGGCATTTAAATCCTCAATTTCATTTGAAAGCCGTGGGTAGTGATGCCCGAGA
ACCCATTCATGTCTTCGCAACACTTACTCAGGGTGTGCAGTCCACATCTCGATCTACAGC
CAGCTTTGGGAATTACCAATTAGTTGGAGATGCACCTCGATTTTATATCGGCATGCGTGT
CATAAAAACCGGTGGTAGGCGATCTGGAAAGAATATCTTTATGCATGAAGCAGTGAATG
GAACAGATTATGTGAATGCAAGAGAAATATCTTGCGAAATGGTACTTGATCCTGATCCT
AAAGGTTACACGATTGTTCCAACCACACATGCGCCGGGTGAAGAATGCCAGTTCTTGCT
CTCAGTTTTCACGAAGGCTTCTATCATCTTGGAACCGCTTTAG

C. 5’ targeting sequence (5’_TGS)

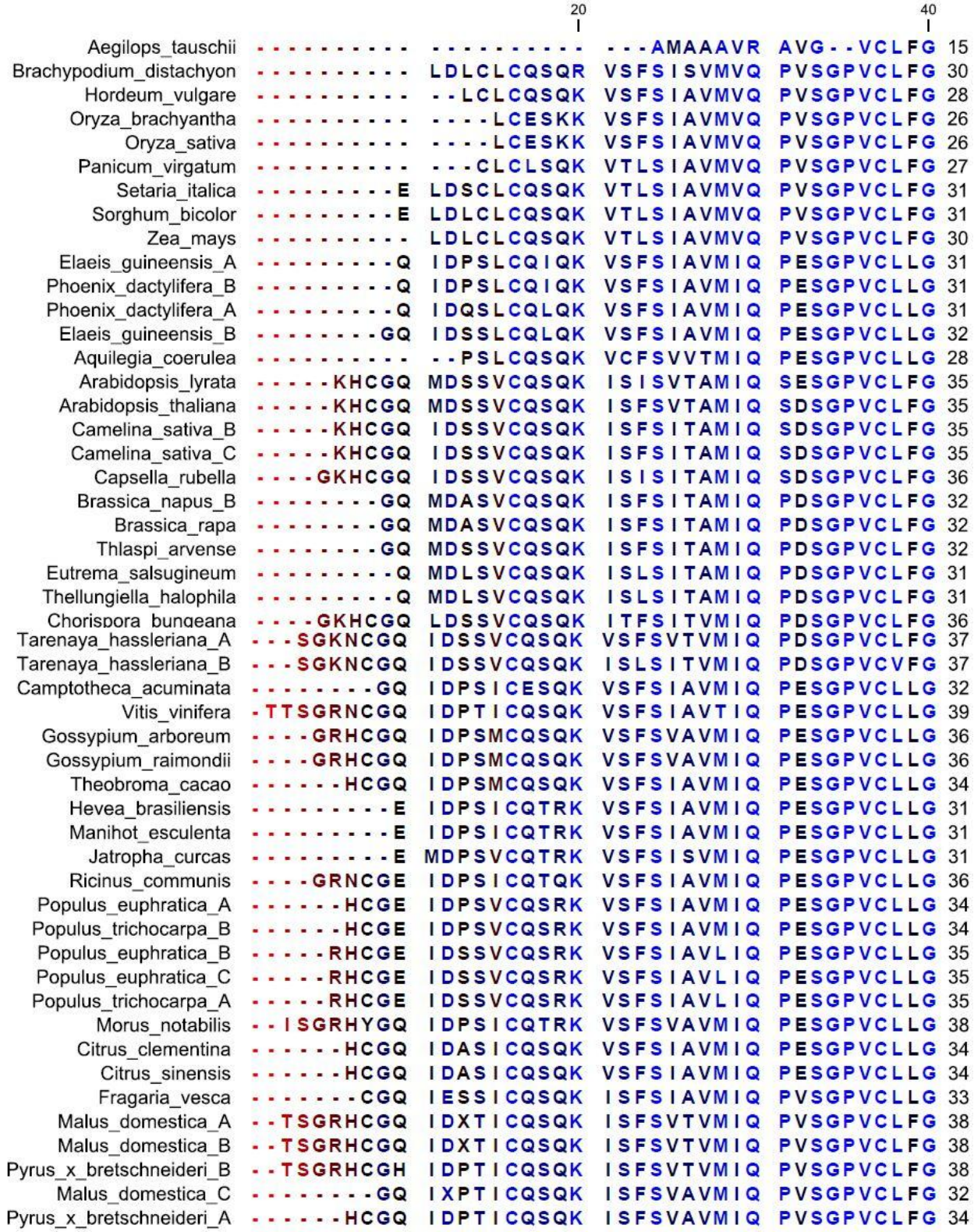
TGCATGATTAATATGATCTTCAGTTTTGGGCATAGATGAAGTATACCCTTTCCAAGTTAG
TAGATCATAGAGTTATCATTCGGTGGTATTATTCCGAACTAGAGCACATATAACCGATTC
CTCTGAAGTGTTTGATATCGTAAATTTAGTACATTAAGTCGTAATCTTAAGTCCACTATT
CCTGATTTATGGTGTCTAAAACTTTTATTTAAAATTAATTATAAGTGTTGGTGGGTTTCTC
TGCACTTGCGTCTCCAAGTGTTTGGTCCTAAAGTTCTTTATAAATGACGTGTCTTGCATC
GGAACAAGAATCTAGTGTACAGTCATCGATGTTGGACAACATTTCATATAGTACACAAT
GTTTTGATGCCGAATTTAGTTGCATACTGTGTTGCATTCCTACTGTACTAAGTAGATGAT
GGTGGGTGCAATAGTCTCGTTTCAAGGTGCTTCAAAAGTGGAATTTCTGAATATCTGATT
CAAGAATGTGGCACATGACTACGTATGTTATTCAGTTATATGCGGGAGAAGGAAACAGA
AGAGGAGCTTTTGCAGCAAAGGCGTGAGGAGGAAGGAAAAGGTCGTGAGAGGCGCAG
GGCTTTGCTCGAAAAGGAAGAACGACGACGTCACGAAATTGAGGCGTCAGTTATTTCAA
CTGTTCCTGATGCAGGTACCAAAGAAGCAGCAGCGATGGCAGCTGCTGTTCGTGCAGTT
GGAAGAGATATTTTACTCGATGAGAACTCTGCCAGTGATCAGGTGTCGAGTTTGGCCCG
GCGAATATTAGTTGCTCAAAGAGCCCAACGTGCGCAGCAGGTCTGAGAACGAAGGGCTT
ATTTTCCGTGCGATGCAGGACTGCTTTTTAAAATTGTTTAAATTTATGAATAGTTCTGCGT
TAAAGCTGGAAATTGTTTTACAAACCTTTTTAATATATTGATTGGCACATTTCAGACCGG
AGTCCATGGAACTGTTTGTGTGATGGACAATGAGCCTCGTAATAGTGGTCGCTGCTGTG
GACTTATTGACCCAAGTGTTTGCGGGAGCCAGAAAGTGAGCTTTTCTGCAGCAGTGATG
ATTCAACCAGATTCTGGACCAGTCTGCATATTGGGCGTTGAAGCTCACAATAAAAAATG
TTTGGAATTTCTAGTCGCGGGTGCTGAGCAGGGTTTAGAAGCAGGGCAAGTTGGACTGC
GGCTTGTATCGAAAGGAGCTGGCCAGTCAACACATAATAGAGAGTGGTTTATCGGACAC
ACCTGTGTAGCTGATGGAAGGTGATTTGACTCGAACACATTCTTAGCCGACAGAAATAC
ACATTTTAAAGATGTTGATATTTCTCTATGTTTAAGACAAATGCCTCAACCTCTCCACTTT

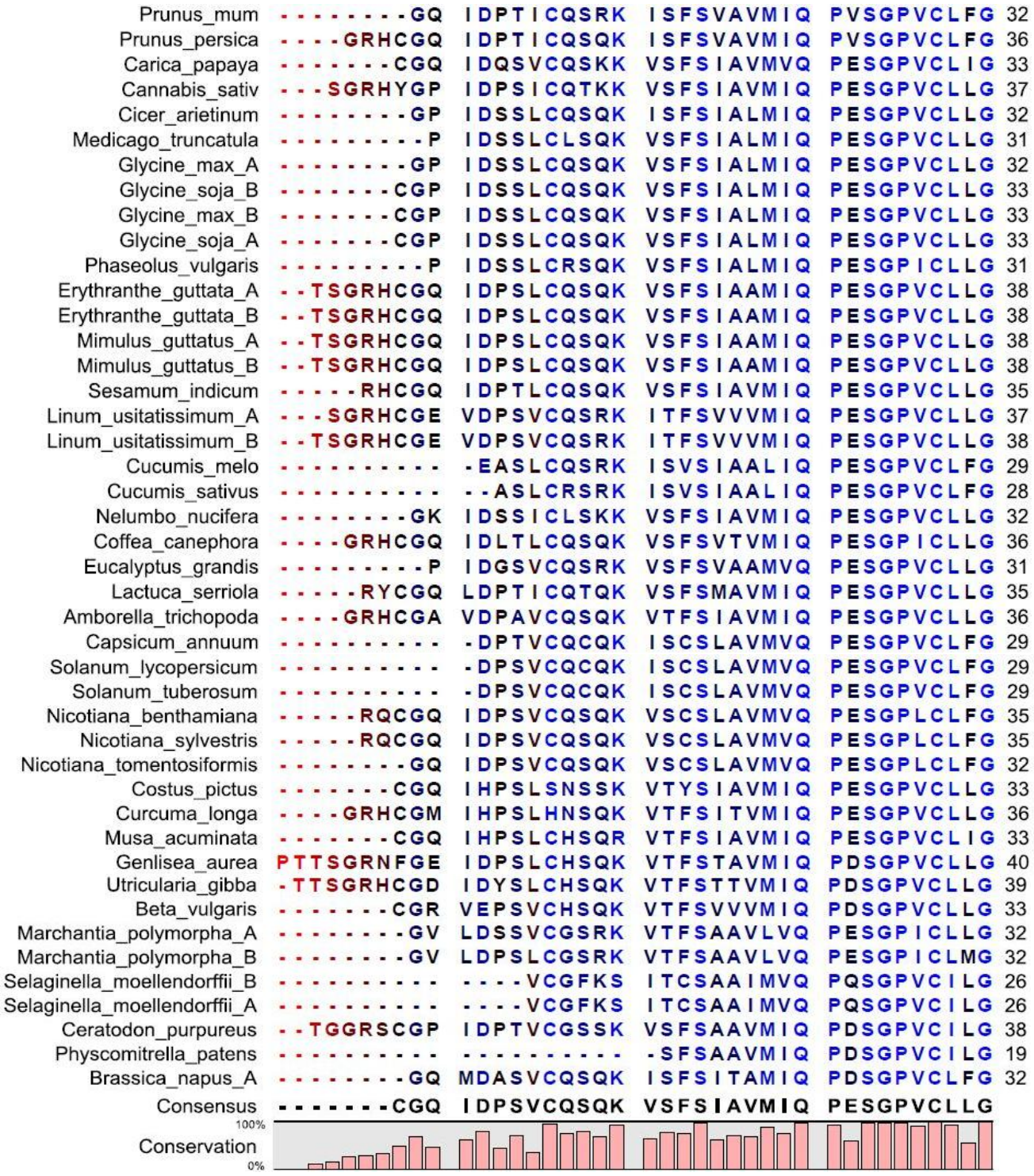
TGAATGCAGATGGCACTCAGTTACTATAACGGTTGATGCTTCTACTGGCGAGGCTTCTGC
TTTCATAGACGGAGGTTTCGATGGAGACAACAGATGTGACTTGTTGCTGCTGGGAGAGG
GTGGAGTATGGGAAGAAGGGATGGAAGTATGGGTTGGCATCAGGCCGCCTATGGATCTT
GATGCCTTTGGTCGTTCCGATAGTGAAGGAGCGGAATCCAAAATGCATGCCATGGATGT
TTTCATGTGGGGTCGCTGTTTAACCGAAGATGAAATTTTAATGGTGCATAATTGTGCTAA
TCTGGAGGAGCAGAATGAGCTGGATATTATGGATGACTATTGGCAGGGATCTCCAATGG
AAACGCCAATGAGAGTAAGTTGGGCATATCACGAAGCACATG

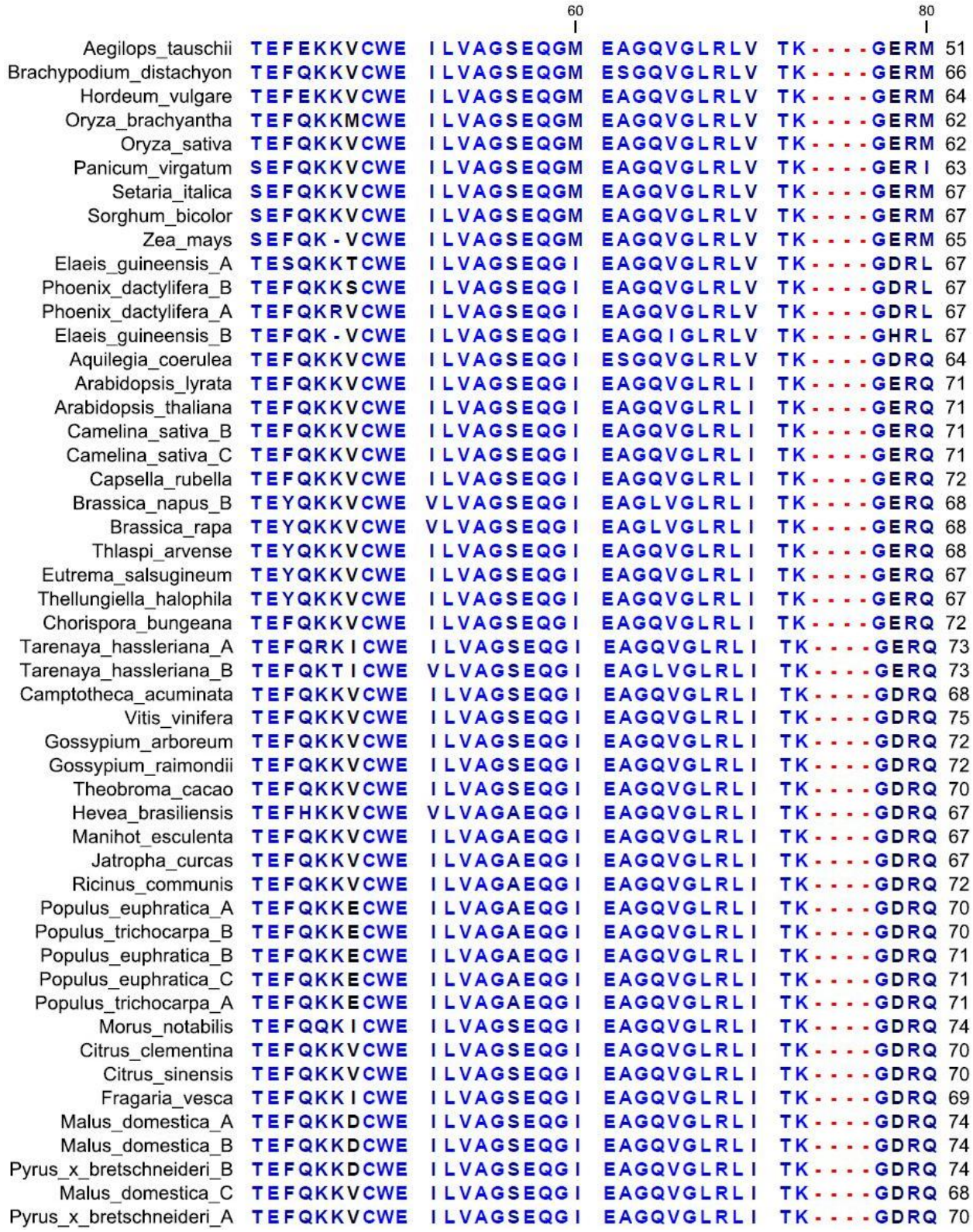
D. 3’ targeting sequence (3’_TGS)

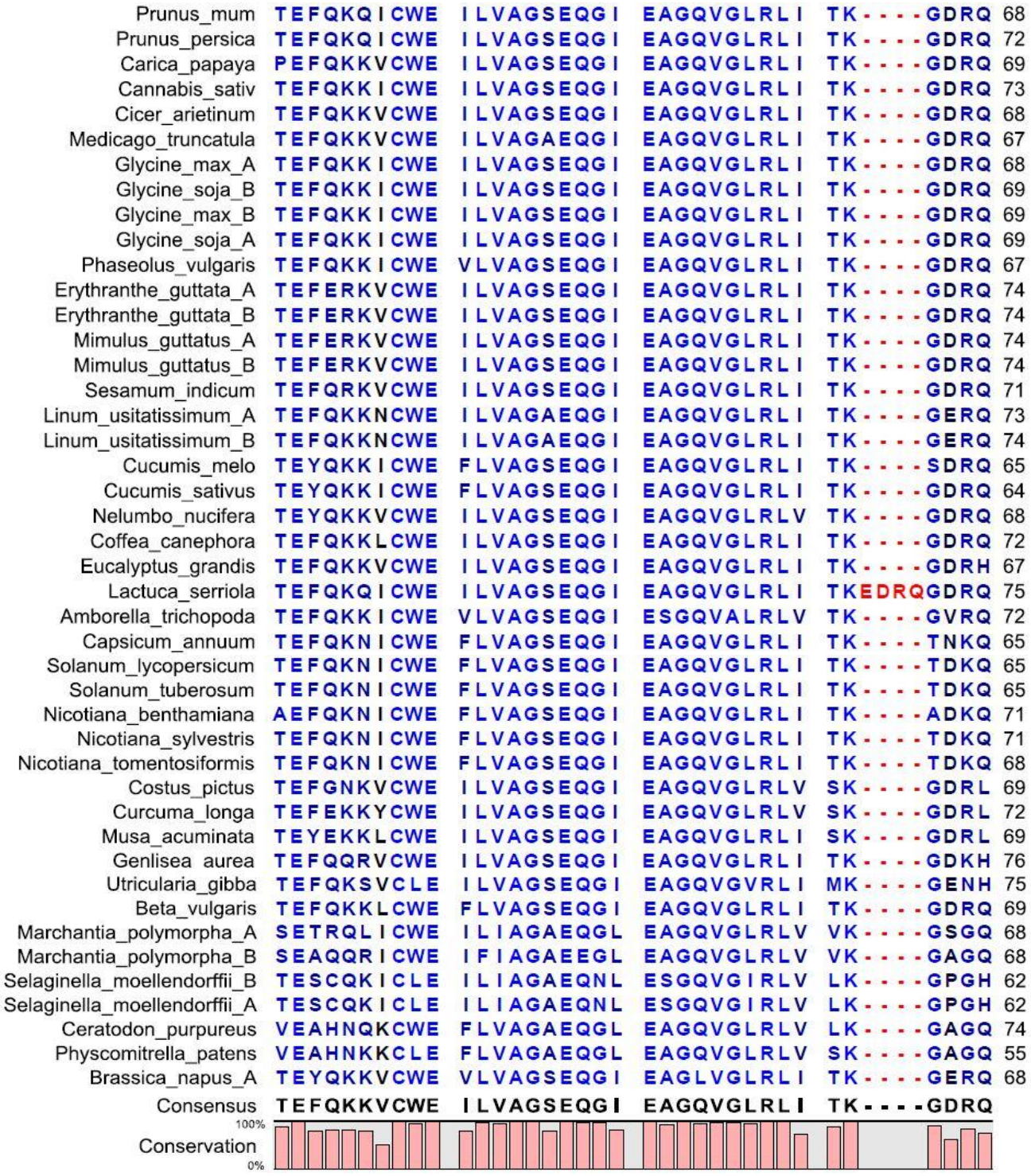
GTTTAATTGAGTCAGTAATTAGATGTGGGGAATTGTAAGCTTGGAGTCTGCATGTTCTTT
AGATTGCATGTCAAAACAACCCTCATGCTTCTAAGTTACTTATGTTTGATGTATTCATCC
GAAGTGGTCTGACCATTTGGTTTTGTTGAAGGTGCTTAAGTTTTTATAACTCATACTGTG
ATTCTCATGTGTGTGTGTTTTGATGTCCTGCCGGTCGTGTATCTATCTTGTTCTCTAGTTT
GAGGATTGGAGATATGATCCGGATCTCAGCTACGAGCGAGATGAAATTTTTTGGGATGA
ACAACAATCCTCAGGAGCAAAGAGGAAGGCAGCAGACATCGAGAATGTGGCGGTAGAT
ATTGAATATTTATCCAGGAAGATCCGACGCCCCAAAATTGAAACGAAAGAGGAGGTCTT
GCAGCGCATGCAGGCTGTTGAATTGGCTATTAAGGAGGCTTTGGTAGCAAGAGGCGAGA
AGAGATTTACCGACCAAGAGTTTCCGCCATCAGATCAGTCGCTTTTTGTTGATACTGATC
ACCCTGCTGCTAAGCTTCAGGTAAGCTAATATCCTTAACTCGTCTTTTAGTGTCCAGGAA
TATTTAGATGTAATACATACAGGCTATTTTGGTGTGGTGACTGCCAATTAAAAGTGCATT
TTTATGTTTGCCAGGTTGTTAACAAGTGGTCGCGCCCAGAAGAGGTTTTGCGCGAACAA
GGGCAAGATTCTCGGCCGTGCCTGTTTGCAGGCACTGCCAACCCTTCAGACGTCTGCCA
GGTTCGAAGTGAACTTTATTTGTCTCCTTCAAGTTTTACTCAACCCGTTCATTGATTCATT
CATGAACACCATTTGAGCGCTTTCGTGTTAAGCTGCATGTAGATAGAATACTTAAACGA
GCACCATTTTCTGAGTCCTTCACCAGGGCTACATACGCGGTCATTGAAGCAACTTGGAA
ACATTTTCGGTCATATCAAGCTAACAGCTTTTTAATAGTAATAGTAGTAACACTTTCGCC
AAATTCTTAGGGGCATAGCACGTTATAGGGTAAGAATTATGTCTAATCGCAGACACAGA
GAATCTTTACCCACATCTCATGGAAACAAAGCATATTAAAAATTGCAGTAAAGGAAGGT
TGAGCACCTTCGAATCAGCTCCTGTTTATCTTCGTTTACGTATCAGT

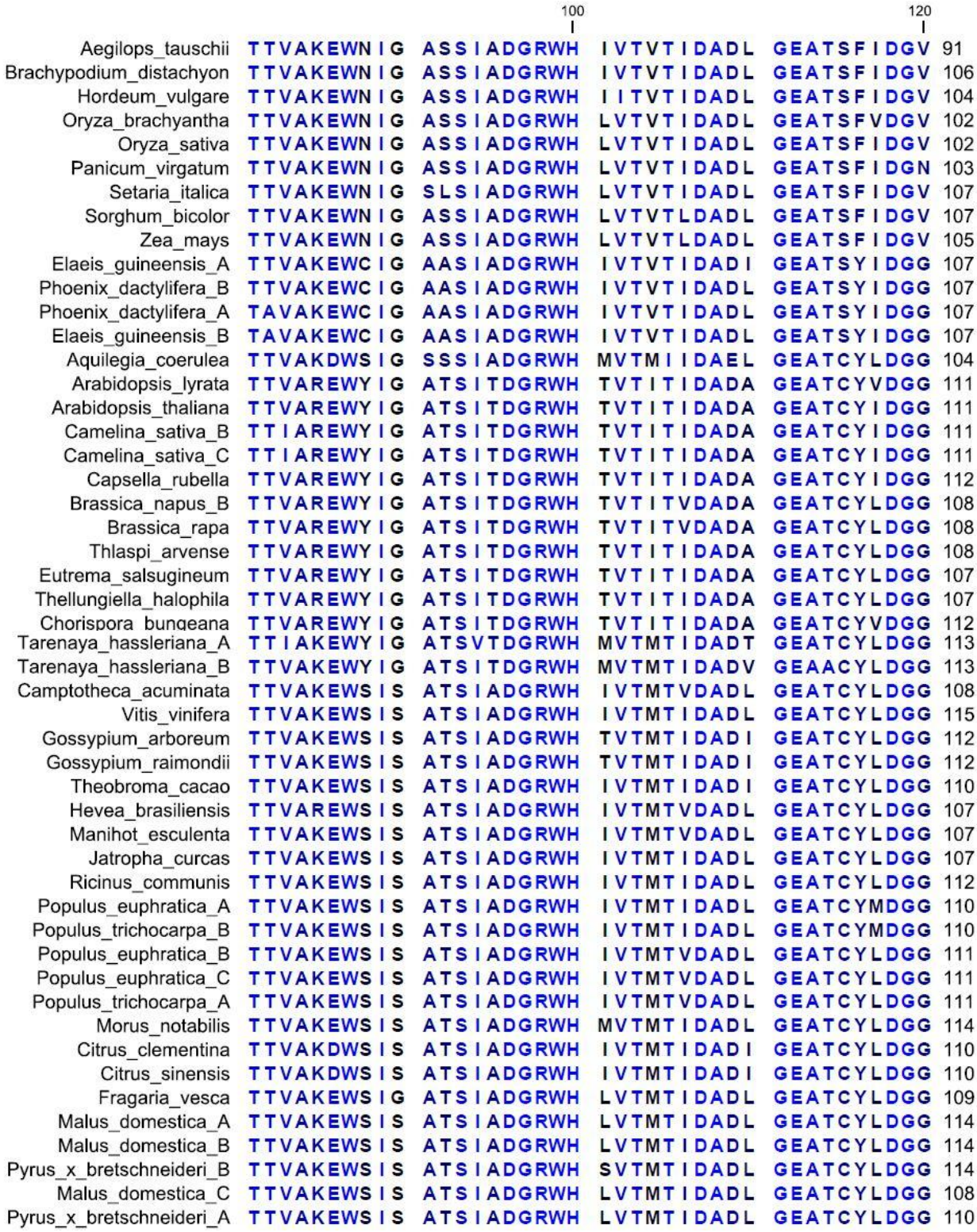
Appendix 5: Multiple sequence alignment of land plants DEK1-LG3 sequences

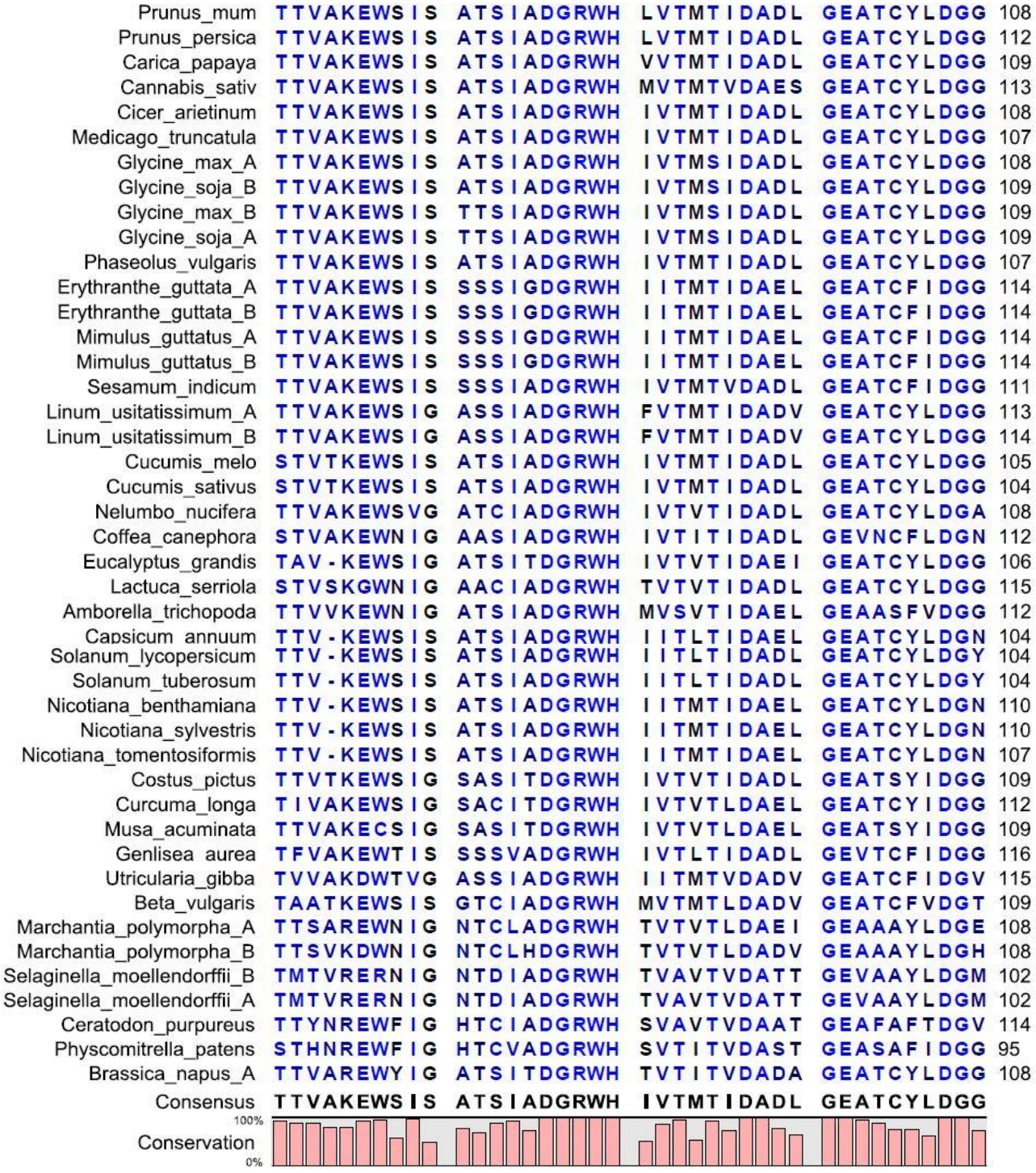


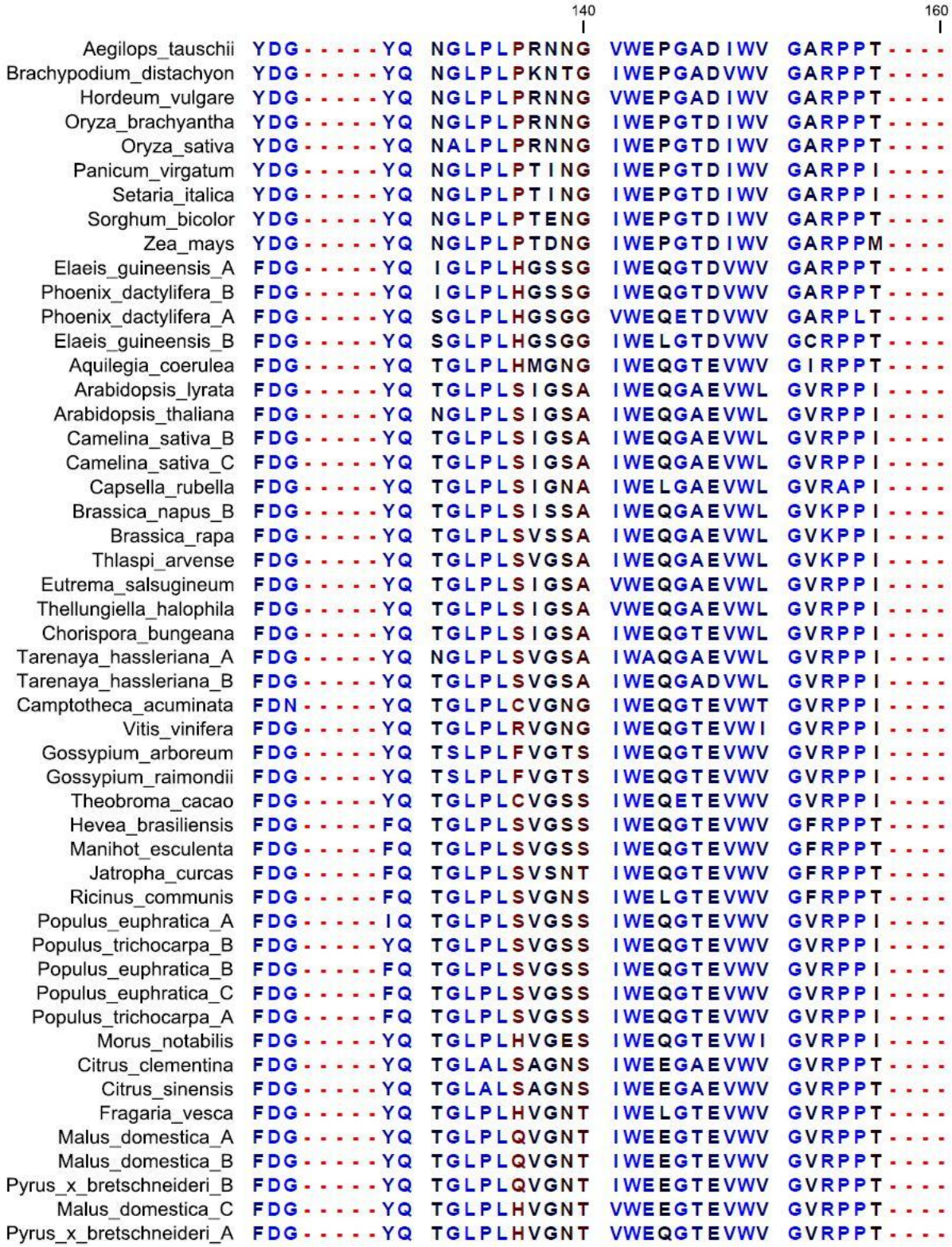

Prunus_mum --------GQ IDPTICQSRK ISFSVAVMIQ PVSGPVCLFG 32
Prunus_persica ----GRHCGQ IDPTICQSQK ISFSVAVMIQ PVSGPVCLFG 36
Carica_papaya -------CGQ IDQSVCQSKK VSFSIAVMVQ PESGPVCLIG 33
Cannabis_sativ ---SGRHYGP IDPSICQTKK VSFSIAVMIQ PESGPVCLLG 37
Cicer_arietinum --------GP IDSSLCQSQK ISFSIALMIQ PESGPVCLLG 32
Medicago_truncatula ---------P IDSSLCLSQK VSFSIALMIQ PESGPVCLLG 31
Glycine_max_A --------GP IDSSLCQSQK VSFSIALMIQ PESGPVCLLG 32
Glycine_soja_B -------CGP IDSSLCQSQK VSFSIALMIQ PESGPVCLLG 33
Glycine_max_B -------CGP IDSSLCQSQK VSFSIALMIQ PESGPVCLLG 33
Glycine_soja_A -------CGP IDSSLCQSQK VSFSIALMIQ PESGPVCLLG 33
Phaseolus_vulgaris ---------P IDSSLCRSQK VSFSIALMIQ PESGPICLLG 31
Erythranthe_guttata_A --TSGRHCGQ IDPSLCQSQK VSFSIAAMIQ PESGPVCLLG 38
Erythranthe_guttata_B --TSGRHCGQ IDPSLCQSQK VSFSIAAMIQ PESGPVCLLG 38
Mimulus_guttatus_A --TSGRHCGQ IDPSLCQSQK VSFSIAAMIQ PESGPVCLLG 38
Mimulus_guttatus_B --TSGRHCGQ IDPSLCQSQK VSFSIAAMIQ PESGPVCLLG 38
Sesamum_indicum -----RHCGQ IDPTLCQSQK VSFSIAVMIQ PESGPVCLLG 35
Linum_usitatissimum_A ---SGRHCGE VDPSVCQSRK ITFSVVVMIQ PESGPVCLLG 37
Linum_usitatissimum_B --TSGRHCGE VDPSVCQSRK ITFSVVVMIQ PESGPVCLLG 38
Cucumis_melo ---------- -EASLCQSRK ISVSIAALIQ PESGPVCLFG 29
Cucumis_sativus ---------- --ASLCRSRK ISVSIAALIQ PESGPVCLFG 28
Nelumbo_nucifera --------GK IDSSICLSKK VSFSIAVMIQ PESGPVCLLG 32
Coffea_canephora ----GRHCGQ IDLTLCQSQK VSFSVTVMIQ PESGPICLLG 36
Eucalyptus_grandis ---------P IDGSVCQSRK VSFSVAAMVQ PESGPVCLLG 31
Lactuca_serriola -----RYCGQ LDPTICQTQK VSFSMAVMIQ PESGPVCLLG 35
Amborella_trichopoda ----GRHCGA VDPAVCQSQK VTFSIAVMIQ PESGPVCLLG 36
Capsicum_annuum ---------- -DPTVCQCQK ISCSLAVMVQ PESGPVCLFG 29
Solanum_lycopersicum ---------- -DPSVCQCQK ISCSLAVMVQ PESGPVCLFG 29
Solanum_tuberosum ---------- -DPSVCQCQK ISCSLAVMVQ PESGPVCLFG 29
Nicotiana_benthamiana -----RQCGQ IDPSVCQSQK VSCSLAVMVQ PESGPLCLFG 35
Nicotiana_sylvestris -----RQCGQ IDPSVCQSQK VSCSLAVMVQ PESGPLCLFG 35
Nicotiana_tomentosiformis --------GQ IDPSVCQSQK VSCSLAVMVQ PESGPLCLFG 32
Costus_pictus -------CGQ IHPSLSNSSK VTYSIAVMIQ PESGPVCLLG 33
Curcuma_longa ----GRHCGM IHPSLHNSQK VTFSITVMIQ PESGPVCLLG 36
Musa_acuminata -------CGQ IHPSLCHSQR VTFSIAVMIQ PESGPVCLIG 33
Genlisea_aurea PTTSGRNFGE IDPSLCHSQK VTFSTAVMIQ PDSGPVCLLG 40
Utricularia_gibba -TTSGRHCGD IDYSLCHSQK VTFSTTVMIQ PDSGPVCLLG 39
Beta_vulgaris -------CGR VEPSVCHSQK VTFSVVVMIQ PDSGPVCLLG 33
Marchantia_polymorpha_A --------GV LDSSVCGSRK VTFSAAVLVQ PESGPICLLG 32
Marchantia_polymorpha_B --------GV LDPSLCGSRK VTFSAAVLVQ PESGPICLMG 32
Selaginella_moellendorffii_B ---------- ----VCGFKS ITCSAAIMVQ PQSGPVCILG 26
Selaginella_moellendorffii_A ---------- ----VCGFKS ITCSAAIMVQ PQSGPVCILG 26
Ceratodon_purpureus --TGGRSCGP IDPTVCGSSK VSFSAAVMIQ PDSGPVCILG 38
Physcomitrella_patens ---------- ---------- -SFSAAVMIQ PDSGPVCILG 19
Brassica_napus_A --------GQ MDASVCQSQK ISFSITAMIQ PDSGPVCLFG 32
Consensus -------CGQ IDPSVCQSQK VSFSIAVMIQ PESGPVCLLG
100%
Conservation
0%

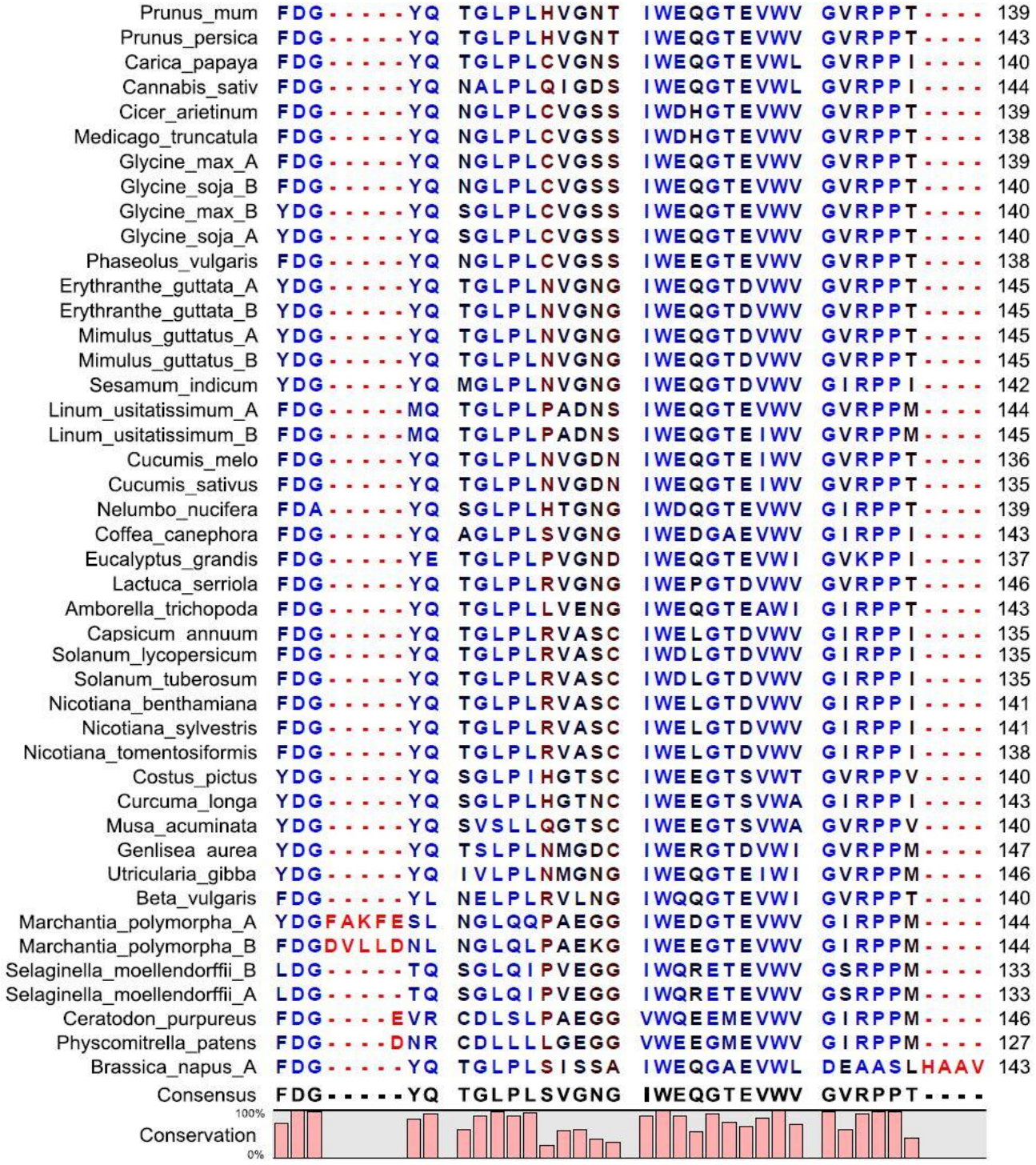

60
80
Aegilops_tauschii TEFEKKVCWE ILVAGSEQGM EAGQVGLRLV TK----GERM 51
Brachypodium_distachyon TEFQKKVCWE ILVAGSEQGM ESGQVGLRLV TK----GERM 66
Hordeum_vulgare TEFEKKVCWE ILVAGSEQGM EAGQVGLRLV TK----GERM 64
Oryza_brachyantha TEFQKKMCWE ILVAGSEQGM EAGQVGLRLV TK----GERM 62
Oryza_sativa TEFQKKVCWE ILVAGSEQGM EAGQVGLRLV TK----GERM 62
Panicum_virgatum SEFQKKVCWE ILVAGSEQGM EAGQVGLRLV TK----GERI 63
Setaria_italica SEFQKKVCWE ILVAGSEQGM EAGQVGLRLV TK----GERM 67
Sorghum_bicolor SEFQKKVCWE ILVAGSEQGM EAGQVGLRLV TK----GERM 67
Zea_mays SEFQK-VCWE ILVAGSEQGM EAGQVGLRLV TK----GERM 65
Elaeis_guineensis_A TESQKKTCWE ILVAGSEQGI EAGQVGLRLV TK----GDRL 67
Phoenix_dactylifera_B TEFQKKSCWE ILVAGSEQGI EAGQVGLRLV TK----GDRL 67
Phoenix_dactylifera_A TEFQKRVCWE ILVAGSEQGI EAGQVGLRLV TK----GDRL 67
Elaeis_guineensis_B TEFQK-VCWE ILVAGSEQGI EAGQIGLRLV TK----GHRL 67
Aquilegia_coerulea TEFQKKVCWE ILVAGSEQGI ESGQVGLRLV TK----GDRQ 64
Arabidopsis_lyrata TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GERQ 71
Arabidopsis_thaliana TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GERQ 71
Camelina_sativa_B TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GERQ 71
Camelina_sativa_C TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GERQ 71
Capsella_rubella TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GERQ 72
Brassica_napus_B TEYQKKVCWE VLVAGSEQGI EAGLVGLRLI TK----GERQ 68
Brassica_rapa TEYQKKVCWE VLVAGSEQGI EAGLVGLRLI TK----GERQ 68
Thlaspi_arvense TEYQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GERQ 68
Eutrema_salsugineum TEYQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GERQ 67
Thellungiella_halophila TEYQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GERQ 67
Chorispora_bungeana TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GERQ 72
Tarenaya_hassleriana_A TEFQRKICWE ILVAGSEQGI EAGQVGLRLI TK----GERQ 73
Tarenaya_hassleriana_B TEFQKTICWE VLVAGSEQGI EAGLVGLRLI TK----GERQ 73
Camptotheca_acuminata TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 68
Vitis_vinifera TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 75
Gossypium_arboreum TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 72
Gossypium_raimondii TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 72
Theobroma_cacao TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 70
Hevea_brasiliensis TEFHKKVCWE VLVAGAEQGI EAGQVGLRLI TK----GDRQ 67
Manihot_esculenta TEFQKKVCWE ILVAGAEQGI EAGQVGLRLI TK----GDRQ 67
Jatropha_curcas TEFQKKVCWE ILVAGAEQGI EAGQVGLRLI TK----GDRQ 67
Ricinus_communis TEFQKKVCWE ILVAGAEQGI EAGQVGLRLI TK----GDRQ 72
Populus_euphratica_A TEFQKKECWE ILVAGAEQGI EAGQVGLRLI TK----GDRQ 70
Populus_trichocarpa_B TEFQKKECWE ILVAGAEQGI EAGQVGLRLI TK----GDRQ 70
Populus_euphratica_B TEFQKKECWE ILVAGAEQGI EAGQVGLRLI TK----GDRQ 71
Populus_euphratica_C TEFQKKECWE ILVAGAEQGI EAGQVGLRLI TK----GDRQ 71
Populus_trichocarpa_A TEFQKKECWE ILVAGAEQGI EAGQVGLRLI TK----GDRQ 71
Morus_notabilis TEFQQKICWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 74
Citrus_clementina TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 70
Citrus_sinensis TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 70
Fragaria_vesca TEFQKKICWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 69
Malus_domestica_A TEFQKKDCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 74
Malus_domestica_B TEFQKKDCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 74
Pyrus_x_bretschneideri_B TEFQKKDCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 74
Malus_domestica_C TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 68
Pyrus_x_bretschneideri_A TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 70

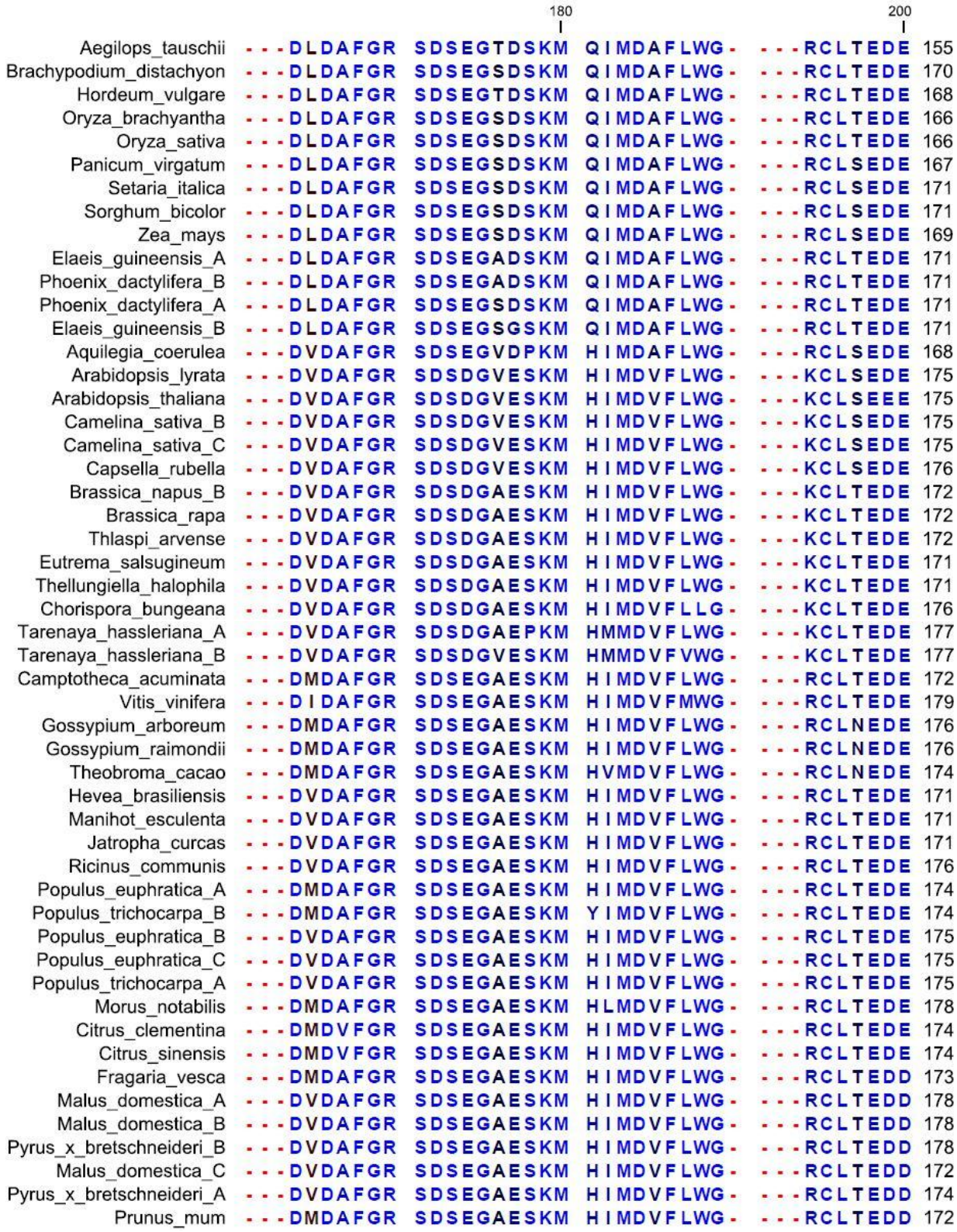

Prunus_mum TEFQKQICWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 68
Prunus_persica TEFQKQICWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 72
Carica_papaya PEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 69
Cannabis_sativ TEFQKKICWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 73
Cicer_arietinum TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 68
Medicago_truncatula TEFQKKVCWE ILVAGAEQGI EAGQVGLRLI TK----GDRQ 67
Glycine_max_A TEFQKKICWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 68
Glycine_soja_B TEFQKKICWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 69
Glycine_max_B TEFQKKICWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 69
Glycine_soja_A TEFQKKICWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 69
Phaseolus_vulgaris TEFQKKICWE VLVAGSEQGI EAGQVGLRLI TK----GDRQ 67
Erythranthe_guttata_A TEFERKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 74
Erythranthe_guttata_B TEFERKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 74
Mimulus_guttatus_A TEFERKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 74
Mimulus_guttatus_B TEFERKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 74
Sesamum_indicum TEFQRKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 71
Linum_usitatissimum_A TEFQKKNCWE ILVAGAEQGI EAGQVGLRLI TK----GERQ 73
Linum_usitatissimum_B TEFQKKNCWE ILVAGAEQGI EAGQVGLRLI TK----GERQ 74
Cucumis_melo TEYQKKICWE FLVAGSEQGI EAGQVGLRLI TK----SDRQ 65
Cucumis_sativus TEYQKKICWE FLVAGSEQGI EAGQVGLRLI TK----GDRQ 64
Nelumbo_nucifera TEYQKKVCWE ILVAGSEQGI EAGQVGLRLV TK----GDRQ 68
Coffea_canephora TEFQKKLCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ 72
Eucalyptus_grandis TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRH 67
Lactuca_serriola TEFQKQICWE ILVAGSEQGI EAGQVGLRLI TKEDRQGDRQ 75
Amborella_trichopoda TEFQKKICWE VLVAGSEQGI ESGQVALRLV TK----GVRQ 72
Capsicum_annuum TEFQKNICWE FLVAGSEQGI EAGQVGLRLI TK----TNKQ 65
Solanum_lycopersicum TEFQKNICWE FLVAGSEQGI EAGQVGLRLI TK----TDKQ 65
Solanum_tuberosum TEFQKNICWE FLVAGSEQGI EAGQVGLRLI TK----TDKQ 65
Nicotiana_benthamiana AEFQKNICWE FLVAGSEQGI EAGQVGLRLI TK----ADKQ 71
Nicotiana_sylvestris TEFQKNICWE FLVAGSEQGI EAGQVGLRLI TK----TDKQ 71
Nicotiana_tomentosiformis TEFQKNICWE FLVAGSEQGI EAGQVGLRLI TK----TDKQ 68
Costus_pictus TEFGNKVCWE ILVAGSEQGI EAGQVGLRLV SK----GDRL 69
Curcuma_longa TEFEKKYCWE ILVAGSEQGI EAGQVGLRLV SK----GDRL 72
Musa_acuminata TEYEKKLCWE ILVAGSEQGI EAGQVGLRLI SK----GDRL 69
Genlisea aurea TEFQQRVCWE ILVAGSEQGI EAGQVGLRLI TK----GDKH 76
Utricularia_gibba TEFQKSVCLE ILVAGSEQGI EAGQVGVRLI MK----GENH 75
Beta_vulgaris TEFQKKLCWE FLVAGSEQGI EAGQVGLRLI TK----GDRQ 69
Marchantia_polymorpha_A SETRQLICWE ILIAGAEQGL EAGQVGLRLV VK----GSGQ 68
Marchantia_polymorpha_B SEAQQRICWE IFIAGAEEGL EAGQVGLRLV VK----GAGQ 68
Selaginella_moellendorffii_B TESCQKICLE ILIAGAEQNL ESGQVGIRLV LK----GPGH 62
Selaginella_moellendorffii_A TESCQKICLE ILIAGAEQNL ESGQVGIRLV LK----GPGH 62
Ceratodon_purpureus VEAHNQKCWE FLVAGAEQGL EAGQVGLRLV LK----GAGQ 74
Physcomitrella_patens VEAHNKKCLE FLVAGAEQGL EAGQVGLRLV SK----GAGQ 55
Brassica_napus_A TEYQKKVCWE VLVAGSEQGI EAGLVGLRLI TK----GERQ 68
Consensus TEFQKKVCWE ILVAGSEQGI EAGQVGLRLI TK----GDRQ
100%
Conservation
0%

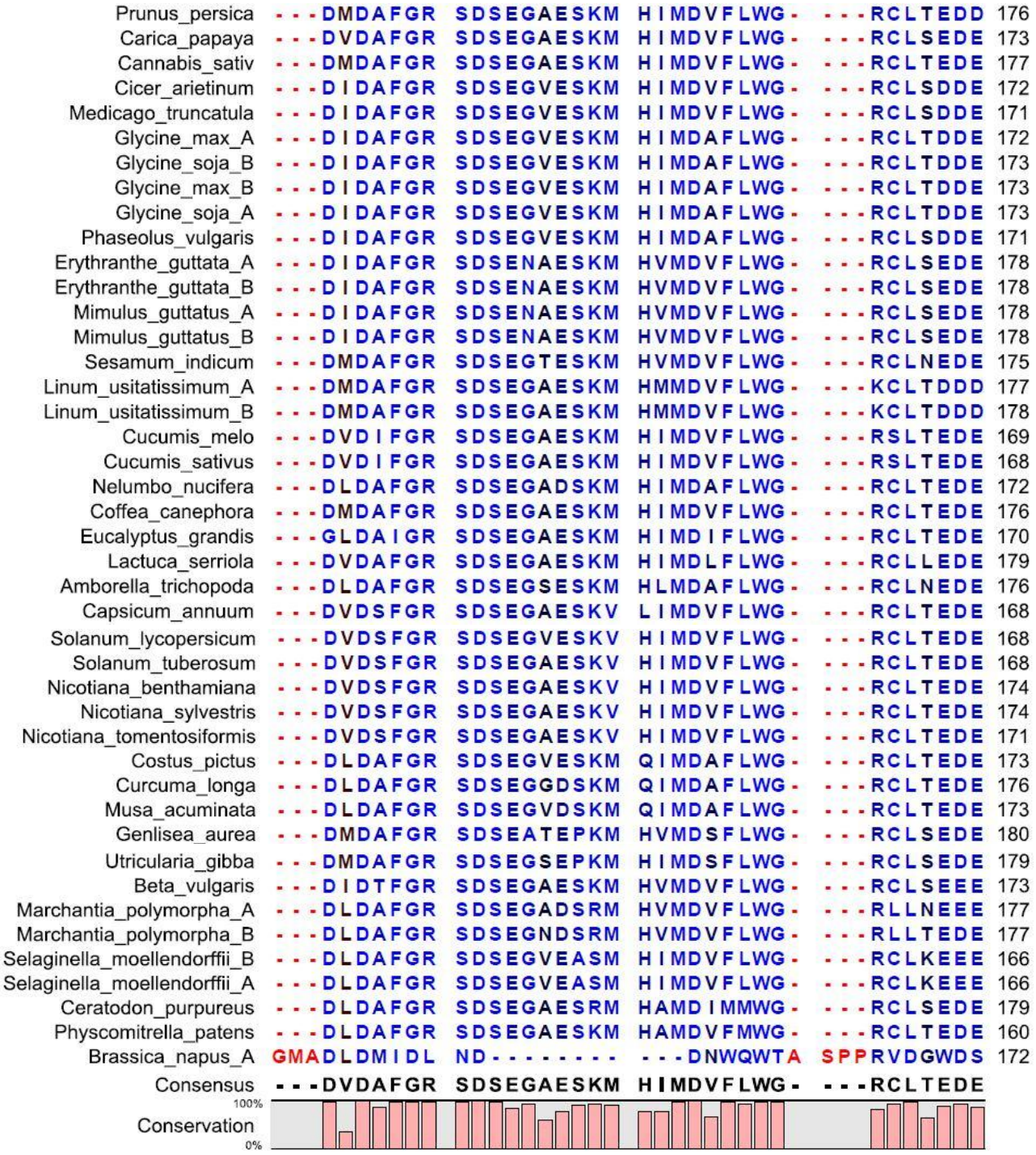

100
120
Aegilops_tauschii TTVAKEWNIG ASSIADGRWH IVTVTIDADL GEATSFIDGV 91
Brachypodium_distachyon TTVAKEWNIG ASSIADGRWH IVTVTIDADL GEATSFIDGV 106
Hordeum_vulgare TTVAKEWNIG ASSIADGRWH IITVTIDADL GEATSFIDGV 104
Oryza_brachyantha TTVAKEWNIG ASSIADGRWH LVTVTIDADL GEATSFVDGV 102
Oryza_sativa TTVAKEWNIG ASSIADGRWH LVTVTIDADL GEATSFIDGV 102
Panicum_virgatum TTVAKEWNIG ASSIADGRWH LVTVTIDADL GEATSFIDGN 103
Setaria_italica TTVAKEWNIG SLSIADGRWH LVTVTIDADL GEATSFIDGV 107
Sorghum_bicolor TTVAKEWNIG ASSIADGRWH LVTVTLDADL GEATSFIDGV 107
Zea_mays TTVAKEWNIG ASSIADGRWH LVTVTLDADL GEATSFIDGV 105
Elaeis_guineensis_A TTVAKEWCIG AASIADGRWH IVTVTIDADI GEATSYIDGG 107
Phoenix_dactylifera_B TTVAKEWCIG AASIADGRWH IVTVTIDADL GEATSYIDGG 107
Phoenix_dactylifera_A TAVAKEWCIG AASIADGRWH IVTVTIDADL GEATSYIDGG 107
Elaeis_guineensis_B TAVAKEWCIG AASIADGRWH IVTVTIDADL GEATSYIDGG 107
Aquilegia_coerulea TTVAKDWSIG SSSIADGRWH MVTMIIDAEL GEATCYLDGG 104
Arabidopsis_lyrata TTVAREWYIG ATSITDGRWH TVTITIDADA GEATCYVDGG 111
Arabidopsis_thaliana TTVAREWYIG ATSITDGRWH TVTITIDADA GEATCYIDGG 111
Camelina_sativa_B TTIAREWYIG ATSITDGRWH TVTITIDADA GEATCYIDGG 111
Camelina_sativa_C TTIAREWYIG ATSITDGRWH TVTITIDADA GEATCYIDGG 111
Capsella_rubella TTVAREWYIG ATSITDGRWH TVTITIDADA GEATCYIDGG 112
Brassica_napus_B TTVAREWYIG ATSITDGRWH TVTITVDADA GEATCYLDGG 108
Brassica_rapa TTVAREWYIG ATSITDGRWH TVTITVDADA GEATCYLDGG 108
Thlaspi_arvense TTVAREWYIG ATSITDGRWH TVTITIDADA GEATCYLDGG 108
Eutrema_salsugineum TTVAREWYIG ATSITDGRWH TVTITIDADA GEATCYLDGG 107
Thellungiella_halophila TTVAREWYIG ATSITDGRWH TVTITIDADA GEATCYLDGG 107
Chorispora_bungeana TTVAREWYIG ATSITDGRWH TVTITIDADA GEATCYVDGG 112
Tarenaya_hassleriana_A TTIAKEWYIG ATSVTDGRWH MVTMTIDADT GEATCYLDGG 113
Tarenaya_hassleriana_B TTVAKEWYIG ATSITDGRWH MVTMTIDADV GEAACYLDGG 113
Camptotheca_acuminata TTVAKEWSIS ATSIADGRWH IVTMTVDADL GEATCYLDGG 108
Vitis_vinifera TTVAKEWSIS ATSIADGRWH IVTMTIDADL GEATCYLDGG 115
Gossypium_arboreum TTVAKEWSIS ATSIADGRWH TVTMTIDADI GEATCYLDGG 112
Gossypium_raimondii TTVAKEWSIS ATSIADGRWH TVTMTIDADI GEATCYLDGG 112
Theobroma_cacao TTVAKEWSIS ATSIADGRWH IVTMTIDADI GEATCYLDGG 110
Hevea_brasiliensis TTVAREWSIS ATSIADGRWH IVTMTVDADL GEATCYLDGG 107
Manihot_esculenta TTVAKEWSIS ATSIADGRWH IVTMTVDADL GEATCYLDGG 107
Jatropha_curcas TTVAKEWSIS ATSIADGRWH IVTMTIDADL GEATCYLDGG 107
Ricinus_communis TTVAKEWSIS ATSIADGRWH IVTMTIDADL GEATCYLDGG 112
Populus_euphratica_A TTVAKEWSIS ATSIADGRWH IVTMTIDADL GEATCYMDGG 110
Populus_trichocarpa_B TTVAKEWSIS ATSIADGRWH IVTMTIDADL GEATCYMDGG 110
Populus_euphratica_B TTVAKEWSIS ATSIADGRWH IVTMTVDADL GEATCYLDGG 111
Populus_euphratica_C TTVAKEWSIS ATSIADGRWH IVTMTVDADL GEATCYLDGG 111
Populus_trichocarpa_A TTVAKEWSIS ATSIADGRWH IVTMTVDADL GEATCYLDGG 111
Morus_notabilis TTVAKEWSIS ATSIADGRWH MVTMTIDADL GEATCYLDGG 114
Citrus_clementina TTVAKDWSIS ATSIADGRWH IVTMTIDADI GEATCYLDGG 110
Citrus_sinensis TTVAKDWSIS ATSIADGRWH IVTMTIDADI GEATCYLDGG 110
Fragaria_vesca TTVAKEWSIG ATSIADGRWH LVTMTIDADL GEATCYLDGG 109
Malus_domestica_A TTVAKEWSIS ATSIADGRWH LVTMTIDADL GEATCYLDGG 114
Malus_domestica_B TTVAKEWSIS ATSIADGRWH LVTMTIDADL GEATCYLDGG 114
Pyrus_x_bretschneideri_B TTVAKEWSIS ATSIADGRWH SVTMTIDADL GEATCYLDGG 114
Malus_domestica_C TTVAKEWSIS ATSIADGRWH LVTMTIDADL GEATCYLDGG 108
Pyrus_x_bretschneideri_A TTVAKEWSIS ATSIADGRWH LVTMTIDADL GEATCYLDGG 110


Prunus_mum TTVAKEWSIS ATSIADGRWH LVTMTIDADL GEATCYLDGG 108
Prunus_persica TTVAKEWSIS ATSIADGRWH LVTMTIDADL GEATCYLDGG 112
Carica_papaya TTVAKEWSIS ATSIADGRWH VVTMTIDADL GEATCYLDGG 109
Cannabis_sativ TTVAKEWSIS ATSIADGRWH MVTMTVDAES GEATCYLDGG 113
Cicer_arietinum TTVAKEWSIS ATSIADGRWH IVTMTIDADL GEATCYLDGG 108
Medicago_truncatula TTVAKEWSIS ATSIADGRWH IVTMTIDADL GEATCYLDGG 107
Glycine_max_A TTVAKEWSIS ATSIADGRWH IVTMSIDADL GEATCYLDGG 108
Glycine_soja_B TTVAKEWSIS ATSIADGRWH IVTMSIDADL GEATCYLDGG 109
Glycine_max_B TTVAKEWSIS TTSIADGRWH IVTMSIDADL GEATCYLDGG 109
Glycine_soja_A TTVAKEWSIS TTSIADGRWH IVTMSIDADL GEATCYLDGG 109
Phaseolus_vulgaris TTVAKEWSIS ATSIADGRWH IVTMTIDADL GEATCYLDGG 107
Erythranthe_guttata_A TTVAKEWSIS SSSIGDGRWH IITMTIDAEL GEATCFIDGG 114
Erythranthe_guttata_B TTVAKEWSIS SSSIGDGRWH IITMTIDAEL GEATCFIDGG 114
Mimulus_guttatus_A TTVAKEWSIS SSSIGDGRWH IITMTIDAEL GEATCFIDGG 114
Mimulus_guttatus_B TTVAKEWSIS SSSIGDGRWH IITMTIDAEL GEATCFIDGG 114
Sesamum_indicum TTVAKEWSIS SSSIADGRWH IVTMTVDADL GEATCFIDGG 111
Linum_usitatissimum_A TTVAKEWSIG ASSIADGRWH FVTMTIDADV GEATCYLDGG 113
Linum_usitatissimum_B TTVAKEWSIG ASSIADGRWH FVTMTIDADV GEATCYLDGG 114
Cucumis_melo STVTKEWSIS ATSIADGRWH IVTMTIDADL GEATCYLDGG 105
Cucumis_sativus STVTKEWSIS ATSIADGRWH IVTMTIDADL GEATCYLDGG 104
Nelumbo_nucifera TTVAKEWSVG ATCIADGRWH IVTVTIDADL GEATCYLDGA 108
Coffea_canephora STVAKEWNIG AASIADGRWH IVTITIDADL GEVNCFLDGN 112
Eucalyptus_grandis TAV-KEWSIG ATSITDGRWH IVTVTIDAEI GEATCYLDGG 106
Lactuca_serriola STVSKGWNIG AACIADGRWH TVTVTIDADL GEATCYLDGG 115
Amborella_trichopoda TTVVKEWNIG ATSIADGRWH MVSVTIDAEL GEAASFVDGG 112
Capsicum_annuum TTV-KEWSIS ATSIADGRWH IITLTIDAEL GEATCYLDGN 104
Solanum_lycopersicum TTV-KEWSIS ATSIADGRWH IITLTIDADL GEATCYLDGY 104
Solanum_tuberosum TTV-KEWSIS ATSIADGRWH IITLTIDADL GEATCYLDGY 104
Nicotiana_benthamiana TTV-KEWSIS ATSIADGRWH IITMTIDAEL GEATCYLDGN 110
Nicotiana_sylvestris TTV-KEWSIS ATSIADGRWH IITMTIDAEL GEATCYLDGN 110
Nicotiana_tomentosiformis TTV-KEWSIS ATSIADGRWH IITMTIDAEL GEATCYLDGN 107
Costus_pictus TTVTKEWSIG SASITDGRWH IVTVTIDADL GEATSYIDGG 109
Curcuma_longa TIVAKEWSIG SACITDGRWH IVTVTLDAEL GEATCYIDGG 112
Musa_acuminata TTVAKECSIG SASITDGRWH IVTVTLDAEL GEATSYIDGG 109
Genlisea_aurea TFVAKEWTIS SSSVADGRWH IVTLTIDADL GEVTCFIDGG 116
Utricularia_gibba TVVAKDWTVG ASSIADGRWH IITMTVDADV GEATCFIDGV 115
Beta_vulgaris TAATKEWSIS GTCIADGRWH MVTMTLDADV GEATCFVDGT 109
Marchantia_polymorpha_A TTSAREWNIG NTCLADGRWH TVTVTLDAEI GEAAAYLDGE 108
Marchantia_polymorpha_B TTSVKDWNIG NTCLHDGRWH TVTVTLDADV GEAAAYLDGH 108
Selaginella_moellendorffii_B TMTVRERNIG NTDIADGRWH TVAVTVDATT GEVAAYLDGM 102
Selaginella_moellendorffii_A TMTVRERNIG NTDIADGRWH TVAVTVDATT GEVAAYLDGM 102
Ceratodon_purpureus TTYNREWFIG HTCIADGRWH SVAVTVDAAT GEAFAFTDGV 114
Physcomitrella_patens STHNREWFIG HTCVADGRWH SVTITVDAST GEASAFIDGG 95
Brassica_napus_A TTVAREWYIG ATSITDGRWH TVTITVDADA GEATCYLDGG 108
Consensus TTVAKEWSIS ATSIADGRWH IVTMTIDADL GEATCYLDGG
100%
Conservation
0%

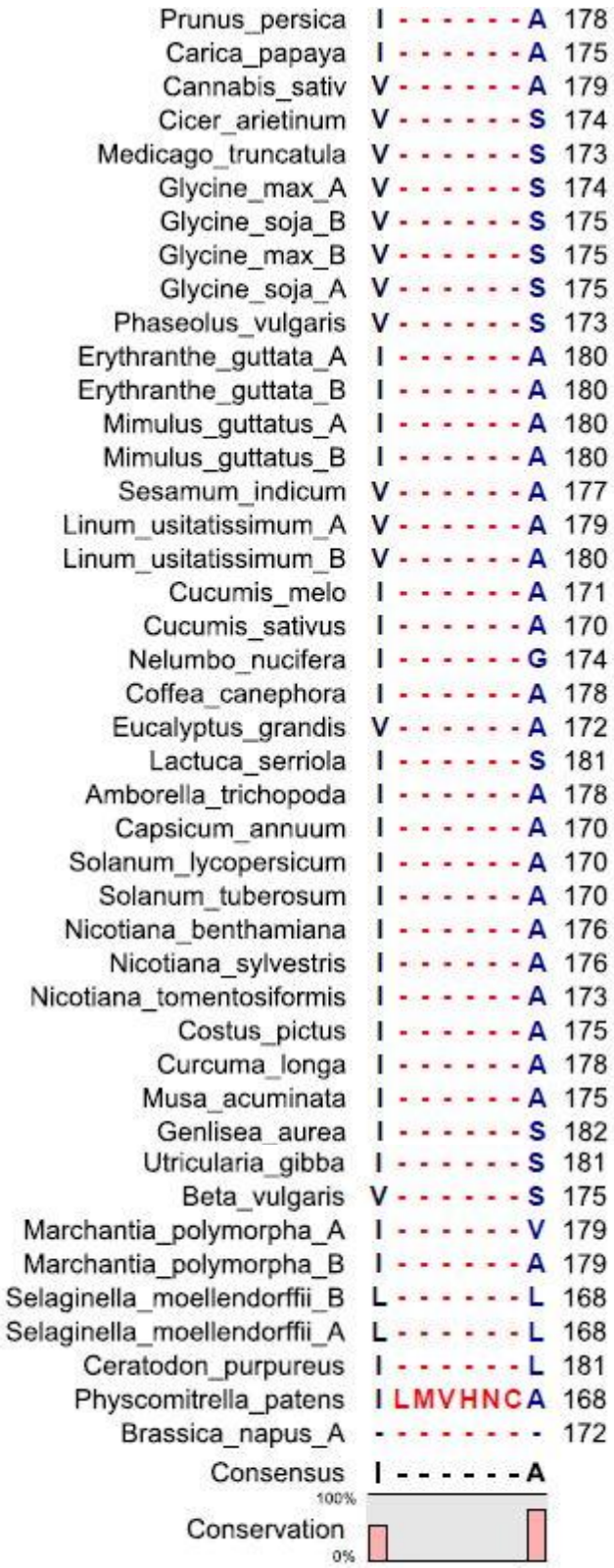

140
160
Aegilops_tauschii YDG-----YQ NGLPLPRNNG VWEPGADIWV GARPPT---- 122
Brachypodium_distachyon YDG-----YQ NGLPLPKNTG IWEPGADVWV GARPPT---- 137
Hordeum_vulgare YDG-----YQ NGLPLPRNNG VWEPGADIWV GARPPT---- 135
Oryza_brachyantha YDG-----YQ NGLPLPRNNG IWEPGTDIWV GARPPT---- 133
Oryza_sativa YDG-----YQ NALPLPRNNG IWEPGTDIWV GARPPT---- 133
Panicum_virgatum YDG-----YQ NGLPLPTING IWEPGTDIWV GARPPI---- 134
Setaria_italica YDG-----YQ NGLPLPTING IWEPGTDIWV GARPPI---- 138
Sorghum_bicolor YDG-----YQ NGLPLPTENG IWEPGTDIWV GARPPT---- 138
Zea_mays YDG-----YQ NGLPLPTDNG IWEPGTDIWV GARPPM---- 136
Elaeis_guineensis_A FDG-----YQ IGLPLHGSSG IWEQGTDVWV GARPPT---- 138
Phoenix_dactylifera_B FDG-----YQ IGLPLHGSSG IWEQGTDVWV GARPPT---- 138
Phoenix_dactylifera_A FDG-----YQ SGLPLHGSGG VWEQETDVWV GARPLT---- 138
Elaeis_guineensis_B FDG-----YQ SGLPLHGSGG IWELGTDVWV GCRPPT---- 138
Aquilegia_coerulea FDG-----YQ TGLPLHMGNG IWEQGTEVWV GIRPPT---- 135
Arabidopsis_lyrata FDG-----YQ TGLPLSIGSA IWEQGAEVWL GVRPPI---- 142
Arabidopsis_thaliana FDG-----YQ NGLPLSIGSA IWEQGAEVWL GVRPPI---- 142
Camelina_sativa_B FDG-----YQ TGLPLSIGSA IWEQGAEVWL GVRPPI---- 142
Camelina_sativa_C FDG-----YQ TGLPLSIGSA IWEQGAEVWL GVRPPI---- 142
Capsella_rubella FDG-----YQ TGLPLSIGNA IWELGAEVWL GVRAPI---- 143
Brassica_napus_B FDG-----YQ TGLPLSISSA IWEQGAEVWL GVKPPI---- 139
Brassica_rapa FDG-----YQ TGLPLSVSSA IWEQGAEVWL GVKPPI---- 139
Thlaspi_arvense FDG-----YQ TGLPLSVGSA IWEQGAEVWL GVKPPI---- 139
Eutrema_salsugineum FDG-----YQ TGLPLSIGSA VWEQGAEVWL GVRPPI---- 138
Thellungiella_halophila FDG-----YQ TGLPLSIGSA VWEQGAEVWL GVRPPI---- 138
Chorispora_bungeana FDG-----YQ TGLPLSIGSA IWEQGTEVWL GVRPPI---- 143
Tarenaya_hassleriana_A FDG-----YQ NGLPLSVGSA IWAQGAEVWL GVRPPI---- 144
Tarenaya_hassleriana_B FDG-----YQ TGLPLSVGSA IWEQGADVWL GVRPPI---- 144
Camptotheca_acuminata FDN-----YQ TGLPLCVGNG IWEQGTEVWT GVRPPI---- 139
Vitis_vinifera FDG-----YQ TGLPLRVGNG IWEQGTEVWI GVRPPI---- 146
Gossypium_arboreum FDG-----YQ TSLPLFVGTS IWEQGTEVWV GVRPPI---- 143
Gossypium_raimondii FDG-----YQ TSLPLFVGTS IWEQGTEVWV GVRPPI---- 143
Theobroma_cacao FDG-----YQ TGLPLCVGSS IWEQETEVWV GVRPPI---- 141
Hevea_brasiliensis FDG-----FQ TGLPLSVGSS IWEQGTEVWV GFRPPT---- 138
Manihot_esculenta FDG-----FQ TGLPLSVGSS IWEQGTEVWV GFRPPT---- 138
Jatropha_curcas FDG-----FQ TGLPLSVSNT IWEQGTEVWV GFRPPT---- 138
Ricinus_communis FDG-----FQ TGLPLSVGNS IWELGTEVWV GFRPPT---- 143
Populus_euphratica_A FDG-----IQ TGLPLSVGSS IWEQGTEVWV GVRPPI---- 141
Populus_trichocarpa_B FDG-----YQ TGLPLSVGSS IWEQGTEVWV GVRPPI---- 141
Populus_euphratica_B FDG-----FQ TGLPLSVGSS IWEQGTEVWV GVRPPI---- 142
Populus_euphratica_C FDG-----FQ TGLPLSVGSS IWEQGTEVWV GVRPPI---- 142
Populus_trichocarpa_A FDG-----FQ TGLPLSVGSS IWEQGTEVWV GVRPPI---- 142
Morus_notabilis FDG-----YQ TGLPLHVGES IWEQGTEVWI GVRPPI---- 145
Citrus_clementina FDG-----YQ TGLALSAGNS IWEEGAEVWV GVRPPT---- 141
Citrus_sinensis FDG-----YQ TGLALSAGNS IWEEGAEVWV GVRPPT---- 141
Fragaria_vesca FDG-----YQ TGLPLHVGNT IWELGTEVWV GVRPPT---- 140
Malus_domestica_A FDG-----YQ TGLPLQVGNT IWEEGTEVWV GVRPPT---- 145
Malus_domestica_B FDG-----YQ TGLPLQVGNT IWEEGTEVWV GVRPPT---- 145
Pyrus_x_bretschneideri_B FDG-----YQ TGLPLQVGNT IWEEGTEVWV GVRPPT---- 145
Malus_domestica_C FDG-----YQ TGLPLHVGNT VWEEGTEVWV GVRPPT---- 139
Pyrus_x_bretschneideri_A FDG-----YQ TGLPLHVGNT VWEQGTEVWV GVRPPT---- 141


Prunus_mum FDG-----YQ TGLPLHVGNT IWEQGTEVWV GVRPPT---- 139
Prunus_persica FDG-----YQ TGLPLHVGNT IWEQGTEVWV GVRPPT---- 143
Carica_papaya FDG-----YQ TGLPLCVGNS IWEQGTEVWL GVRPPI---- 140
Cannabis_sativ FDG-----YQ NALPLQIGDS IWEQGTEVWL GVRPPI---- 144
Cicer_arietinum FDG-----YQ NGLPLCVGSS IWDHGTEVWV GVRPPT---- 139
Medicago_truncatula FDG-----YQ NGLPLCVGSS IWDHGTEVWV GVRPPT---- 138
Glycine_max_A FDG-----YQ NGLPLCVGSS IWEQGTEVWV GVRPPT---- 139
Glycine_soja_B FDG-----YQ NGLPLCVGSS IWEQGTEVWV GVRPPT---- 140
Glycine_max_B YDG-----YQ SGLPLCVGSS IWEQGTEVWV GVRPPT---- 140
Glycine_soja_A YDG-----YQ SGLPLCVGSS IWEQGTEVWV GVRPPT---- 140
Phaseolus_vulgaris FDG-----YQ NGLPLCVGSS IWEEGTEVWV GVRPPT---- 138
Erythranthe_guttata_A YDG-----YQ TGLPLNVGNG IWEQGTDVWV GVRPPT---- 145
Erythranthe_guttata_B YDG-----YQ TGLPLNVGNG IWEQGTDVWV GVRPPT---- 145
Mimulus_guttatus_A YDG-----YQ TGLPLNVGNG IWEQGTDVWV GVRPPT---- 145
Mimulus_guttatus_B YDG-----YQ TGLPLNVGNG IWEQGTDVWV GVRPPT---- 145
Sesamum_indicum YDG-----YQ MGLPLNVGNG IWEQGTDVWV GIRPPI---- 142
Linum_usitatissimum_A FDG-----MQ TGLPLPADNS IWEQGTEVWV GVRPPM---- 144
Linum_usitatissimum_B FDG-----MQ TGLPLPADNS IWEQGTEIWV GVRPPM---- 145
Cucumis_melo FDG-----YQ TGLPLNVGDN IWEQGTEIWV GVRPPT---- 136
Cucumis_sativus FDG-----YQ TGLPLNVGDN IWEQGTEIWV GVRPPT---- 135
Nelumbo_nucifera FDA-----YQ SGLPLHTGNG IWDQGTEVWV GIRPPT---- 139
Coffea_canephora FDG-----YQ AGLPLSVGNG IWEDGAEVWV GIRPPI---- 143
Eucalyptus_grandis FDG-----YE TGLPLPVGND IWEQGTEVWI GVKPPI---- 137
Lactuca_serriola FDG-----YQ TGLPLRVGNG IWEPGTDVWV GVRPPT---- 146
Amborella_trichopoda FDG-----YQ TGLPLLVENG IWEQGTEAWI GIRPPT---- 143
Capsicum_annuum FDG-----YQ TGLPLRVASC IWELGTDVWV GIRPPI---- 135
Solanum_lycopersicum FDG-----YQ TGLPLRVASC IWDLGTDVWV GIRPPI---- 135
Solanum_tuberosum FDG-----YQ TGLPLRVASC IWDLGTDVWV GIRPPI---- 135
Nicotiana_benthamiana FDG-----YQ TGLPLRVASC IWELGTDVWV GIRPPI---- 141
Nicotiana_sylvestris FDG-----YQ TGLPLRVASC IWELGTDVWV GIRPPI---- 141
Nicotiana_tomentosiformis FDG-----YQ TGLPLRVASC IWELGTDVWV GIRPPI---- 138
Costus_pictus YDG-----YQ SGLPIHGTSC IWEEGTSVWT GVRPPV---- 140
Curcuma_longa YDG-----YQ SGLPLHGTNC IWEEGTSVWA GIRPPI---- 143
Musa_acuminata YDG-----YQ SVSLLQGTSC IWEEGTSVWA GVRPPV---- 140
Genlisea_aurea YDG-----YQ TSLPLNMGDC IWERGTDVWI GVRPPM---- 147
Utricularia_gibba YDG-----YQ IVLPLNMGNG IWEQGTEIWI GVRPPM---- 146
Beta_vulgaris FDG-----YL NELPLRVLNG IWQQGTEVWI GVRPPT---- 140
Marchantia_polymorpha_A YDGFAKFESL NGLQQPAEGG IWEDGTEVWV GIRPPM---- 144
Marchantia_polymorpha_B FDGDVLLDNL NGLQLPAEKG IWEEGTEVWV GIRPPM---- 144
Selaginella_moellendorffii_B LDG-----TQ SGLQIPVEGG IWQRETEVWV GSRPPM---- 133
Selaginella_moellendorffii_A LDG-----TQ SGLQIPVEGG IWQRETEVWV GSRPPM---- 133
Ceratodon_purpureus FDG----EVR CDLSLPAEGG VWQEEMEVWV GIRPPM---- 146
Physcomitrella_patens FDG----DNR CDLLLLGEGG VWEEGMEVWV GIRPPM---- 127
Brassica_napus_A FDG-----YQ TGLPLSISSA IWEQGAEVWL DEAASLHAAV 143
Consensus FDG-----YQ TGLPLSVGNG IWEQGTEVWV GVRPPT----
100%
Conservation
0%


180
200
Aegilops_tauschii ---DLDAFGR SDSEGTDSKM QIMDAFLWG- ---RCLTEDE 155
Brachypodium_distachyon ---DLDAFGR SDSEGSDSKM QIMDAFLWG- ---RCLTEDE 170
Hordeum_vulgare ---DLDAFGR SDSEGTDSKM QIMDAFLWG- ---RCLTEDE 168
Oryza_brachyantha ---DLDAFGR SDSEGSDSKM QIMDAFLWG- ---RCLTEDE 166
Oryza_sativa ---DLDAFGR SDSEGSDSKM QIMDAFLWG- ---RCLTEDE 166
Panicum_virgatum ---DLDAFGR SDSEGSDSKM QIMDAFLWG- ---RCLSEDE 167
Setaria_italica ---DLDAFGR SDSEGSDSKM QIMDAFLWG- ---RCLSEDE 171
Sorghum_bicolor ---DLDAFGR SDSEGSDSKM QIMDAFLWG- ---RCLSEDE 171
Zea_mays ---DLDAFGR SDSEGSDSKM QIMDAFLWG- ---RCLSEDE 169
Elaeis_guineensis_A ---DLDAFGR SDSEGADSKM QIMDAFLWG- ---RCLTEDE 171
Phoenix_dactylifera_B ---DLDAFGR SDSEGADSKM QIMDAFLWG- ---RCLTEDE 171
Phoenix_dactylifera_A ---DLDAFGR SDSEGSDSKM QIMDAFLWG- ---RCLTEDE 171
Elaeis_guineensis_B ---DLDAFGR SDSEGSGSKM QIMDAFLWG- ---RCLTEDE 171
Aquilegia_coerulea ---DVDAFGR SDSEGVDPKM HIMDAFLWG- ---RCLSEDE 168
Arabidopsis_lyrata ---DVDAFGR SDSDGVESKM HIMDVFLWG- ---KCLSEDE 175
Arabidopsis_thaliana ---DVDAFGR SDSDGVESKM HIMDVFLWG- ---KCLSEEE 175
Camelina_sativa_B ---DVDAFGR SDSDGVESKM HIMDVFLWG- ---KCLSEDE 175
Camelina_sativa_C ---DVDAFGR SDSDGVESKM HIMDVFLWG- ---KCLSEDE 175
Capsella_rubella ---DVDAFGR SDSDGVESKM HIMDVFLWG- ---KCLSEDE 176
Brassica_napus_B ---DVDAFGR SDSDGAESKM HIMDVFLWG- ---KCLTEDE 172
Brassica_rapa ---DVDAFGR SDSDGAESKM HIMDVFLWG- ---KCLTEDE 172
Thlaspi_arvense ---DVDAFGR SDSDGAESKM HIMDVFLWG- ---KCLTEDE 172
Eutrema_salsugineum ---DVDAFGR SDSDGAESKM HIMDVFLWG- ---KCLTEDE 171
Thellungiella_halophila ---DVDAFGR SDSDGAESKM HIMDVFLWG- ---KCLTEDE 171
Chorispora_bungeana ---DVDAFGR SDSDGAESKM HIMDVFLLG- ---KCLTEDE 176
Tarenaya_hassleriana_A ---DVDAFGR SDSDGAEPKM HMMDVFLWG- ---KCLTEDE 177
Tarenaya_hassleriana_B ---DVDAFGR SDSDGVESKM HMMDVFVWG- ---KCLTEDE 177
Camptotheca_acuminata ---DMDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 172
Vitis_vinifera ---DIDAFGR SDSEGAESKM HIMDVFMWG- ---RCLTEDE 179
Gossypium_arboreum ---DMDAFGR SDSEGAESKM HIMDVFLWG- ---RCLNEDE 176
Gossypium_raimondii ---DMDAFGR SDSEGAESKM HIMDVFLWG- ---RCLNEDE 176
Theobroma_cacao ---DMDAFGR SDSEGAESKM HVMDVFLWG- ---RCLNEDE 174
Hevea_brasiliensis ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 171
Manihot_esculenta ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 171
Jatropha_curcas ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 171
Ricinus_communis ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 176
Populus_euphratica_A ---DMDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 174
Populus_trichocarpa_B ---DMDAFGR SDSEGAESKM YIMDVFLWG- ---RCLTEDE 174
Populus_euphratica_B ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 175
Populus_euphratica_C ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 175
Populus_trichocarpa_A ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 175
Morus_notabilis ---DMDAFGR SDSEGAESKM HLMDVFLWG- ---RCLTEDE 178
Citrus_clementina ---DMDVFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 174
Citrus_sinensis ---DMDVFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 174
Fragaria_vesca ---DMDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDD 173
Malus_domestica_A ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDD 178
Malus_domestica_B ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDD 178
Pyrus_x_bretschneideri_B ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDD 178
Malus_domestica_C ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDD 172
Pyrus_x_bretschneideri_A ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDD 174
Prunus_mum ---DMDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDD 172


Prunus_persica ---DMDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDD 176
Carica_papaya ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLSEDE 173
Cannabis_sativ ---DMDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 177
Cicer_arietinum ---DIDAFGR SDSEGVESKM HIMDVFLWG- ---RCLSDDE 172
Medicago_truncatula ---DIDAFGR SDSEGVESKM HIMDVFLWG- ---RCLSDDE 171
Glycine_max_A ---DIDAFGR SDSEGVESKM HIMDAFLWG- ---RCLTDDE 172
Glycine_soja_B ---DIDAFGR SDSEGVESKM HIMDAFLWG- ---RCLTDDE 173
Glycine_max_B ---DIDAFGR SDSEGVESKM HIMDAFLWG- ---RCLTDDE 173
Glycine_soja_A ---DIDAFGR SDSEGVESKM HIMDAFLWG- ---RCLTDDE 173
Phaseolus_vulgaris ---DIDAFGR SDSEGVESKM HIMDAFLWG- ---RCLSDDE 171
Erythranthe_guttata_A ---DIDAFGR SDSENAESKM HVMDVFLWG- ---RCLSEDE 178
Erythranthe_guttata_B ---DIDAFGR SDSENAESKM HVMDVFLWG- ---RCLSEDE 178
Mimulus_guttatus_A ---DIDAFGR SDSENAESKM HVMDVFLWG- ---RCLSEDE 178
Mimulus_guttatus_B ---DIDAFGR SDSENAESKM HVMDVFLWG- ---RCLSEDE 178
Sesamum_indicum ---DMDAFGR SDSEGTESKM HVMDVFLWG- ---RCLNEDE 175
Linum_usitatissimum_A ---DMDAFGR SDSEGAESKM HMMDVFLWG- ---KCLTDDD 177
Linum_usitatissimum_B ---DMDAFGR SDSEGAESKM HMMDVFLWG- ---KCLTDDD 178
Cucumis_melo ---DVDIFGR SDSEGAESKM HIMDVFLWG- ---RSLTEDE 169
Cucumis_sativus ---DVDIFGR SDSEGAESKM HIMDVFLWG- ---RSLTEDE 168
Nelumbo_nucifera ---DLDAFGR SDSEGADSKM HIMDAFLWG- ---RCLTEDE 172
Coffea_canephora ---DMDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE 176
Eucalyptus_grandis ---GLDAIGR SDSEGAESKM HIMDIFLWG- ---RCLTEDE 170
Lactuca_serriola ---DVDAFGR SDSEGAESKM HIMDLFLWG- ---RCLLEDE 179
Amborella_trichopoda ---DLDAFGR SDSEGSESKM HLMDAFLWG- ---RCLNEDE 176
Capsicum_annuum ---DVDSFGR SDSEGAESKV LIMDVFLWG- ---RCLTEDE 168
Solanum_lycopersicum ---DVDSFGR SDSEGVESKV HIMDVFLWG- ---RCLTEDE 168
Solanum_tuberosum ---DVDSFGR SDSEGAESKV HIMDVFLWG- ---RCLTEDE 168
Nicotiana_benthamiana ---DVDSFGR SDSEGAESKV HIMDVFLWG- ---RCLTEDE 174
Nicotiana_sylvestris ---DVDSFGR SDSEGAESKV HIMDVFLWG- ---RCLTEDE 174
Nicotiana_tomentosiformis ---DVDSFGR SDSEGAESKV HIMDVFLWG- ---RCLTEDE 171
Costus_pictus ---DLDAFGR SDSEGVESKM QIMDAFLWG- ---RCLTEDE 173
Curcuma_longa ---DLDAFGR SDSEGGDSKM QIMDAFLWG- ---RCLTEDE 176
Musa_acuminata ---DLDAFGR SDSEGVDSKM QIMDAFLWG- ---RCLTEDE 173
Genlisea_aurea ---DMDAFGR SDSEATEPKM HVMDSFLWG- ---RCLSEDE 180
Utricularia_gibba ---DMDAFGR SDSEGSEPKM HIMDSFLWG- ---RCLSEDE 179
Beta_vulgaris ---DIDTFGR SDSEGAESKM HVMDVFLWG- ---RCLSEEE 173
Marchantia_polymorpha_A ---DLDAFGR SDSEGADSRM HVMDVFLWG- ---RLLNEEE 177
Marchantia_polymorpha_B ---DLDAFGR SDSEGNDSRM HVMDVFLWG- ---RLLTEDE 177
Selaginella_moellendorffii_B ---DLDAFGR SDSEGVEASM HIMDVFLWG- ---RCLKEEE 166
Selaginella_moellendorffii_A ---DLDAFGR SDSEGVEASM HIMDVFLWG- ---RCLKEEE 166
Ceratodon_purpureus ---DLDAFGR SDSEGAESRM HAMDIMMWG- ---RCLSEDE 179
Physcomitrella_patens ---DLDAFGR SDSEGAESKM HAMDVFMWG- ---RCLTEDE 160
Brassica_napus_A GMADLDMIDL ND--------- ---DNWQWTA SPPRVDGWDS 172
Consensus ---DVDAFGR SDSEGAESKM HIMDVFLWG- ---RCLTEDE
100%
Conservation
0%

Aegilops_tauschii	I - - - - - - A	157
Brachypodium_distachyon	V - - - - - - A	172
Hordeum_vulgare	I - - - - - - A	170
Oryza_brachyantha	V - - - - - - A	168
Oryza_sativa	V - - - - - - A	168
Panicum_virgatum	V - - - - - - A	169
Setaria_italica	V - - - - - - A	173
Sorghum_bicolor	V - - - - - - T	173
Zea_mays	V - - - - - - T	171
Elaeis_guineensis_A	I - - - - - - A	173
Phoenix_dactylifera_B	I - - - - - - A	173
Phoenix_dactylifera_A	I - - - - - - A	173
Elaeis_guineensis_B	I - - - - - - A	173
Aquilegia_coerulea	I - - - - - - V	170
Arabidopsis_lyrata	A - - - - - - A	177
Arabidopsis_thaliana	A - - - - - - A	177
Camelina_sativa_B	A - - - - - - A	177
Camelina_sativa_C	A - - - - - - A	177
Capsella_rubella	A - - - - - - A	178
Brassica_napus_B	A - - - - - - A	174
Brassica_rapa	A - - - - - - A	174
Thlaspi_arvense	A - - - - - - A	174
Eutrema_salsugineum	A - - - - - - A	173
Thellungiella_halophila	A - - - - - - A	173
Chorispora_bungeana	A - - - - - - A	178
Tarenaya_hassleriana_A	I - - - - - - A	179
Tarenaya_hassleriana_B	I - - - - - - A	179
Camptotheca_acuminata	I - - - - - - A	174
Vitis_vinifera	I - - - - - - A	181
Gossypium_arboreum	V - - - - - - A	178
Gossypium_raimondii	V - - - - - - A	178
Theobroma_cacao	I - - - - - - A	176
Hevea_brasiliensis	I - - - - - - A	173
Manihot_esculenta	I - - - - - - A	173
Jatropha_curcas	I - - - - - - A	173
Ricinus_communis	I - - - - - - A	178
Populus_euphratica_A	I - - - - - - A	176
Populus_trichocarpa_B	I - - - - - - A	176
Populus_euphratica_B	I - - - - - - A	177
Populus_euphratica_C	I - - - - - - A	177
Populus_trichocarpa_A	I - - - - - - A	177
Morus_notabilis	I - - - - - - A	180
Citrus_clementina	I - - - - - - A	176
Citrus_sinensis	I - - - - - - A	176
Fragaria_vesca	I - - - - - - A	175
Malus_domestica_A	V - - - - - - A	180
Malus_domestica_B	V - - - - - - A	180
Pyrus_x_bretschneideri_B	V - - - - - - A	180
Malus_domestica_C	V - - - - - - A	174
Pyrus_x_bretschneideri_A	V - - - - - - A	176
Prunus_mum	I - - - - - - A	174


Prunus_persica I - - - - - - A 178
Carica_papaya I - - - - - - A 175
Cannabis_sativ V - - - - - - A 179
Cicer_arietinum V - - - - - - S 174
Medicago_truncatula V - - - - - - S 173
Glycine_max_A V - - - - - - S 174
Glycine_soja_B V - - - - - - S 175
Glycine_max_B V - - - - - - S 175
Glycine_soja_A V - - - - - - S 175
Phaseolus_vulgaris V - - - - - - S 173
Erythranthe_guttata_A I - - - - - - A 180
Erythranthe_guttata_B I - - - - - - A 180
Mimulus_guttatus_A I - - - - - - A 180
Mimulus_guttatus_B I - - - - - - A 180
Sesamum_indicum V - - - - - - A 177
Linum_usitatissimum_A V - - - - - - A 179
Linum_usitatissimum_B V - - - - - - A 180
Cucumis_melo I - - - - - - A 171
Cucumis_sativus I - - - - - - A 170
Nelumbo_nucifera I - - - - - - G 174
Coffea_canephora I - - - - - - A 178
Eucalyptus_grandis V - - - - - - A 172
Lactuca_serriola I - - - - - - S 181
Amborella_trichopoda I - - - - - - A 178
Capsicum_annuum I - - - - - - A 170
Solanum_lycopersicum I - - - - - - A 170
Solanum_tuberosum I - - - - - - A 170
Nicotiana_benthamiana I - - - - - - A 176
Nicotiana_sylvestris I - - - - - - A 176
Nicotiana_tomentosiformis I - - - - - - A 173
Costus_pictus I - - - - - - A 175
Curcuma_longa I - - - - - - A 178
Musa_acuminata I - - - - - - A 175
Genlisea_aurea I - - - - - - S 182
Utricularia_gibba I - - - - - - S 181
Beta_vulgaris V - - - - - - S 175
Marchantia_polymorpha_A I - - - - - - V 179
Marchantia_polymorpha_B I - - - - - - A 179
Selaginella_moellendorffii_B L - - - - - - L 168
Selaginella_moellendorffii_A L - - - - - - L 168
Ceratodon_purpureus I - - - - - - L 181
Physcomitrella_patens I LMVHNC A 168
Brassica_napus_A - - - - - - - - 172
Consensus I - - - - - - A
100%
Conservation
0%

Appendix 6: Chemicals and kits

Name	Supplier	Country
Phusion® High-Fidelity PCR kit	Themo Scientific	Lithuania
QIAquick gel extraction kit	QIAGEN	Germany
Zero Blunt® PCR Cloning Kit	Invitrogen	USA
In-Fusion HD cloning kit	TaKaRa Clontech	USA
GeneArt®Site-Directed Mutagenesis PLUS kit	Invitrogen	USA
HOT FIREPol DNA polymerase	Solis BioDyne	Estonia
PureYield™ Plasmid Minprep System kit	Promega	USA
BigDye® terminator v3.1 cycle sequencing kit	Applied Biosystem	USA
QIAGEN Plasmid mid kit	QIAGEN	Germany
Phire Plant Direct PCR Kit	Thermo Scientific	Lithuania
PhytoPure Genomic DNA Extraction Kit	Illustra	UK
RNAqueous®- Micro kit	Life Techologies	Lithuania
DNase I-Amplification Grade	Invitrogen	USA
Superscript III® Reverse Transcriptase	Invitrogen	USA
RNaseOUT™	Invitrogen	USA
PCR DIG Probe Synthesis Kit	Roche	Germany
Micribiology LB (MILLER)	Merck	Germany
Kanamycin	Sigma-Aldrich	USA
Ampicillin	Sigma-Aldrich	USA
Hygromycin	Sigma-Aldrich	USA
Carbanicillin	Invitrogen	USA
Driselase	Sigma-Aldrich	USA
Manitol	Sigma-Aldrich	USA

Agar	Sigma-Aldrich	USA
$FeSO_4.7H_2O$	Merck	Germany
AccuPrime *Pfx*	Invitrogen	USA
SeaKem® LEA Agarose	Lonza	USA
Deammonia tertrate	Alfa Aesar	Germany
Tween 20	VWR	France
CSPD	Roche	Germany
Ant-DIG-AP	Roche	Germany

Appendix 7: *P. patens* growth media

A. BCD medium

Reagent	1 Litre	Final concentration
Agar	7g	0.7 % (w/v)
$FeSO_4.7H_2O$	12.5 mg	45 µM
Solution B	10 ml	1 mM $MgSO_4$
Solution C	10 ml	1.84 Mm KH_2PO_4
Solution D	10 ml	10 Mm KNO_3
TES	1 ml	Trace

For BCDA media, add 920 mg of Diammonium tertrate to the medium to make final concentration of 5 mM. Add dH_2O to 1 L and autoclave at 121 °C for 40 min. After autoclaving add $CaCl_2$ to final concentration of 1 mM.

B. Protoplast Regeneration Medium for the Bottom layer (PRMB)

Reagent	1 Litre	Final concentration
Agar	7g	0.7 % (w/v)
Diammonium tertrate	920 mg	5 mM
D-Mannitol	60g	6 % (w/v)
BCD medium, liquid	10 ml each	

Add dH_2O to 1 L and autoclave at 121 °C for 40 min. After autoclaving add $CaCl_2$ to the final concentration of 10 mM.

C. Protoplast Regeneration Medium for the Top layer (PRMT)

Reagent	1 Litre	Final concentration
Agar (Sigma-Aldrich)	4 g	0.4 % (w/v)
Diammonium tertrate	920 mg	5 mM
D-Mannitol	80g	8 % (w/v)
BCD medium, liquid	10 ml each	

Add dH_2O to 1 L and autoclave at 121 °C for 40 min. After autoclaving add $CaCl_2$ to the final concentration of 10 mM.

Appendix 8: Solutions

A. Driselase solution

Reagent	100 ml	Final concentration
Driselase	2g	2%
D-Mannitol solution (8.5% w/v)	to 100 ml	

Stir the mixture for 15 min. Centrifuge at 2500g for 5 min. and filter sterilize the clear supernatant.

B. Hoagland's A-Z trace element solution (TES)

Reagent	1 Litre	Final concentration
$CoCl_2.6H_20$	55 mg	0.006% (w/v)
$CuSO_4.5H_2O$	55 mg	0.006% (w/v)
H_3BO_3	614 mg	0.061% (w/v)
KI	28 mg	0.003% (w/v)
$MnCl_2.4H_2O$	389 mg	0.039% (w/v)
$ZnSO_4.7H_2O$	55 mg	0.006% (w/v)
Add dH_2O to 1 litre.		

C. MMM solution

Reagent	10 ml	Final concentration
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
dH_2O	8.85	
Dissolve D-mannitol in the dH_2O and sterilize by autoclaving. On the day of use, add the MES and $MgCl_2$ and filter sterilize.		

D. PEG solution for protoplast fusion (PEG/F)

Reagent	Quantity
$CaCl_2.6H_20$	109 mg
dH_2O	10 ml
Polyethylene glycol (PEG) (MW 6000)	5g

Dissolve $CaCl_2.6H_20$ in dH_2O and then mix the solution with the melted PEG

E. Solution B

Reagent	Quantity (for 1 litre)	Final concentration
$MgSO_4.7H_2O$	25 g	0.1 M
Add dH_2O to 1 litre and sterilize by autoclaving for 20 min. at 120 °C.		

F. Solution C

Reagent	Quantity (for 1 litre)	Final concentration
KH_2PO_4	25 g	184 mM
Adjust the pH to 6.5 with 4 M KOH. Add dH_2O to 1 litre and sterilize by autoclaving for 20 min. at 120 °C.		

G. Solution D

Reagent	Quantity (for 1 litre)	Final concentration
KNO_3	101 g	1 M
Add dH_2O to 1 litre and sterilize by autoclaving for 20 min. at 120 °C.		

H. SOC medium

Reagent	Quantity (1 litre)	Final concentration
Tryptone	20 g	2 %
Yeast Extract	5 g	0.5%
1M NaCl	10 g	10 mM

1M KCl	2.5 ml	2.5 mM
1M $MgCl_2$	10 ml	10 mM
1M $MgSO_4$	10 ml	10 mM
1M Glucose	20 ml	20 mM

Adjust pH to 7 with NaOH, add dH_2O to 1 litre and autoclave for 20 min.

I. Antibiotics

Antibiotics	Stock concentration	Working concentration
Amplicillin	100 mg/ml	100 µg/ml
Kanamycin	50 mg/ml	50 µg/ml
Hygromycin	50 mg/ml	20 µg/ml
Carbenicillin	250 mg/ml	100 µg/ml

J. Southern Blot Buffers

1. Depurination solution

Component	1 litre
0.25M (37%) HCl	19.2 ml

Add dH_2O until 1 litre, then autoclave and store at RT

2. Denaturation solution

Component	1litre
1.5M NaCl	80g
0.5M NaOH	20g

Add dH_2O to 1 litre, autoclave and store at 4 °C

3. Neutralization solution

Component	1 litre
0.5M Tris-HCl pH 7.0	66.6g
1.5M NaCl	88g
Add dH_2O, adjust pH to 7.5, autoclave and store at 4 °C	

4. Southern blot transfer buffer (20X SSC)

Component	1 litre
NaCl	175.3g
Trisodium citrate	88.2g
Add dH_2O, adjust pH to 7, then autoclave and store at RT	

5. 10X DIG1 buffer

Component	1 litre
Maleic acid	116.1g
NaOH	87.66g
Add dH_20, adjust to pH 7.5 with NaOH pellets, autoclave and store at RT	

J. Stock solutions

Reagent	Quantity (100 mL)	Sterilization
1M NaCl	5.80 g	Autoclave
1M KCl	7.50 g	Autoclave
1M $MgCl_2 \cdot 6H_20$	20.30 g	Autoclave
1M $MgSO_4$	12.00g	Filter

1M $CaCl_2$	11.10 g	Filter
3M NaOAc.$3H_20$, pH 5.2	40.8 g	Autoclave
5M EDTA pH 8	186.1g	Autoclave
1M Tris	12.114g	Autoclave