INVESTIGATING THE ROLE OF ARMADILLO-RELATED PROTEINS IN EARLY LAND PLANTS

by

LAURA ALISON MOODY

A thesis submitted to

The University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences

The University of Birmingham

September 2011

ABSTRACT

Mosses evolved approximately 500 million years ago and were among the earliest plants to make the transition from water to land. Mosses are therefore placed at an ideal evolutionary position in which to understand how plant physiology and development has evolved from simple unicellular aquatic organisms to generate the huge diversity of complex modern day flowering plants. The moss *Physcomitrella* has the unique ability among known land plants to carry out homologous recombination at a similar efficiency to the yeast *Saccharyomyces cerevisiae*.

Armadillo-related proteins play important roles in cellular processes both in animals and plants. In *Arabidopsis*, ARABIDILLO1 and ARABIDILLO2 control root system architecture. ARABIDILLO-like proteins have been identified extensively throughout the plant kingdom, including early-evolving moss and agriculturally important crops such as rice and maize.

Three *Physcomitrella* ARABIDILLO homologues have been identified; *PHYSCODILLO1A*, *PHYSCODILLO1B* and *PHYSCODILLO2*. Cloning, sequencing and Southern blotting approaches confirmed that *PHYSCODILLO2* was a single copy gene, whereas full-length *PHYSCODILLO1A* and *PHYSCODILLO1B* genes were 100% identical and exist in a tail-to-tail orientation with 8kb separating their stop codons. A number of *physcodillo* deletion mutants have been generated. Phenotypic analyses revealed that PHYSCODILLO proteins appear to play important roles during early developmental processes, including growth of filaments from protoplasts and spore germination.

ACKNOWLEDGEMENTS

I would like to thank my fantastic supervisor Juliet Coates for agreeing to take me on as her second PhD student and for all her support and encouragement along the way. Thank you to Younousse Saidi and Susan Bradshaw for help and support in the lab and for providing many giggles. Thanks also to past members of the Coates lab, particularly Daniel Gibbs and Emma Smiles, for helping me get started and for your continued support.

I would like to thank all friends and colleagues from Franklin, Franklin-Tong, Armstong, Sanchez-Moran and Pritchard labs for advice and encouragement and for supplying us with the odd rare chemical! Thanks also to the many friendly faces in the office (S208).

Thanks to Andrew Cuming, Jane Langdale and the Reski lab for helping us to get the moss tissue culture up and running at the University of Birmingham.

I would like to acknowledge the BBSRC for funding my PhD studentship, The Leverhulme Trust for funding my many experiments and the SEB, SIVB and New Phytologist for travel grants that enabled me to present my research in the UK, Finland and Japan.

Finally, thanks to my parents and sisters for their immense support and to Chris for putting up with me at home throughout my PhD! Thanks also to all of my friends, who have now dispersed far and wide, for providing giggles along the way.

CONTENTS

Chapter I: Introduction

LIST OF FIGURES

LIST OF TABLES

Chapter I: Introduction

1.1 Project Introduction

The bryophytes were the earliest plants to successfully transition from water to land approximately 500 million years ago and they comprise the liverworts, mosses and hornworts. Movement to the terrestrial environment exposed plants to harsh conditions such as drought, UV stress and salinity. The acquisition of desiccation tolerance was therefore one of the key evolutionary events that enabled plants to successfully survive and reproduce on land. To enable plants to move from moist to progressively drier habitats, plants evolved a number of complex morphological adaptations. Vascular tissues evolved to provide mechanical support against the pull of gravity and a means of transporting water and nutrients throughout the plant. Roots evolved to allow plants to anchor to substrata and provide them with better access to water and nutrients. Seeds and flowers evolved most recently to improve the reproductive success of plants.

A number of evolutionarily ancient protein families are found extensively throughout land plants, from early-evolving bryophytes to modern day flowering plants, such as the tetraspanins, Glygogen Synthase Kinase 3s (GSK3) and Armadillo-repeat proteins. They are therefore likely to be important for plant development and are probably involved in evolutionarily conserved developmental processes.

The protein family investigated in this thesis is the evolutionarily ancient Armadillo-repeat proteins and these are important for multicellular development in animals, amoebae and plants. Animal and amoebal Armadillo play dual roles in cell adhesion and cell signalling. In the nucleus, they switch on genes important for multicellular development and at cellular junctions, they interact indirectly with the actin cytoskeleton in order to maintain tissue architecture (Oda *et al*., 1993; Peifer *et al*., 1993; Oda *et al*., 1997; van Noort *et al*., 2002; Grimson *et al*., 2000; Coates *et al*., 2002).

ARABIDILLO1 and ARABIDILLO2 are Armadillo-related proteins that promote root branching in *Arabidopsis*. ARABIDILLO1 and ARABIDILLO2 contain an F-box domain, a linker region of leucine-rich repeats and an Armadillo (ARM) repeat region (Coates *et al*., 2006). The ARABIDILLOs are *bona fide* F-box proteins and may target a lateral root inhibitor for ubiquitin-mediated degradation via the 26S proteasome (Nibau *et al*., 2011; Daniel Gibbs, PhD thesis).

ARABIDILLO homologues are found extensively throughout the land plants and their protein sequences are all highly conserved (Nibau *et al*., 2011). Armadillo-related proteins are also found in the bryophytes, the most basal land plant lineage. However, bryophytes lack the sophisticated rooting structures found in higher plants and instead, produce filamentous rhizoids that are suggested to be equivalent to root hairs (Menand *et al*., 2007). The aims of this PhD project were to clone and functionally characterise Armadillo-related genes in the moss *Physcomitrella*, in order to determine the ancestral function of Armadillo-related proteins in land plants.

1.2 The origins of multicellularity

Animals and plants have independently acquired multicellularity from a common unicellular ancestor and this has resulted in the enormous morphological diversity that we observe today (Meyerowitz, 1999). Molecular phylogeny has revealed that each multicellular lineage came from a unicellular ancestor and the presence of unicellular sister groups confirms this: The Choanoflagellates, a group of flagellated protozoa, are believed to be the closest living ancestor of animals; Fungi evolved from nucleariids, a group of protists found predominantly in soil and freshwater environments; the social amoeba, the Dictyostelids, which have both unicellular and multicellular phases, can be traced back to free-living

unicellular lobose amoeba (Abedin and King, 2010); Algae such as *Volvox carteri* acquired simple multicellularity in water and are thought to have arisen from unicellular and flagellated relatives of Volvocine algae such as *Chlamydomonas reinhardtii* (Bonner, 1998).

Land plants and red and green algae all descended from a single-celled ancestor that acquired a photosynthetic cyanobacterium. This subsequently gave rise to three distinct groups; the glaucophytes (freshwater algae), rhodophytes (red algae) and green plants. The green plants comprise the chlorophytes (freshwater and marine algae) and streptophytes (Charophytes and land plants; Bowman *et al*., 2007). It is widely accepted that the embryophytes or land plants form a monophyletic group that is most closely related to the Charophycean algae (Kenrick and Crane, 1997). However, it is unsure whether the last common ancestor of the Charophycean algae and land plants was unicellular or multicellular. Fossil evidence supports the antithetic theory whereby a green algal ancestor had a gametophyte-dominant haplontic life cycle with no sporophyte generation (Gerrienne and Gonez, 2010). The sporophyte generation of the algal ancestor was independent of the gametophyte and confined to the unicellular diploid zygote, which underwent meiosis to produce haploid spores (Bennici, 2008). It is likely that meiosis was delayed until after the mitotic division of the diploid zygote to create a full multicellular diploid stage. The life cycle therefore acquired a novel diploid sporophyte generation (Niklas and Kutschera, 2009; Dolan, 2009).

1.3 Colonisation of the terrestrial environment

One of the most important events in the history of life was the colonisation of the terrestrial environment by green, multicellular land plants during the mid-Ordovician (460mya). In order to successfully transition to land, plants evolved gametophytes bearing

sexual organs and a life cycle with heteromorphic alternation of multicellular generations.

The earliest pioneering group of land plants to successfully colonise the land were the bryophytes. The bryophytes consequently pioneered responses to stresses that secured their subsequent survival and enabled them to colonise diverse habitats from arid desert plains to subzero Antarctica (Minami *et al*., 2005; Wang *et al*., 2009). The three extant bryophyte lineages are the liverworts, the mosses and the hornworts. They all have a dominant multicellular gametophyte and a strongly reduced sporophyte generation (Kenrick and Crane, 1997). It is widely accepted that the liverworts are the most basal land plant lineage and that genes present in both *Marchantia polymorpha* (liverwort) and *Physcomitrella patens* (moss) are likely to be present in a common ancestor of all land plants (Hedges, 2002).

Evolving desiccation tolerance was one of the key survival strategies of the earliest land plants and this provided them with resistance to rapid drying as a result of unfavourable environmental conditions, such as heat, UV radiation and salinity (Farrant and Moore, 2011). Vegetative tissues of the bryophytes are therefore extremely desiccation tolerant and can recover from up to 92% water loss (Frank *et al*., 2005; Cuming *et al*., 2007). Desiccation tolerance in vegetative tissues has been lost in vascular plants, but has been retained in a few specialised structures, such as seeds, pollen and spores (Khandelwal *et al*., 2010).

Plants evolved a number of adaptations that enabled them to move from moist to progressively drier habitats (Becker and Marin, 2009). One notable adaptation was the switch in the alternation of generations to a dominant sporophyte generation with a reduced gametophyte generation. Consequently, there was an explosion of morphological diversity during the Devonian period (420-350mya) (Dolan, 2009). As plants evolved, sporophytes acquired branching and vascular tissues to provide mechanical support and a means of transporting water and nutrients throughout the plant (Langdale, 2008).

Small, low-lying bryophytes rely upon simple diffusion to transport both water and solutes in and out of the plant. Some species transport water and sugars using primitive conductive tissues called hydroids and leptoids respectively. Bryophytes lack true roots for efficient uptake of water and nutrients from soil substrata. Liverworts possess single-celled filamentous rhizoids on ventral sides of their gametophytes whereas mosses and hornworts possess multicellular rhizoids for nutrient uptake and support (Dolan, 2009). Later evolving lycophytes such as *Selaginella moellendorffii* possess rhizomes and these differentiate into root-like structures that have root hairs and a root cap like those found in angiosperms (Banks, 2009). Roots are thought to have evolved from ancestral shoots and independently at least twice in both the lycophytes and euphyllophytes, which comprise the ferns and horsetails (Dolan, 2009). Vascular plants evolved extensively branched rooting systems that provided them with efficient means of anchorage and enabled them to explore larger volumes of soil for water and nutrients. Interestingly, some genes controlling the development of cells with a rooting function are evolutionarily ancient and are conserved between the bryophytes and angiosperms (Menand *et al*., 2007).

Basal land plants have simple leaf-like structures that are borne on gametophytes and these are only composed of a single cell layer. The earliest leaves evolved in the lycophyte lineage independently of shoots and these were microphylls that possessed a single vascular trace. Leaves with complex venation patterns known as megaphylls evolved more recently and at least twice within the euphyllophyte lineage. True leaves are observed in the dominant sporophytes of vascular plants and these have multiple cell layers and lignified vasculature. They also possess an epidermal cell layer that is coated with a waxy cuticle and covered in stomata, to permit gas exchange and protect the plant against water loss (Acharya *et al*., 2009; Dolan, 2009). Mosses and some hornworts (but not liverworts) possess stomata that function

in a similar way to those of vascular plants, but they are restricted to the sporophyte generation (Chater *et al*., 2011).

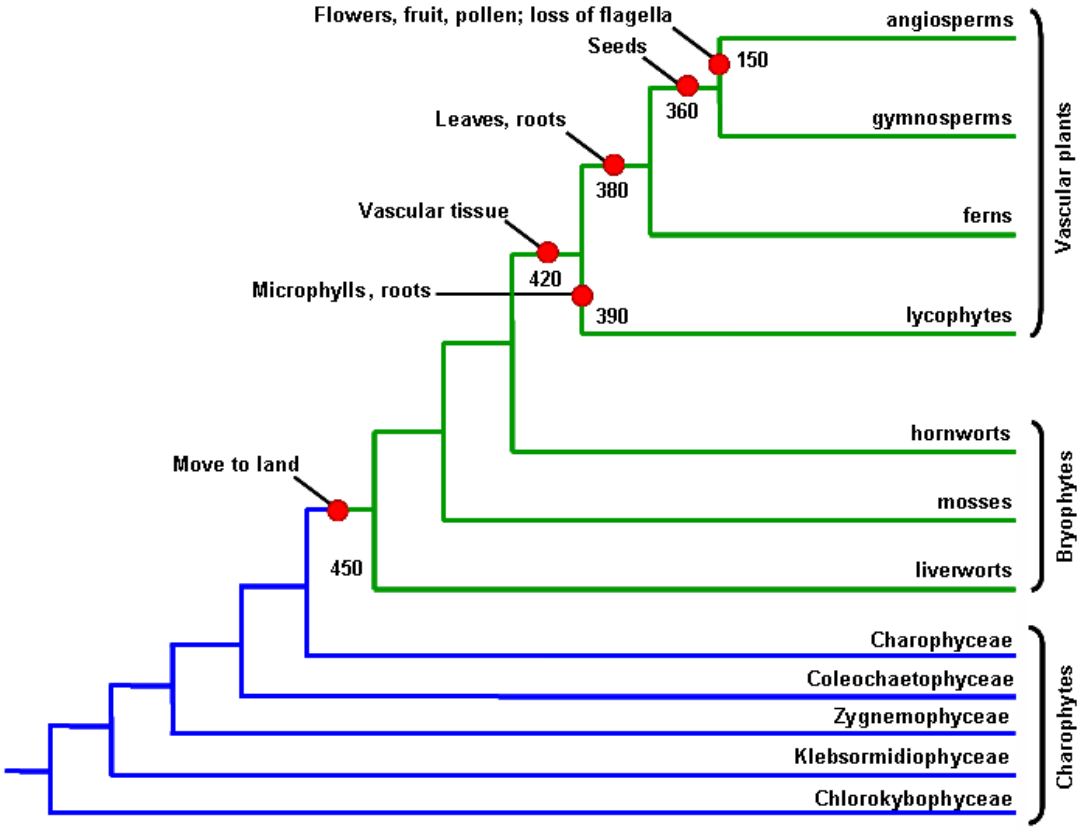
The bryophytes, lycophytes and euphyllophytes have biflagellated sperm and are therefore dependent on water to fertilise egg cells. Among the gymnosperms, the cycads and gingko are the most recently evolving plants to possess flagellated sperm. However, these differ from those observed in earlier-evolving plant lineages in that they only possess single flagella. All other gymnosperms and angiosperms rely upon transfer of sperm to ovules via pollen tubes and the advantage of this is that water is not required (Southworth and Cresti, 1997; Prigge and Bezanilla, 2010).

Bryophytes have a predominantly haploid life cycle and reproduce by germination of haploid spores that are produced by meiosis within a diploid sporophyte. Thousands of spores can be produced and these are dispersed far and wide by aerial means (Bennici, 2008). Uniquely, bryophyte spores are protected from predators, mechanical damage and UV light by an extremely resistant polymer known as sporopollenin. Sporopollenin is believed to have evolved within the Charophycean algae prior to the emergence of land plants and has subsequently been retained by plants to confer desiccation and decay tolerance to both spores and pollen (Langdale, 2008; Becker and Marin, 2009).

In seed plants, sexual organs are borne on cones or flowers in gymnosperms and angiosperms respectively. These structures sit at an elevated position permitting greater aerial distribution of seeds into uncolonised and drier habitats (Prigge and Bezanilla, 2010). Flowering structures evolved to permit insect or wind-mediated transfer of pollen between flowers, promoting outbreeding and creating more genetically diverse populations. Insect pollinated plants usually have large, colourful flowers designed to attract insect pollinators,

whereas those dependent on wind for pollination are usually less elaborate (Langdale, 2008; **Figure 1.1**).

Figure 1.1 Phylogenetic tree showing the relationships between the major groups of land plants and the sister group to land plants, the Charophyte algae. Aquatic lineages of Charophyte algae are highlighted in blue, whereas all land plant lineages are highlighted in green. Land plant lineages have been subdivided into the bryophytes and vascular plants. Red circles mark the transition to land and the acquisition of various morphological adaptations throughout land plant evolution. Approximate timings of divergence (millions of years ago) are indicated as appropriate.


Flowers, fruit, pollen; loss of flagella
Seeds
150
360
Leaves, roots
380
Vascular tissue
420
Microphylls, roots
390
Move to land
450
angiosperms
gymnosperms
ferns
lycophytes
Vascular plants
hornworts
mosses
liverworts
Bryophytes
Charophyceae
Coleochaetophyceae
Zygnemophyceae
Klebsormidiophyceae
Chlorokybophyceae
Charophytes

1.4 *Physcomitrella patens* as a model organism

The bryophyte *Physcomitrella patens* is an ephemeral moss that is found predominantly in temperate zones. *Physcomitrella* recently became the first non-flowering land plant to have its genome sequenced, making it an ideal model organism in which to answer fundamental questions about plant evolution at the molecular level. The genome consists of 460Mb distributed across 27 chromosomes and the sequencing project involved the compilation of a large number of scaffold sequences via a whole-genome shotgun strategy. The genome is believed to encode over 35,000 predicted gene models, compared to approximately 25,000 in the *Arabidopsis thaliana* genome (125Mb) and over 50,000 in *Oryza sativa* (466Mb; The *Arabidopsis* Genome Initiative; Yu *et al*., 2002; Rensing *et al*., 2008). Kamisugi *et al*. (2008) have generated a partial linkage map for *Physcomitrella*, which will assist in forward genetic approaches to identify genes based on their mutagenic phenotypes (Kamisugi *et al*., 2008).

Physcomitrella was the first land plant shown to carry out somatic homologous recombination at a similar efficiency to the yeast, *Saccharomyces cerevisiae*. It is therefore an extremely useful model organism in which to study gene function by gene targeting. *Physcomitrella* is placed at an ideal evolutionary position in which to understand how plant physiology and development has evolved from simple unicellular aquatic organisms to more complex modern day angiosperms (Schaefer and Zrÿd, 2001). Many important signalling pathways are conserved between the bryophytes and angiosperms. Understanding the origins of these pathways in a simple, early-evolving moss will highlight their ancestral functions (Prigge and Bezanilla, 2010).

The *Physcomitrella* life cycle consists of a dominant haploid dimorphic gametophyte and a reduced sporophyte. Unlike angiosperms, *Physcomitrella* uses free-swimming flagellated sperm for fertilisation, a process entirely dependent on water (Prigge and Bezanilla, 2010).

Throughout the evolution of land plants, a shift of the alternation of generations occurred to a dominant diploid sporophyte and a reduced haploid gametophyte (Gerrienne and Gonez, 2011). The major advantage of *Physcomitrella* possessing a dominant haploid gametophyte means that backcrossings are not required in order to generate homozygous mutants, which saves both time and labour (Reski and Frank, 2005). However, a major disadvantage is that knockouts causing embryo lethality cannot be recovered, but this can be resolved by using the technique of RNAi (Prigge and Bezanilla, 2010). RNAi constructs can be designed to reduce levels of mRNA transcripts, resulting in knockdown rather than a knockout of specific genes or even whole gene families simultaneously. Double stranded RNA stem-loop precursors are processed by DICER proteins into single stranded small interfering RNAs (siRNA) or microRNAs (miRNA). These are then recruited into an RNA induced silencing complex (RISC), where they bind to complementary mRNA sequences and promote their direct cleavage (Bezanilla *et al*., 2003; Khraiwesh *et al*., 2008).

The life cycle of *Physcomitrella* commences with either spore germination or artificially, with the regeneration of protoplasts; both processes involve polar outgrowth and the production of filamentous networks of protonemata. Protonemata are composed of two types of filament; chloronemata and caulonemata. The first filament types to emerge are the chloronemata and these are extremely chloroplast-rich, containing 50-100 chloroplasts per cell. The cell plates that form between dividing chloronemal cells are always transverse to the long axis of the cell (Menand *et al*., 2007). Caulonemata arise from division of chloronemal apical cells in an auxin-dependent manner. Caulonemal cells grow faster and typically possess fewer chloroplasts than chloronemal cells. The cell plates that form between dividing caulonemal cells are always oblique to the long axis of the cell (Prigge and Bezanilla, 2010). Caulonemata side branch initials differentiate into buds, which develop into gametophores. A

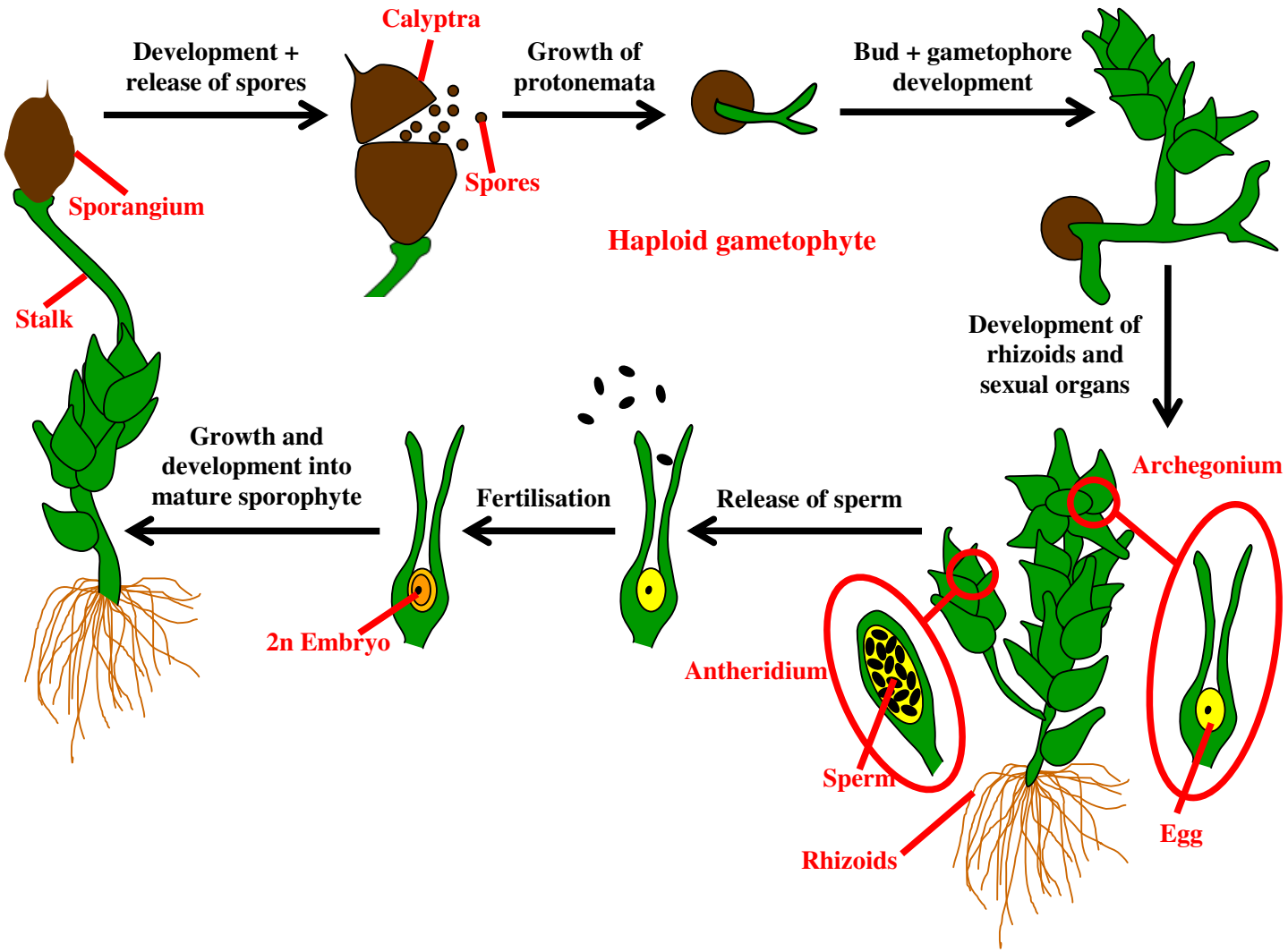
diagonal cell division produces an apical cell that gives rise to a leafy gametophore and a basal cell that gives rise to rhizoids. The bryophytes do not have what are defined as 'true leaves', since they lack vascular tissues and stomata. Gametophores initially consist of only juvenile leaves, which are unistratose and midribless. Subsequently, adult leaves form and these possess a single midrib, a 6-cell cylinder "vein" spanning the entire length of the leaf. Adult gametophores are characterised by the presence of adult leaves or are otherwise juvenile gametophores. Gametophores continue to differentiate and eventually produce the reproductive organs, male antheridia and female archegonia. Since *Physcomitrella* is monoecious, both male and female gametes are produced on the same plant. In the presence of water, antheridia form antherozoids, which subsequently swim towards archegonia, which contain egg cells. After fertilisation takes place, a zygote forms within the sporophyte generation, which then grows and develops into a mature diploid sporophyte, composed of a short seta topped with a spore capsule. Following meiosis, several thousand spores are produced within the spore capsule and dispersed into the surrounding environment. Subsequent germination of these spores recommences the life cycle, which is normally completed within three months (**Figure 1.2**; Schaefer and Zrÿd, 2001; Prigge and Bezanilla, 2010).

Physcomitrella vegetative tissues can be continuously propagated *in vitro*, which requires very little laboratory space. Cells are totipotent, and accordingly can regenerate from protoplasts or after mechanical disruption of tissues. Transformation of *Physcomitrella* is carried out by either particle bombardment of protonemata or by Polyethylene glycol (PEG)-mediated transformation of protoplasts, which yields stable transformants within 4-6 weeks (Schaefer *et al.*, 1991; Prigge and Bezanilla, 2010). Compared with later-evolving angiosperms, *Physcomitrella* is composed of relatively few tissue types and consequently has

an extremely simple developmental pattern. This makes phenotypic observations via simple live microscopy extremely straightforward. Furthermore, studies of protoplast regeneration and spore germination provide valuable insights into the control of pattern formation and cell polarity (Cove *et al*., 2006). Like angiosperms, mosses respond to light, gravity, environmental stresses (salt, mechanical, cold and drought) and the plant hormones auxin, cytokinin and abscisic acid (ABA; Cho *et al*., 2007).

Figure 1.2 The life cycle of *Physcomitrella patens*.

The dominant haploid gametophyte phase of the life cycle begins with the release and subsequent germination of spores. Polar outgrowths then form protonemata, composed of chloronemal and caulonemal filaments. Simple buds that form from filament subapical cells differentiate into leafy gametophores and these later form rhizoids. Gametophores bear male and female reproductive organs, the antheridia and archegonia respectively. Zygotes form following fertilisation, forming the diploid sporophyte phase of the life cycle. Zygotes grow and develop into mature sporophytes bearing sporangia. These subsequently undergo a series of meiotic divisions to produce several thousand spores, contained within a spore capsule. Their subsequent release and germination recommences the dominant haploid generation of the life cycle.


Development + release of spores
Calyptra
Spores
Growth of protonemata
Bud + gametophore development
Sporangium
Haploid gametophyte
Stalk
Development of rhizoids and sexual organs
Archegonium
Growth and development into mature sporophyte
Fertilisation
Release of sperm
2n Embryo
Antheridium
Sperm
Egg
Rhizoids

1.5 *Selaginella moellendorffii* as a model organism

There are three extant families of lycophytes that evolved during the Silurian/early Devonian period; the *lycopodiaceae* (club mosses), *isoeteaceae* (quillworts) and *selaginellaceae* (spike mosses). The model lycophyte species *Selaginella moellendorffii* is a perennial and deciduous plant, native to eastern Asia. *Selaginella* shares traits with both nonseed and seed plants and therefore represents an important link between the earliest land plants and more recently evolving angiosperm species. *Selaginella* has naked sporangia like those of early land plants and has bifurcating shoot and root meristems that form Y-shaped junctions instead of extensive branching networks like those observed in angiosperms. Like *Physcomitrella* and other non-seed plants, *Selaginella* is dependent on water for the dispersal of flagellated sperm. Interestingly though, the gametophyte generation is not buried in maternal tissues but is instead independent of the sporophyte. It is therefore a model organism in which to understand how the alternation of generations is regulated (Banks *et al*., 2011). Like angiosperms, the lycophytes possess vascular tissues with lignified cells and have a dominant sporophyte generation. *Selaginella* is therefore an ideal model organism in which to study early development of processes specific to vascular plants (Banks, 2009). *Selaginella* produces two different types of spores and is thus termed heterosporous. The production of spores results from a switch from vegetative development to reproductive development and the subsequent growth of strobili, sporangia-bearing structures along leaf-like structures. Strobili with microsporangia produce haploid microspores whereas strobili with megasporangia produce haploid megaspores.

The *Selaginella* genome was recently sequenced using a whole-genome shotgun strategy. It is the smallest land plant genome to be sequenced so far and accordingly has smaller genes, due to smaller intron sizes. The genome assembly currently contains two sequenced 105MBp

haplotypes that are approximately 98.5% identical. The *Selaginella* genome encodes approximately 22285 protein-coding genes and roughly one third are supported by EST evidence. Interestingly and unlike any other sequenced land plant genome, *Selaginella* shows no evidence of any whole-genome duplication events. Banks *et al*. (2011) compared the proteomes of *Chlamydomonas*, *Physcomitrella*, *Selaginella* and *Arabidopsis* in order to identify which genes were acquired and lost throughout land plant evolution. Over 3000 gene families are conserved in all land plant lineages and represent the minimum number of genes present in the common ancestor of all four organisms. Functional studies using *Selaginella* will therefore help to determine which genes were essential for the successful adaptation to land by vascular land plants.

1.6 *Arabidopsis thaliana* as a model organism

Arabidopsis thaliana is a dicotyledonous angiosperm from the *Brassicaceae* family. In 2000, *Arabidopsis* became the first plant genome to be sequenced and encodes approximately 25,000 protein-encoding genes (The *Arabidopsis* genome initiative, 2000). The use of reverse genetics techniques in *Arabidopsis* has brought about huge advances in the understanding of plant molecular genetics, development, physiology and development. Previously, forward genetics approaches began by identifying mutant phenotypes and then determining the genotype of a specific plant. Nowadays it is possible to use reverse genetics to knockout individual genes of interest in order to determine resulting mutant phenotypes and consequently gene functions. The generation of gene knockouts using large non-coding fragments of DNA known as transposable elements or T-DNA is possible in *Arabidopsis* (Krysan *et al*., 1999). Consequently, there is now a huge database of T-DNA insertion lines that are available for use in research. It is possible to obtain individual knockout lines for

specific genes of interest from the *Arabidopsis* Biological Resource Centre (ABRC) or the European *Arabidopsis* Stock Centre (NASC; Alonso *et al*., 2003). Many *Arabidopsis* genes have been fully characterised since genome-wise insertional mutagenesis was carried out and T-DNA insertion lines made available to the public. Consequently, *Arabidopsis* is now a model angiosperm for comparisons with earlier-evolving organisms such as the bryophytes and lycophytes, in so called 'evo-devo' studies.

Arabidopsis has a short generation time and often takes only six weeks from germination to producing large numbers of mature seed. It is not expensive to grow *Arabidopsis* since it grows readily in small pots in a greenhouse or on minimal medium within a growth room. Furthermore, *Arabidopsis* is self-fertile so it is therefore possible to generate homozygous plants by selfing heterozygous individuals. Since *Arabidopsis* is so small, it is possible to visualise whole seedlings using simple microscopy (Meyerowitz, 1987; Meinke *et al*., 1998). Using *Agrobacterium tumefaciens*, it is possible to transform genetic constructs into *Arabidopsis* using the floral dip method (Clough and Bent, 1998).

1.7 The evolution of roots

Plants developed sophisticated rooting structures that enabled them to transition from moist to progressively drier habitats. The evolution of heavily branched rooting systems allowed plants to explore larger areas of soil for essential water and nutrients. Plant roots are positively gravitropic and provide support and anchorage to soil substrata, enhancing their survival in unfavourable environmental conditions (Nibau *et al*., 2008).

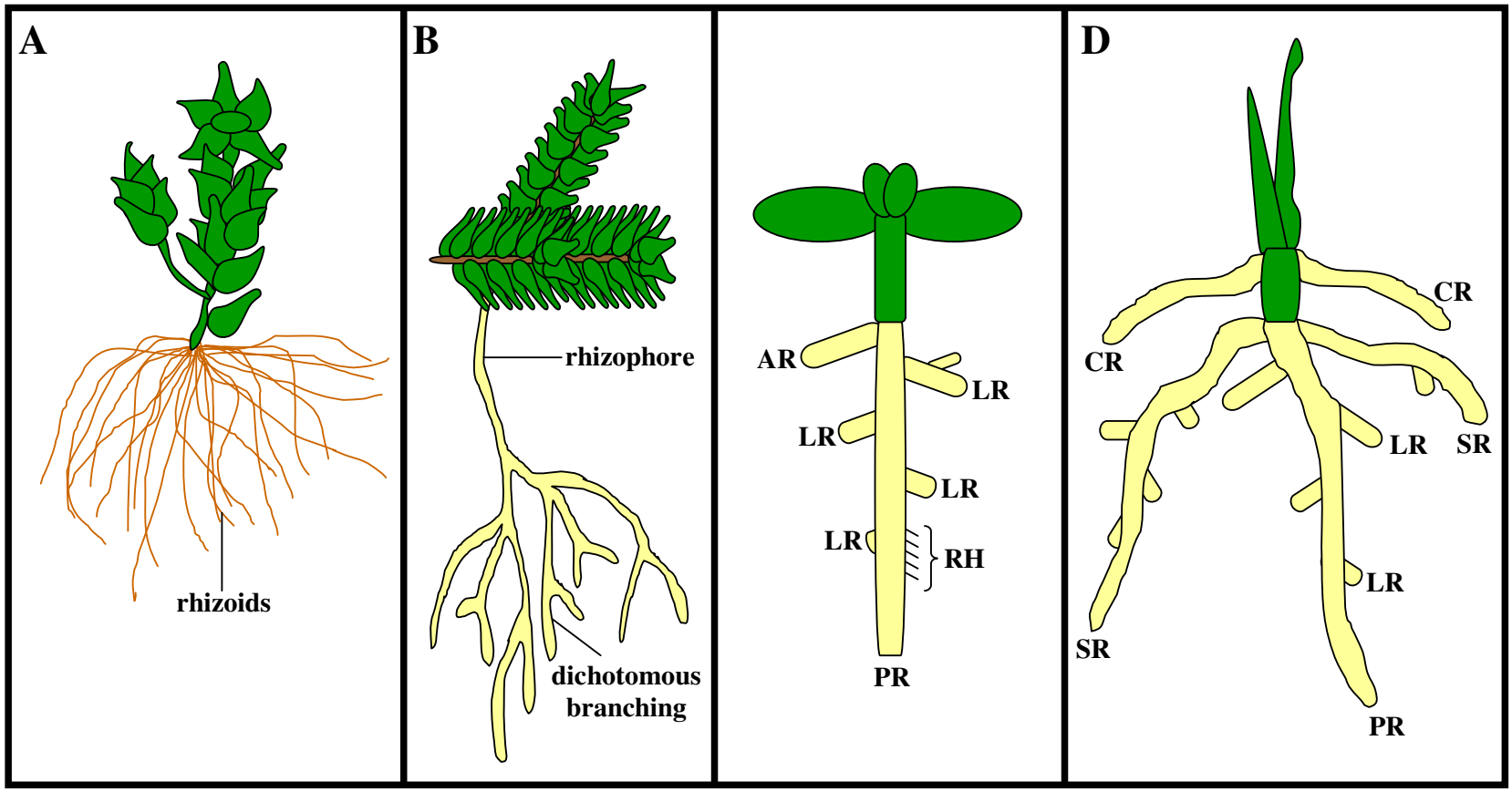
1.7.1 Rhizoids

Bryophytes such as *Physcomitrella* do not possess roots, but instead produce structures known as rhizoids. Rhizoids, like roots, allow plants to anchor to soil and rock surfaces, so that they can acquire nutrients from the environment. Caulonemata are also believed to play roles in nutrient acquisition, but published literature in this area is limited (Menand *et al*., 2007). *Physcomitrella* rhizoids grow by tip growth processes and develop from the leafy structures (gametophores; **Figure 1.3A**). Basal rhizoids arise from gametophyte tissues beneath juvenile leaves. Midstem rhizoids are formed beneath adult leaves of gametophores composed of greater than eight adult leaves (Sakakibara *et al*., 2003). Several interesting studies have demonstrated the genetic and hormonal control of both rhizoid and caulonemata development (Ashton *et al*., 1979; Sakakibara *et al*., 2003).

1.7.2 Rhizophores

Lycophytes such as *Selaginella moellendorffii* possess rhizophore structures, which emerge from so-called angle meristems. Rhizophores are leafless cylindrical structures that lack root hairs and a root cap, characteristic features of higher plant roots. Like roots, rhizophores are positively gravitropic and grow downwards to locate water and nutrients. Eventually, rhizophores differentiate and give rise to more clearly defined rooting structures that possess both root hairs and a root cap. These then bifurcate at the root tip but do not form lateral roots, a characteristic of higher plants (Banks, 2008; **Figure 1.3B**).

Figure 1.3 The evolution of root development (adapted from Nibau *et al*., 2008). Diagrams showing rooting structures of *Physcomitrella patens* (A), *Selaginella moellendorffii* (B), *Arabidopsis thaliana* (C) and *Oryza sativa* (rice, D). **A)** *Physcomitrella patens* possesses basal filamentous rhizoids as well as midstem rhizoids (not shown). **B)** *Selaginella moellendorffii* produces rhizophores that differentiate into multicellular rooting structures that branch dichotomously and possess root hairs and a root cap. **C)** *Arabidopsis thaliana* produces a primary root (PR), which undergoes higher order branching to form lateral roots (LRs). They possess root hairs (RH) and adventitious roots (AR) form at shoot-root junctions, arising from the stem. **D)** Rice produces a primary root (PR), crown roots (CR) and seminal roots (SR) and have lateral roots (LR).


A
rhizoids
B
rhizophore
dichotomous
branching
AR
LR
LR
LR
LR
RH
PR
D
CR
CR
LR
SR
LR
SR
PR

1.7.3 'True roots'

Gymnosperm and angiosperm roots initially consist of a primary root and as seedlings develop, the primary root gives rise to new branches known as lateral roots. Lateral roots initiate in specialised meristematic tissues within the primary root, known as the pericycle. The origin of lateral roots differs slightly between the dicots and the monocots, such as the important cereal crops. Dicot lateral roots form from the xylem pole pericycle cells, which are adjacent to the xylem. Monocot lateral roots form from the endodermis and the phloem pole pericycle cells, which are adjacent to the phloem. Pericycle founder cells undergo a number of asymmetric cell divisions to give rise to lateral root primordia. Lateral root meristems are subsequently formed and these give rise to lateral outgrowths of the primary root. The initiation of lateral root primordia is promoted by auxin and repressed by ABA.

Dicots such as *Arabidopsis* have a primary root that undergoes higher order branching to form lateral roots. These in turn are able to undergo higher order branching to form a heavily branched rooting system. Epidermal cells within the primary root give rise to root hairs, which are composed of a single cell and grow by tip growth like pollen tubes and fungal hyphae. Dicot roots also have adventitious roots at shoot-root junctions and these usually arise from the stem (**Figure 1.3C**).

Cereal roots are morphologically more complex than those found in dicots. In addition to a primary root, they also produce seminal roots and crown roots. Seminal roots originate postembryonically at the top of the primary root and these undergo higher-order branching. Crown roots originate from the stem and also undergo higher-order branching (**Figure 1.3D**; Nibau *et al*., 2008).

1.7.4 Root hairs

Root hairs (RH) are thin, tubular outgrowths of the epidermis of angiosperm roots that acquire nutrients and interact with soil-borne symbionts (Sieberer *et al*., 2005). RH architecture can be altered by a variety of biological factors such as auxin treatment and phosphate starvation, which promote RH formation (Savage, 2007). In *Arabidopsis*, RHs derive from root epidermal tissue that is arranged into alternate rows of atrichoblasts (non-hair cells) and trichoblasts (hair cells; Carol and Dolan, 2002; Menand *et al*., 2007). Fates of both hair and non-hair cells are dependent on their position within the epidermis; hair cells form from cells in contact with two cortical cells whereas non-hair cells form from cells in contact with single cortical cells (Savage, 2007). Like rhizoids, RH cells grow by apical tip growth processes, regulated by calcium ions and also microtubules, which lie parallel to axes of growing hair cells. Cell walls of RHs are redistributed during tip growth, to allow vesicle fusion to plasma membranes and new cell wall material to be deposited. Essential 'tip growth machinery' assembles at apices of elongating RHs and this includes the endoplasmic reticulum, cytoskeletal components, endocytotic vesicles, mitochondria and nuclei (Sieberer *et al*., 2005).

1.7.5 Evolution of genes with rooting functions

Menand *et al*. investigated the roles of *Physcomitrella* homologues of two *Arabidopsis* basic-helix-loop-helix (bHLH) transcription factors, ROOT HAIR DEFECTIVE 6 (AtRHD6) and RHD SIX-LIKE1 (AtRSL1). When these genes are knocked out in *Arabidopsis*, seedlings do not produce RHs (Masucci and Schiefelbein, 1994). Seven RHD SIX-LIKE genes (PpRSL1-7) were identified in the *Physcomitrella* genome; PpRSL1 and PpRSL2 were selected for closer examination since they shared greatest homology to both *Arabidopsis*

genes. *Pprsl1* and *Pprsl2* double mutants produced small, dark green protonemal colonies, consisting of only chloronemata and this occurred even in the presence of auxin (Menand *et al*., 2007; Jang and Dolan, 2011). This absence of caulonemata caused gametophores to develop from chloronemata instead of from caulonemata and these possessed very short rhizoids. Given that these chloronemal filaments developed normally, PpRSL1 and PpRSL2 do not play roles in the general regulation of tip growth, but instead control caulonemata and rhizoid development, tissues which both possess rooting functions. Interestingly, complementation studies carried out by introducing PpRSL1 into an *Atrhd6-3* mutant background restored normal RH development, demonstrating that their functions have been conserved throughout land plant evolution. Liam Dolan's group have therefore demonstrated that *Physcomitrella* rhizoids are equivalent to *Arabidopsis* RHs (Menand *et al*., 2007). Subsequent work showed that PpRSL1 and PpRSL2 are sufficient for the development of both basal and mid-stem rhizoids. In addition, constitutive overexpression of PpRSL1 and PpRSL2 promotes rhizoid formation from developing gametophores (Jang and Dolan, 2011; Jang *et al*., 2011). Intriguingly, RSL genes do not promote growth and development of tip-growing pollen tubes in *Arabidopsis* and have been recruited specifically for RH development. PpRSL genes only appear to control growth and development of caulonemata and rhizoids and these, like RHs, also play key roles in nutrient assimilation. RSL genes have been identified in all land plant genomes sequenced to date. It is therefore reasonable to assume that RSL genes were co-opted to control development of filamentous RHs in roots after they evolved from the bryophyte lineage (Jang *et al*., 2011).

1.8 Stress responses in plants and the conservation of hormone signalling

Early plants were subjected to harsh conditions when they made the transition from water to land and accordingly, they have evolved a number of stress response pathways to ensure their survival. Intrinsic growth regulators or hormones are known to play critical roles in plant growth and development and adapting to changing environmental conditions. There are eight classes of plant hormones: Auxin, Brassinosteroids, Cytokinins, Ethylene, Gibberellins, Jasmonates, Strigolactones and abscisic acid (ABA; Depuydt and Hardtke, 2011). Auxin, cytokinin, ABA and ethylene have been detected in mosses and liverworts, the most basal land plant lineages (Anterola and Shanle, 2008). Some of the roles of auxin, cytokinin, ABA and ethylene in *Physcomitrella* development have been documented in recent years. However, it is not clear whether functional brassinosteroid, gibberellin and jasmonate signalling pathways exist in early land plants. The following sections summarise our current understanding of hormone signalling pathways in plants.

1.8.1 Auxin

Auxin is synthesised by a wide number of species, including multicellular Charophycean algae and all land plant lineages (Ross and Reid, 2010). Auxin plays pivotal roles in fruit and root development. The auxin response pathway is mediated by the TRANSPORT INHIBITOR RESPONSE 1 (TIR1) receptor, F-box protein and component of a Cullin-Ring E3 ubiquitin ligase (Kepinski and Leyser, 2005; Dharmasiri *et al*., 2005). When auxin is bound to the TIR1 receptor, it promotes the degradation of Aux/IAA proteins that prevent transcription factors known as auxin response factors (ARF), from activating the transcription of auxin-responsive genes (Guilfoyle and Hagen, 2007). Homologues of TIR1, AUX/IAA and ARF appear to be absent from the unicellular green alga *Chlamydomonas*

reinhardtii. However, orthologs of TIR1, AUX/IAA and ARF are present in the bryophytes and all other later-evolving land plant lineages (Rensing *et al*., 2008). It is therefore believed that auxin signalling was present in a common ancestor of bryophytes and vascular plants, but evolved after algae diverged from that ancestral lineage (DeSmet *et al*., 2010). Polar auxin transport is mediated by polar localised transmembrane proteins known as PINs that rapidly distribute auxin when required (Leyser, 2005). Polar auxin transport occurs in basal land plants; the liverwort *Marchantia polymorpha* and in the moss *Funaria hygrometrica* (Cooke *et al*., 2002). However, polar auxin transport appears to be restricted to the *Physcomitrella* sporophyte and mediated by a PIN protein that localises to the ER membrane and not to the plasma membrane (Fujita *et al*., 2008). *Physcomitrella* has two distinct types of rhizoids; basal rhizoids and mid-stem rhizoids. Auxin promotes the development of both basal rhizoids, for which auxin alone is sufficient and mid-stem rhizoids, requiring auxin in addition to an unknown co-factor (Sakakibara *et al*., 2003). Auxin also promotes caulonema differentiation and subsequently controls the caulonema/chloronema ratio by regulating the expression of ROOT-HAIR DEFECTIVE SIX-LIKE1 gene homologues (Eklund *et al*., 2010; Jang and Dolan, 2011). In liverworts, auxin promotes rhizoid formation and controls the initiation of gametangia (Kumra and Chopra, 1984).

1.8.2 Brassinosteroids

Brassinosteroids have been identified in a large number of plant species, ranging from the bryophytes to the more recent angiosperms (Ross and Reid, 2010). In angiosperms, brassinosteroids appear to play important roles in stem elongation, xylem differentiation and fruit development. The brassinosteroid pathway is highly conserved between the monocots and dicots (Bai *et al*., 2007; Ross and Reid, 2010). BRASSINOSTEROID INSENSITIVE 1

(BRI1) is a putative transmembrane brassinosteroid receptor that has kinase activity when bound to brassinosteroid (Clouse *et al*., 1996; He *et al*., 2000; Li and Chory, 1997). The binding of brassinosteroid to BRI1 induces its dimerisation with the co-receptor BAK1 and a number of regulators, such as BIN2. BIN2 negatively regulates BRASSINAZOLE RESISTANT 1 (BZR1) and bri1-EMS-SUPPRESSOR1 (BES1), two redundant transcription factors that target genes involved in brassinosteroid biosynthesis (Ye *et al*., 2011). Brassinosteroid signalling has been well characterised in *Arabidopsis* but little is known about brassinosteroid signalling in early-evolving land plants. BRI1 receptor homologues have been identified in genomes of algae, *Physcomitrella* and all other land plants (Depuydt and Hardtke, 2011). However, it has recently been reported that fully functional brassinosteroid signalling components only exist in vascular plants and it is unknown whether exogenous application of brassinosteroids induces physiological changes in basal land plants (Prigge and Bezanilla, 2010).

1.8.3 Cytokinin

Cytokinin has been detected in algae and all land plant lineages (Stirk *et al*., 2003). Cytokinin is perceived by CRE1/AHK4 and AHK3, plasma membrane-localised His kinase receptors that contain cyclases/histidine kinases-associated sensory extracellular (CHASE) domains (Spichal *et al*., 2004). Receptor autophosphorylation promotes phosphorylation of ARR1 transcription factors and these activate transcription of negative regulators of cytokinin signalling. No CHASE domain-containing proteins have been identified in algal genomes but at least two have been identified in *Physcomitrella*, the most basal land plant known to possess all components of a functional cytokinin signalling pathway. In *Physcomitrella* and *Funaria hygrometrica*, similarly to later-evolving land plant lineages, cytokinin promotes bud

development when plants switch from filamentous growth to meristematic growth (Schumaker and Dietrich, 1998). Cytokinin has also been detected in liverworts where it enhances the production of archegonia and reduces the production of antheridia in *Riccia gangetica* (Kumra and Chopra, 1984). Pils and Heyl (2009) hypothesise that cytokinin was required by plants to successfully make the transition from water to land (Pils and Heyl, 2009).

1.8.4 Ethylene

Ethylene is a simple gaseous hormone that is found ubiquitously throughout the plant kingdom. Ethylene plays important roles in plant development, particularly in leaf and flower senescence, flower opening and fruit ripening (Zhu and Guo, 2008). Ethylene is also involved in adaptive responses to stresses, most notably the submergence response. In deepwater varieties of rice, ethylene-regulated genes SNORKEL1 and SNORKEL2 stimulate plant elongation in response to being submerged and this enhances their chances of survival during periods of flooding (Nagai *et al*., 2010). The ethylene receptor, Ethylene Resistant 1 (ETR1) was the first phytohormone receptor to be identified. ETR1 contains three transmembrane domains and an ethylene binding domain. ETR1 has four close relatives (ETR2, EIN4, ERS1 and ERS2), which may function redundantly (Zhu and Guo, 2008). When ethylene is absent, ETR1 binds to a Raf-like protein kinase (CTR1) and maintains it in an active state. This results in the repression of downstream signalling components, such as EIN2, EIN3 and EIN3-like (EIL) transcription factors. The EIN3-binding F-box protein (EBF) negatively regulates ethylene signalling by binding to EIN3/EIL proteins and targeting them for ubiquitination and subsequent degradation via the 26S proteasome. In the presence of ethylene, the receptor-CTR1 complex is inactivated and this subsequently allows EIN3 and

EILs to accumulate in the nucleus, where they activate transcription of a large number of ethylene-responsive genes (Stepanova and Alonso, 2009; Yang *et al*., 2010).

Endogenous ethylene has been detected in the liverwort, *Pellia epiphylla* (Thomas *et al*., 1983). In addition, homologues of ethylene signalling components are encoded by the *Physcomitrella* genome and have been implicated in the moss submergence response (Yasumura *et al*., unpublished data).

1.8.5 Gibberellins

Gibberellins (GA) are found in fungi, bacteria and throughout vascular land plants. A soluble predominantly nuclear-localised gibberellin receptor called GIBBERELLIN INSENSITIVE DWARF 1 (GID1) has been identified (Ueguchi-Tanaka *et al*., 2005). GID1 binds to bioactive gibberellins and this binding event induces a conformational change. This conformational change promotes interactions between GID1 and the DELLA proteins, GIBBERELLIC ACID INSENSITIVE 1 (GA1) and REPRESSOR OF ga1-3 (RGA), which function as transcriptional repressors. The F-box protein SLEEPY1 targets DELLA proteins for ubiquitination and subsequent degradation via the 26S proteasome (Strada *et al*., 2004; Ueguchi-Tanaka *et al*., 2005).

Although described extensively in vascular land plants, there is little evidence to support the presence of bioactive gibberellins in either algae or the bryophytes (MacMillan, 2001; Ross and Reid, 2010). *Physcomitrella* and the lycophyte species *Selaginella kraussiana* do not exhibit growth responses to exogenously applied gibberellins (Yasumura *et al*., 2007). *ent*-kaurene synthase is inhibited by AMO-1618 in angiosperms, preventing production of the gibberellin precursor *ent*-kaurene and consequently gibberellin signalling. AMO-1618 also inhibits *ent*-kaurene biosynthesis in *Physcomitrella* and consequently inhibits

spore germination even in the absence of a functional gibberellin signalling pathway (Anterola *et al*., 2008). *Physcomitrella* mutants impaired in their ability to synthesise *ent*-kaurene had defects in the differentiation of protonemal tissues. This suggests that gibberellin-like responses are present in all land plants (Hayashi *et al*., 2010).

Interestingly, all of the components of a gibberellin signalling pathway are encoded by the *Physcomitrella* genome. Unlike vascular plants, bryophytes do not appear to have a functional GID1-DELLA-GA response (Yasumura *et al*., 2007; Hirano *et al*., 2007).

Yeast 2-hybrid approaches have been used to determine whether GID1-like proteins and DELLA partners from early land plants are able to interact. *Selaginella kraussiana* GID1 (SkGID1) and DELLA (SkDELLA) are able to interact and their interaction is enhanced by GA_3 like it is in *Arabidopsis*. Although they are present, *Physcomitrella* GID1 (PpGLP1) and DELLA (PpDELLAa) are not able to interact even in the presence of GA_3. Interestingly, PpGLP1 is able to interact with SkDELLA. However, PpDELLAa cannot interact with neither SkGID1 nor *Arabidopsis* GID1c (AtGID1c). DELLA proteins probably evolved an affinity for GID1 proteins after the bryophytes and before the lycophytes (Yasumura *et al*., 2007). Similarly to *Selaginella kraussiana*, GID1 proteins (SmGID1a and 1b) can interact with DELLA proteins (SmDELLA1 and 2) in *Selaginella moellendorffii* and their interaction is enhanced by GA. Remarkably, SmGID1a, 1b and 2a are able to rescue dwarf rice *gid1* and *gid2* mutants. In wild type rice plants, SmDELLA1 overexpression results in semi-dwarf phenotypes since accumulated DELLA proteins enhance the degradation of GID1 proteins. Strangely, overexpression of SmDELLA2 is unable to induce a semi-dwarf phenotype. Similarly, PpDELLA cannot carry out a cross-species function when overexpressed in wild type rice plants and this supports findings by Yasumura *et al* (Hirano *et al*., 2007).

It appears that plants acquired a fully functional GID1-DELLA-GA response after the divergence of the bryophytes from the lineage leading to the lycophytes (Yasumura *et al*., 2007). However, gibberellins appear to have no physiological effects on the lycophytes and it is argued that the growth inhibitory effects of GA may have arisen after the divergence from lycophytes (Depuydt and Hardtke, 2011).

1.8.6 Jasmonates

Jasmonates are fatty acid-derived hormones that share structural and functional properties with animal prostaglandins. Jasmonates play important roles in protecting plants against attack by herbivores and pathogens but also against exposure to ozone, UV light and other abiotic stresses. In angiosperms, jasmonates control important developmental processes, which include fertility, root elongation and fruit ripening (Wastermack and Kombrink, 2009). The F-box protein CORONATINE INSENSITIVE 1 (COI1) is critical for all jasmonate responses and targets JAsmonate ZIM-domain (JAZ) repressor proteins for Ubiquitin/26S proteasome-mediated degradation. The transcription factor MYC2 is subsequently released from repression and is able to enter the nucleus where it activates transcription of jasmonate-regulated genes. COI1 binds directly to an amino acid conjugate of JA (JA-Ile) and therefore functions as a jasmonate receptor (Fonseca *et al*., 2009). Unlike later-evolving plants, *Physcomitrella* is unable to synthesise jasmonates. However, it can convert octadecatrienoic acid into a precursor of jasmonic acid using the enzyme Allene Oxide Cyclise (AOC; Stumpe *et al*., 2010). *Physcomitrella aoc* knockout mutants have fertility defects, a characteristic feature of higher plant JA-deficient mutants (Feys *et al*., 2004; Devoto *et al*., 2005; Stumpe *et al*., 2010). This suggests that JA precursors, not

jasmonates themselves, play roles in reproductive development and that this function is evolutionarily conserved in plants (Stumpe *et al*., 2010).

1.8.7 Strigolactones

Strigolactones are the least well characterised group of plant hormones. Strigolactones were originally identified as germination stimulants in root exudates of parasitic plants. They function as recognition signals that enable symbiotic arbuscular mycorrhizal (AM) fungi to successfully colonise host plant roots. However, they are also found in non-hosts of AM-fungi, such as *Arabidopsis* and are widely distributed throughout the plant kingdom (Xi *et al*., 2010). Strigolactones have been shown to play critical roles in regulating shoot branching in a number of different plants, such as *Arabidopsis*, pea, rice and *Chrysanthemum*. Strigolactone-deficient plants are significantly more branched than their wild type counterparts (Gomez-Roldan *et al*., 2008; Umehara *et al*., 2008). Strigolactones have also been implicated in root development (Koltai, 2011). In *Arabidopsis* and pea, CAROTENOID CLEAVAGE DIOXYGENASES (CCDs), CCD7 and CCD8, catalyse strigolactone biosynthesis and repress branching (Booker *et al*., 2004; Johnson *et al*., 2006). *Physcomitrella* homologues of CCD8 are involved in strigolactone biosynthesis in an evolutionarily ancient organism. PpCCD8 regulates filament branching and colony extension, suggesting that strigolactones have a conserved branching function (Proust *et al*., 2011).

There is currently speculation that the F-box protein MORE AXILLARY GROWTH2 (MAX2; also known as ORESARA9 or ORE9) could be a contender for the strigolactone receptor. MAX2 shares structural characteristics with both TIR1 and COI1 and all contain an F-box domain and a series of leucine-rich repeats (LRR). It also functions as a *bona fide* F-box protein since it forms an SCF complex with ASK1 and AtCul1 *in planta* (Stirnberg *et al*.,

2007). *max2* mutant plants resemble those of strigolactone-deficient plants and this adds further support to this theory (Nelson *et al*., 2011) .

1.9 ABA

1.9.1 ABA signalling: A snapshot

ABA is a terpenoid phytohormone that is found extensively throughout all kingdoms of life except for the archea (Hauser *et al*., 2011). ABA signalling comprises three layers: (i) ABA biosynthesis, catabolism and transport via ATP-dependent ABA transporters; (ii) perception and signal transduction through SNF1-related protein kinases (SnRK2), ABA receptors (PYR/PYL/RCARs) and group A protein phosphatase 2Cs (PP2C); (iii) ABA signal response and modulation (involving ABRE-binding transcription factors that modulate ABA-responsive gene expression). Although not identified in the alga *Chlamydomonas*, group A PP2C proteins, subclass III SnRK2s and ABF/AREB transcription factors are all found in the basal bryophyte, *Physcomitrella*, suggesting that the PP2C-SnRK2-mediated signal transduction mechanism is conserved from the bryophytes to the angiosperms and that it was important for the successful colonisation of land by plants (Hauser *et al*., 2011).

1.9.2 ABA: the receptor saga

The hunt for the *bona fide* ABA receptor has been full of controversy. In 2006, Razem *et al*. claimed that a nuclear-localised RNA-binding protein called FLOWERING CONTROL LOCUS A (FCA) was an ABA receptor (Razem *et al*., 2006). A rival group published work contradicting these findings and stated that FCA did not in fact bind ABA at all (Risk *et al*., 2008). Consequently, the original publication was severely criticised and subsequently retracted. The second protein to be proposed as an ABA receptor was the

Arabidopsis chloroplast-localised Magnesium-cheletase subunit H (CHLH) and this was shown to bind ABA directly. RNAi to knockdown the expression of CHLH generated ABA insensitivity phenotypes, suggested that in fact CHLH was involved in ABA signalling in some way (Shen *et al*., 2006). Once again, a contradictory publication argued that CHLH was not a *bona fide* ABA receptor. Muller and Hansson (2009) carried out studies on the barley CHLH homologue, which could not bind ABA and had no ABA-related phenotypes (Muller and Hansson, 2009). Liu *et al*. (2007) proposed a third putative ABA receptor, the plasma membrane-localised seven-transmembrane G protein-coupled receptor, GCR2 (Liu *et al*., 2007). Subsequent studies by rival groups showed that GCR2 was not required for all ABA responses, showing that it was not an ABA receptor (Gao *et al*., 2007; Risk *et al*., 2009).

Pandey *et al*. (2009) proposed two GPCR-type G-proteins, GTG1 and GTG2, which are plasma membrane-localised and contain nine predicted transmembrane domains. They showed that GTG1 and GTG2 functioned redundantly and both were able to bind ABA. In addition, *gtg1gtg2* double knockout mutants were ABA insensitive compared to wild type. However, *gtg1gtg2* knockout phenotypes did not abolish ABA responses and they were unable to show the interaction with ABA *in vivo* (Pandey *et al*., 2009). Even though GCR2, GTG1/2 and CHLH respond to exogenous ABA, Muschietti and McCormick (2010) argue that ABA-binding assays for the three putative membrane-bound receptors should have been carried out in a membranous environment within an organism that lacks a functional ABA signalling pathway, such as yeast or *Xenopus* (Muschiette and McCormick, 2010).

The fifth and final group of ABA receptors was independently proposed by two different research groups (Ma *et al*., 2009; Park *et al*., 2009). Ma *et al*. (2009) used a yeast 2-hybrid approach to screen for proteins that interacted with the PP2Cs, ABI2 and HAB1. They identified an *Arabidopsis* protein with unknown function that they named Regulatory

Component of ABA Receptor 1 (RCAR1). In the presence of RCAR1, ABA blocked ABI2 phosphatase activity *in vitro*. When expressed in *Arabidopsis* protoplasts, RCAR1 enhanced the transcription of a number of genes involved in the ABA response. In addition, the ABA signalling pathway was stimulated even in the absence of exogenous ABA. Transgenic plants expressing high levels of RCAR1 exhibited impaired regulation of stomatal aperture and were hypersensitive to ABA, which affected both seed germination and root elongation. RCAR1 belongs to a 14-member superfamily (RCAR1-14), which all share structural similarities. Interestingly, the RCARs are all soluble, which contrasts to previous studies of putative membrane-localised receptors (Ma *et al*., 2009).

Park *et al*. (2009) used a chemical approach to identify mutants resistant to the synthetic seed germination inhibitor, Pyrabactin. Using this strategy, they identified PYRABACTIN RESISTANCE 1 (PYR1) and 13 PYRABACTIN RESISTANCE 1-LIKE genes (PYL1-13) that were all conserved and contained ligand-binding StAR-related lipid transfer (START) domains. Pyr and Pyl genes were responsive to ABA and expressed highly in both guard cells and seeds. They obtained both triple (pyr1, pyl2, pyl4) and quadruple (pyr1, pyl1, pyl2, pyl4) knockout mutants and these exhibited ABA insensitivity. Park *et al*. (2009) showed that as more PYR/PYL genes were disrupted, the severity of ABA insensitivity increased. However, they did not analyse phenotypes of PYR/PYL overexpression lines (Park *et al*., 2009).

Using PYR1 and ABI1 as bait in yeast 2-hybrid assays, Park *et al*. determined that PYR1 and ABI1 only interact in the presence of ABA. However, the nature of this interaction remained uncertain (Park *et al*., 2009). Santiago *et al*. (2009) determined that when co-expressed in transgenic *Arabidopsis* plants, PYL5 antagonised HAB1 function. Consequently, plants exhibited hypersensitive responses and were more drought tolerant than

wild type. Similarly to ABI1 and ABI2, the PYR/PYL proteins appeared to also inhibit HAB1 phosphatase activity (Santiago *et al*., 2009).

Ma *et al*. and Park *et al*. independently identified both RCAR1-14 and PYR1/PYL1-13 as the same soluble ABA receptor family. They are therefore commonly known as the soluble PYR/PYL/RCAR ABA receptor family.

All members of the receptor family utilise a gate-latch-lock mechanism of ligand sensing and signal transduction. The gate and latch reside in an open conformation when ABA is not bound. When ABA is bound, the ß-loops of both the gate and latch close and create a gate/latch interface to which PP2Cs are able to bind. Melcher *et al*. (2009) published the crystal structures of the PYL1 and PYL2 receptors in the absence of ABA, ABA-bound PYL2 and ABA-bound PYL2 in complex with the PP2C protein HAB1, demonstrating that the PP2Cs can indeed function as co-receptors that facilitate the interaction between receptors and ABA (Melcher *et al*., 2009). Miyazono *et al*. (2009) independently presented the crystal structures of ABA-bound PYL1 and ABA-bound PYL1 in complex with the PP2C protein ABI1. PYL1 binds to ABA via a specific ligand-binding site and consequently forms a hydrophobic pocket on the surface of a closed lid. ABA-bound PYL1 associates with ABI1 by using the hydrophobic pocket of the lid to conceal the ABI1 active site (Miyazono *et al*., 2009). Work from both groups confirms that the PYR/PYL/RCARs inhibit PP2C proteins in ABA signalling in an ABA-dependent manner (Melcher *et al*., 2009; Miyazono *et al*., 2009).

1.9.3 The SNF1-related protein kinases (SnRK2s)

Subclass III SnRK2 proteins comprise SnRK2.2, SnRK2.3 and SnRK2.6/OPEN STOMATA 1 (OST1). SnRK2s are activated by ABA and are therefore positive regulators of ABA signalling. Triple knockouts of all subclass III SnRK2 genes impair ABA responses

and mutants exhibit ABA insensitivity (Fujii *et al*., 2009; Umezawa *et al*., 2009). Yeast 2-hybrid analyses demonstrated that SnRK2.6/OST1 interacts with PP2C proteins within guard cells both in the presence and absence of ABA (Umezawa *et al*., 2009). Nishimura *et al.* (2010) showed that PYR/PYL/RCARs interact with SnRK2.3 in a co-immunoprecipitation experiment (Nishimura *et al*., 2010). The minimal components required for ABA signalling are therefore PYR/PYL/RCAR receptors, PP2C proteins, SnRK2s and the transcription factors ABF2/AREB1 (Fujii *et al*., 2009).

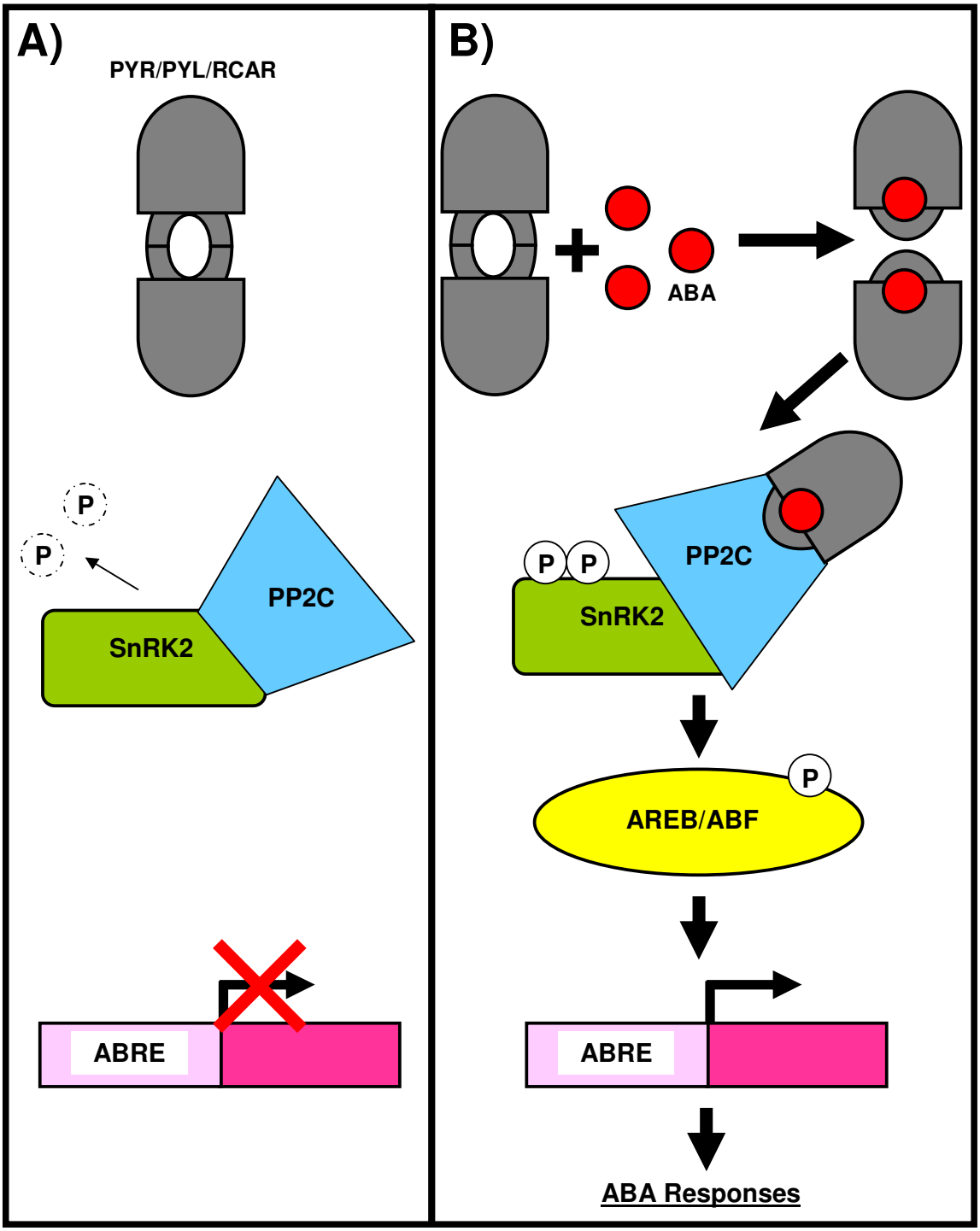
1.9.4 Signalling via the soluble ABA receptor family: The current understanding

Under normal conditions, PP2Cs bind to and dephosphorylate SnRK2s and consequently maintain SnRK2s in an inactive state. Exposure to biotic and abiotic stress causes ABA levels to accumulate. ABA binds to PYR/PYL/RCAR receptors and this receptor-ABA complex is able to interact with PP2Cs. This renders PP2Cs unable to dephosphorylate target proteins. The receptor-PP2C-SnRK2 complex consequently forms the ABA signalosome (Raghavendra *et al*., 2010). SnRK2s are released from PP2C-dependent regulation and directly phosphorylate key regulators of ABA signalling. In guard cells, SnRK2s specifically target the ion channels SLAC1 and KAT1. In the nucleus, SnRK2 specifically targets transcription factors for phosphorylation. Phosphorylated transcriptional regulators, such as ABA-responsive Element Binding Factor (ABF/AREBs) and bZIP-type transcriptional regulators (ABI5) bind to ABA-responsive cis-elements (ABRE) and activate transcription of ABA-responsive genes (**Figure 1.4**; Fujii *et al*., 2009; Umezawa *et al*., 2010).

Figure 1.4 Current model of the major ABA signalling pathway.

A) Under normal conditions, PP2Cs bind to and dephosphorylate SnRK2s. SnRK2s are therefore not able to phosphorylate AREB/ABF transcription factors in order to activate gene transcription.

B) ABA levels increase following exposure to biotic and abiotic stresses. ABA (red circles) binds to PYR/PYL/RCAR (grey) and the ABA-receptor complex binds to PP2Cs (blue). PP2C phosphatase activity is blocked, which restores SnRK2 (green) kinase activity. SnRK2s are released from PP2C-dependent regulation and directly phosphorylate AREB/ABF transcription factors (yellow), which activate the transcription of ABA-responsive genes (Adapted from Umezawa *et al*., 2010).


A)
PYR/PYL/RCAR
P
P
PP2C
SnRK2
ABRE
B)
ABA
P
P
PP2C
SnRK2
P
AREB/ABF
ABRE
ABA Responses

In order to determine whether ABA signalling via PP2C proteins is conserved, Komatsu *et al* (2009) generated transgenic *Physcomitrella* lines expressing the *Arabidopsis* abi1-1 gene. Transgenic lines were more susceptible to freezing and osmotic stresses compared to wild type. The expression of two *Physcomitrella* ABI1 homologues, PpABI1A and PpABI1B, was induced both by exogenously applied ABA and osmotic stress. Cold treatment also enhanced the expression of PpABI1A but not PpABI1B. PpABI1A can suppress GUS expression in a line expressing the ABA-inducible Em promoter::GUS reporter construct. Elements within the wheat Em gene promoter are induced by both ABA and osmotic stress. PpABI1A and *Arabidopsis* ABI1 therefore have a conserved regulatory function, which was present in their common ancestor (Komatsu *et al*., 2009). Tougane *et al*. (2010) were the first to show this conserved regulatory function in liverworts. They identified an ABI1 homologue from *Marchantia polymorpha* (MpABI1), which had high similarity to *Physcomitrella* homologues of ABI1 (Tougane *et al*., 2010). *Marchantia polymorpha* is a dioecious and thallose liverwort species with extensive distribution across the globe. The gametophyte of *Marchantia* consists of a thallus with filamentous rhizoids that attach the plant to substrata. On the upper surface of the thallus are gemma cups, which contain tissue fragments known as gemmae that are genetically identical to the parent and provide a means of asexual reproduction. *Marchantia* also reproduces sexually by producing gametophores that contain the male antheridia, which produce the sperm and female archegonia, which produce ova (Raven *et al*., 1998). ABA has a growth inhibitory effect on gemmae but not in the thallus of *Marchantia* and accordingly, the expression of nuclear-localised MpABI1 is induced in the gemmae by exogenous ABA but not in the thallus. When co-expressed in *Marchantia* suspension cells with the Em-promoter::GUS construct by particle bombardment, MpABI1 can suppress ABA-induced gene expression. MpABI1 can also suppress ABA-

induced gene expression in *Physcomitrella* protonemata. Transgenic *Physcomitrella* lines expressing MpABI1 are hypersensitive to freezing and osmotic stress. Following ABA treatment, non-transgenic lines accumulated the LEA protein 17B9 to higher levels that lines expressing MpABI1. In addition, transgenic lines did not produce brood cells, a normal characteristic of cells resulting from ABA treatment (Tougane *et al*., 2010).

1.9.5 An 'evo-devo' view: The role of ABA in plant growth and development

The role of ABA is well understood within the plant kingdom where it is required by plants for correct timing of seed germination, control of stomatal aperture and to successfully adapt to various biotic and abiotic stress conditions. ABA accumulates up to 30-fold during periods of drought, salinity and cold stress and is important for the protection of plants against pathogen attack and UV radiation (Acharya *et al*., 2009; Rodríguez-Gacio *et al*., 2009; Wang *et al*., 2010). When stresses are perceived by plants, ABA is synthesised in the vascular tissues and rapidly distributed throughout the plant with the assistance of ATP-dependent transporters (Kuromori *et al*., 2009).

It is known that ABA responses are evolutionarily ancient and are present in bryophytes, including the liverworts, the oldest extant land plant lineage (Li *et al*., 1994). There is strong evidence to suggest that ABA responses have been conserved since the emergence of land plants and that the utilisation of ABA by early land plants was essential for their successful colonisation of land (Tougane *et al*., 2010).

Ruszala *et al*. (2011) demonstrated that the most basal vascular plant lineages, the lycophytes, possess stomata that respond to ABA in a similar way to that of higher plants. In the lycophyte *Selaginella uncinata*, stomatal aperture was controlled by both ABA and fluctuations in CO_2 concentrations. The S*elaginella moellendorffii* OST1 homologue was

able to rescue aberrant stomatal aperture phenotypes of *Atost1-4* mutants. Rather excitingly, these results showed that OST1 genes fulfilled a conserved function across 400 million years of land plant evolution (Ruszala *et al*., 2011). This finding disproved claims that dependency of stomatal responses on ABA and CO_2 appeared after the euphyllophytes (Brodribb *et al*., 2009).

In bryophytes, stomata are absent from liverworts and are restricted to the sporophyte generation of both mosses and hornworts (Ruszala *et al*., 2011). Until recently, the function of stomata in mosses, the most basal land plant that possesses stomata) was unknown. Chater *et al*. (2011) demonstrated that turgor-driven stomatal aperture of both *Physcomitrella* and angiosperm guard cells is controlled by fluctuations in ABA and CO_2. This discovery confirms that the ABA-mediated control of stomatal aperture is evolutionarily more ancient than originally thought (Chater *et al*., 2011). Histochemical staining of transgenic *Physcomitrella* lines expressing the ABA-inducible Em promoter::GUS reporter construct highlighted specific expression within the stomatal ring at the base of the diploid sporophyte structure. A number of ABA-signaling components are expressed in the sporophyte generation and this includes OPEN STOMATA 1 (OST1) (Chater *et al*., 2011). A *Physcomitrella* OST1 homologue, PpOST1, is able to carry out the guard-cell specific function of OST1 in *Arabidopsis* and restore ABA-induced stomatal closure in *Atost1* mutants. These findings suggested that OST1 function has been conserved since the mosses and vascular plants diverged (Chater *et al*., 2011).

Stomata are likely to have been a key adaptation that enabled plants to survive on land. Mosses evolved approximately half a billion years ago when atmospheric levels of CO_2 were very high. Restricting stomata to the sporophyte therefore probably had many advantages. In CO_2-rich environments, very few stomata would be required for the purposes

of gas exchange and this would also limit water loss from the plant. This would have been hugely beneficial in plants lacking vascular tissues and true multicellular roots. When atmospheric CO_2 levels became progressively reduced throughout evolution, plants are likely to have developed more of a dependency on stomata to gain access to vital CO_2 more effectively. Stomata therefore became more abundant and coated entire leaf surfaces of plants.

1.9.6 The role of ABA in abiotic stress tolerance

1.9.6.1 The role of ABA in freezing tolerance

Plants can acclimatise well to cold weather conditions. ABA is known to confer freezing tolerance to a number of mosses and three freezing-sensitive species of liverworts, *Marchantia polymorpha*, *Riccia fluitans* and *Helicodontium cappelare* (Pence, 1998). Several *Physcomitrella* ABA-responsive genes are upregulated in response to cold treatment and these include Late Embryogenesis Abundant (LEA)-like genes. Treatment of tissues with ABA prior to freezing enhances their survival; as the concentration of ABA increases, the percentage of surviving tissues increases proportionately (Minami *et al*., 2005). During this ABA-mediated cold acclimation response, levels of sucrose increase and levels of the trisaccharide theanderose increase more than 30-fold (Nagao *et al*., 2006).

1.9.6.2 The role of ABA in osmotic stress tolerance

Group 1 ABA-inducible Late Embryogenesis Abundant (LEA) proteins are expressed highly within angiosperm seeds. The Em protein is the single most abundant protein within the mature wheat embryo and after seed germination, Em levels decline rapidly (Cuming, 1984). In the presence of ABA, the wheat Em promoter drives the expression of a GUS

reporter gene in *Physcomitrella* since it contains an ABA-responsive element (ABRE). Stable transgenic lines expressing the ABA-inducible Em-GUS gene show clear ABA-responsiveness in all tissues (Knight *et al*., 1995). Transcript levels of a *Physcomitrella* group 1 LEA gene homologue (PpLEA1) accumulate rapidly in response to osmotic stress and exogenous application of ABA. The PpLEA1 gene promoter contains a characteristic ABRE sequence (AACGTG) upstream of the putative TATA box (Kamisugi and Cuming, 2005). Huge transcriptional changes occur when *Physcomitrella* protonemal tissues are treated with exogenous ABA or subjected to drought, osmotic and salt stresses. PpLEA2 transcript is particularly abundant during extreme drought stress and accumulates extremely rapidly after encountering stress (Cuming *et al*., 2007). Similarly to *Arabidopsis*, *Physcomitrella* transcription-associated proteins (TAPs) are rapidly induced by ABA and when exposed to high salinity. These include *Physcomitrella* homologues of Dehydration-responsive element-binding (DREB) proteins, ABA-responsive transcription factors that play roles in seed germination and drought tolerance. There is a significant overlap between genes induced by salt stress in both *Physcomitrella* and *Arabidopsis*. Since a number of ABA-biosynthetic genes were also upregulated, this suggested that salt stress responses are mediated by ABA and are conserved among the land plants (Richardt *et al*., 2009). ABA-treatment of *Physcomitrella* protonemal tissues results in upregulation of a number of proteins involved in transcriptional regulation and defence (Wang *et al*., 2010).

1.9.6.3 The role of ABA in desiccation tolerance

The earliest pioneering land plants lacked sophisticated water transport and retention mechanisms, which rendered them extremely desiccation prone. During their transition from water to land, plants acquired desiccation tolerance in their vegetative tissues. This trait was

lost in later-evolving plant lineages, which only retained desiccation tolerance in specialised structures such as seeds (Oliver *et al*., 2000; Khandelwal *et al*., 2010). ABA confers desiccation tolerance to the moss *Funaria hygrometrica* (Werner, 1991) and in the liverworts *Exormotheca holstii* (Hellwege *et al*., 1994) and *Riccia fluitans* (Hellwege *et al*., 1996). Uniquely among the land plants, bryophytes such as *Physcomitrella* can survive up to 80-92% water loss within vegetative tissues (Frank *et al*., 2005; Cuming *et al*., 2007). Oldenhof *et al*. (2006) claimed that *Physcomitrella* could even recover from 99.8% water loss, but this result was enhanced by exogenous ABA treatment (Oldenhof *et al*., 2006). Koster *et al*. (2010) claimed that *Physcomitrella* cannot be termed 'desiccation tolerant' since it is unable to survive in atmospheric water potentials of less than 13MPa (Koster *et al*., 2010).

Arabidopsis ABA-insensitive mutants were originally identified based on their ability to germinate in the presence of ABA, which normally inhibits seed germination. Among those identified were *abi1*, *abi2* and *abi3* (Koornneef *et al*., 1984; Finkelstein and Somerville, 1990). *Arabidopsis ABI3* encodes a transcription factor with roles in seed development. *Arabidopsis abi3* mutants are desiccation intolerant and have green seeds at maturity, as a result of a failure to degrade chlorophyll. ABI3 regulates seed dormancy, storage capability and chlorophyll degradation during seed maturation (Giraudat *et al*., 1992; Parcy *et al*., 1994; Nambura *et al*., 1995). Although *abi3* mutants do not appear to have defects in ABA-mediated stomatal closure, ABI3 is able to rescue the stomatal closure function of *abi1-1* mutants (Müller-Röber *et al*., 1998).

The *Physcomitrella* genome contains at least three *ABI3* gene homologues; *PpABI3A*, *PpABI3B* and *PpABI3C*. *abi3* mutant lines expressing *PpABI3A* no longer had green seeds and appeared morphologically normal. Wild type *Arabidopsis* seeds in addition to *AtABI3* complemented lines were not able to germinate at concentrations of ABA above 10µm

whereas *abi3* mutants could germinate at 100µm ABA. Only 60% of *abi3* mutants expressing *PpABI3A* were able to germinate at 100µm ABA compared to 100% *abi3* mutants alone. *PpABI3A* can therefore only partially complement the ABA-insensitivity phenotype of *abi3* mutants and partially restores ABI3 target gene expression. This suggests that some ABI3 functions are highly conserved whereas other functions have diverged throughout land plant evolution (Marella *et al*., 2006). In order to determine whether the regulatory pathways controlling desiccation tolerance in angiosperm seeds and the vegetative tissues of *Physcomitrella* are conserved, Khandelwal *et al*. (2010) examined triple *ppabi3* deletion mutants, generated by sequential gene targeting. When pre-treated with either 10µm or 100µm ABA, wild type protonemal tissue is able to survive extreme desiccation. Triple *ppabi3* deletion mutants are desiccation intolerant even when pre-treated with ABA prior to desiccation. Both ABA and ABI3 are therefore required to ensure survival during periods of drought. However, throughout evolution the regulatory pathways involving ABA and ABI3 have been co-opted from providing desiccation tolerance to vegetative tissues to a role in seed desiccation tolerance. Interestingly, Khandelwal *et al*. did not examine the effects of ABA on germination of triple *ppabi3* deletion mutants. This would determine whether spore germination, like seed germination, was controlled by ABA (Khandelwal *et al*., 2010).

1.10 Armadillo and Armadillo-related proteins

Proteins that are the focus of this project are the Armadillo-related proteins. They are found in animals, social amoebae, fungi and plants and are therefore evolutionarily very ancient. Animal and amoebal Armadillos have a nuclear function where they activate the transcription of a number of genes controlling multicellular development. They also localise to cellular junctions where they interact indirectly with the actin cytoskeleton in order to

maintain tissue architecture and consequently control cell polarity and migration (Coates, 2003; Tewari *et al*., 2010).

Armadillo-related proteins are highly conserved in land plants and are over 60% identical in *Physcomitrella* and *Arabidopsis*, plants that belong to two different lineages that diverged over 400 million years ago. *Arabidopsis* Armadillos (ARABIDILLOs) promote root branching (Coates *et al*., 2006). They also play important roles in ABA responses within germinating seeds (Gibbs and Coates, unpublished data).

The major goal of this research was to determine whether plant Armadillos were part of evolutionarily conserved signalling pathways. The following sections summarise functions of previously characterised Armadillo and Armadillo-related proteins.

1.10.1 Armadillo repeats

Armadillo (ARM) repeat proteins share evolutionarily conserved structures (Tewari *et al*., 2010). They contain tandem arrays of an imperfect 42 amino acid sequence motif, known as an ARM repeat (Peifer *et al*., 1994). ARM repeats associate to form alpha-helices, which pack together to form elongated superhelices (Huber *et al*., 1997). Superhelical structures form surfaces suitable for interacting with a number of different protein partners. ARM repeat proteins are structurally related to HEAT repeat proteins, such as importin-ß, which form similar helical structures, also possessing ligand binding sites (Andrade *et al*., 2001).

1.10.2 Armadillo

ARM repeats were originally identified in the *Drosophila* Armadillo protein and its name was derived from the appearance of mutant embryos that exhibited a segment polarity phenotype (**Figure 1.5**; Riggleman *et al*., 1989).

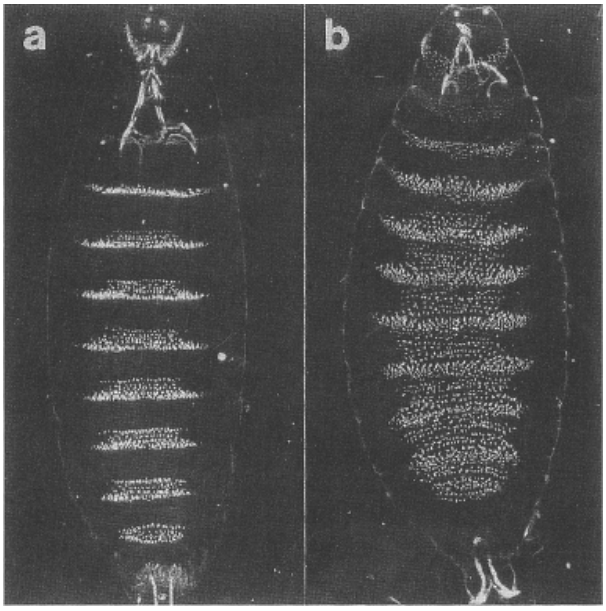


Figure 1.5 *armadillo* mutants have segment polarity phenotypes.

Dark-field micrographs of ventral cuticle preparations of **(a)** wild-type *Drosophila* embryo **(b)** arm^{25B} mutant embryo. *arm* mutants exhibit a segment polarity phenotype (from Riggleman *et al*., 1989).

Riggleman *et al*. (1989) cloned the *Armadillo* gene and determined the transcribed regions using transposon-tagging and northern blotting approaches respectively. Armadillo protein contained 12.5 ARM repeats and no known membrane-spanning sequences (confirmed by Peifer *et al*., 1994; **Figure 1.8**). Klingensmith *et al*. (1989) demonstrated that both *wingless* (*wg*) and *arm* mutants exhibited identical segment polarity phenotypes (Klingensmith *et al*., 1989). Peifer *et al*. (1991) went further and showed that the severity of the *arm* mutant phenotype corresponded to the extent of the armadillo mutation, that

armadillo null mutants were embryonic lethal and that Arm and Wg act together during anterior-posterior patterning (Peifer *et al*., 1991). It has also been shown that Arm plays roles in cell migration and cell adhesion, by interacting with α-catenin and E-cadherin at cellular adherens junctions (Oda *et al*., 1993; Peifer *et al*., 1993; Cox *et al*., 1996; Oda *et al*., 1997). Cox *et al*. (1996) examined the effects on expressing mutant versions of Arm, containing different truncations of the Arm repeat region, on *Drosophila* embryonic development. The extent of the truncation correlated with the severity of the phenotype and thus they determined that the Arm repeat region was critical for Arm function. They also determined that when Arm was absent, E-cadherin and α-catenin did not localise at the plasma membrane but accumulated in the cytosol instead (Cox *et al*., 1996). Sanson *et al*. generated *Drosophila* stocks that were expressing variable levels of Armadillo and determined that different sets of genes regulated the cell cycle, cell signalling and cell adhesion (Sanson *et al*., 1999).

1.10.3 ß-catenin

ß-catenin is the mammalian homologue of Armadillo, which plays pivotal roles in cell signalling and cell adhesion. ß-catenin maintains tissue architecture by interacting indirectly with the actin cytoskeleton via the adhesive transmembrane protein E-cadherin and α-catenin, in a similar way to Armadillo (**Figure 1.6**).

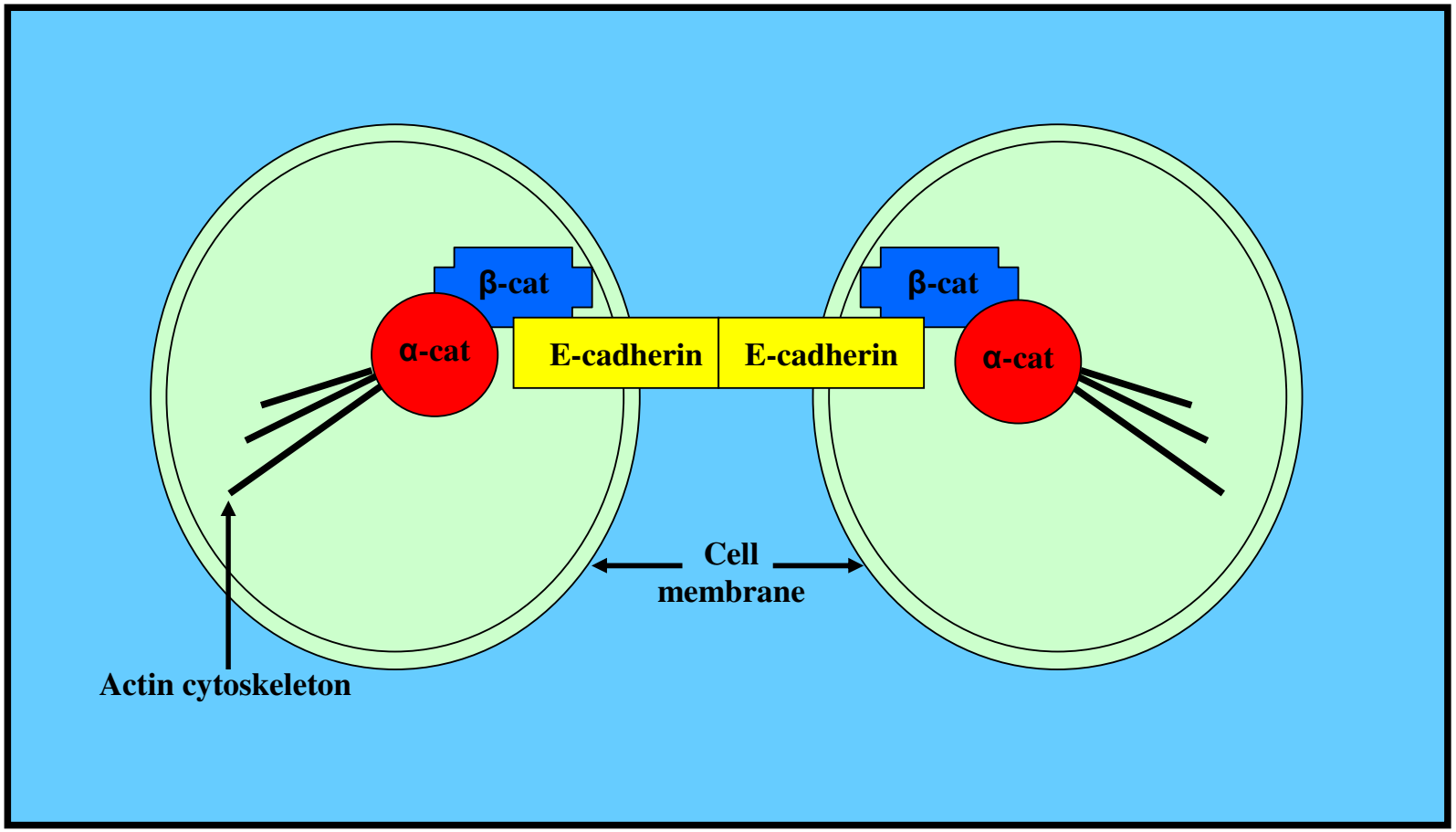


Figure 1.6 ß-catenin and cell adhesion. ß-catenin (blue) interacts indirectly with the actin cytoskeleton (indicated) by associating with the adhesive transmembrane protein E-cadherin (yellow) and α-catenin (red) at cellular junctions.

ß-catenin also plays a central role in the Wnt signalling pathway, which is equivalent to the *Drosophila* Wg signalling pathway (Shapiro, 1997). The ß-catenin (and Armadillo) protein contains a central region of ARM repeats, flanked on one side by amino-terminal phosphorylation sites for tyrosine kinases and glycogen synthase kinase 3 (GSK3) and, on the other side, by an acidic carboxyl-terminus that participates in gene activation in conjunction with Lymphoid Enhancer Factor (LEF)/T-cell transcription factor (Tcf). ß-catenin ARM repeats are known to mediate protein interactions with cadherins, Adenomatous polyposis coli (APC), Tcf transcription factors and the epidermal growth factor (EGF) receptor tyrosine kinase domain (Sethi and Vidal-Puig, 2010). The structure of murine ß-catenin (ß59) has been determined. It is composed of 12 ARM repeats, which interact with their adjacent neighbours and pack together to form an elongated, cylindrical right-handed superhelix of helices (Huber *et al*., 1997).

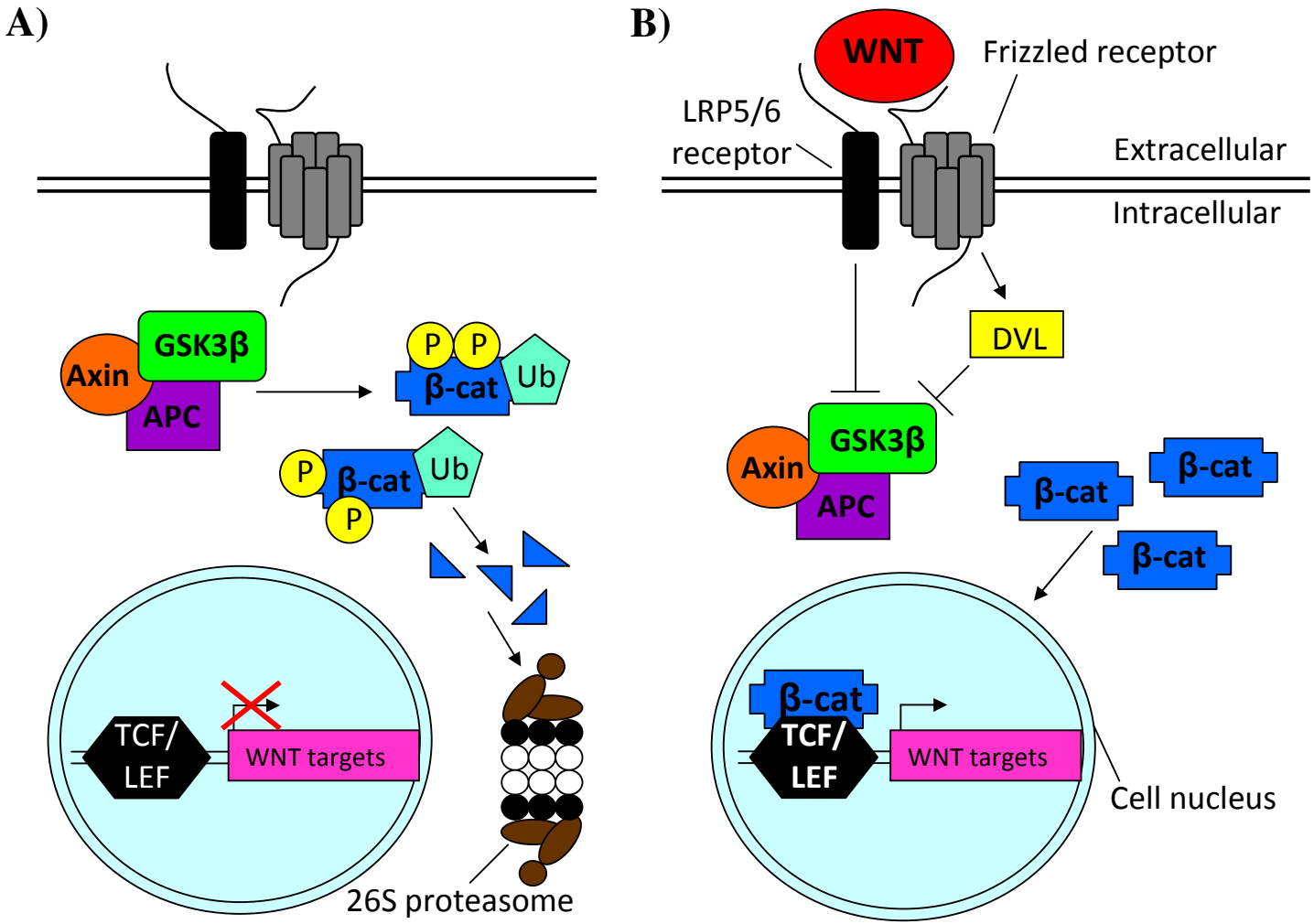
1.10.4 ß-catenin and its role in the Wnt signalling pathway

The role of ß-catenin during the Wnt signalling pathway has been well characterised. Essentially, the cytosolic pool of ß-catenin is dependent on a large family of cysteine-rich glycoproteins known as Wnt ligands. Wnt ligands bind to the Frizzled family of serpentine cell-surface receptors, which inactivates the Axin/GSK3/APC destruction complex that normally promotes phosphorylation and subsequent proteasome-mediated degradation of ß-catenin (van Noort *et al*., 2002). However, in the presence of Wnt signals, the cytosolic pool of ß-catenin accumulates, which enables it to migrate to cell nuclei, where it acts as a transcriptional activator with the assistance of the Tcf/LEF transcription factors (**Figure 1.7**; van Noort *et al*., 2002; Sethi and Vidal-Puig, 2010).

Figure 1.7 The Wnt signalling pathway.

A) When Wnt ligands are not bound to Frizzled cell-surface receptors, ß-catenin is degraded by a destruction complex containing Axin, GSK3 and APC, which phosphorylates ß-catenin and targets it for ubiquitin/26S proteasome-mediated degradation.

B) When Wnt ligands are bound to Frizzled receptors, Dishevelled (DVL) promotes downstream Wnt signalling and the inactivation of the destruction complex, allowing ß-catenin to accumulate in the cytosol. ß-catenin then enters the nucleus where it is able to interact with Tcf/LEF transcription factors and activate transcription of Wnt-target genes (Sethi and Vidal-Puig, 2010).


A)
GSK3β
Axin
APC
P
P
β-cat
Ub
P
β-cat
Ub
P
TCF/
LEF
WNT targets
26S proteasome
B)
WNT
Frizzled receptor
LRP5/6
receptor
Extracellular
Intracellular
DVL
GSK3β
Axin
APC
β-cat
β-cat
β-cat
β-cat
TCF/
LEF
WNT targets
Cell nucleus

1.10.5 Aardvark

An Armadillo/ß-catenin-related protein has also been identified outside of the animal kingdom in the unicellular amoeba, *Dictyostelium discoideum*. In periods of starvation, single-celled *Dictyostelium* aggregate forming a multicellular structure, which eventually becomes a slug. This is both light and temperature responsive and therefore able to migrate to an area containing ample food supplies. Slug structures differentiate into fruiting bodies, which are composed of a spore head and a stalk (Schaap, 2011). The Dictyostelium ß-catenin related protein, Aardvark is a 757 amino acid protein composed of ten ARM repeats and shares 50% homology with ß-catenin (Grimson *et al*., 2000; **Figure 1.8**). Like ß-catenin and Armadillo, Aardvark contains putative phosphorylation sites for GSK3 and a putative binding site for α-catenin (Juliet Coates, PhD thesis). Aardvark is present throughout the cellular cytosol and associates with cellular junctions of developing fruiting bodies and these resemble adherens junctions. *aardvark* mutants consequently have weak, collapsible fruiting bodies and defective actin-containing junctions and produce ectopic stalks. The *aardvark* mutant phenotype can be restored by expressing Aardvark cDNA (Grimson *et al*., 2000; Coates *et al*., 2002).

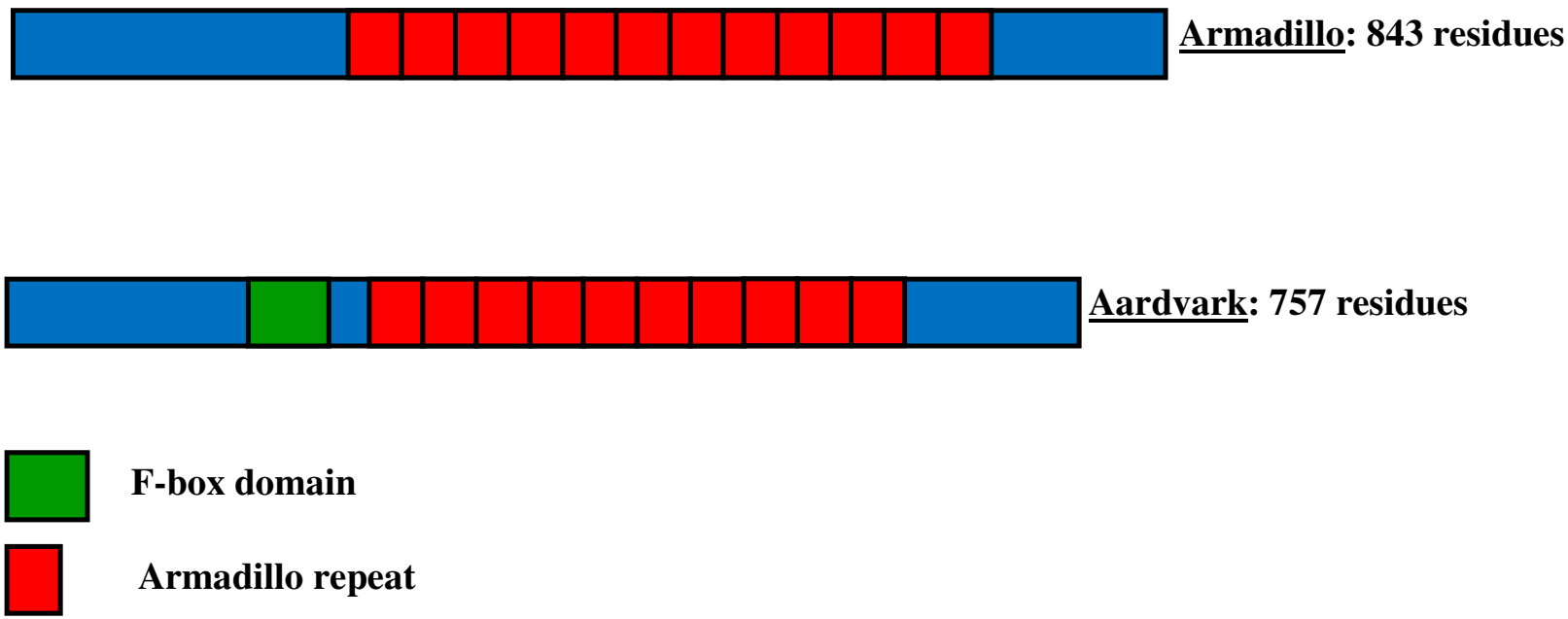


Figure 1.8 Animal and amoebal Armadillo protein domains.

Armadillo/ß-catenin has 12 ARM repeats; Aardvark has 10 ARM repeats (indicated by red boxes) and an N-terminal F-box domain (indicated by a green box). Lengths of each protein are indicated to the right of each diagram.

Dickinson *et al*. (2011) showed that a *Dictyostelium* α-catenin (Dd α-catenin) localised at cell-cell contacts in an Aardvark dependent manner. Aardvark and Dd α-catenin are believed to be required for the organisation and polarisation of the tip epithelium during culmination. *Dictyostelium* has no true cadherins and therefore does not possess junctions equivalent to metazoan cadherin-containing adherens junctions. It appears that Aardvark plays roles in maintaining cell polarity and that this function has been conserved from early-evolving *Dictyostelium* to metazoa (Dickinson *et al*., 2011). However, *Dictyostelium* does not possess all of the components of a functional Wnt signalling pathway (Harwood, 2008).

1.10.6 The ARABIDILLOs

Armadillo-related proteins have also been discovered in the plant kingdom. ARABIDILLO1 and ARABIDILLO2 are two *Arabidopsis* Armadillos that have been functionally characterised. ARABIDILLO1 and ARABIDILLO2 proteins are 80% identical and contain a nuclear localisation signal (NLS), an F-box domain, eight leucine-rich repeats (LRR) and nine ARM repeats. ARABIDILLO1 and ARABIDILLO2 are the only *Arabidopsis* proteins that have this unique protein domain architecture (Coates *et al*., 2006; Nibau *et al*., 2010; **Figure 1.9**).

Both *ARABIDILLO1* and *ARABIDILLO2* gene transcripts are expressed in roots and shoots, as detected by *pARABIDILLO::GUS* reporter fusions. *ARABIDILLO1* and *ARABIDILLO2* mRNA transcripts were detected in all organs, as indicated by RT-PCR. *arabidillo1* and *arabidillo2* single knockout mutants displayed no obvious phenotypes, whereas *arabidillo1/arabidillo2* double knockout mutants possessed fewer lateral roots when compared to wild type seedlings (Coates *et al*., 2006). Furthermore, ARABIDILLO1 overexpressing mutants produced significantly greater numbers of lateral roots compared to

wild type. Both *arabidillo1/arabidillo2* double knockout and ARABIDILLO overexpressing mutants exhibited wild type responses to auxin and nutrient deficiency, suggesting that the ARABIDILLOs do not participate in these signalling pathways (Coates *et al*., 2006; Nibau *et al*., 2011). ARABIDILLO-YFP protein fusions localised to nuclei of *Arabidopsis* root cells but were not detected elsewhere (Coates *et al*., 2006). This suggested that ARABIDILLO stability could be dependent on posttranslational control mechanisms (Nibau *et al*., 2010).

Using a yeast 2-hybrid approach, Nibau *et al*. demonstrated an interaction between the ARABIDILLO1-F-box domain and the ASK proteins, ASK1 and ASK2, known components of *Arabidopsis* SCF complexes that mediate Ubiquitin/26S proteasome-mediated degradation. This interaction could be abolished by mutating key residues within the F-box motif via site-directed mutagenesis (Nibau *et al*., 2010). It is now known that a subset of three angiosperm-specific MYB transcription factors (Myb53, Myb92 and Myb93) interact with the ARABIDILLO1 ARM repeat region in a yeast 2-hybrid screen (Daniel Gibbs, PhD thesis; Gibbs *et al*., manuscript in preparation).

In addition to their role in root branching, ARABIDILLOs appear to play important roles in the regulation of seed germination. *arabidillo1/2* double knockout mutants exhibit insensitivity to ABA-mediated inhibition of seed germination whereas ARABIDILLO1/2 overexpressing mutants exhibit hypersensitive responses (Gibbs and Coates, unpublished data). Real-time RT-PCR also revealed that *ARABIDILLO* genes are significantly upregulated by exogenous ABA treatment (Daniel Gibbs, PhD thesis).

ARABIDILLO homologues have been identified in all land plant genomes sequenced to date and they are all highly conserved, particularly within the F-box and ARM repeat regions (Nibau *et al*., 2011). Apart from the Charophycean algae, clear homologues of the ARABIDILLO proteins (i.e. that share the same novel F-box-LRR-ARM domain

architecture) have not been identified in any other sequenced algal genomes (e.g. *Chlamydomonas reinhardtii*, *Volvox carteri*, *Cyanidioschyzon merolae*), suggesting that they were important during the transition of plants from water to land (Nibau *et al*., 2011).

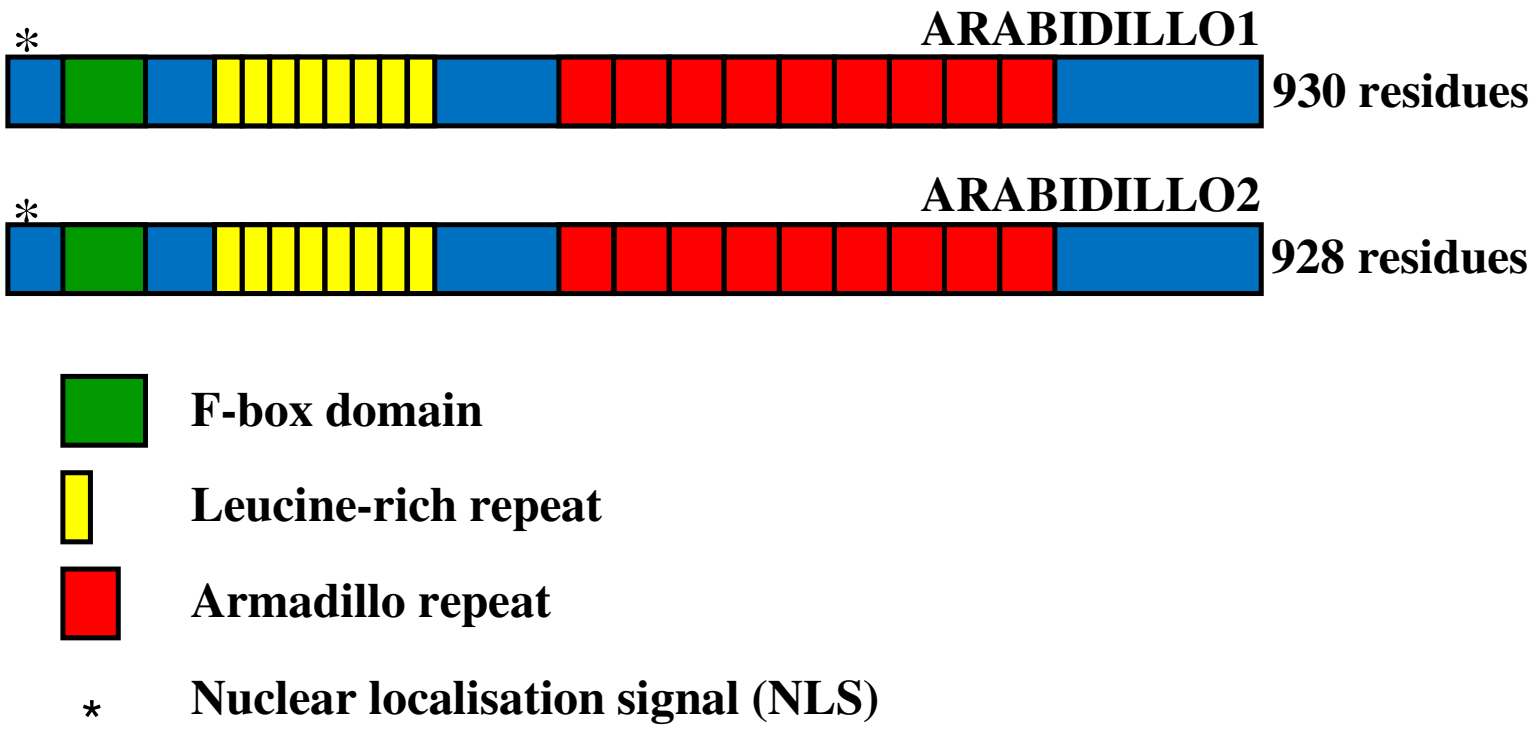


Figure 1.9 ARABIDILLO protein domains.

ARABIDILLO1 and ARABIDILLO2 are 80% identical at the amino acid level. Both proteins contain a nuclear localisation signal (NLS; *), an N-terminal F-box domain (indicated by a green box), a linker region composed of eight leucine-rich repeats (LRR; indicated by yellow boxes) and nine ARM repeats (indicated by red boxes; Coates *et al*., 2006; Nibau *et al*., 2011).

1.10.7 Diverse roles of Armadillo repeat proteins in plants

Armadillo-repeat proteins are found extensively throughout the plant kingdom and appear to play diverse and important roles in plant development. A few examples are: Leaf and Flower Related (LFR), which regulates the development of aerial organs, leaves and flowers in *Arabidopsis* (Wang *et al*., 2008); PHOTOPERIOD REGULATED 1 (PHOR1) forms an important component of the GA signalling pathway and is imported into the nucleus when gibberellins are present (Amador *et al*., 2001; Monte *et al*., 2003); ARMADILLO REPEAT ONLY (ARO) proteins play important roles in polar tip growth of pollen tubes and F-actin organisation (Gebert *et al*., 2008); ARIA positively regulates ABA responses in plants, particularly during seed development and osmotic stress (Lee *et al*., 2010).

1.11 F-box proteins and their role in Ubiquitin/26S-mediated proteasome degradation pathway

ARABIDILLO proteins are unique among plant Armadillo-related proteins as they possess an F-box domain/ARM combination that is only shared with *Dictyostelium* Aardvark (Coates *et al*., 2006). F-box proteins play important roles in protein degradation and therefore have important roles in plant development.

1.11.1 F-box proteins

F-box proteins are present in a wide variety of organisms and play significant roles in the Ub/26S degradation pathway. F-box proteins usually contain additional motifs that allow protein substrates to bind. One notable example are the *Arabidopsis* F-box proteins, ARABIDILLO1 and ARABIDILLO2, which contain a series of nine ARM repeats, believed to bind ligands (Coates *et al*., 2006; Ho *et al*., 2008). F-box motifs are able to engage with a

number of additional proteins to form SCF complexes, which are E3 ubiquitin-protein ligases. SCF complexes are composed of **S**UPPRESSOR OF KINETOCHORE PROTEIN 1 (SKP1), **C**ullin 1 (CUL1), **F**-box and RING-BOX 1 (RBX1) proteins. CUL1 can regulate the assembly of SCF complexes by interacting with RBX1 at its carboxyl terminus forming the catalytic core and by also interacting with SKP1 at its amino terminus and thus connecting the incomplete complex to F-box proteins. F-box proteins are able to interact with a variety of target proteins and it is this interaction that provides 'specificity' to SCF complexes. The *Arabidopsis* genome encodes over 700 F-box proteins (Lechner *et al*., 2006; Xu *et al*., 2008).

1.11.2 Roles of F-box proteins in plant growth and development

A number of plant growth and developmental responses rely upon posttranslational regulation. TRANSPORT INHIBITOR RESPONSE 1 (TIR1), an auxin receptor, is able to bind to a variety of AUX/IAA proteins and target them for Ub/26S proteasome mediated degradation. This means that these proteins can no longer bind to the AUXIN RESPONSE FACTOR (ARF) transcription factors and effect the transcription of auxin-inducible genes (Kepinski and Leyser, 2005; Dharmasiri *et al*., 2005). Subsequent analyses in rice and *Physcomitrella* showed that TIR1 was able to perform an evolutionarily conserved function in auxin signalling responses (Hayashi *et al*., 2008).

In *Arabidopsis*, the F-box protein CORONATINE INSENSITIVE 1 (COI1) participates in the Jasmonate signalling pathway (Fonseca *et al*., 2009).

SLEEPY1 (SLY1) plays roles in the *Arabidopsis* gibberellin signalling pathway, for which there are homologues in *Physcomitrella* and *Selaginella* (McGinnis *et al*., 2003; Vandenbussche *et al*., 2007). SLY1 encodes an F-box protein that forms a functional SCF complex to stimulate DELLA protein degradation via ubiquitin/26S proteasome-mediated

degradation. *sly1-2* and *sly1-10* mutants exhibit dwarfism, delayed germination and overaccumulation of the DELLA proteins RGA, GA1 and RGL2 since they are no longer able to degrade them. Another F-box protein known as SNEEZY can partially compensate for the loss of SLY1 (Strader *et al*., 2004; Ariizumi *et al*., 2011).

EIN3-binding F-box proteins (EBF) mediate Ubiquitin/26S proteasome-mediated degradation of EIN3/EIN3-like (EIL) proteins and consequently negatively regulate the ethylene signalling pathway. Tomato SlEBF1 and SlEBF2 are functionally redundant and are required for the development of reproductive organs, flower opening and fruit ripening (Yang *et al*., 2010).

F-box proteins have also been implicated in the formation and development of organ structures. Mutations in TIR1, as well as ARABIDILLO1 and ARABIDILLO2 and CEGENDUO give rise to defective lateral root formation. CEGENDUO mutants possess more lateral roots than wild type seedlings whereas mutations in both TIR1 and the ARABIDILLOs result in significantly reduced lateral root formation. Since their mRNA transcripts are expressed relatively broadly throughout the plant, it is likely that root architecture is controlled, to some extent, by the Ub/26S proteasome pathway. Coates *et al*. postulate that ARABIDILLO proteins may function redundantly to target inhibitors of lateral root formation for degradation (Coates *et al*., 2006).

1.11.3 The Ubiquitin/26S proteasome degradation pathway

The Ubiquitin/26S proteasome pathway is a crucial posttranslational control process that ensures that either incorrectly folded or excessive cytosolic pools of protein are targeted for degradation. Ubiquitin (Ub) is a 76 amino acid protein that binds to target proteins and 'tags' them for breakdown by the 26S proteasome. The 26S proteasome is entirely ATP-

dependent and composed of a 20S core protease (CP) and a 19S regulatory particle (RP). Each end of a CP is capped by an RP, which directs entry of proteins into the CP lumen via a narrowgated channel whilst removing attached Ubs (Vierstra, 2003).

The process initially involves activation of Ub by an ATP-dependent Ub-activating enzyme (E1) forming an E1-Ub intermediate. Activated Ub is then transferred to a Ub-conjugating enzyme (E2) by transesterification. Once target proteins have bound to E3 protein ligases, these are connected to C-termini of Ub molecules, contained within the E2 complex. Multiple Ubs are subsequently assembled and these are recognised and degraded by the 26S proteasome (Vierstra, 2003; Petroski and Deshaies, 2005).

There are many types of E3-Ubiquitin ligases. The vast majority of E3s in *Arabidopsis* are Ring-box and U-box E3s and relatively few other types are found. Ring-box E3s possess N-terminal Zinc-binding Ring-finger motifs that carry additional domains that are capable of recognising target proteins. U-box E3s are similar to Ring-box proteins but do not utilise zinc as a co-factor. HECT E3s are single peptides that possess characteristic HECT domains and these function as binding sites for Ub-E2 complexes. Motifs adjacent to HECT domains allow interactions with protein binding partners (Vierstra, 2003).

1.12 Project Objectives

The aim of the project was to determine the ancestral roles of ARABIDILLO genes in land plants by characterising their function in the moss *Physcomitrella*, which lacks multicellular roots and seeds. Determining whether *Physcomitrella* ARABIDILLOs can function in *Arabidopsis* and vice versa will determine whether ARABIDILLO proteins have previously unidentified and conserved functions across the land plants.

Chapter II: Materials and Methods

2.1 *Physcomitrella* tissue culture

2.1.1 Preparation of growth medium

BCD minimal medium contained 250mg/l $MgSO_4.7H_2O$, 250mg/l KH_2PO_4 (pH6.5), 1010mg/l KNO_3, 12.5mg/l $FeSO_4.7H_2O$, 0.001% Trace Element Solution (TES – 0.614mg/l H_3BO_3, 0.055mg/l $AlK(SO_4)_2.12H_2O$, 0.055mg/l $CuSO_4.5H_2O$, 0.028mg/l KBr, 0.028mg/l LiCl, 0.389mg/l $MnCl_2.4H_2O$, 0.055mg/l $CoCl_2.6H_2O$, 0.055mg/l $ZnSO_4.7H_2O$, 0.028mg/l KI and 0.028mg/l $SnCl_2.2H_2O$) with or without 8g/l agar, depending on whether liquid or solid cultures desired.

Gametophores were maintained for one month on solid BCD minimal medium supplemented with 1mM $CaCl_2$. Protonemata were maintained for one week on solid BCD minimal medium supplemented with 1mM $CaCl_2$ and 5mM ammonium tartrate. In order to select for stable transformants, medium was supplemented with either 20µg/ml Hygromycin B, 50µg/ml G418 or 50µg/ml Zeocin.

Protoplasts were regenerated using two types of solid regeneration medium. Solid protoplast regeneration medium bottom layer (PRMB) was composed of BCD minimal medium supplemented with 5mM ammonium tartrate, 60g/l mannitol, 10mM $CaCl_2$, 0.5% glucose and 8g/l agar. Solid protoplast regeneration medium top layer (PRMT) was composed of BCD minimal medium supplemented with 5mM ammonium tartrate, 80g/l mannitol, 10mM $CaCl_2$, 0.5% glucose and 5g/l agar. Protoplast liquid regeneration medium was composed of minimal BCD medium supplemented with 80g/l mannitol, 1mM $CaCl_2$ and 0.5% glucose.

Spore germination medium was composed of minimal BCD supplemented with 5mM ammonium tartrate, 10mM $CaCl_2$ and 8g/l agar.

All medium was autoclaved for 15 minutes at 121°C and poured/inoculated under sterile conditions in a laminar flow hood.

2.1.2 Culture conditions

The 'Gransden' wild type laboratory strain of *Physcomitrella* was gratefully received from Andy Cuming's group and cultured on either solid medium or in liquid suspension at 22°C with 16h light and 8h dark. Petri dishes were sealed with micropore tape (3M Healthcare, Germany) and culture flasks sealed using silicon stoppers (VWR). In order to induce sporulation, gametophytes were grown on sterile peat plugs for 6-8 weeks at 22°C, grown at 15°C for 2-3 weeks (under short day conditions; 8h light and 16h dark) and returned to 22°C to grow for an additional 2-3 weeks. Archegonia and antheridia were harvested immediately after peat plugs had been transferred to 22°C. Sporophytes were harvested after maturation using forceps under a SMZ645 light dissecting microscope (Nikon).

In order to propagate leafy tissues, gametophores were removed from established *Physcomitrella* plants and transferred to fresh medium using sterile forceps.

In order to prepare protonemata from either protonemata or gametophores, tissue was harvested and homogenised in sterile water for 1 minute at 19,000rpm using a polytron tissue tearer (IKA® T25 digital Ultra-Turrax). Protonemata were maintained on cellophane overlaid BCD. Cellophane discs (A.A. PACKAGING limited) were autoclaved for 15 minutes at 121°C, prior to their use.

2.1.3 Spore germination

Sporangia were sterilised in 20% parozone solution for 15 minutes in an orbital incubator at room temperature. Sporangia were then washed 3-4 times using sterile water. Spores were released from sporangia by perforating them with a sterile pipette tip in a final volume of 1ml sterile water. 200µl was spread onto each of 5 petri dishes containing cellophane overlaid spore regeneration medium.

2.2 *Arabidopsis thaliana* plant cultivation, growth and harvesting

2.2.1 Growth of *Arabidopsis* on ½ Murashige and Skoog (MS) agar plates

2.2.1.1 Preparation of ½ MS plates

2.2g MS basal salts (Sigma-Aldrich) were added to 1L sterile water (pH5.6, 8g/l agar).

2.2.1.2 Plating of *Arabidopsis* seeds

Arabidopsis seed was sterilised in 20% Parozone^{TM} bleach on a turning wheel for 10-15 minutes and then washed 2-3 times in sterile water. Seeds were then plated onto 1/2MS plates, cold treated for 48h and then transferred to the growth room for 7-10 days (22^{o}C, 16h light and 8h dark). Plates were supplemented with 50µg/ml kanamycin to select for transformants.

2.2.2 Growth of *Arabidopsis* plants in soil

Seeds were sown in Levington M3 compost/vermiculite mix in the greenhouse at 20-22^{o}C with 16h light and 8h dark.

2.3 Microscopy

Physcomitrella images were captured using a Nikon Digital Sight DS-Fil camera on a Nikon SMZ 1000 stereomicroscope. Measurements of length and area were carried out using the NIS-Elements BR software (version BR3.0). PHYSCODILLO1A-GFP and SELAGIDILLO-GFP images were captured using a Leica SP2 inverted confocal microscope. GFP fluorescence and chloroplast red autofluorescence were revealed using the following excitation/emission wavelengths: 488/530 for mGFP and 634/696 for TRITC (chloroplasts). The images obtained were projections of 20-25 image stacks. A light microscope was used to count protoplast number during moss transformations and to determine the percentage of germinating spores.

2.4 Bioinformatics

Theoretical genomic DNA, cDNA and protein sequences for *PHYSCODILLO1A*, *PHYSCODILLO1B* and *PHYSCODILLO2* were obtained from the *Physcomitrella patens* genome resource (version 1.1) at either JGI (http://genome.jgi-psf.org/Phypa1_1/) or COSMOSS (www.cosmoss.org). Theoretical genomic DNA, cDNA and protein sequences for *SELAGIDILLO* were obtained from the *Selaginella moellendorffii* genome resource (version 1.1) at JGI (http://genome.jgi-psf.org/Selmo1/Selmo1.home.html). Protein sequences for ARABIDILLO homologues from *Populus trichocarpa*, *Ricinis communis*, *Manihot esculenta*, *Glycine max*, *Vitis vinifera*, *Mimulus guttatus*, *Brachypodium distachyon*, *Zea mays*, *Aquilegia coerulea*, *Arabidopsis lyrata*, *Eucalyptus grandis*, *Citrus sinensis*, *Citrus clementina*, *Cucumis sativa*, *Prunus persica*, *Setaria italica*, *Carica papaya* and *Sorghum bicolor* were obtained from Phytozome (http://www.phytozome.net/). Protein sequences for ARABIDILLO homologues from *Oryza*

sativa and *Hordeum vulgare* were obtained following a protein BLAST at NCBI (www.**ncbi**.nlm.nih.gov/**BLAST**/). Gene structure diagrams were illustrated with the assistance of the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/help.php). ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html) was used to align all genomic DNA and protein sequences. Multiple alignments were shaded using BOXSHADE 3.21 (http://www.ch.embnet.org/software/BOX_form.html). Primers were designed using FastPCR 4.0 (http://www.biocenter.helsinki.fi/bi/Programs/download.htm). Protein motifs were detected with the assistance of pfam (http://pfam.sanger.ac.uk/), SMART (http://smart.embl-heidelberg.de/) and InterProScan (http://www.ebi.ac.uk/Tools/InterProScan/). The phylogenetic tree of ARABIDILLO homologues was constructed using GeneiousPro version 5.4.

2.5 Nucleic acid procedures

2.5.1 Genomic DNA isolation

Tissue from *Physcomitrella* and *Selaginella* was harvested and dried using filter paper, frozen rapidly in liquid nitrogen and stored at -80^{o}C until required. Tissue was homogenised by hand using a polypropylene pellet pestle (SIGMA-ALDRICH), firstly in liquid nitrogen and prior to thawing, in 0.6ml (per 500mg tissue) extraction buffer (100mM Tris-HCl pH8.0, 1.42M NaCl, 2% CTAB, 20mM EDTA, 2% PVP-40, 0.07% β-mercaptoethanol and 1mg/ml ascorbic acid), preheated to 65^{o}C. 10µl (per 500mg tissue) 10mg/ml RNase A was then added to the suspension and incubated at 65^{o}C for 5 minutes. An equal volume of chloroform:isoamyl alcohol (24:1) was then added, the samples vortexed for 30 seconds and then centrifuged for 10 minutes at 9,300xg, after which the resulting upper phase was transferred to a fresh eppendorf tube. An equal volume of isopropanol was added to the samples and these were then centrifuged for an additional 5

minutes at 9,300xg. The pellet was washed in 70% ethanol and left to air dry, redissolved in sterile water, quantified on a 0.8% agarose gel alongside a 1kb marker ladder (Invitrogen) and stored at -20°C.

2.5.2 RNA isolation

Tissue from *Physcomitrella* and *Selaginella* was harvested and dried using filter paper, frozen rapidly in liquid nitrogen and stored at -80°C until required. Tissue was homogenised using a polypropylene pellet pestle (SIGMA-ALDRICH), firstly in liquid nitrogen and prior to thawing, in 0.5ml (per 500mg tissue) extraction buffer (0.1M Tris-HCl pH9.0, 0.5% SDS, 2% PVP-40, 5mM EDTA and 0.07% β-mercaptoethanol). An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added and the resulting tissue suspensions vortexed for 30 seconds to produce an emulsion. Samples were centrifuged for 2 minutes at 12,000xg and the upper aqueous phase removed and transferred to a fresh tube. 0.1 volumes of 3M sodium acetate pH5.2 and two volumes of absolute ethanol were added to the tube, samples were flash frozen in liquid nitrogen for 5-10 seconds and then centrifuged for 5 minutes at 12,000xg. Pellets were washed using 70% ethanol, allowed to air dry and then resuspended in 20µl sterile water. 10µl 5M NaCl was added, samples placed on ice for 20 minutes and centrifuged for 2 minutes at 12,000xg. Supernatants were transferred to fresh tubes, 190µl 5M NaCl added and samples incubated overnight (4-12 hours) at 4°C. RNA was recovered following centrifugation at 4°C for 10 minutes at 12,000xg, resuspension in 400µl 2.5M NaCl and subsequent centrifugation for 5 minutes at 12,000xg. Pellets were washed using 70% ethanol, allowed to air dry, resuspended in sterile water and stored at -80°C.

2.5.3 Reverse-transcriptase PCR (RT-PCR)

RNA was isolated as described in section 2.4.2. RNA was DNaseI treated according to the manufacturer's instructions (Fermentas) and then converted to cDNA using Superscript™II reverse transcriptase according to the manufacturer's instructions (Invitrogen). Tissue-specific *PHYSCODILLO1A/1B* expression patterns were determined using P1-RT.GSP.F and P1-RT.GSP.R. Tissue-specific *PHYSCODILLO2* expression patterns were determined using P2-RT.GSP.F and P2-RT.GSP.R (see section **2.6.6** for primer sequences).

2.5.4 Agarose gel electrophoresis

0.8% agarose was made using 1xTBE (90mM Tris, 90mM boric acid, 2.5mM EDTA) and poured into gel trays containing 0.5µg/ml ethidium bromide. DNA samples were mixed with 6xDNA loading buffer (40% glycerol, 0.25% bromophenol blue) prior to loading into wells and running alongside a 1kb marker ladder (Invitrogen/NEB).

2.5.5 Gel extractions

The following products were gel extracted using the GeneJET™ Gel Extraction Kit (Fermentas): full length *PHYSCODILLO1A/1B* and *PHYSCODILLO2* genomic DNA and cDNA products; all PCR products amplified during the characterisation of the *PHYSCODILLO1A/1B* locus; components of PHYSCODILLO deletion, *PHYSCODILLO*-promoter::GUS and GFP fusion constructs; construct components digested from cloning vectors; *Selaginella* and *Arabidopsis* sequences; digested and CIP treated vector sequences; purification of Southern blot hybridisation probes.

2.5.6 Ligations

Ligations were carried out using T4 DNA ligase using a 3:1 molar ratio of insert:vector, according to the manufacturer's directions (New England Biolabs).

2.5.7 Plasmid DNA preparation

Mini cultures were set up by inoculating 2ml LB liquid cultures with single transformed colonies using sterile cocktail sticks (stock plates were also made by streaking the same colonies onto LB petri dishes containing appropriate antibiotics). These were then incubated in an orbital incubator overnight at 37^{o}C and plasmid DNA purified using a GeneJet^{TM} Plasmid Miniprep kit, according to the manufacturer's instructions (Fermentas Life Sciences).

Midi cultures were set up from either glycerol stocks or streak plates (once it had been verified that plasmid minipreps contained correct inserts). Midi cultures were set up by inoculating 30-200ml LB liquid cultures with streaked colonies using sterile cocktail sticks. These were then incubated in an orbital incubator overnight at 37^{o}C and plasmid DNA purified using a QIAGEN Plasmid Midi Kit, according to the manufacturer's instructions (Qiagen).

2.5.8 Restriction digests

All genomic and plasmid DNA was digested using restriction enzymes obtained from New England Biolabs according to the manufacturer's instructions.

2.6 Genomic DNA and cDNA cloning

2.6.1 Cloning of full-length *PHYSCODILLO1A/1B* and *PHYSCODILLO2* genomic DNA sequences

Genomic DNA was prepared from 7-day old protonemal tissue and used as a template in a PCR reaction using Long PCR enzyme mix (Fermentas Life Sciences). All primers were used at a final concentration of 0.3µM. The following thermal cycling conditions were used: 94°C for 5m, 10 cycles: 94°C 45s; 55°C 45s; 68°C 8m, 25 cycles: 94°C 45s; 55°C 45s; 68°C 8m+10s/cycle, followed by a final extension for 10m at 68°C. Full-length *PHYSCODILLO1A/1B* was amplified using P1-5'-1 and P1-3'-1. *PHYSCODILLO2* was amplified using P2-5'-1 and P2-3'-1 (see section **2.6.6** for primer sequences).

2.6.2 Cloning of full-length *PHYSCODILLO1A/1B* and *PHYSCODILLO2* cDNA sequences

RNA was isolated from 7-day old protonemata or 3-week old gametophores and reverse-transcribed into cDNA using SuperScript™ II Reverse Transcriptase, according to the manufacturer's instructions (Invitrogen). 2pmol gene-specific primers (P1-3'-1 and P2-3'-1) were used in each first-strand cDNA synthesis reaction. Full-length *PHYSCODILLO1A/1B cDNA* was amplified using P1-5'-1 and P1-3'-1 primers. Full-length *PHYSCODILLO2* cDNA was amplified using P2-5'-1 and P2-3'-1 primers. PCR products were amplified using Long PCR Mix (Fermentas Life Sciences) using the following thermal cycling conditions: 94°C for 5m, 10 cycles: 94°C 45s; 55°C 45s; 68°C 6m, 25 cycles: 94°C 45s; 55°C 45s; 68°C 6m+10s/cycle, followed by a final extension for 10m at 68°C.

2.6.3 Cloning the *PHYSCODILLO1A/1B* locus

Full-length *PHYSCODILLO1A* sequences (including gene and promoter) were amplified using P1downstream1R and P1-3'-1. The *PHYSCODILLO1A* intergenic region was amplified using primers P1/3D and P1/2seq5. Full-length *PHYSCODILLO1B* sequences (including gene and promoter) were amplified using P3upstream3F and P1-3'-1. The *PHYSCODILLO1B* intergenic region was amplified using primers P1/3D_RC and P1/2seq5RC (see section **2.6.6** for primer sequences). PCR products were then sequenced using a number of primers listed in 2.6.6 and sequences compiled by aligning overlapping reads using ClustalW2.

2.6.4 Cloning the full-length *SELAGIDILLO* genomic DNA sequence

Genomic DNA was prepared from *Selaginella* leafy tissue and used as a template in a PCR reaction using Long PCR enzyme mix (Fermentas Life Sciences). Full-length *SELAGIDILLO* was amplified using Sel_5'Start and Sel_3'-1 (see section **2.6.6** for primer sequences) using the following thermal cycling conditions: 94^{o}C 5m, 10 cycles: 94^{o}C 45s; 55^{o}C 45s; 68^{o}C 8m, 25 cycles: 94^{o}C 45s; 55^{o}C 45s; 68^{o}C 8m+10s/cycle, followed by a final extension for 10m at 68^{o}C.

2.6.5 Cloning the full-length *SELAGIDILLO* cDNA sequence

RNA was isolated from *Selaginella moellendorffii* and reverse-transcribed into cDNA using SuperScript^{TM} II Reverse Transcriptase, according to the manufacturer's instructions (Invitrogen). 2pmol gene-specific primer (Sel_3'-1) was used during the first-strand cDNA synthesis reaction. Full-length *SELAGIDILLO* cDNA was amplified using Sel_5'Start and Sel_3'-1 primers and Long PCR Mix (Fermentas Life Sciences) using the following thermal

cycling conditions: 94°C for 5m, 10 cycles: 94°C 45s; 55°C 45s; 68°C 6m, 25 cycles: 94°C 45s; 55°C 45s; 68°C 6m+10s/cycle, followed by a final extension for 10m at 68°C.

2.6.6 Primer list

Primer name	Primer sequence
P3upstream3F	5’-AGACTCCATGTCCACAGCCT
P3upstream4R	5’-GAAGTTGGGGCATAGATATGAGGGG
P3upstream5R	5’-TCCCTCTCGCTTTCGTGATTGATGT
P3upstream6R	5’-AGGGGGTGAAGCAAACGATTG
P3upstream7R	5’-CCGTCGTTCGTCGCGATGAGG
P3upstream8R	5’-GGAATCAATGCCACCACCGCG
P3upstream9R	5’-CGGAAGAGCGGAAGCAGGAAG
P1downstream1R	5’-TTGTTTGACAGAAGCAGCTGAC
P1/3A	5’-TTGGAAGTGAAGAATGATGCCGCA
P1/3B	5’-TGTGGCATGATCCGCACAGCT
P1/3C	5’-GGATGCGGTTGGGATTGACGT
P1/3D	5’-GCATTCACCATAGCTTTCGTCCAA
P1/3E	5’-GTCATAGCCAAGTTTTGGACGAA
P1/3F	5’-TGCAATGCACCCAAAGGTCCTGAT
P1/3G	5’-TTTGGCGTGATGGATTGGGAT
P1/3H	5’-AAGGCGACGTGGCGAATGGT
P1/3I	5’-GGGTTTGTGAGGAATGAGGCTGAT
P1/3J	5’-GTCAAAAGGCACCCAAGGAT
P1/3K	5’-GACCCAATGTAACCTCTAGGGGGA
P1/3L	5’-GGGGATATTGCATCAGTGGTGT
P1/3M	5’-GAATGCGGTTTCACACCAATGC
P1-5’-1	5’-GATGGTGCATGCGCGGTGCG
P2-5’-1	5’-GCGGAATGGGATCGAGTCCTGG
P1/2-5’Start	5’-ATGTCCAACAAGCGGCGGCG
P1-3’-1	5’-CTCATCCCAACCTCTTACGACAG
P2-3’-1	5’-CGCGCTGCCACTGGCTTCACC
P1/2seq1	5’-GCTGGGTTCGGATTGCGAGAG
P1/2Sseq1RC	5’-TCTCGCAATCCGAACCCAGCAGC
P1/2seq2	5’-GGGCGGAAGCTGTGATGAACGG
P1seq3	5’-CTGGCGCAGGCGCAATCGAAG
P1seq3RC	5’-CTTCGATTGCGCCTGCGCCAG
P2seq3	5’-CCGAGGCAGGCGCAATTGAAGC
P2seq3RC	5’-GCTTCAATTGCGCCTGCCTCGG
P1/2seq4	5’-CTGGTGGCGTGGAGGCACTG
P1/2seq4RC	5’-CAGTGCCTCCACGCCACCAG
P1/2seq5	5’-GCGTACATGTTCGATGGCAGG
P1/2seq5RC	5’-CCTGCCATCGAACATGTACGC
P1/2seq1B	5’-TAGGGAGCTGAGCGGCGATTG
P1/2seq2B	5’-GTGTTTTGCTGCGAATTGGTGGT
P1/2seq3B	5’-TGAATGGTGTGGTGCGGCGAG
P1/2seq3B.RC	5’-CTCGCCGCACCACACCATTCA
P1/2seq4B	5’-AGGTGGTGTTGTTGGCGGAAGCG

Primer name	Primer sequence
P1/2seq5B	5’-ATGGGACAGTGGCGGTGATGAGG
P1/2seq5B.RC	5’-CCTCATCACCGCCACTGTCCCAT
P1/2seq4C	5’-GGGCTGCATGACGTTGAGCGT
P1/2seq4C.RC	5’-ACGCTCAACGTCATGCAGCCC
P1/2seq5C	5’-GGAACCTGGTTTTCAATCCTGG
P1/3C_RC	5’-ACGTCAATCCCAACCGCATCC
P1-3’-1.RC	5’-CTGTCGTAAGAGGTTGGGATGAG
P1.promA	5’-TCCCACGCTCTCTTGCTCTTGT
P3.prB	5’-GGGCACAACGTGCACAAAACAGA
P1/3.7A.R	5’-TGTTGGGACCGCGAGCACGC
P2.promA	5’-TTGACCGGGACCTTGCGGCA
P1-RT.GSP.F	5’-GGCGCAATCGAAGCACTGGTGG
P1-RT.GSP.R	5’-TGTACGTCCTCAAAATCAGAGTGC
P2-RT.GSP.F	5’-CGCAATTGAAGCACTGGTGGATCT
P2-RT.GSP.R	5’-ACGTCCTCAGAGTTCGAGTGTGC
P1/3D_RC	5’-TTGGACGAAAGCTATGGTGAATGC
Sel_5’Start	5’-ATGCGTCGCGTTCGCCGGAAATGCG
Sel_3’-1	5’-ATTGGCCGCTAAAGATCG
Physco1-3F-XbaI	5’-AAATCTAGAGTGGAGCCATGGTTGGGAGATGG
Physco1-3R-NotI	5’-AAAGCGGCCGCCCTCATCCCAACCTCTTACGACAG
P2.5F.KpnI	5’-AAAGGTACCCGTCGTCCGCTTGTAGTGCCC
P2.5R.XhoI	5’-AAACTCGAGCACGCCTACACTCGAGGGAGG
P2.3F.XbaI	5’-AAATCTAGAGGCTGGTGCTTCGTTGGGATGC
P2.3R.NotI	5’-AAAGCGGCCGCCGCGCTGCCACTGGCTTCACC
P1+3KO_5’F-2.XbaI	5’-AAATCTAGAACTGGCTCCGGTCGGAACGAG
P1+3KO_5’R-2.BamHI	5’-AAAGGATCCGTTCTCGGCTGTCACAAGCTGG
P1+3KO_3’F.SphI	5’-AAAGCATGCGACTCTTGTCGTTTTTCTGCATC
P1+3KO_3’R.EcoRV	5’-AAAGATATCCCTGAAACTTCCAAAATGCCAG
P1_prF.NotI	5’-AAGCGGCCGCACGGTGACAAGTGCCGGACGAA
P1_promoter_R_BamHI	5’-AAAGGATCCGTCATCAGCTGCTCCTCCGAT
P2_prF_NotI	5’-AAGCGGCCGCTTCACAAGCTGAGATGTTCGGAGC
P2_prom_R_BamHI	5’-AAAGGATCCATCATCATCACCTCCTCCCACAATC
P3prF.NotI	5’-AAGCGGCCGCTACGGCGAAAAAGTTCCTGGC
P1_promoter_R_BamHI	5’-AAAGGATCCGTCATCAGCTGCTCCTCCGAT
P1/3.prom.F.Not	5’-AAGCGGCCGCATCCCACAGAGAACATGTTAAACA
P1_promoter_R_BamHI	5’-AAAGGATCCGTCATCAGCTGCTCCTCCGAT
G418.R.319	5’-TACTTTCTCGGCAGGAGCAAG
G418.F.341	5’-CCATCATGGCTGATGCAATGC
P1/2-5'Start_BamHI	5’-AAAGGATCCATGTCCAACAAGCGGCGGCG
P1_3'end_no_STOP_NotI	5'-AAAGCGGCCGCCGACAGTGCCACCACCGTGTTGC
Sel.5'St.XbaI	5’-AAATCTAGAATGCGTCGCGTTCGCCGGAAATGCG
Sel-St.NotI	5’-AAGCGGCCGCGCTTTTCTCGCTCTGGCACAGCTCCA

2.7 Southern blotting

2.7.1 Southern blotting procedure

15µg *Physcomitrella* genomic DNA was digested to completion using restriction enzymes overnight at 37°C, according to the manufacturer's instructions (NEB). An equal volume of phenol was added to the completed digest, centrifuged at maximum speed for 1 minute and the aqueous phase removed. An equal volume of chloroform:isoamylalcohol (24:1) was added to the aqueous phase and centrifuged at maximum speed for 1 minute. A one-tenth volume of 3M sodium acetate (pH5.3) and 2 volumes of absolute ethanol were added to the resulting aqueous phase and incubated for 30 minutes at -20°C to precipitate DNA. DNA was harvested by centrifugation at 4°C for 20 minutes. The pellet was then washed using 70% ethanol, allowed to air-dry and resuspended in 30µl sterile water (to ensure entire contents loaded into a single well). Digested genomic DNA samples were run overnight on a 0.8% agarose gel at 50V and visualised on a transilluminator alongside a ruler (in order to calculate fragment size at a later date).

The gel was depurinated for 20 minutes in 0.25M HCl, denatured for 1 hour in denaturing buffer (1.5M NaCl, 0.5M NaOH) and neutralised for 1 hour in neutralising buffer (1M Tris.Cl pH8.0, 1.5M NaCl). The gel was then incorporated into the blotting apparatus (as shown in figure 2.1) and the DNA transferred to N^+ Hybond membrane (Amersham) overnight by capillarity using transfer buffer (1.5M NaCl, 0.25M NaOH). The blotting apparatus was partially dismantled, leaving the gel as a template for marking the lanes on the membrane. The membrane was then washed using 6xSSC at room temperature for 5 minutes and allowed to air-dry between two sheets of 3MM filter paper. The DNA was then UV-crosslinked to the membrane.

2.7.2 Generation of probes for labelling

To screen for the presence of *PHYSCODILLO2*, the *PHYSCODILLO2* gene-specific P2-5' probe was digested out of the *PHYSCODILLO2* deletion construct using KpnI and XhoI (see section **2.11**).

To screen for the presence of either *PHYSCODILLO1A* or *PHYSCODILLO1B*, the *PHYSCODILLO1A/1B* gene-specific probe P1/3-3' probe was generated by PCR using Physco1-3F-XbaI and Physco1-3R-NotI (see section **2.6.6** for primer sequences). Products were then cloned into the plasmid pSC-B, sequenced and digested out of the vector using XbaI and NotI.

To distinguish between *PHYSCODILLO1A* and *PHYSCODILLO1B*, the *PHYSCODILLO1A/1B* gene promoter-specific probe pP1/3 was removed from the *pPHYSCODILLO1A/1B*::GUS construct using NotI and BamHI (see section **2.12**).

2.7.3 Hybridisation using radioactive probes

The nitrocellulose membrane was incubated for 2 hours at 65^{o}C in hybridisation buffer (1% BSA, 1mM EDTA pH8.0, 0.5M Na_2HPO_4, 7% SDS). Hybridisation probes were labelled with dCTP, [α-^{32}P]-3000Ci/mmol (Perkin Elmer) using the Prime-It® II Random Primer Labeling Kit, according to the manufacturer's instructions (Stratagene) and then added to the hybridization buffer. Blots were incubated with radiolabelled probes overnight at 65^{o}C.

Blots were washed twice for 15 minutes in 2xSSC (+0.1% SDS), 15 minutes in 1xSSC and 15 minutes in 0.5xSSC. All wash steps were carried out at 42^{o}C. After each wash step, counts were monitored and further wash steps terminated if detectable radioactivity was low. Blots were wrapped in cling film and exposed to Hyperfilm^{TM} ECL (Amersham) for 3-7 days in an autoradiography cassette at -80^{o}C (Southern, 1975).

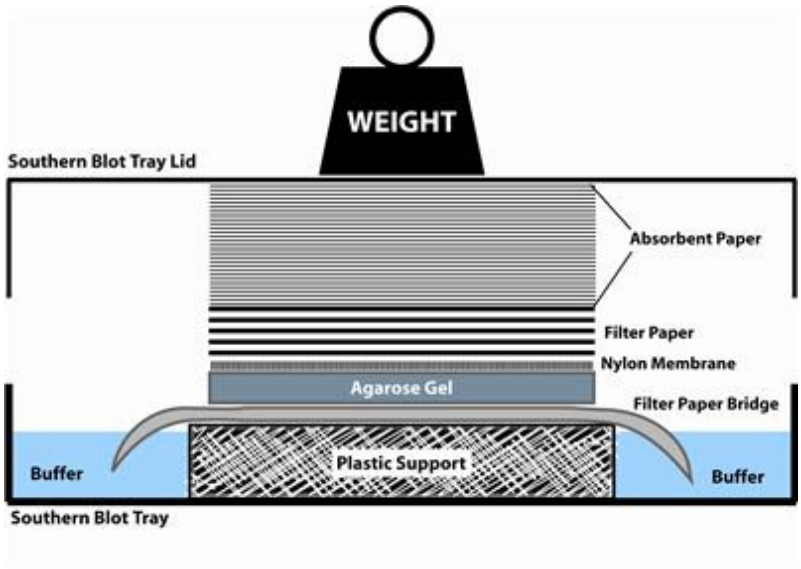


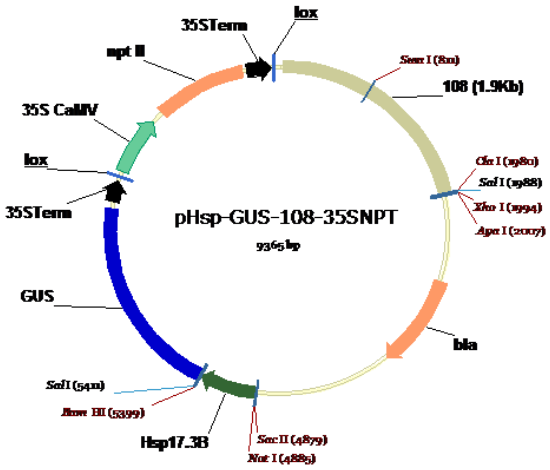
Figure 2.1 Southern blotting apparatus. *Physcomitrella* genomic DNA was digested to completion using restriction enzymes and run on a 0.8% agarose gel. The gel was then depurinated, denatured and neutralised prior to assembling into the transfer apparatus shown in the diagram. A bottom layer of one sheet of filter paper acted as a wick in order to take up the transfer buffer in the reservoir by capillarity. The blot was then compiled as follows: gel (overturned), nitrocellulose membrane, three sheets of 3MM paper, two stacks of paper towels and a weight. The digested DNA was transferred to the membrane at room temperature overnight (Southern, 1975).

2.8 Vectors

pHSP-GUS-108-35SNPT (Generated by Younousse Saidi)

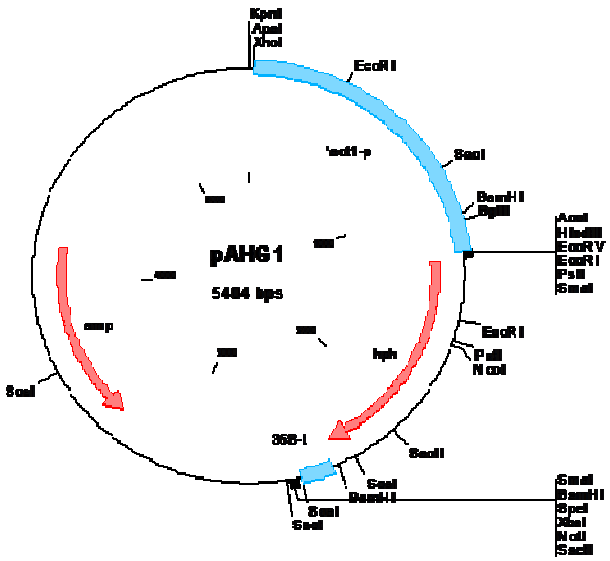
A plasmid adapted for use in *PHYSCODILLO*-promoter::GUS analyses. The 9365bp plasmid contains a heat-shock promoter, a GUS gene, a neomycin phosphotransferase cassette

(NPTII, G418) and a 1.9kb sequence homologous to the 108 locus. The heat-shock promoter is flanked on either side by NotI and BamHI sites. These were exploited in order to replace the heat-shock promoter with *PHYSCODILLO* gene promoters.



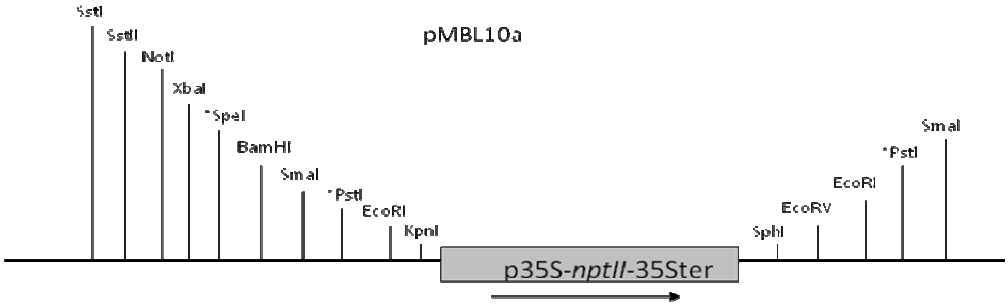
pAHG1 (kindly provided by Andy Cuming)

A plasmid used to generate the *PHYSCODILLO2* single deletion construct. The plasmid contains a Hygromycin resistance cassette to allow screening of stable transformants using Hygromycin B.



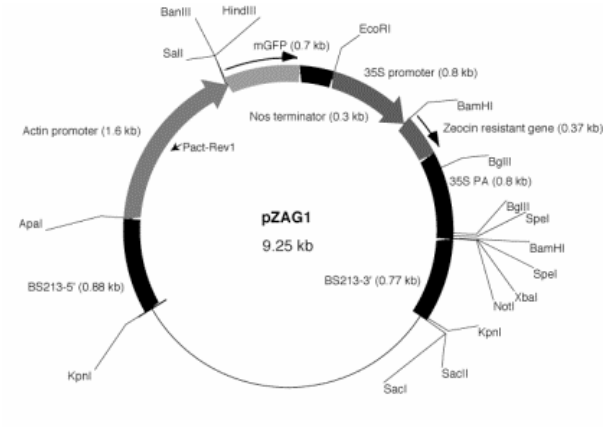
pMBL10a (kindly provided by Andy Cuming)

A plasmid used to generate the *PHYSCODILLO1A/1B* double deletion construct. The plasmid contains a neomycin phosphotransferase cassette to allow screening of stable transformants using G418 in addition to Hygromycin B.



pZAG1 (kindly provided by Jean-Pierre Zryd)

A plasmid used to generate the *ppdillo1a/1b/2* mutant complementation lines. The plasmid contains a Zeocin resistance cassette to allow screening of stable transformants using Zeocin in addition to G418 and Hygromycin B.



pGreenII 0029 and pSoup

4.6kb disarmed T-DNA vectors that confer kanamycin resistance in *E. coli*. The pSoup plasmid encodes the RecA protein, which allows pGreen (Kan^R) to replicate in *Agrobacterium*. pSoup has gentamycin resistance and encodes transfer genes required for T-DNA to successfully integrate into the *Arabidopsis* genome (Hellens *et al*., 2000).

2.9 Bacterial methods

2.9.1 Bacterial strains

***Escherichia coli* DH5α**

supE44, lacU169 (Φ80lacZΔM15), hsdR17, recA1, endA1, gyrA96, thi-1, relA1.

***Agrobacterium tumefaciens* GV3101**

A disarmed *Agrobacterium* strain (C58 background) with a rifampicin resistance gene on chromosome 1 and a pMP90 Ti virulence plasmid (pTiC5ΔT-DNA) carrying a gentamycin resistance gene (Koncz *et al*., 1992).

2.9.2 Bacterial culture medium

E. coli

E. coli were cultured on petri dishes containing solid LB medium (10g/l bacto-tryptone, 5g/l bacto-yeast extract, 10g/l NaCl and 15g/l bacto-agar) or in liquid LB cultures (10g/l bacto-tryptone, 5g/l bacto-yeast extract, 10g/l NaCl). Medium was additionally supplemented with either 50µg/µl ampicillin or 50µg/µl kanamycin. In the case of blue/white screening, 40µl 2% 5-bromo-4-chloro-3-indolyl- beta-D-galactopyranoside (X-gal) was added to LB agar surfaces prior to plating out transformation mixtures.

A. tumefaciens

A.tumefaciens were cultured on Petri dishes containing solid LB medium (as above) or in liquid LB cultures. Medium was additionally supplemented with 50µg/µl kanamycin (to screen for the presence of pGreen/pSoup) and 20µg/µl rifampicin (to screen for the presence of *A. tumefaciens*).

2.9.3 Bacterial growth conditions

E. coli

E. coli liquid cultures were incubated in an orbital incubator at 200rpm, for approximately 16 hours at 37^{o}C. *E. coli* agar plates were grown inverted for approximately 16 hours at 37^{o}C.

A. tumefaciens

A. tumefaciens liquid cultures were incubated for either 24h or 48h at 28°C in an orbital incubator at 200rpm. *A. tumefaciens* agar plates were grown inverted for 48 hours at 28°C.

2.9.4 Preparation of competent cells

2.9.4.1 Preparation of chemically-competent *E. coli* DH5α cells

5ml LB broth (without antibiotics) was inoculated with a single colony of competent *E. coli* DH5α competent cells and grown overnight at 37°C/200rpm in an orbital incubator. 200µl inoculated culture was then added to 100ml LB within a 1L flask and grown for approximately 3 hours at 37°C/200rpm until the culture had reached an OD_{600nm} of 0.35-0.45. The culture was subsequently spun at 4°C for 5 minutes at 1300 x *g*. The supernatant was discarded and pellet resuspended in 40ml TfbI buffer (30mM potassium acetate, 10mM rubidium chloride, 10mM calcium chloride, 50mM manganese chloride, 15% (v/v) glycerol, pH5.8). Cells were left on ice for 3 hours and then spun at 4°C for 5 minutes at 1300 x *g*. The supernatant was discarded and pellet resuspended in 4ml TfbII buffer (10mM MOPS, 75mM calcium chloride, 10mM rubidium chloride, 15% (v/v) glycerol, pH6.5) before overnight incubation on ice. Competent cells were then aliquoted, snap frozen in liquid nitrogen and stored at -80°C.

2.9.4.2 Preparation of electrocompetent *A. tumefaciens* GV3101 cells

5ml LB-rifampicin broth was inoculated with a single colony of *A. tumefaciens* GV3101 cells and grown at 28°C for 48h at 200rpm. The inoculated 5ml culture was then used to inoculate 200ml LB-rifampicin broth and grown overnight until the OD_{600nm} was 0.5-0.8. Cells were placed on ice to cool and then spun at 4°C for 5 minutes at 5000 x *g*. The supernatant was discarded. The cells were then washed successively with 1, 0.5, 0.2 and 0.02 culture volumes of

cold 10% glycerol by repeated suspension and centrifugation of the cells. 100µl aliquots were prepared and these were immediately snap-frozen in liquid nitrogen and subsequently stored at -80°C.

2.9.5 Bacterial transformations

2.9.5.1 Transformation of *E. coli* by heat-shock

1-2µl ligated DNA product was added to 50µl chemically competent DH5α *E. coli* cells (thawed on ice) and incubated on ice for 20 minutes. The transformation mixture was then heat-shocked for 45 seconds at 42°C and returned to ice for an additional 2 minutes. 250µl sterile LB medium (no antibiotics) was then added to the mixture and incubated for 1 hour at 37°C on an orbital incubator at 200rpm. 50µl and 250µl transformation mixture were then spread onto LB-agar plates containing the appropriate antibiotics, using a sterile glass spreader. Plates were then inverted and incubated overnight at 37°C.

2.9.5.2 Transformation of *A. tumefaciens* by electroporation

100ng pGreen (containing construct of interest) and 100ng pSoup were added to an aliquot of *A. tumefaciens* electrocompetent cells in an electroporation cuvette (0.2cm, BioRad). The DNA/cell mixture was incubated on ice for approximately 20 minutes and then electroporated using a BioRad Micropulser at 2.5kV for 5-6ms. Cells were then recovered by adding 1ml LB broth (without antibiotics) and incubating at 28°C for 3-4 hours. Cells were then spread onto plates containing the appropriate antibiotics, using a sterile glass spreader. Plates were then inverted and incubated at 28°C for 2 days.

2.10 Generation of constructs for use in *Physcomitrella* transformations

2.10.1 Ligation of inserts into target vectors

The following detail describes the mechanism by which flanking sequences used for targeted gene replacement via homologous recombination, were ligated into moss transformation vectors: PCR products were gel extracted, purified and ligated into the pSC-B cloning vector using the StrataClone™ Blunt PCR Cloning Kit and subsequently transformed into SoloPack competent cells, according to the manufacturer's instructions (Stratagene). Transformation mixtures were plated out onto LB agar petri dishes containing 50µg/ml ampicillin and 40µl X-gal for blue/white selection of transformants and incubated overnight at 37°C. 4-6 white colonies were selected and used to inoculate individual 2ml LB+50µg/ml Amp liquid cultures, which were subsequently incubated in an orbital incubator overnight at 37°C. Plasmid DNA was then isolated and screened for correct integration of the PCR product by digesting the resulting plasmids using *EcoRI* (sites present at either side of the pSC-B MCS). Inserts were then sequenced from both ends using the universal primers M13F and M13R.

5µg destination plasmids and 10µg plasmid product containing the correct insert sequence were digested using an appropriate pair of restriction enzymes (e.g. *BamHI* and *XbaI* were used to digest pMBL10a and the pSC-B plasmid containing the 5' flanking sequence for the *PHYSCODILLO1A/1B* double deletion construct) for approximately 4 hours at 37°C. After 3 hours, 0.5U/µg Calf Intestinal Phosphatase (NEB) was added to the destination plasmid digest in order to prevent it recircularising during the ligation step. Digested products were gel extracted and purified. Inserts were ligated into corresponding destination vectors using T4 DNA ligase, according to the manufacturer's instructions (NEB). Ligated products were transformed into chemically competent DH5α *E. coli* as described in section 2.7.4. Colonies were used to

inoculate LB+Amp liquid cultures and plasmids isolated as described previously. Plasmid DNA was isolated and screened for correct integration of insert by digesting it using restriction enzymes compatible to both ends of the insert. If two inserts per vector were required (e.g. *PHYSCODILLO* deletion and *PHYSCODILLO*-GFP knock-in constructs), sequential ligations were carried out.

2.11 Generation of *PHYSCODILLO* deletion constructs

The *PHYSCODILLO2* single deletion construct was generated by cloning 5' and 3' homologous flanking sequences from *PHYSCODILLO2* genomic DNA and inserting them into the pAHG1 vector either side of a Hygromycin resistance cassette. The 5' flanking sequence was amplified using P2.5F.KpnI and P2.5R.XhoI and ligated into KpnI/XhoI cut pAHG1 to create pP2KO-5'. The 3' flanking sequence was amplified using P2.3F.XbaI and P2.3R.NotI and ligated into XbaI/NotI cut pP2KO-5'.

The *PHYSCODILLO1A/1B* double deletion construct was generated by cloning 5' and 3' homologous flanking sequences from regions either side of the *PHYSCODILLO1A/1B* locus and inserting them into pMBL10a either side of a G418 resistance cassette. The 5' flanking sequence was amplified using P1+3KO_5'F-2.XbaI and P1+3KO_5'R-2.BamHI and ligated into XbaI/BamHI cut pMBL10a to create pP1+3dblKO-5'. The 3' flanking sequence was amplified using P1+3KO_3'F.SphI and P1+3KO_3'R.EcoRV and ligated into SphI/EcoRV cut pP1+3dblKO-5' (see section **2.6.6** for primer sequences).

All PCR products required for the generation of deletion constructs were amplified using Pfu Turbo (Stratagene) according to the manufacturer's instructions and the following thermal

cycling conditions used: 95°C for 5m, 35 cycles: 95°C 1m; 52°C 1m; 72°C 4m, followed by a final extension for 10m at 72°C.

2.12 Generation of *pPHYSCODILLO*::GUS constructs

The *PHYSCODILLO1A* promoter sequence for the *pPHYSCODILLO1A*::GUS construct was amplified using P1_prF.NotI and P1_promoter_R_BamHI. The *PHYSCODILLO2* promoter sequence for the *pPHYSCODILLO2*::GUS construct was amplified using P2_prF_NotI and P2_prom_R_BamHI. The *PHYSCODILLO1B* promoter sequence for the *pPHYSCODILLO1B*::GUS construct was amplified using P3prF.NotI and P1_promoter_R_BamHI. The *PHYSCODILLO1A/1B* promoter sequence for the *pPHYSCODILLO1A/1B*::GUS construct was amplified using P1/3.prom.F.Not and P1_promoter_R_BamHI (see section **2.6.6** for primer sequences). Promoter sequences were ligated into BamHI/NotI cut pHSP-GUS-108-35SII (cut vector was gel purified to ensure complete removal of the existing HSP promoter). All PCR products required for the generation of deletion constructs were amplified using long PCR enzyme mix according to the manufacturer's instructions (Fermentas).

2.13 Transformation of moss protoplasts

2g polyethylene glycol 6000 (polyethylene glycol BIOULTRA 6000, Sigma-Aldrich#81255) was melted in a flat-bottomed autoclavable vial, added to 5ml mannitol/$Ca(NO_3)_2$ solution (0.8% mannitol, 0.1M $Ca(NO_3)_2$, 10mM Tris pH8.0) and incubated for 2-3 hours at room temperature. 60mg Driselase® (Driselase® Basidiomycetes sp. Plant cell culture, Sigma-Aldrich#D8037) was dissolved in 6ml 8% mannitol in a round bottom snap cap

tube (BD Biosciences) for 15-20 minutes at room temperature. The enzyme solution was gently inverted to mix and centrifuged for 3 minutes at 3,300 xg. After the supernatant was filter sterilised through a 0.2µM syringe filter, 1.5-2 plates of 7-day protonemal tissue was added. The tissue was digested for 30 minutes to 1 hour at room temperature and gently mixed at 5-10 minute intervals. The tissue suspension was then filtered through a sterile 100µM nylon-fitted funnel and incubated at room temperature for an additional 5-10 minutes. The solution was decanted into a fresh snap cap tube and centrifuged for 3 minutes at 120xg. Supernatants were removed carefully in order to avoid disruption of cells, which were subsequently washed in 6ml 8% mannitol and centrifuged for 3 minutes at 120xg. Following a repeated wash step, cells were resuspended in 6-10ml 8% mannitol and total protoplast number determined using a haemocytometer.

Pellets were gently resuspended in an appropriate volume of MMM solution (91g/l mannitol, 0.15M $MgCl_2$, 0.1% MES pH5.6), in order to obtain a cell density of 1.2-1.62x10^6/ml. 15µg linearised construct was added to a series of snap cap tubes prior to the addition of 300µl protoplasts (in MMM solution). Controls included empty vector and no DNA controls. 300µl PEG solution was added to the protoplast mixture in drops and swirled to mix after each addition. The samples were then heat-shocked for 5 minutes at 45^{o}C and incubated for an additional 5 minutes at room temperature.

300µl 8% mannitol was added to each tube, 5 times at 4-6 minute intervals and tilted gently to mix after each addition. 1ml 8% mannitol was then added to each tube, 5 times at 4-6 minutes and tilted gently to mix after each addition. Cells were centrifuged for 4 minutes at 140xg, supernatants discarded and pellets resuspended in 5ml liquid BCD, containing 10mM

$CaCl_2$, 8% mannitol and 0.5% glucose. Tubes were wrapped in aluminium foil, to protect them from light, and incubated overnight at 25°C.

Following overnight incubation, cells were centrifuged for 4 minutes at 120xg and resuspended in 8% mannitol; 0.5ml per 3 petri dishes or 0.8ml per 4 petri dishes. Prior to solidification, 2.5ml (for 3 plates) or 3.2ml (for 4 plates) PRMT, supplemented with 10mM $CaCl_2$ and maintained at 37°C, was then added to cell suspensions and 1ml immediately plated onto cellophane disc overlaid PRMB medium.

Plates were incubated in continuous light for 5 days at 22°C after which the cellophane discs were transferred to selection plates (BCD supplemented with Hygromycin B, G418 or Zeocin) and incubated in continuous light for an additional 2 weeks (If colonies began to overlap after 2 weeks (14 days), they were subcultured onto non-selection plates). On day 21, cellophane discs were transferred to non-selective plates and incubated at 22°C in continuous light for an additional 2 weeks. On day 35, colonies were transferred from non-selective plates to selective plates, but this time, cellophane discs were turned upside down so that the aerial parts were in direct contact with the agar. The discs were smoothed using a spatula, to ensure that all of the colonies had been transferred. The plates were incubated for at least a week in continuous light, to select for 'stable' transformants. Colonies were then screened for targeted integration by PCR using GoTaq (Promega).

2.14 Treatment of regenerating protoplasts with Abscisic acid (ABA)

Protoplasts were isolated from wild type and *ppdillo1a/1b/2* triple deletion mutants (as described in section **2.11**). They were then plated onto PRMB medium (see section **2.1.1**),

supplemented with 1, 10, 25, 50 or 100µm ABA (dissolved in methanol solvent) and grown for 1-2 weeks under standard growth conditions (see section **2.1.2**).

2.15 Treatment of germinating spores with Abscisic acid (ABA)

Sporophytes were harvested from wild type and *ppdillo1a/1b/2* triple deletion mutants (as described in section **2.1.2**). Sporophytes were then sterilised and germinated by plating them onto spore germination medium (see section **2.1.1**), supplemented with 1, 10, 25, 50 and 100µm ABA and grown for one week under standard growth conditions (see section **2.1.2**).

2.16 Desiccation and freezing tolerance assays

Wild type and *ppdillo1a/1b/2* mutant protonemal tissues were grown on cellophane-overlaid BCD under standard conditions for one week. Cellophanes were then transferred to BCD medium, supplemented with ammonium tartrate and different concentrations of ABA and grown under standard conditions overnight. Cellophanes were then subjected to either desiccation or freezing for a period of one week. To test for desiccation tolerance, cellophanes were transferred into an empty petri dish for one week. To test for freezing tolerance, protonemal plates were incubated at -20^{o}C for one week. After one week, cellophanes were transferred to BCD medium, supplemented with ammonium tartrate (without ABA) and grown under standard conditions for one week.

2.17 Genotyping of *ppdillo1a/1b/2* triple deletion mutants

To detect the presence of a G418 resistance cassette within the *PHYSCODILLO1A/1B* locus and confirm the generation of *ppdillo1a/1b/2* triple deletion mutants, PCR was carried out.

To check for 5' integration into the *PHYSCODILLO1A/1B* locus, P3_upstr_3F and G418.R.319 primers were used. To check for 3' integration into the *PHYSCODILLO1A/1B* locus, P1_downstr_1R and G418.F.341 primers were used (see section **2.6.6** for primer sequences).

2.18 Staining of *PHYSCODILLO*-promoter::GUS lines

Tissue was harvested and submerged in an X-gluc solution (0.5mg/ml X-gluc, 0.2% Triton X-100, 0.5mM $K_4Fe(CN)_6.3H_2O$, 0.5mM $K_3Fe(CN)_6$, 50mM sodium phosphate buffer pH7.0) for 1-24 hours at 37°C. After staining, the tissue was cleared using the following solutions; 70% ethanol for 24 hours (to eliminate chlorophyll), 50% ethanol/10% glycerol for 2 hours, 30% ethanol/30% glycerol for 2 hours, 50% glycerol for 2 hours. Stained tissues were then stored at 4°C in 50% glycerol. Tissues were mounted on slides in 50% glycerol and visualised using a Nikon SMZ 1000 stereomicroscope (Nikon).

2.19 *Arabidopsis thaliana* transformations

2.19.1 Generation of the PHYSCODILLO1A-GFP construct

The PHYSCODILLO1A-GFP cDNA sequence (minus the stop codon) was amplified using P1/2-5'Start_BamHI and P1_3'end_no_STOP_NotI and ligated into BamHI/NotI cut pGreen (containing 35S-GFP-ter) to create 35S-PHYSCODILLO1A-GFP-ter.

2.19.2 Generation of the SELAGIDILLO-GFP construct

The SELAGIDILLO-GFP cDNA sequence (minus the stop codon) was amplified using Sel.5'St.XbaI and Sel-St.NotI and ligated into XbaI/NotI cut pGreen (containing 35S-GFP-ter) to create 35S-SELAGIDILLO-GFP-ter (see section **2.6.6** for primer sequences).

2.19.3 Transformation of *Arabidopsis thaliana* by floral dip

A. thaliana plants were grown for approximately 4-6 weeks until the first buds on the main inflorescence spike had begun to open.

A single colony of *A. tumefaciens* containing the construct of interest was used to inoculate a 5ml LB culture supplemented with kanamycin and rifampicin and incubated at 28^{o}C for 48h at 200rpm in an orbital incubator. This culture was subsequently used to inoculate 400ml LB culture (containing kanamycin and rifampicin) and incubated for an additional 24h at 28^{o}C/200rpm. The bacterial pellet was harvested by spinning the culture at 5000rpm for 15 minutes. The pellet was then resuspended in 200ml 5% sucrose. Silwet L-77 (100µl) (wetting agent) was added after the pellet had been thoroughly resuspended.

The bacterial solution was transferred to a beaker in the greenhouse. The plants were then inverted, dipped into the solution, gently agitated and then removed after approximately 20s. Plants were then placed in a clear autoclave bag and left in low light for approximately 24h. Plants were then transferred to normal growth conditions until seeds were ready for harvesting.

2.19.4 Screening of *A. thaliana* transformants

T1 seeds were harvested and sterilised as described in section **2.2.1.2**. Seeds were then incubated in PPM solution (1% PPM, 50g/l $MgSO_4$) on a turning wheel overnight at room temperature. Seeds were washed 2-3 times in sterile water, resuspended in 0.7% agarose and

poured onto ½ MS plates containing 0.7% agar and 50µg/ml kanamycin. Plates were sealed with micropore tape, cold treated for 48h and then grown under normal growth conditions for 7-10 days (see section **2.2.1.2**). Successfully transformed seedlings (i.e. those that remained green after selection) were transferred to soil and allowed to set T2 seed.

T2 segregation analysis was carried out by sterilising T2 seed and plating it out onto ½ MS containing 50µg/ml kanamycin. The ratio of resistant seedlings to susceptible seedlings (to kanamycin treatment) was then determined.

Chapter III:

Cloning and expression analyses of ARABIDILLO homologues in lower land plants

3.1 Introduction

ARABIDILLO1 and ARABIDILLO2 promote lateral root development in *Arabidopsis*. These proteins have unique domain architectures and possess an N-terminal F-box domain, a linker region composed of leucine-rich repeats and an Armadillo-repeat domain (Coates *et al.,* 2006; Nibau *et al.,* 2011). ARABIDILLO homologues are found extensively throughout the land plants, including the early bryophytes, lycophytes and angiosperms (Nibau *et al*., 2011).

In order to determine the function of ARABIDILLOs early in land plant evolution, their function will be characterised in an early-evolving moss species that lacks roots. The moss *Physcomitrella patens* recently became the first bryophyte with a sequenced genome and represents the most basal land plant lineage in studies of ARABIDILLO function (Rensing *et al*., 2008). The *Physcomitrella* genome is particularly amenable to genetic manipulation since it is predominantly haploid and carries out homologous recombination (HR) at a similar efficiency to *Saccharyomyces cerevisiae* (Prigge and Bezanilla, 2011). Using *Physcomitrella* to determine ancestral functions of ARABIDILLO proteins will highlight whether they have conserved functions throughout land plants.

The genome of *Selaginella moellendorffii* has more recently been completed. *Selaginella moellendorffii* belongs to the lycophytes, which evolved before the angiosperms and after the bryophytes. The lycophytes have vascular tissues and simple rooting structures that bifurcate at their tips (Banks, 2009; Banks *et al*., 2011). *Selaginella moellendorffii* therefore represents a 'link' between the non-vascular bryophytes and angiosperms, which have vascular tissues, seeds and roots.

The aim of the work in this chapter was to identify and clone bryophyte and lycophyte ARABIDILLO homologues from *Physcomitrella* and *Selaginella* respectively. This would determine whether Armadillo protein architecture is conserved among all land plant lineages. Work also includes a more in-depth analysis of *Physcomitrella* ARABIDILLOs in order to define ancestral roles of ARABIDILLO proteins in land plants.

3.2 ARABIDILLO homologues are present in the bryophyte species, *Physcomitrella patens*

In order to identify ARABIDILLO homologues in *Physcomitrella*, BLAST searches were carried out at the *Physcomitrella patens* genome resource version 1.1 (http://www.cosmoss.org/ and http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html) using either *ARABIDILLO1* genomic DNA (blastn) or ARABIDILLO1 protein (tblastn and blastp) sequences. The blastn search resulted in no hits whereas the tblastn and blastp searches identified three putative *Physcomitrella* ARABIDILLO homologues, which were subsequently named *PHYSCODILLO1A*, *PHYSCODILLO1B* and *PHYSCODILLO2*. Accession numbers for *PHYSCODILLO1A*, *PHYSCODILLO1B* and *PHYSCODILLO2* were available from the COSMOSS database and these were Phypa_186256, Phypa_131917 and Phypa_175550 respectively. When the three putative *PHYSCODILLO* sequences were used in a blastp search against the *Arabidopsis* proteome, ARABIDILLO1 was returned as a reciprocal best hit, confirming that they were true homologues.

In the published genome sequence, the full-length *PHYSCODILLO2* gene resided alone on scaffold 13 and lay adjacent to unrelated genes, defined in this thesis as ‘a’, ‘b’ and ‘c’

(**Figure 3.1A**). Gene 'a' (protein ID: 115450) encoded a putative glutathione reductase, gene 'b' (protein ID: 159475) encoded a G-protein coupled receptor and gene 'c' (protein ID: 115455) encoded a callose synthase. The annotated *PHYSCODILLO2* gene was 5576 base pairs long and contained 11 exons and 10 introns. The corresponding transcript was 2802 base pairs long and encoded a 934 amino acid protein.

A full-length copy of the *PHYSCODILLO1A* gene and a partial *PHYSCODILLO1B* gene were located on scaffold 91 and surrounded by genes 'd' and 'e'. Gene 'd' (protein ID: 230042) encoded a protein of unknown function and gene 'e' (protein ID: 131941) encoded a putative metallopeptidase. The *PHYSCODILLO1B* gene sequence was truncated by a 4kb gap in the contig sequence but was otherwise 100% identical to the equivalent portion of the *PHYSCODILLO1A* gene (**Figure 3.1B**). The identity extended to 1.6kb of both *PHYSCODILLO1A* and *PHYSCODILLO1B* promoter regions, which were also 100% identical. The gap in the contig sequence was followed by a 4kb intergenic region, which shared no homology with any of the *PHYSCODILLO* genes (**Figure 3.1B**). The annotated *PHYSCODILLO1A* gene was 6251 base pairs long and contained 12 exons and 11 introns. The corresponding transcript was 2823 base pairs long and encoded a 941 amino acid protein.

Interestingly, following a blastn search at both JGI and Phytozome, a partial sequence of a truncated pseudogene was also detected. It was located upstream of the putative *PHYSCODILLO1B* gene on scaffold 91: 474945-475467. The pseudogene was 522 base pairs long and was 95% identical to the first 534 base pairs of both *PHYSCODILLO1A* and *PHYSCODILLO1B* genes. The start codon of the pseudogene had been mutated from methionine

(ATG) to isoleucine (ATA). Beyond the 522 base pair sequence, there was no further homology. The promoter region also shared no similarity with any of the *PHYSCODILLO* gene promoters (see **Appendix 8.1**).

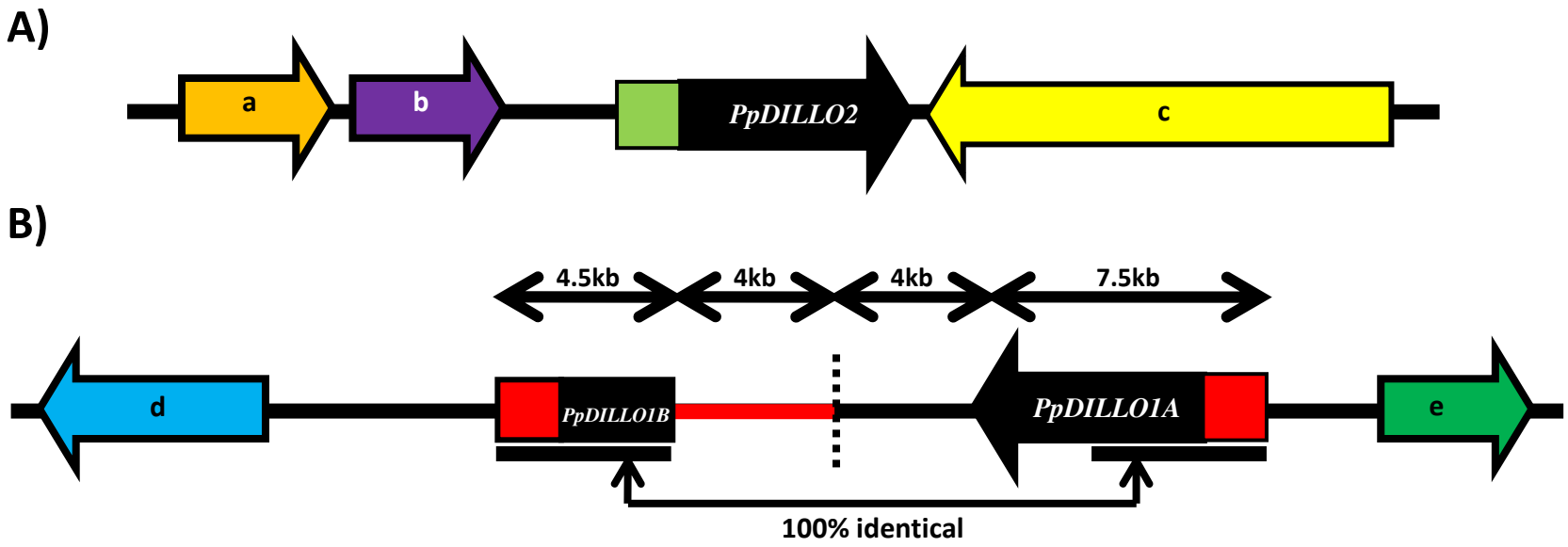


Figure 3.1 Three putative ARABIDILLO homologues in *Physcomitrella patens*.

Using the ARABIDILLO1 protein sequence in a tblastn search at the *Physcomitrella* genome resource, three putative *PHYSCODILLO* genes (*Physcomitrella* ARABIDILLOs) were identified. **A)** Full-length *PHYSCODILLO2* was located on scaffold 13 of the published genome sequence and had an intact putative promoter sequence (indicated by a green box). The *PHYSCODILLO2* gene lay adjacent to genes 'a', 'b' and 'c' as indicated. **B)** A full-length copy of the *PHYSCODILLO1A* gene and a partial *PHYSCODILLO1B* gene were located on scaffold 91 of the published genome sequence. *PHYSCODILLO1B* was truncated by a 4kb gap in the contig sequence (indicated by a red line) and this was followed by a 4kb intergenic sequence. *PHYSCODILLO1B* was identical to the equivalent portion of the *PHYSCODILLO1A* gene and this included their putative 1.6kb promoter regions (red boxes and arrows). *PHYSCODILLO1A* and *PHYSCODILLO1B* were adjacent to genes 'e' and 'd' respectively and these are indicated on the diagram. *PHYSCODILLO1A*, *PHYSCODILLO1B* and *PHYSCODILLO2* have been shortened to *Ppdillo1A*, *Ppdillo1B* and *Ppdillo2* respectively.

3.3 Determining the correct number and sequence of *PHYSCODILLO* genes

In order to verify the published *PHYSCODILLO2* sequences, full-length genomic DNA and cDNAs were amplified by PCR and sequenced. Alignment of the amplified/cloned gDNA and cDNA demonstrated that the *PHYSCODILLO2* genomic DNA sequence contained an additional 250 base pairs that was not annotated in version 1.1 of the *Physcomitrella* genome. The sequence contained an additional exon and consequently, the *PHYSCODILLO2* gene was 5860 base pairs long and contained 12 exons and 11 introns (**Figure 3.2A**). The corresponding transcript was 2823 base pairs long and encoded a 941 amino acid protein that was 85% identical to sequences predicted for PHYSCODILLO1A and PHYSCODILLO1B.

To confirm the presence of either *PHYSCODILLO1A* or *PHYSCODILLO1B* genes, full-length genomic DNA and cDNA sequences were amplified by PCR using primers specific to the full-length *PHYSCODILLO1A* gene annotated in the genome. Subsequent sequencing revealed that both genomic DNA and cDNA sequences obtained were identical to theoretical sequences retrieved from the *Physcomitrella* genome database. These data confirmed that either *PHYSCODILLO1A* or *PHYSCODILLO1B* or both were indeed present, and that either or both genes were expressed since full-length cDNA was cloned with relative ease. However, using this approach, it was not possible to distinguish between *PHYSCODILLO1A* and *PHYSCODILLO1B*.

In order to determine whether *PHYSCODILLO1A* and *PHYSCODILLO1B* genes were both present within the same locus as an inverted repeat, as the draft genomic sequence suggested, long PCR products were generated using primers specific to the known full-length *PHYSCODILLO1A* (or *-1B*) gene and within divergent sequences either side of the

PHYSCODILLO1A/1B locus (within genes 'd' and 'e'; **Figure 3.3A**). This ensured that *PHYSCODILLO1A* and *PHYSCODILLO1B* could be distinguished from one another from sequencing data produced (**Figure 3.3A**). Both full-length *PHYSCODILLO1A* and *PHYSCODILLO1B* gene and 1.6kb promoter sequences were amplified and sequencing revealed that they were 100% identical. *PHYSCODILLO1A* and *PHYSCODILLO1B* were in a tail-to-tail orientation with 8kb intergenic sequence separating their stop codons. The intergenic region was made up of two duplicated 4kb intergenic regions that also sat in a tail-to-tail orientation (**Figure 3.3A**). The *PHYSCODILLO1A* and *PHYSCODILLO1B* intergenic regions were 100% identical apart from seven individual nucleotide differences towards the ends that were furthest from the two stop codons of *PHYSCODILLO1A* and *PHYSCODILLO1B* genes. An additional 111 nucleotide sequence separated the two intergenic regions. This sequence was completely unique and consequently aided assembly of the *PHYSCODILLO1A/1B* locus by overlapping sequences of cloned PCR products (**Figure 3.3B**).

Figure 3.2 *PHYSCODILLO2* was misannotated in the *Physcomitrella* genome. *PHYSCODILLO2* genomic and cDNA sequences were cloned and sequenced. **A)** Experimental sequences were aligned to theoretical sequences obtained from the *Physcomitrella* genome resource using ClustalW2. Nucleotides that corresponded to the presence of an additional exon have been underlined. Identical nucleotides are denoted by an asterisk (*). Numbers to the right of the alignment correspond to the relative position of nucleotides from the start codon, ATG (nucleotides 1-3). **B)** Gene structure diagrams for the version of *PHYSCODILLO2* present in the draft genome (theoretical) and of that confirmed by sequencing (experimental). Exons are numbered and indicated by black boxes and introns denoted by an adjoining horizontal line. An asterisk (*) indicates the presence of a new exon present in the *PHYSCODILLO2* genomic DNA sequence (Emma Smiles, unpublished).

A)

```
PHYSCODILLO2.theoretical       GTTGAGGCTTGGACAGTGGGTAACGACAGCAGTGGTTCCCAATGCAA--- 5347
PHYSCODILLO2.experimental      GTTGAGGCTTGGACAGTGGGTAACGACAGCAGTGGTTCCCAATGCAATTG 5350
                               *********************************************

PHYSCODILLO2.theoretical       ---------------------------------------------------
PHYSCODILLO2.experimental      GACGTGGACAAAATTTGAGAGTTCGGCGTGGTTGGTTGCAGTGGAGCGGA 5400

PHYSCODILLO2.theoretical       ---------------------------------------------------
PHYSCODILLO2.experimental      GCTTGGACGGTTTGTGGCCATGTTGAGGAATGGTTCTGCAGTTCTCCGAA 5450

PHYSCODILLO2.theoretical       ---------------------------------------------------
PHYSCODILLO2.experimental      CATGTGCTGCGTTTGCTCTTCTGCAGGTAATCACTGGTGCTGATCGACAG 5500

PHYSCODILLO2.theoretical       ---------------------------------------------------
PHYSCODILLO2.experimental      GACAGAGTTGACGGAGCGATGGGGTGGTTGTATCATTTTTAGGTTCGTTT 5550

PHYSCODILLO2.theoretical       ---------------------------------------------------
PHYSCODILLO2.experimental      TGTTACGGGGAGTGTTGGACTGAGGTTTAAGGTACCTTTCGACTTAGGTT 5600

PHYSCODILLO2.theoretical       --------------------------------GGTTTACTGACCCGTAGGG 5366
PHYSCODILLO2.experimental      TACTGACCCGTAGGGTGTGGGTGTGTGGAACGGTTTACTGACCCGTAGGG 5650
                                                               ******************

PHYSCODILLO2.theoretical       TGTGGGTGTGTGGAACAGTTTACTATGCCTGGAGGTAGGCATGCGATGCA 5416
PHYSCODILLO2.experimental      TGTGGGTGTGTGGAACAGTTTACTATGCCTGGAGGTAGGCATGCGATGCA 5700
                               *************************************************
```

B)

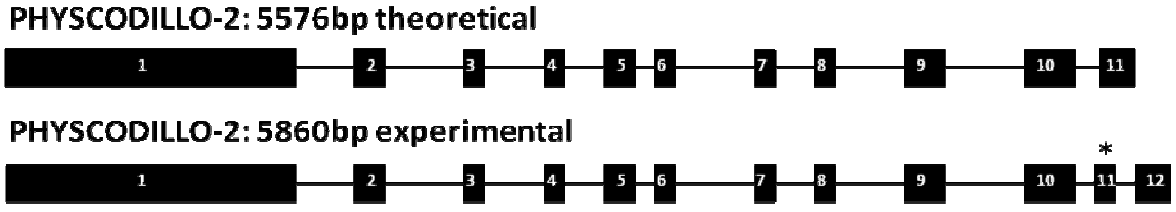
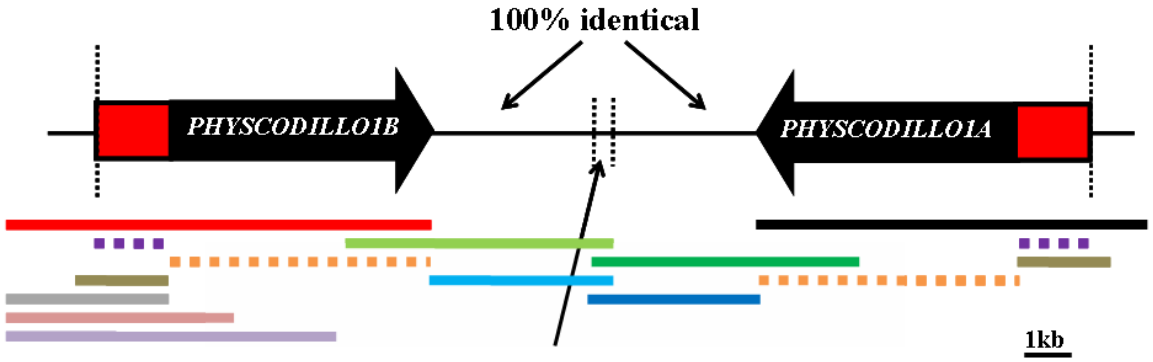


Figure 3.3 *PHYSCODILLO1A* and *PHYSCODILLO1B* exist as an identical inverted repeat in a tail-to-tail orientation.

A) Different primer pairs were used to amplify various portions of the *PHYSCODILLO1A/1B* locus. These are indicated by a series of coloured lines below the diagram. Sequences that distinguish between *PHYSCODILLO1A* and *PHYSCODILLO1B* are indicated by solid lines. Sequences that confirm the presence of either *PHYSCODILLO1A* or *PHYSCODILLO1B*, but do not distinguish between them, are represented by hashed lines. Sequencing across the previously annotated gap sequence enabled re-annotation of the *PHYSCODILLO1A/1B* locus. Both *PHYSCODILLO1A* and *PHYSCODILLO1B* genes existed as an inverted repeat and were 100% identical (black arrow) and this includes their promoter regions (red box). The 4kb intergenic regions were also duplicated, although a few changes were identified. Limits of the inverted repeat are indicated by the two outermost black hashed lines and symmetry indicated by the innermost black hashed lines.

B) *PHYSCODILLO1A* and *PHYSCODILLO1B* intergenic regions were sequenced and aligned to the intergenic sequence (annotated in the *Physcomitrella* genome resource) using ClustalW2. P3.I/G.I, II and III are the sequenced products of three independent PCR reactions utilising a forward *PHYSCODILLO1A/1B* gene-specific primer and the reverse primers underlined in the alignment. P1.I/G.I represents the sequenced data collected from three independent PCR reactions utilising the same gene-specific primer and the reverse complement of the underlined primers. I/G.RC represents the reverse complement of the intergenic sequence, annotated in the draft version of the *Physcomitrella* genome. Identical nucleotides are denoted by an asterisk (*). Numbers to the right of the alignment correspond to the relative position of nucleotides from the stop codon (TAA) of either *PHYSCODILLO1A* or *PHYSCODILLO1B*. Nucleotides that differ between the two intergenic sequences have been highlighted in black. Additional sequences that were present in the *PHYSCODILLO1A* intergenic region and NOT in the *PHYSCODILLO1B* intergenic region have been indicated in red type.

A)



B)

```
P3.I/G.I        CTATATGCCTTAGTACCTTCACAACAATCAAGAAATATACACTTTGTACCTTTTGCATCA 3660
P3.I/G.II       CTATATGCCTTAGTACCTTCACAACAATCAAGAAATATACACTTTGTACCTTTTGCATCA 3660
P3.I/G.III      CTATATGCCTTAGTACCTTCACAACAATCAAGAAATATACACTTTGTACCTTTTGCATCA 3660
P1.I/G.I        CTATATGCCTTAGTACCTTCACAACAATCAAGAAATATACACTTTGTACCTTTTGCATCA 3660
I/G.RC          CTATATGCCTTAGTACCTTCACAACAATCAAGAAATATACACTTTGTACCTTTTGCATCA 3660
                ************************************************************

P3.I/G.I        AGTTTGGACCTCATACAATCTAGTATTTTTTTATATGCAATGCACCCAAAGGTCCTGATA 3720
P3.I/G.II       AGTTTGGACCTCATACAATCTAGTATTTTTTTATATGCAATGCACCCAAAGGTCCTGATA 3720
P3.I/G.III      AGTTTGGACCTCATACAATCTAGTATTTTTTTATATGCAATGCACCCAAAGGTCCTGATA 3720
P1.I/G.I        AGTTTGGACCTCATACAATCTAGTATTTTTTTATATGCAATGCACCCAAAGGTCCTGATA 3720
I/G.RC          AGTTTGGACCTCATACAATCTAGTATTTTTTTATATGCAATGCACCCAAAGGTCCTGATA 3720
                ************************************************************

P3.I/G.I        TGAGAAATGCAAAGTTGTCTCTCACTCCAAGCTTCTTGTGGTGTCATTGATGCAAGGGCA 3780
P3.I/G.II       TGAGAAATGCAAAGTTGTCTCTCACTCCAAGCTTCTTGTGGTGTCATTGATGCAAGGGCA 3780
P3.I/G.III      TGAGAAATGCAAAGTTGTCTCTCACTCCAAGCTTCTTGTGGTGTCATTGATGCAAGGGCA 3780
P1.I/G.I        TGGAAAATGCAAAGTTGTCTCTCACTCCAAGCTTCTTGTGGTGTCATTGATGCAAGAGCA 3780
I/G.RC          TGGAAAATGCAAAGTTGTCTCTCACTCCAAGCTTCTTGTGGTGTCATTGATGCAAGAGCA 3780
                **  ********************************************************* ***

P3.I/G.I        ATTGTTGGATATCTATTTCTTGTGTACAT------------------------------- 3809
P3.I/G.II       ATTGTTGGATATCTATTTCTTGTGTACAT------------------------------- 3809
P3.I/G.III      ATTGTTGGATATCTATTTCTTGTGTACAT------------------------------- 3809
P1.I/G.I        ATTGTTGGACACGTATTTCATGTGTACATTGGATTCAGGACAACTTTCATCCAAGAGTCA 3840
I/G.RC          ATTGTTGGACACGTATTTCATGTGTACATTGGATTCACGACAACTTTCATCCAAGAGTCA 3840
                ********* *  ****** *********

P3.I/G.I        ------------------------------------------------------------
P3.I/G.II       ------------------------------------------------------------
P3.I/G.III      ------------------------------------------------------------
P1.I/G.I        TAGCCAAGTTTTGGACGAAAGCTTAATCCAAGAGTCAAAAACCTTAGCTATAAGTTTTGGA 3900
I/G.RC          TAGCCAAGTTTTGGACGAAAGCTTAATCCAAGAGTCAAAAACCTTAGCTATAAGTTTTGGA 3900

P3.I/G.I        --------------------
P3.I/G.II       --------------------
P3.I/G.III      --------------------
P1.I/G.I        CGAAAGCTATGGTGAATGCA 3920
I/G.RC          CGAAAGCTATGGTGAATGCA 3920
```

3.4 Confirming the correct number of *PHYSCODILLO* genes by Southern blotting

In order to confirm the presence of three *Physcomitrella ARABIDILLO* homologues in the *Physcomitrella* genome, Southern blotting (Southern, 1975) was carried out using a variety of hybridisation probes specific to regions both within the *PHYSCODILLO2* gene and the *PHYSCODILLO1A/1B* locus.

To confirm the presence of the *PHYSCODILLO2* gene, *Physcomitrella* genomic DNA was digested to completion using *EcoRI* and *BglII* (**Figure 3.4A**). Using a radiolabelled probe that specifically hybridised to the 5' end of the *PHYSCODILLO2* gene, readily distinguishable bands were observed after a 3 day exposure to film; 7.6kb for *EcoRI*, 3.5kb and 4.1kb for *BglII* (**Figure 3.4B** and **3.4C**). The presence of these bands confirmed that a single copy of the *PHYSCODILLO2* gene was present in the *Physcomitrella* genome.

Southern blotting was also carried out in order to confirm the presence of either a full-length *PHYSCODILLO1A* or *PHYSCODILLO1B* gene. The presence of restriction sites within the *PHYSCODILLO1A/1B* locus had already been determined by sequencing. *Physcomitrella* genomic DNA was digested to completion using *EcoRI*, *HindIII* and *BclI* (**Figure 3.5A**). Using a radiolabelled probe specific to the 3' end of either *PHYSCODILLO1A* or *PHYSCODILLO1B* genes and a set of restriction enzymes that could cut both within the genes themselves and within the two 4kb intergenic regions, readily distinguishable band sizes were expected. These bands (*EcoRI* - 6.5kb; *HindIII* – 10.8kb; *BclI* – 4.4kb) confirmed the presence of a full-length *PHYSCODILLO1A* OR *PHYSCODILLO1B* gene (**Figure 3.5B** and **3.5C**). However, it was not

possible to determine whether both genes were present in the same genetic locus since they produced identical banding patterns using the probe and restriction enzymes selected.

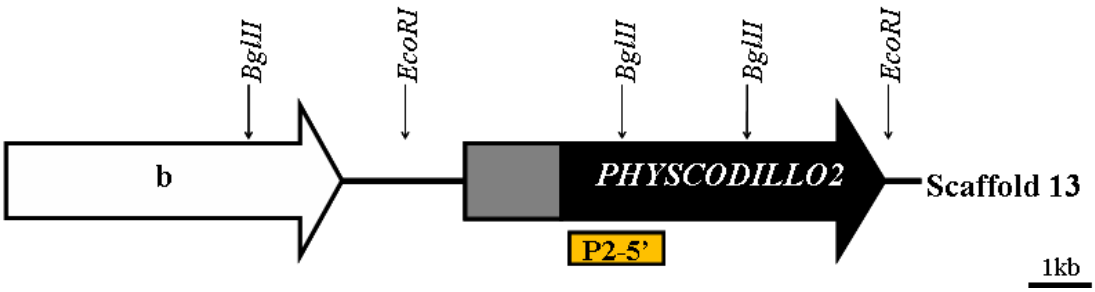
Figure 3.4 Southern blot analyses to confirm the presence of the *PHYSCODILLO2* gene.

A) Southern blots were carried out using a radiolabelled probe homologous to the 5' end of the *PHYSCODILLO2* gene (P2-5'). The *PHYSCODILLO2* gene and its promoter sequence are indicated by a black arrow and grey box respectively. The probe is indicated by a yellow box and *EcoRI* and *BglII* sites both within the *PHYSCODILLO2* gene and in adjacent sequences are indicated by arrows.

B) Using *EcoRI* and *BglII*, readily distinguishable fragment sizes were expected. *BglII* produces two different fragments since it cuts within the sequence homologous to the P2-5' probe sequence.

C) 15µg *Physcomitrella* genomic DNA was digested to completion using *EcoRI* and *BglII*, transferred to nitrocellulose membrane and hybridised using the P2-5' probe overnight at 65^{o}C. Blots were subsequently washed at 42^{o}C and exposed to film for 3 to 7 days. Correctly sized fragments were observed in both lanes. Bands have been indicated by arrows.

A)



B)

Restriction enzyme	Expected fragment size corresponding to *PHYSCODILLO2* (kb)
EcoRI	7.6
BglII	3.5 and 4.1

C)

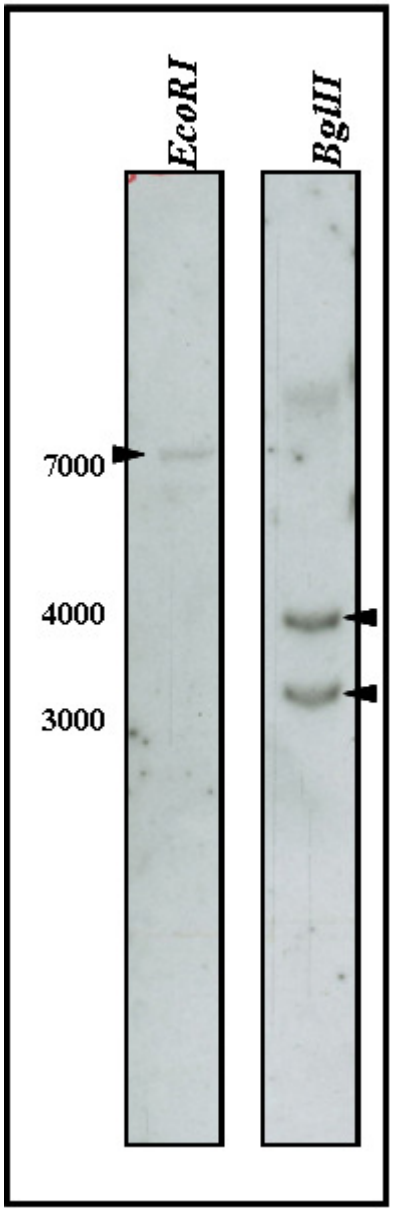


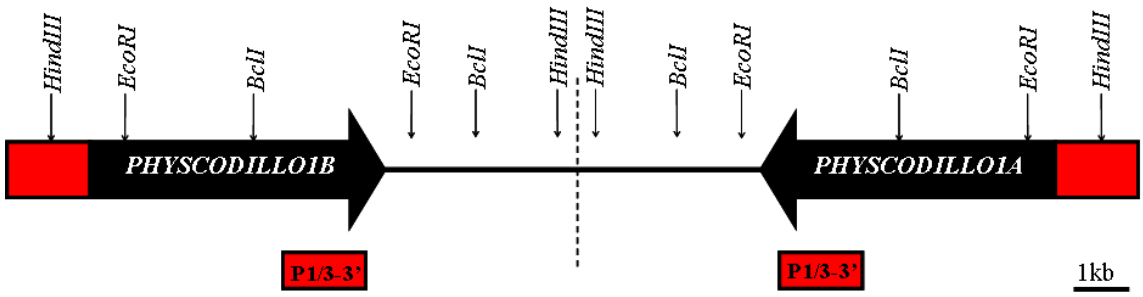
Figure 3.5 Southern blot analyses to confirm the presence of either *PHYSCODILLO1A* or *PHYSCODILLO1B* genes.

A) Southern blots were carried out using a radiolabelled probe homologous to the 3' end of either *PHYSCODILLO1A* or *PHYSCODILLO1B* genes (P1/3-3'). Full-length *PHYSCODILLO1A* and *PHYSCODILLO1B* genes and their promoter sequences are indicated by black arrows and red boxes respectively. The probe is indicated by a red box. *EcoRI*, *HindIII* and *BclI* sites within the *PHYSCODILLO1A/1B* locus are indicated by arrows.

B) Using *EcoRI*, *HindIII* and *BclI*, readily distinguishable fragment sizes were expected.

C) 15µg *Physcomitrella* genomic DNA was digested to completion using *EcoRI*, *HindIII* and *BclI*, transferred to nitrocellulose membrane and hybridised using the P1/3-3' probe overnight at 65^{o}C. Blots were subsequently washed at 42^{o}C and exposed to film for 3 to 7 days. Correctly sized fragments were observed in all lanes, confirming the presence of a full-length *PHYSCODILLO1A* or *PHYSCODILLO1B* gene. Bands have been indicated by arrows.

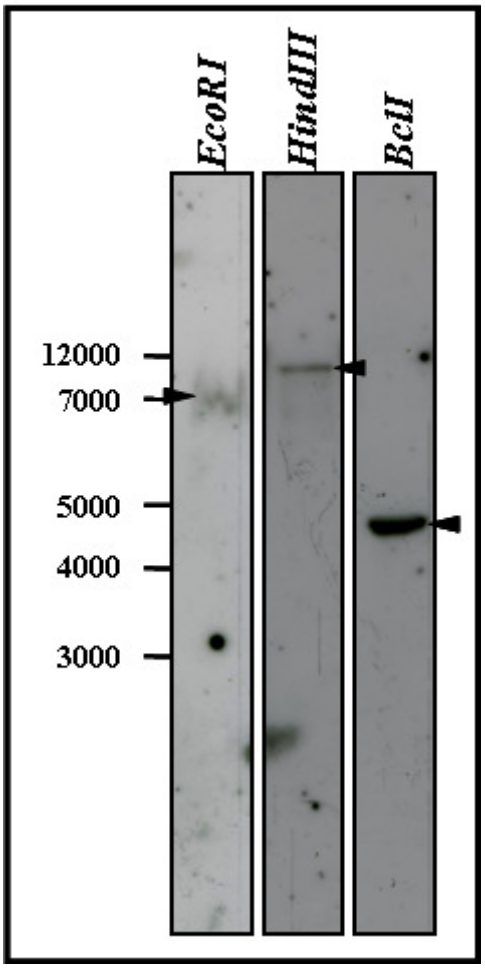
A)



B)

Restriction enzyme	Expected fragment size corresponding to *PHYSCODILLO1A* OR PHYSCODILLO1B (kb)
EcoRI	**6.5**
HindIII	**10.8**
BclI	**4.4**

C)



A different strategy was used to distinguish between *PHYSCODILLO1A* and *PHYSCODILLO1B* genes and verify whether three ARABIDILLO homologues existed in the *Physcomitrella* genome.

The *PHYSCODILLO1A/1B* locus was digested using a different set of restriction enzymes that cut both within the two genes and in divergent sequences outside of the locus. Radiolabelled probes homologous to the 1.6kb promoter region shared by both *PHYSCODILLO1A* and *PHYSCODILLO1B* genes were used (**Figure 3.6A**). Theoretically, using this strategy it would be possible to distinguish between the two identical genes since *EcoRI*, *HindIII* and *EcoRV* sites were located at different distances from the start codons of both *PHYSCODILLO1A* and *PHYSCODILLO1B* genes (**Figure 3.6A** and **B**).

An *EcoRI* digest produced two bands of 2.4kb and 2.7kb and these corresponded to the presence of *PHYSCODILLO1A* and *PHYSCODILLO1B* respectively. A *HindIII* digest produced three bands, two of which were expected (1.5kb – *PHYSCODILLO1A*; 2.5kb – *PHYSCODILLO1B*) and one that was not (approximately 3kb). The presence of the unexpected band could be explained by an incomplete digest at one of the *HindIII* sites outside of the locus, resulting in a larger than expected band. Since the *HindIII* site was within the promoter region itself, a 10.8kb band was also expected, but this may have been too faint to detect, especially above the high background on the film.

An *EcoRV* digest produced a correct band of 3.8kb that corresponded to the *PHYSCODILLO1B* gene. However, the 9.4kb band expected to detect the *PHYSCODILLO1A* gene was not detectable after 7 days exposure to film. Instead, an additional 3kb band was

detected (**Figure 3.6C**). This is likely to be due to non-specific binding of the probe to regions that share some homology with the probe sequence (i.e. 1.6kb promoter region shared by both *PHYSCODILLO1A* and *PHYSCODILLO1B* genes). Insignificant matches were detected when the promoter sequence was used in a blastn search at Phytozome. However, stretches of exact homology rarely exceeded more than eight or nine base pairs. The greatest stretch of continuous homology was detected on scaffold 223, which had 20 base pairs of exact homology. It is uncertain whether the probe would be able to recognise and bind to a sequence with such limited identity.

Although these latter blots have only been semi-informative, it was possible to confirm the presence of *PHYSCODILLO1B* in all cases. However, when combined with experimental sequencing evidence, bands present in *EcoRI* and *HindIII* digests certainly confirm the existence of both genes within a genetic locus containing both *PHYSCODILLO1A* and *-1B*. Nevertheless, Southern blots are in the process of being repeated to obtain firm and conclusive data.

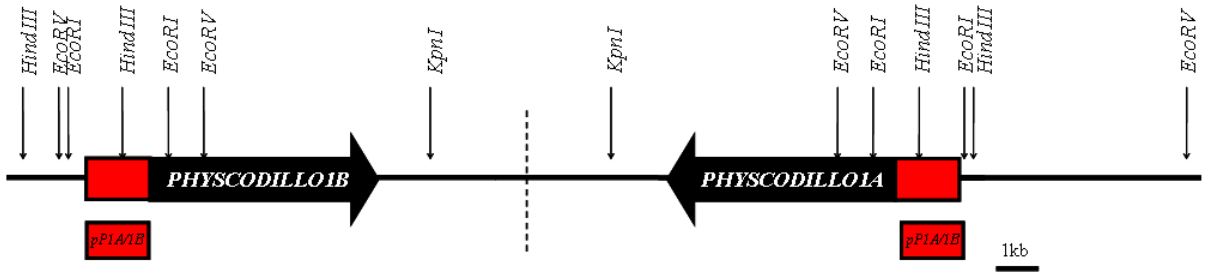
Figure 3.6 Southern blot analyses to confirm the presence of both *PHYSCODILLO1A* and *PHYSCODILLO1B* genes.

A) Southern blots were carried out using a radiolabelled probe homologous to the 1.6kb promoter region shared by both *PHYSCODILLO1A* and *PHYSCODILLO1B.* Full-length *PHYSCODILLO1A* and *PHYSCODILLO1B* genes are indicated by black arrows and red squares indicate the sites to which the promoter probe hybridises. *EcoRI*, *HindIII* and *EcoRV* sites both within and outside of the *PHYSCODILLO1A/1B* locus are indicated by arrows.

B) Using *EcoRI*, *HindIII* and *EcoRV*, readily distinguishable fragment sizes were expected.

C) 15µg *Physcomitrella* genomic DNA was digested to completion using *EcoRI*, *HindIII* and *EcoRV*, transferred to nitrocellulose and hybridised using the *PHYSCODILLO1A/1B* promoter-probe overnight at 65^{o}C. Blots were subsequently washed at 42^{o}C and exposed to film for 3 to 7 days. Readily distinguishable bands were observed in the *EcoRI* lane, *HindIII* lane and *EcoRV* lane. Additional bands were also identified and some larger expected bands were absent. Correctly sized bands have been indicated by arrows.

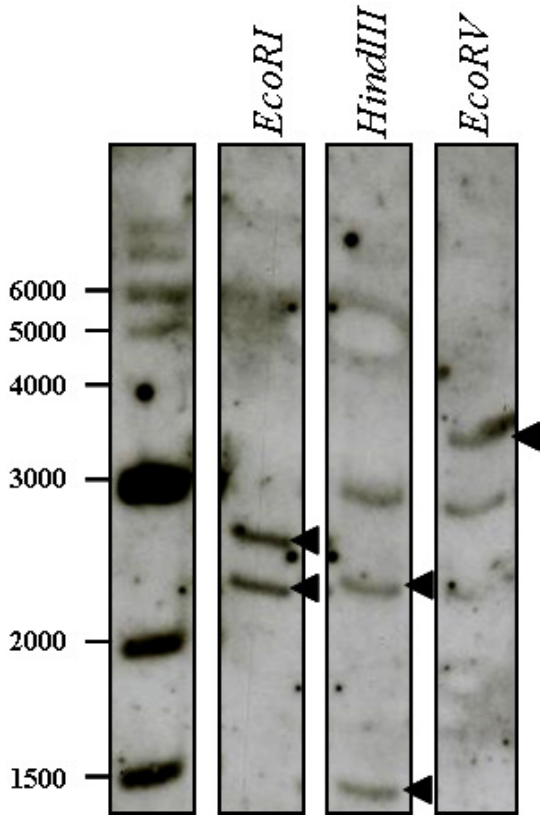
A)



B)

Restriction enzyme	Expected fragment size corresponding to *PHYSCODILLO1A* (kb)	Expected fragment size corresponding to *PHYSCODILLO1B* (kb)
EcoRI	**2.4**	**2.7**
HindIII	**1.5 and 10.8**	**2.5 and 10.8**
EcoRV	**9.4**	**3.8**

C)



3.5 *PHYSCODILLO* gene sequence analysis

Full-length genomic DNA sequences encoding PHYSCODILLO1A, PHYSCODILLO1B and PHYSCODILLO2 proteins were aligned using ClustalW2 (see **Appendix 8.2**). It had already been determined that *PHYSCODILLO1A* and *PHYSCODILLO1B* were 100% identical. *PHYSCODILLO1A* and *PHYSCODILLO1B* shared 70% identity with *PHYSCODILLO2* throughout the entire length of their nucleotide sequences. All three genes contained 12 exons and 11 introns and this was determined by aligning sequenced genomic DNA sequences to their corresponding sequenced cDNA products (**Figure 3.7** and **Appendix 8.3**).

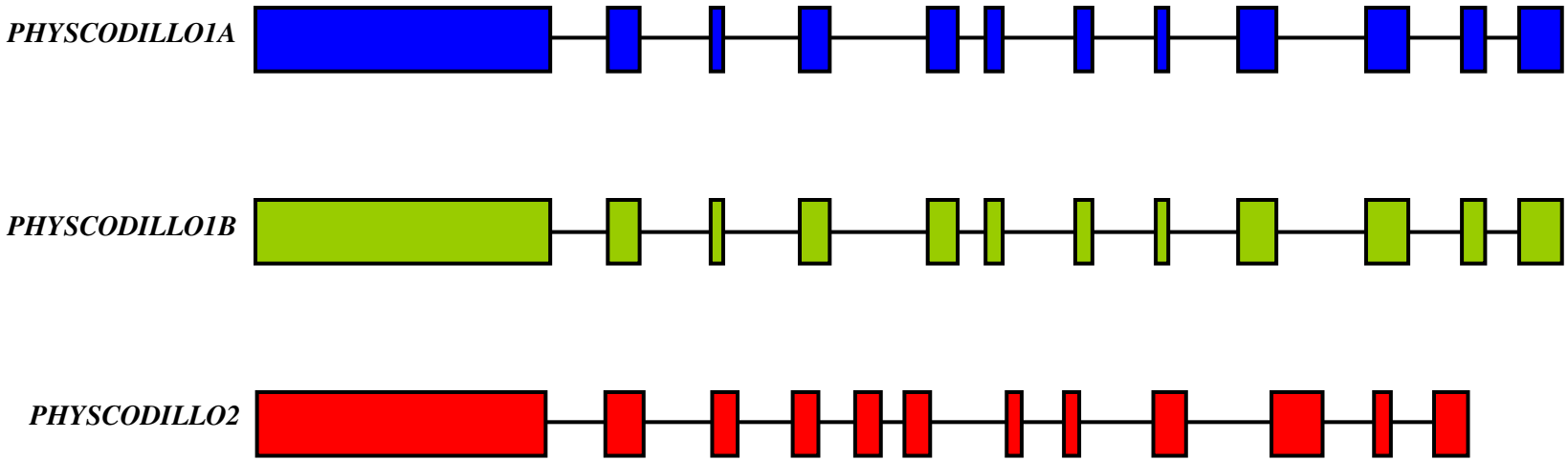


Figure 3.7 Exon-intron structure of the *PHYSCODILLO* genes.

PHYSCODILLO1A and *PHYSCODILLO2* full-length genomic DNA sequences were aligned using ClustalW2. Box diagrams for *PHYSCODILLO1A*, *PHYSCODILLO1B* and *PHYSCODILLO2* were constructed using GSDS (http://gsds.cbi.pku.edu.cn/index.php) from sequences obtained for genomic DNA and cDNA. Exons are indicated by boxes and introns denoted by adjoining horizontal lines. *PHYSCODILLO1A*, *PHYSCODILLO1B* and *PHYSCODILLO2* all have 12 exons and 11 introns.

3.6 PHYSCODILLO proteins are highly related to *Arabidopsis* Armadillos

PHYSCODILLO protein sequences were obtained by translating full-length cDNA sequences at the ExPASy Proteomics Server (http://expasy.org/tools/dna.html). Since it was not possible to differentiate between transcripts produced by either *PHYSCODILLO1A* or *PHYSCODILLO1B* genes, their corresponding proteins have been collectively referred to as PHYSCODILLO1A/1B. PHYSCODILLO1A/1B and PHYSCODILLO2 protein sequences were both 941 amino acid residues long and were 85% identical throughout their entire lengths. This was significantly more similar than the published genome suggested due to misannotation of the *PHYSCODILLO2* gene (**Figure 3.8**). PHYSCODILLO1A/1B was 61% identical to ARABIDILLO1 and 59% identical to ARABIDILLO2. PHYSCODILLO2 was 61% identical to ARABIDILLO1 and 60% identical to ARABIDILLO2. Protein sequences were scanned for the presence of conserved motifs at Pfam (http://pfam.sanger.ac.uk/), SMART (http://smart.embl-heidelberg.de/) and InterProScan (http://www.ebi.ac.uk/Tools/pfa/iprscan/). The ARABIDILLOs and PHYSCODILLOs shared the same structural characteristics. All possessed a nuclear localization signal (NLS; residues 3-8), a C-terminal F-box domain (61-109 in PHYSCODILLO1A/1B; 57-105 in PHYSCODILLO2), a linker region composed of eight leucine-rich repeats (LRR; 109-327 in PHYSCODILLO1A/1B; 105-323 in PHYSCODILLO2) and nine Armadillo (ARM) repeats (388-772 in PHYSCODILLO1A/1B; 384-771 in PHYSCODILLO2; **Figure 3.9**).

Figure 3.8 Alignment of the PHYSCODILLO protein sequences.

PHYSCODILLO1A/1B and *PHYSCODILLO2* cDNA sequences were translated using the ExPASy Translate tool (http://expasy.org/tools/dna.html). PHYSCODILLO protein sequences were then aligned using ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html). Identical amino acids are denoted by an asterisk (*), conserved amino acids denoted by a colon (:) and semi-conserved (similar structure but different chemical properties) amino acids by a period (.). Residues without annotation indicate amino acids that are neither identical nor conserved. Numbers to the right of the alignment correspond to the relative position of amino acid residues from the start residue, methionine.

```
PHYSCODILLO1A      MSNKRRRNVNVAVDDQEQQQAVVYKKARITSSPTSSAASTCGAPAVPATESGVESDAGRD 60
PHYSCODILLO1B      MSNKRRRNVNVAVDDQEQQQAVVYKKARITSSPTSSAASTCGAPAVPATESGVESDAGRD 60
PHYSCODILLO2       MSNKRRRSVNVAVEEQEQQ-PVLFKKARITPS---SSSSACSAPAVSVAESGVGIAGDRD 56
                   *******.*****::**** .*::*******.*   *::*:*.****..:****   ..**
PHYSCODILLO1A      AHWTSLPDETVLGLFNLLNHRDRASLASVCKGWQVLGSSPSLWNSLDLRSHSLNSEMVSA 120
PHYSCODILLO1B      AHWTSLPDETVLGLFNLLNHRDRASLASVCKGWQVLGSSPSLWNSLDLRSHSLNSEMVSA 120
PHYSCODILLO2       AHWTELPDDTVFGLFNLLNYRDRASLASVCRAWRGLGSSTSLWTSLDLRLHRLDSEVVSA 116
                   ****.***:**:*******:**********:.*: ****.***.***** * *:**:***
PHYSCODILLO1A      LAGRCSNLEALKFRRGASASSIVGLQAKGLRELSGDCCSQLSDATLSMVVARHANLESLL 180
PHYSCODILLO1B      LAGRCSNLEALKFRRGASASSIVGLQAKGLRELSGDCCSQLSDATLSMVVARHANLESLL 180
PHYSCODILLO2       LAGRCSNLQCLKFRGGAFANSIVGLHARELRELSGDWCSLLSDATLSMVVARHGNLESLQ 176
                   ********:.**** ** *.*****:*: ******* ** *************.*****
PHYSCODILLO1A      LGSDCERVTSEALKVIAVCCPKLRRLCVSGVLKVERDAIQALFQHCKGLTELGFLDSHTI 240
PHYSCODILLO1B      LGSDCERVTSEALKVIAVCCPKLRRLCVSGVLKVERDAIQALFQHCKGLTELGFLDSHTI 240
PHYSCODILLO2       LGSDCERVTSEALKVVAVCCPKLRRLCISGIRDVDRDAIQAMFQHCQGLTELGFLDSHTI 236
                   ***************:***********:**: .*:******:****:*************
PHYSCODILLO1A      DEGAFGGASSLRFLSVAGCRCIVWSTAAHWWSKLPNLAGLDVSRTDITPTALMQVLAGPE 300
PHYSCODILLO1B      DEGAFGGASSLRFLSVAGCRCIVWSTAAHWWSKLPNLAGLDVSRTDITPTALMQVLAGPE 300
PHYSCODILLO2       DEGAFGAARNLRFLSVAGCRCIAWSTAAQCWSKLPNLSGLDVSRTDITPSTLAQVLACPE 296
                   ******.* .************.*****: *******:***********::* **** **
PHYSCODILLO1A      LRVVCALNCPVLEEGSNPVTLPSSKKTVVLARFTDVMEGLDALLSPSNWKEEAGSCARSR 360
PHYSCODILLO1B      LRVVCALNCPVLEEGSNPVTLPSSKKTVVLARFTDVMEGLDALLSPSNWKEEAGSCARSR 360
PHYSCODILLO2       LKVVCGLNCPLLEEGSNPIPSPSSKTAVLLARFTDLMEGLEVLLNPSNSKEEACSSGRSR 356
                   *:***.****:*******:. ****.:*:******:****:.**.*** **** *..***
PHYSCODILLO1A      CGGRARADSEVAKWTEWMLSHAVLKIAECNAPSL----LKQGIAMMLRLVQSAQEDVQER 416
PHYSCODILLO1B      CGGRARADSEVAKWTEWMLSHAVLKIAECNAPSL----LKQGIAMMLRLVQSAQEDVQER 416
PHYSCODILLO2       YGRRMRVDPEVAKWTERMLSHALLKIAESNAPSLDSFWLKQGTAMMLRLVQSAQEDVQER 416
                    * * *.*.******* *****:*****.*****    **** *****************
PHYSCODILLO1A      AASALATFVVVDDENATVDSARAEAVMNGGGIALLLGLAKSCREGVQSEAAKAIANLSVN 476
PHYSCODILLO1B      AASALATFVVVDDENATVDSARAEAVMNGGGIALLLGLAKSCREGVQSEAAKAIANLSVN 476
PHYSCODILLO2       AAAALAVFVLVDDENATVDSARAEAVMNGGGIALLLGLAKSCGEGVQSEAAKAIANLSVN 476
                   **:***.**:********************************** ******************
PHYSCODILLO1A      TEVAKRVALEGGISILAGLARSRNRWVAEEAAGGLWNLSVGEEHKGAIAGAGAIEALVGL 536
PHYSCODILLO1B      TEVAKRVALEGGISILAGLARSRNRWVAEEAAGGLWNLSVGEEHKGAIAGAGAIEALVGL 536
PHYSCODILLO2       TEVAKRVALEGGISILAALARSPNRWVAEEAAGGLWNLSVGEEHKGAIAEAGAIEALVDL 536
                   *****************.**** ************************** ********.*
PHYSCODILLO1A      AFKWPAGGEGVLERAAGALANLAADDKCSMEVAVAGGVRALVRLARFCNHEGVQEQAARA 596
PHYSCODILLO1B      AFKWPAGGEGVLERAAGALANLAADDKCSMEVAVAGGVRALVRLARFCNHEGVQEQAARA 596
PHYSCODILLO2       AFKWPAGGEGVLERAAGALANLAADDKCSMEVAVAGGVRALVRLAQFCNHEGVQEQAARA 596
                   ********************************************:***************
PHYSCODILLO1A      LANLAAHGDSNGNNAAVGREEGALEALVQLTCSNHEGVRQEAAGALWNLSFDDRNREAIA 656
PHYSCODILLO1B      LANLAAHGDSNGNNAAVGREEGALEALVQLTCSNHEGVRQEAAGALWNLSFDDRNREAIA 656
PHYSCODILLO2       LANLATHGDSNGNNAAVGREAGALEALVRLTGSNHEGVRQEAAGALWNLSFDDRNREAIA 656
                   *****:************** *******:** ****************************
PHYSCODILLO1A      AAGGVEALVALAQGCSSGSQGLQERAAGALWGLSVSEANSIAIGREGGVAPLITLAHSDF 716
PHYSCODILLO1B      AAGGVEALVALAQGCSSGSQGLQERAAGALWGLSVSEANSIAIGREGGVAPLITLAHSDF 716
PHYSCODILLO2       AAGGVEALVALAQDCSSGSQGLQERAAGALWGLSVSEANSIAIGREGGVAPLITLAHSNS 716
                   **************.*********************************************:
PHYSCODILLO1A      EDVHETAVGALWNLVFNPGNALRMVEEEGVPALVHLCSSSRSKMARFMAALALAYMFDGR 776
PHYSCODILLO1B      EDVHETAVGALWNLVFNPGNALRMVEEEGVPALVHLCSSSRSKMARFMAALALAYMFDGR 776
PHYSCODILLO2       EDVHETAVGALWNLAFNPGNALRMA-EEGVPALVHLCSSSRSKMARFMAALALAYMFDGR 775
                   **************.*********. *********************************
PHYSCODILLO1A      MDEVAVGLSSVENNGRTVSLEAARKLALRNIDAFVQTFFDPQSLTAAASSWAGATLNQVA 836
PHYSCODILLO1B      MDEVAVGLSSVENNGRTVSLEAARKLALRNIDAFVQTFFDPQSLTAAASSWAGATLNQVA 836
PHYSCODILLO2       MDEVAVRVSSGENHGRTVNLEAIRKLALRSIDAFVLTFCDQQALTAAASSWAPATLNQVA 835
                   ****** :** **:****.*** ******.***** ** * *:********* *******
PHYSCODILLO1A      ETATIQEAGHLRCS-VAELGRFVAMLRNGHAVLRSCAAFALLQFTMPGGRHALHHAVLLQ 895
PHYSCODILLO1B      ETATIQEAGHLRCS-VAELGRFVAMLRNGHAVLRSCAAFALLQFTMPGGRHALHHAVLLQ 895
PHYSCODILLO2       ETARIQEAGLLRCSGGAELGRFVAMLRNGSAVLRTCAAFALLQFTMPGGRHAMHHADLLQ 895
                   *** ***** ****  ************* ****:*****************:*** ***
PHYSCODILLO1A      KVGASRVLRTVAAAATAAMQAKVFARVVLRNLEHHQAESNTVVALS 941
PHYSCODILLO1B      KVGASRVLRTVAAAATAAMQAKVFARVVLRNLEHHQAESNTVVALS 941
PHYSCODILLO2       KTGASRVLRTVAAAATAAMQAKVFARIVLRNLEHHQAESNAVVALL 941
                   *.************************:*************:****
```

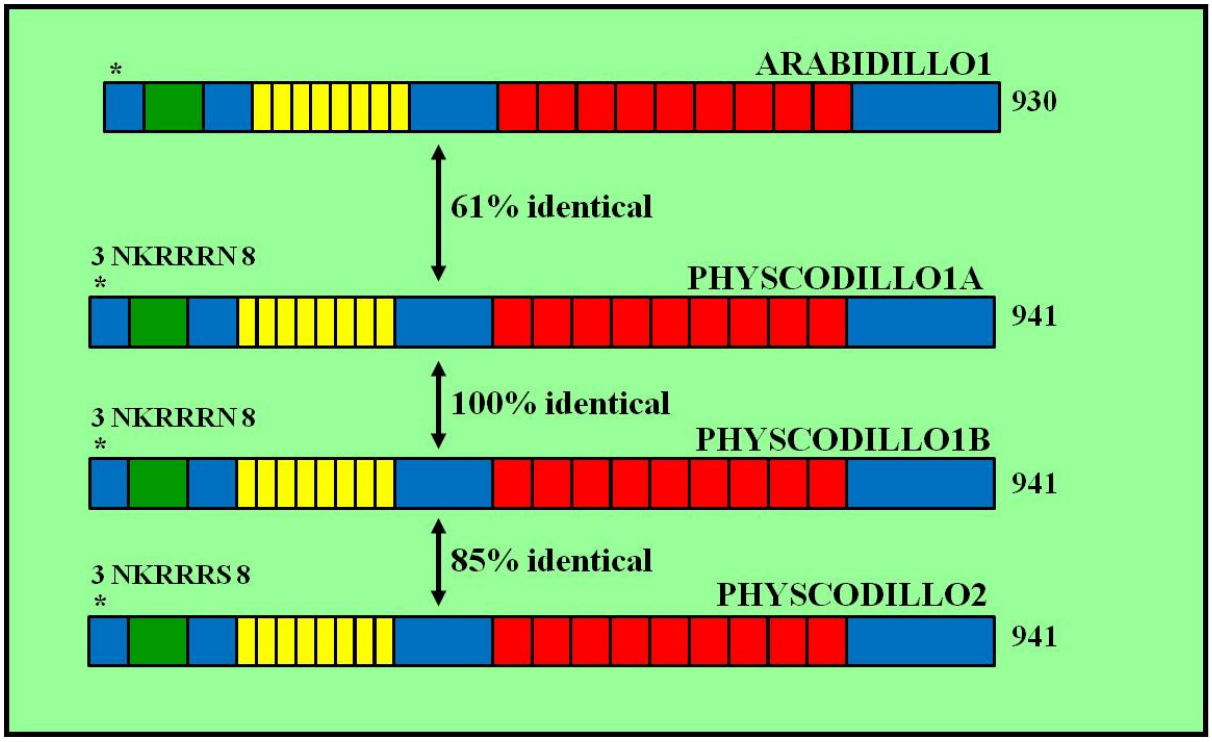


Figure 3.9 PHYSCODILLO1A/1B and PHYSCODILLO2 share the same domain architecture as the ARABIDILLOs. The ARABIDILLOs and PHYSCODILLOs have a nuclear localisation signal (NLS; residues 3-8), an N-terminal F-box domain (residues 61-109 in PHYSCODILLO1A/1B; 57-105 in PHYSCODILLO2), a linker region composed of eight leucine-rich repeats (LRR; residues 109-327 in PHYSCODILLO1A/1B; 105-323 in PHYSCODILLO2) and nine Armadillo (ARM) repeats (residues 388-772 in PHYSCODILLO1A/1B; 384-771 in PHYSCODILLO2). The NLS is indicated by an asterisk (*), F-box domains are indicated by green boxes, LRRs are indicated by yellow boxes and ARM repeats indicated by red boxes. Lengths of corresponding proteins (number of amino acid residues) are indicated to the right of each diagram. ARABIDILLO1 shares 61% identity with PHYSCODILLO1A/1B, PHYSCODILLO1A and PHYSCODILLO1B are 100% identical and these in turn are 85% identical to the PHYSCODILLO2 protein.

3.7 *PHYSCODILLO* gene expression analyses

3.7.1 *PHYSCODILLO* gene promoter sequence analysis

Promoter sequences for *PHYSCODILLO1A*, *PHYSCODILLO2* and *PHYSCODILLO1B* genes were cloned from *Physcomitrella* genomic DNA and their sequences aligned using ClustalW2 (see **Appendix 8.4**). It had already been determined that 1.6kb promoter regions immediately before the start codon of both *PHYSCODILLO1A* and *PHYSCODILLO1B* genes were 100% identical. Additional sequencing beyond this 1.6kb region of identity confirmed that sequences became divergent. The *PHYSCODILLO2* promoter sequence revealed no or very little homology to the promoter region shared by both *PHYSCODILLO1A* and *PHYSCODILLO1B* (**Figure 3.8**).

3.7.2 *PHYSCODILLO*-promoter::GUS analyses

In order to determine expression patterns of the *PHYSCODILLO* genes, *PHYSCODILLO* gene promoter sequences were inserted in-frame with a GUS reporter gene in the pHSP-GUS-108II plasmid (**Figure 3.11A**; Saidi, unpublished). Four different constructs were prepared; *pPHYSCODILLO1A/1B*::GUS (containing the 1.6kb promoter sequence shared by both *PHYSCODILLO1A* and *PHYSCODILLO1B*), *pPHYSCODILLO1A::GUS* (2.1kb promoter sequence containing 1.6kb shared with *PHYSCODILLO1B* and an additional 0.5kb sequence specific to *PHYSCODILLO1A*), *pPHYSCODILLO1B::GUS* (2.1kb promoter sequence containing 1.6kb shared with *PHYSCODILLO1A* and an additional 0.5kb sequence specific to *PHYSCODILLO1B*) and *pPHYSCODILLO2*::GUS (2.1kb of promoter sequence for consistency). Constructs were transformed into *Physcomitrella* protoplasts and targeted to the 108 locus, a site

within the genome where homologous recombination yields no detrimental phenotypes (Schaefer and Zryd, 1997). *PHYSCODILLO*-promoter::GUS lines were generated and plants expressing the transgenes were characterised to reveal *PHYSCODILLO* gene expression patterns. A number of stable *pPHYSCODILLO1A*::GUS and *pPHYSCODILLO2*::GUS lines were generated. Three of each of the lines were compared to a line constitutively expressing the GUS gene, driven from the maize ubiquitin promoter. Both p*PHYSCODILLO1A*::GUS and p*PHYSCODILLO2*::GUS lines showed identical expression patterns throughout both the haploid and diploid phases of the *Physcomitrella* life cycle, in three independent transgenic lines. Strong GUS expression was clearly visible in protonemata and mature gametophores one hour after incubation with the X-gluc substrate (pUbi::GUS expression was detected throughout the plant after approximately 20 minutes). Weak GUS expression was detected in rhizoids after 1 hour, but was clearly visible after 24 hours incubation with the X-gluc substrate. Strong GUS expression was detected throughout male antheridia but only within upper cells of female archegonia. GUS expression was also detected at the foot of the diploid sporophyte, but not elsewhere in the diploid phase of the life cycle (**Figure 3.11** and **Figure 3.12**). Stable lines expressing both *pPHYSCODILLO1B*::*GUS* and *pPHYSCODILLO1A/1B*::GUS were also generated. However, no expression was observed following 24 hour incubation with X-gluc, suggesting that sequences specific to the *PHYSCODILLO1A* gene promoter drive the expression of the GUS gene and that these are absent from the *PHYSCODILLO1B* gene promoter.

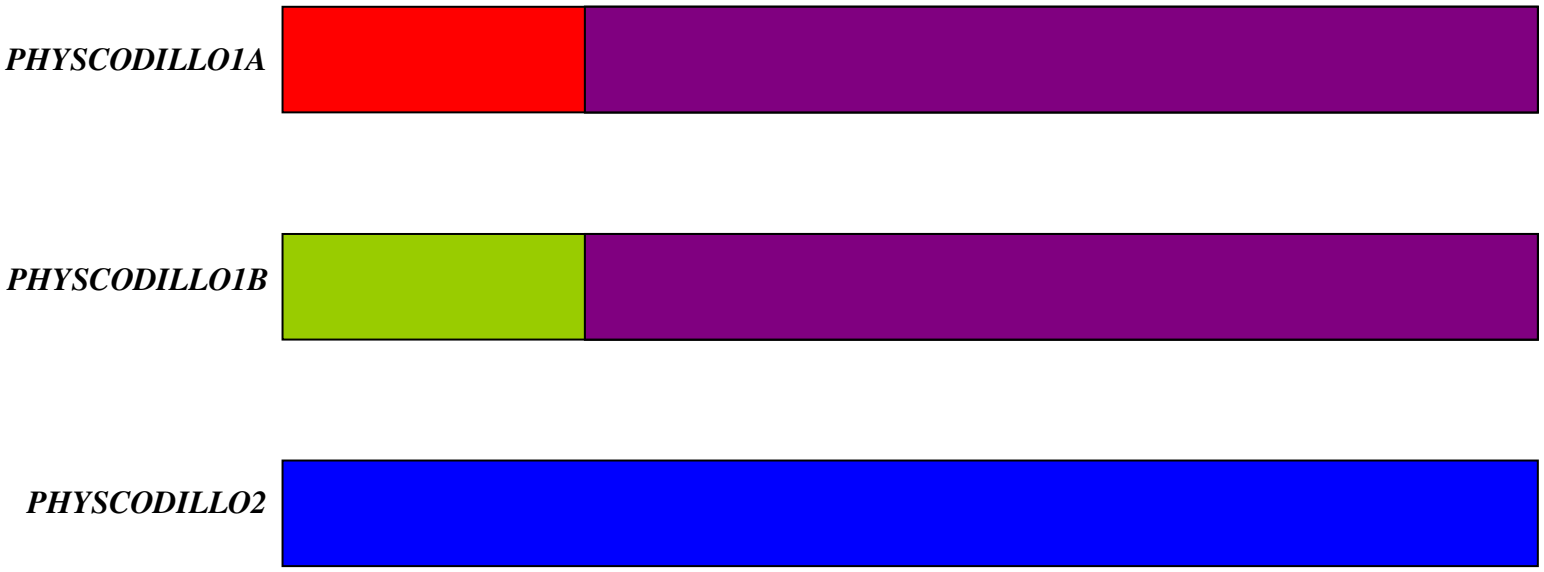


Figure 3.8 *PHYSCODILLO* gene promoters.

PHYSCODILLO1A, *-1B* and *-2* gene promoters (2.1kb) were cloned from *Physcomitrella* genomic DNA and sequenced. *PHYSCODILLO1A* and *PHYSCODILLO1B* gene promoters were 100% identical in the 1.6kb upstream of their start codons (indicated in purple). Beyond this 1.6kb, sequences (0.5kb) were no longer identical. These divergent sequences are indicated in red and green respectively. The *PHYSCODILLO2* gene promoter does not share similarity with either *PHYSCODILLO1A* or *PHYSCODILLO1B* gene promoters and has therefore been indicated in blue.

Figure 3.11 *PHYSCODILLO1A* is expressed throughout the *Physcomitrella* life cycle.

A) The *PHYSCODILLO1A* promoter sequence was amplified by PCR and inserted in-frame with a GUS gene in the pHSP-GUS-108IIB moss transformation vector that also contained a G418 cassette and a 108 targeting sequence (top image). Images show *PHYSCODILLO1A*-promoter::GUS expression in whole gametophores and rhizoids (**B**), leaves of gametophores (**C**), protonemata (**D**), antheridia (**E**), archegonia (**F**) and at the foot of sporophytes (**G**). Arrows in (**F**) and (**G**) highlight localised GUS expression seen in both archegonia and diploid sporophytes.

A)

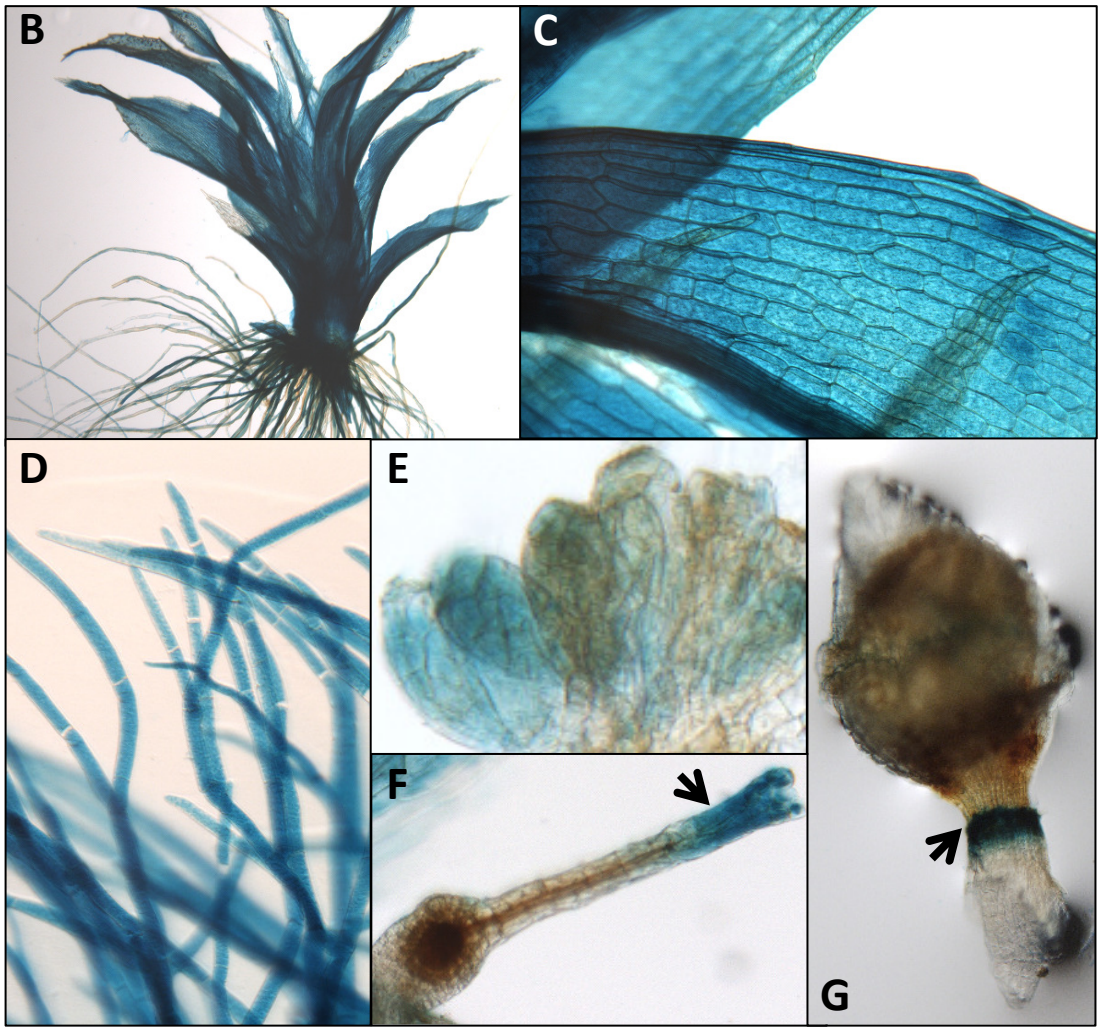
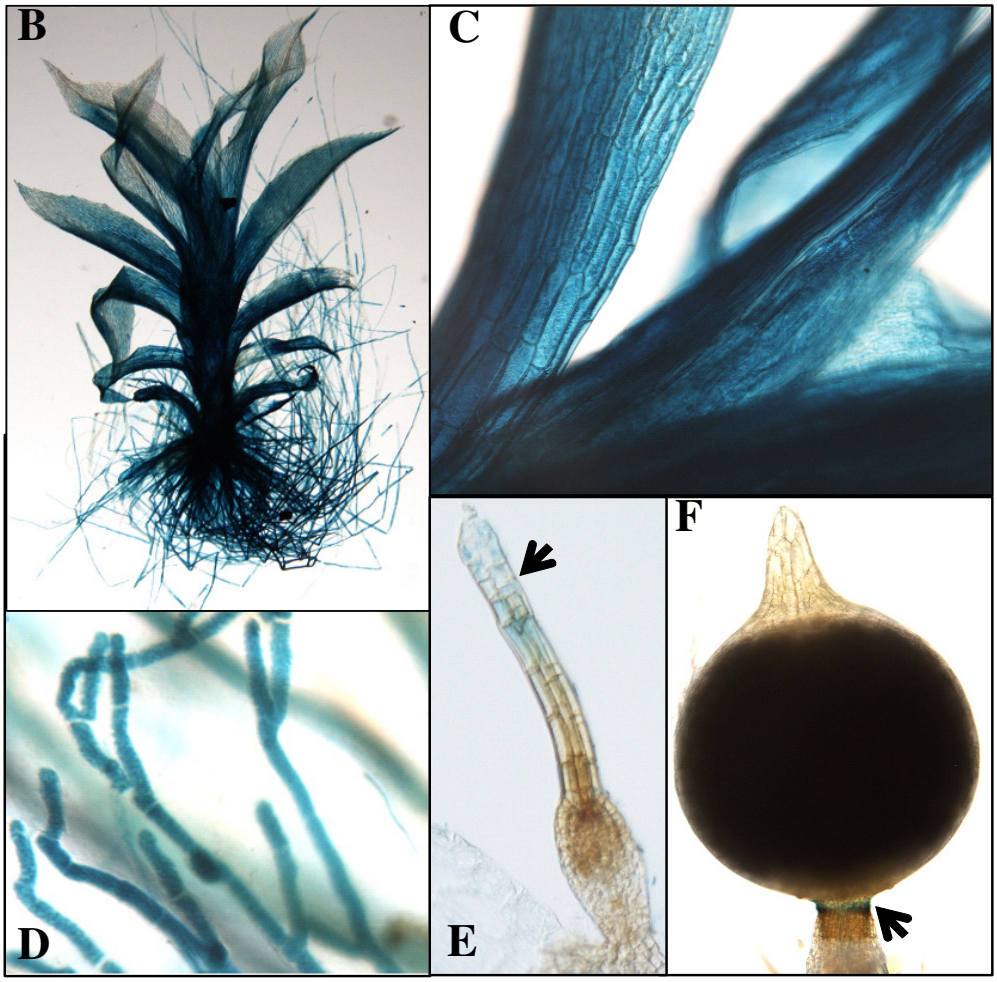


Figure 3.12 *PHYSCODILLO2* is expressed throughout the *Physcomitrella* life cycle.

A) The *PHYSCODILLO2* promoter sequence was amplified by PCR and inserted in-frame with a GUS gene in the pHSP-GUS-108IIB moss transformation vector that also contained a G418 cassette and a 108 targeting sequence (top image). Images show *PHYSCODILLO2*-promoter::GUS expression in whole gametophores and rhizoids (**B**), leaves of gametophores (**C**), protonemata (**D**), antheridia (not shown), archegonia (**E**) and at the foot of sporophytes (**F**). Arrows in (**E**) and (**F**) highlight localised GUS expression seen in both archegonia and diploid sporophytes.

A)



3.7.3 RT-PCR analysis of *PHYSCODILLO* genes

In order to confirm promoter activity by examining gene expression, RT-PCR analyses were carried out using RNA harvested from 7-day old protonemata and 3-week old gametophores. Since it was not possible to differentiate between transcripts produced from either *PHYSCODILLO1A* or *PHYSCODILLO1B* genes, they will be collectively referred to as *PHYSCODILLO1A/1B* throughout the remainder of the thesis. cDNA was prepared from total RNA and PCR carried out using primers specific to either *PHYSCODILLO1A/1B* or *PHYSCODILLO2* transcripts. RT-PCR analyses showed that *PHYSCODILLO1A/1B* and *PHYSCODILLO2* were expressed in protonemata and gametophores and their transcripts were detected with ease (**Figure 3.10** and **Figure 3.11**). *PHYSCODILLO1A/1B* appeared to be more highly expressed than *PHYSCODILLO2*; the abundance of the *PHYSCODILLO1A/1B* transcript was equivalent to the alpha-tubulin control whereas the abundance of the *PHYSCODILLO2* transcript was much lower. Interestingly though, *PHYSCODILLO1A*-promoter::GUS lines did not appear to exhibit stronger GUS expression than *PHYSCODILLO2*-promoter::GUS lines and GUS expression was detected within an hour in both lines.

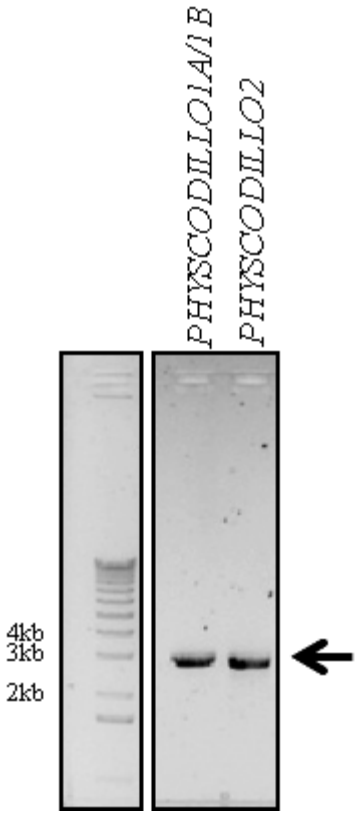


Figure 3.10 *PHYSCODILLO* genes are expressed in protonemata.

RNA was prepared from 7-day old protonemata and cDNA generated. Full-length *PHYSCODILLO1A/1B* and *PHYSCODILLO2* cDNA were successfully amplified by PCR showing that they were expressed in protonemata.

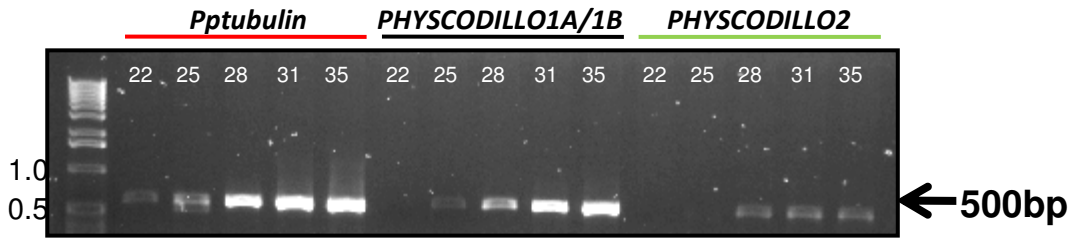


Figure 3.11 *PHYSCODILLO* genes are expressed in gametophytic tissue.

RNA was prepared from 3-week old gametophores and cDNA generated. Partial fragments of cDNA unique to either *Pptubulin* (control), *PHYSCODILLO1A/1B* or *PHYSCODILLO2* were successfully amplified by PCR (approximately 500bp). In order to look at semi-quantitative expression, a 10µl aliquot was removed from a 50µl PCR reaction after 22, 25, 28, 31 and 35 cycles and run on a gel. The presence of a band indicates the relative expression of each of the genes amplified.

3.8 *Selaginella* ARABIDILLO

Selaginella moellendorffii recently became the first lycophyte species to have its genome sequenced. Since the lycophytes evolved after the bryophytes but before the angiosperms, *Selaginella* was therefore an ideal early-evolving plant in which to identify *PHYSCODILLO*-related sequences.

Similarly to the *Physcomitrella* genome assembly, *Selaginella* genomic DNA sequences have not yet been assigned to chromosomes and are also compiled onto a series of scaffolds. The *Selaginella* genome assembly contains sequences from two different haplotypes, so each individual gene has an allelic variant that resides on an alternative scaffold sequence (Banks *et al*., 2011). The *ARABIDILLO1* genomic DNA sequence was used in a blastn search at the *Selaginella moellendorffii* genome resource at JGI (v1.0 assembly). A single putative *Selaginella* ARABIDILLO (*SELAGIDILLO*) was located on scaffold 24 of the genome assembly and its allelic variant was located on scaffold 26. In fact, all of the genes located on scaffold 24 had allelic variants on scaffold 26, suggesting that only one copy of *SELAGIDILLO* was in fact present in the genome. The allelic variants of *SELAGIDILLO* located on scaffold 24 and scaffold 26 have been assigned the protein identifiers 101488 and 102329 respectively. These were 98% identical at the nucleotide level and 99% identical at the amino acid level. The ARABIDILLO1 protein sequence was subsequently used in a blastp search at Phytozome (v7.0) and only the scaffold 24 version (101488) of *SELAGIDILLO* was identified. Cloning primers were therefore designed using the scaffold 24 genomic DNA sequence.

3.8.1 *SELAGIDILLO* genomic DNA

The copy of *SELAGIDILLO* annotated in the genome sequence was 3379 base pairs long and contained 12 exons and 11 introns. The corresponding transcript was 2736 base pairs long and encoded a 911 amino acid protein.

To confirm the presence of an ARABIDILLO homologue in *Selaginella*, full-length *SELAGIDILLO* genomic DNA and cDNA were amplified by PCR using primers specific to the full-length gene annotated in the genome. Subsequent sequencing of two independent clones revealed that the full-length *SELAGIDILLO* gene differed significantly from the two versions of the gene (Sc24 and Sc26) annotated in the genome. There were a total of 61 nucleotide substitutions and 1 extra nucleotide in the sequenced gene. Of these 61 nucleotides; 16 matched the scaffold 24 sequence, 25 matched the scaffold 26 sequence and 20 matched neither scaffold 24 nor scaffold 26 sequences. The additional base pair was identified within an intron sequence. The copy of *SELAGIDILLO* sequenced was therefore 3380 base pairs long and not 3379 as annotated (**Appendix 8.5**). The *SELAGIDILLO* cDNA sequence was 2694 base pairs long and was significantly shorter than the 2736 base pair long transcript annotated in the genome. Some of the sequence belonging to exon 10 of the annotated gene sequence was actually intron sequence (**Figure 3.14**). The *SELAGIDILLO* transcript therefore encoded an 897 amino acid protein and not a 911 amino acid protein as the annotated genome suggested (**Figure 3.15**). Consequently, the SELAGIDILLO protein was 63% identical to ARABIDILLO1 and 69% identical to PHYSCODILLO1A/1B, which was more similar to the annotated version of the protein.

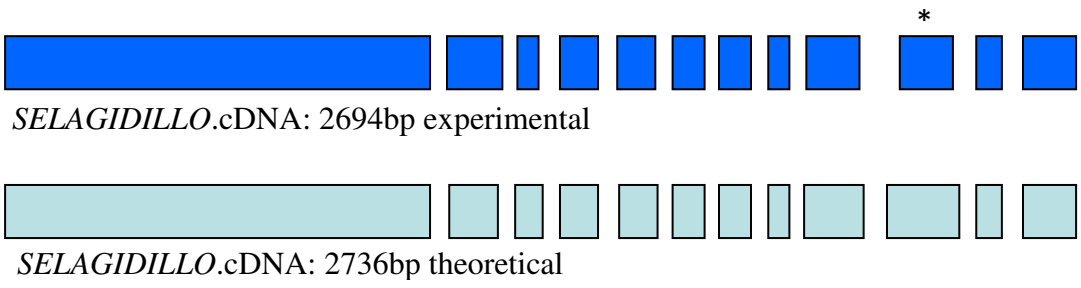
Figure 3.15 *SELAGIDILLO* was misannotated in the *Selaginella* genome.

SELAGIDILLO cDNA sequences were cloned and sequenced. **A)** Experimental sequences were aligned to theoretical sequences obtained from the *Selaginella* genome resource using ClustalW2. SelcDNA.theoretical refers to the cDNA sequence for *SELAGIDILLO* found on Sc24 of the genome sequence. SelgDNA.seq refers to the sequenced *SELAGIDILLO* gene. SelcDNA.seq refers to the sequenced *SELAGIDILLO* cDNA. Nucleotides that are NOT present in the cDNA sequence of cloned *SELAGIDILLO* are highlighted. Identical nucleotides are denoted by an asterisk (*). Numbers to the right of the alignment correspond to the relative position of nucleotides from the start codon, ATG (nucleotides 1-3). **B)** Gene structure diagrams for the version of *SELAGIDILLO* present in the genome (theoretical) and of that confirmed by sequencing (experimental). Exons are numbered and indicated by black white boxes and introns denoted by an adjoining horizontal line. An asterisk (*) indicated the absence of exon 10 sequence from the *SELAGIDILLO* cDNA sequence.

A)

```
SELcDNA.seq           AAGATGGCACGATTCATGGCCGCTCTGGCTCTGGCATACATGTTCGATGGCAG------- 2258
SELgDNA.seq           AAGATGGCACGATTCATGGCCGCTCTGGCTCTGGCATACATGTTCGATGGCAGGTACAAA 2700
SELcDNA.theoretical   AAGATGGCACGATTCATGGCCGCTCTGGCTCTGGCATACATGTTCGATGGCAGG------ 2259
                      **************************************************

SELcDNA.seq           ------------------------------------------------------------
SELgDNA.seq           ACTTTTCTTAAGTGTTCTCGAGCAGCTTTCGTGTATTCCTGGGCCATTTGTCTGAGTGCT 2760
SELcDNA.theoretical   ------------------------------------------------------------

SELcDNA.seq           ------------------------------------------------------------
SELgDNA.seq           CGTTTGCCTGCTTGTCAGGATGGATGAGGTTACTACAAACGAAGTTGTTTACTGTGATAG 2820
SELcDNA.theoretical   ------------------ATGGATGAGGTTACTACAAACGAAGTTGTTTACTGTGATAG 2300

SELcDNA.seq           CATTACCAAAAACGGCGTGGCAAGGCAGTCGGCCATGAAGAACATTGAGGCGTTTGTACA 2318
SELgDNA.seq           CATTACCAAAAACGGCGTGGCAAGGCAGTCGGCCATGAAGAACATTGAGGCGTTTGTACA 2880
SELcDNA.theoretical   CATTACCAAAAACGGCGTGGCAAGGCAGTCGGCCATGAAGAACATTGAGGCGTTTGTACA 2360
                      ************************************************************
```

B)



SELAGIDILLO.cDNA: 2694bp experimental

SELAGIDILLO.cDNA: 2736bp theoretical

3.8.2 SELAGIDILLO protein structure

The SELAGIDILLO protein sequence was obtained by translating the full-length cloned cDNA sequence at the ExPASy Proteomics Server (http://expasy.org/tools/dna.html). The SELAGIDILLO protein was 897 amino acid residues long. SELAGIDILLO was 63% identical to ARABIDILLO1 and 69% identical to PHYSCODILLO1A, PHYSCODILLO1B and PHYSCODILLO2. Protein sequences were scanned for the presence of conserved motifs at Pfam and SMART. SELAGIDILLO shared the same structural characteristics as the ARABIDILLOs and PHYSCODILLOs. SELAGIDILLO possessed a putative NLS (residues 3-8), a C-terminal F-box domain (51-99), a linker region composed of eight LRRs (118-309) and nine ARM repeats (363-751; **Figure 3.16**).

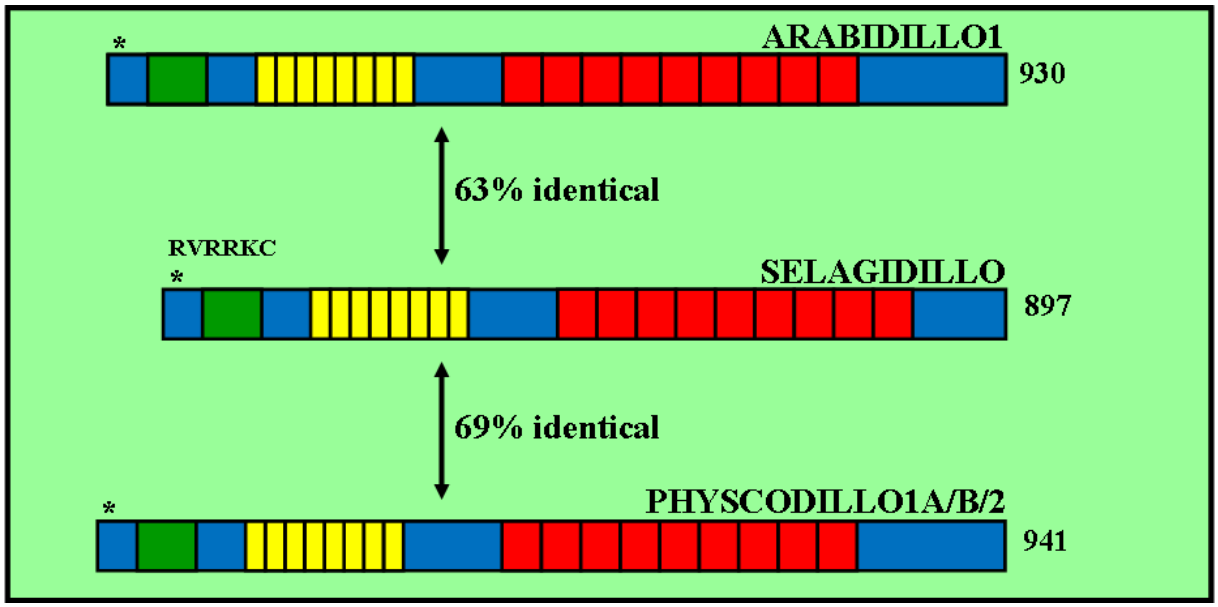


Figure 3.16 SELAGIDILLOs have the same domain architecture as the ARABIDILLO and PHYSCODILLO proteins.

SELAGIDILLO has an NLS (residues 3-8; RVRRKC), an N-terminal F-box domain (residues 51-99), a linker region composed of eight LRRs (residues 118-309) and nine ARM repeats (residues 363-751). The NLS is indicated by an asterisk (*), F-box domains are indicated by green boxes, LRRs indicated by yellow boxes and ARM repeats indicated by red boxes. Lengths of corresponding proteins are indicated to the right of each diagram. SELAGIDILLO shares 63% identity with ARABIDILLO1 and 69% identity with PHYSCODILLO1A, PHYSCODILLO1B and PHYSCODILLO2.

3.9 Plant Armadillos are highly conserved

ARABIDILLO1 and ARABIDILLO2 proteins contain an F-box domain, a linker region of eight LRRs and nine ARM repeats (Nibau *et al*., 2011). SELAGIDILLO and the PHYSCODILLO proteins have the same domain architecture and share 63% and 61% identity with ARABIDILLO1 respectively. The following sections examine the high conservation between protein features of ARABIDILLO homologues in land plants

3.9.1 The PHYSCODILLOs and SELAGIDILLO have conserved F-box domains

Gibbs *et al.* confirmed that ARABIDILLO1 and ARABIDILLO2 were *bona fide* F-box proteins. They showed that the ARABIDILLO1 F-box domain interacted with a number of *Arabidopsis* SKP1-like proteins (ASKs) in a yeast 2-hybrid assay and demonstrated that key residues within the F-box domain (L50, P51, L58 and W77) were critical for this interaction *in vitro* (Daniel Gibbs, PhD thesis; Nibau *et al*., 2011). *In vivo* however, full-length ARABIDILLO proteins carrying these mutations were fully functional and able to rescue lateral root phenotypes of *arabidillo1/2* knockout mutants (Nibau *et al*., 2011).

F-box domains were detected in PHYSCODILLO and SELAGIDILLO sequences using Pfam and Smart and aligned using ClustalW2. The key residues (L50, P51, L58 and W77) required for F-box function in *Arabidopsis* were all conserved in PHYSCODILLO F-box sequences. With the exception of P51, all of the residues were also conserved in the SELAGIDILLO F-box sequence (**Figure 3.17**). Since introducing a P51A mutation (P51 is a key

residue for F-box function in a number of other F-box proteins) into the *ARABIDILLO1* coding sequence has no effect on ARABIDILLO function, neither SELAGIDILLO nor the PHYSCODILLOs or even the ARABIDILLOs may function as F-box proteins *in vivo*.

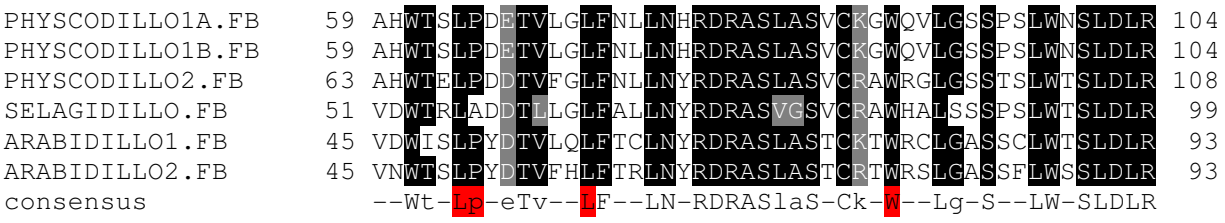


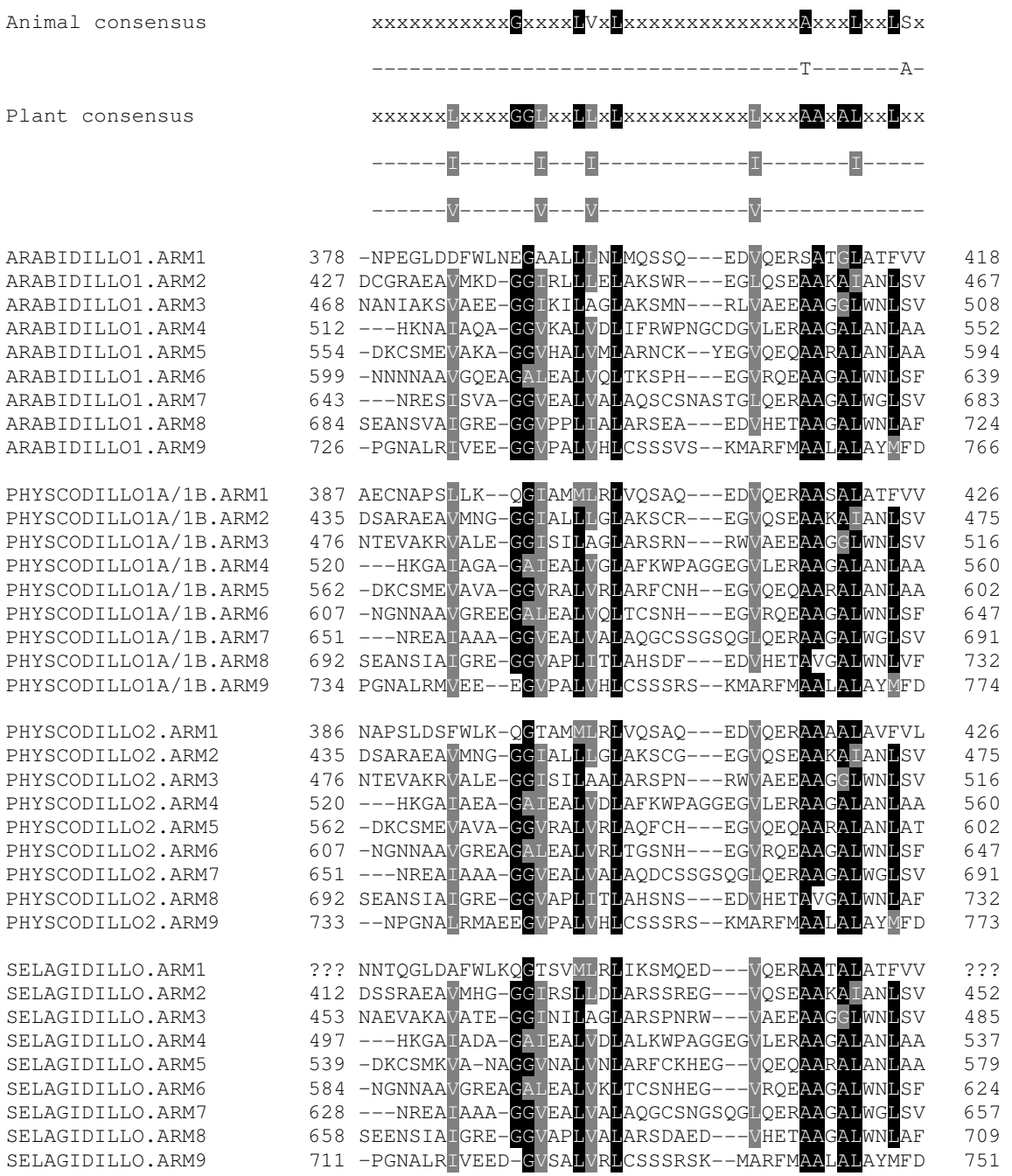
Figure 3.17 Alignment of PHYSCODILLO, SELAGIDILLO and ARABIDILLO F-box sequences.

F-box domains were identified using Pfam, aligned with ClustalW2 and displayed using BOXSHADE3.21 with the fraction of sequences (that must agree for shading) set to 0.9. Key residues (L50, P51, L58 and W77) have been highlighted in red type. Identical amino acids are denoted by black shading and conserved amino acids are denoted by grey shading. Residues without annotation indicate amino acids that are neither identical nor conserved. Numbers to the left and right of the alignment correspond to the relative position of amino acid residues from the start residue, methionine.

3.9.2 PHYSCODILLOs, ARABIDILLOs and SELAGIDILLO have conserved ARM repeats

Armadillo repeats were originally discovered in the founding member of the protein family, Armadillo from *Drosophila melanogaster*. The Armadillo protein contained tandem arrays of an imperfect 41 amino acid sequence (Riggleman *et al*., 1989). The ARM repeat regions of the ARABIDILLO proteins sequences also contained this characteristic repeating pattern (Nibau *et al*., 2011). The presence of ARM repeats in the PHYSCODILLOs and SELAGIDILLO was confirmed by Pfam, Smart and InterproScan and by aligning them to those of ARABIDILLO1. The nine ARM repeats from each of the individual proteins (ARABIDILLO1, ARABIDILLO2, PHYSCODILLO1A/1B, PHYSCODILLO2, SELAGIDILLO) were then aligned to each other. The ARM repeat consensus sequence fits that derived for animal Armadillos (Huber *et al*., 1997; **Figure 3.18**). Sequences of ARM repeats 1-9 of the PHYSCODILLOs and SELAGIDILLO were highly related to ARM repeats 1-9 of the ARABIDILLO proteins. With the exception of the conserved residues (defined by the consensus sequence), there was no significant homology between ARM repeats of the same protein.

Figure 3.18 Alignment of Armadillo (ARM) repeats. The upper sequence shows the alignment of the nine ARM repeats from the ARABIDILLO1 protein sequence. The middle sequences show the alignment of the nine ARM repeats from both PHYSCODILLO1A/1B and PHYSCODILLO2. The lower sequence shows the alignment of the nine ARM repeats from SELAGIDILLO. ARM repeats were detected using Pfam and Smart, aligned using ClustalW2 and displayed using Boxshade 3.21 (http://www.ch.embnet.org/software/BOX_form.html) with the fraction of sequences (that must agree for shading) set to 0.9 and annotation completed manually. Identical residues are denoted by black shading and conserved residues are denoted by grey shading. Residues that are neither identical nor conserved remain unshaded. Numbers to the left and right of the alignment show the relative position of each of the ARM repeats from the start residue, methionine. The consensus sequences of both animal Armadillos and plant Armadillo-related proteins can be found above the alignments. 'x' represents any amino acid. Residues marked in capital letters are always found at that site within the ARM repeat (e.g. L20). Often, different residues are present at a single site and these alternatives are indicated on lines underneath the upper consensus line (e.g. L7 or I7 or V7).


Animal consensus
xxxxxxxxxxxGxxxxLVxLxxxxxxxxxxxxxxxAxxxLxxLSx
----------T-------A-
Plant consensus
xxxxxxLxxxxGGIxxLILxLxxxxxxxxxxLxxxAAxALxxLxx
------I------I---I-----------I-------I-----
------V------V---V-----------V-------------
ARABIDILLO1.ARM1 378 -NPEGLDDFWLNEGAALLLNLMQSSQ---EDVQERSATGLATFVV 418
ARABIDILLO1.ARM2 427 DCGRAEAVMKD-GGIRLLLELAKSWR---EGLQSEAAKAIANLSV 467
ARABIDILLO1.ARM3 468 NANIAKSVAEE-GGIKILAGLAKSMN---RLVAEEAAGGLWNLSV 508
ARABIDILLO1.ARM4 512 ---HKNAIAQA-GGVKALVDLIFRWPNGCDGVLERAAGALANLAA 552
ARABIDILLO1.ARM5 554 -DKCSMEVAKA-GGVHALVMLARNCK--YEGVQEQAARALANLAA 594
ARABIDILLO1.ARM6 599 -NNNNAAVGQEAGAIEALVQLTKSPH---EGVRQEAAGALWNLSF 639
ARABIDILLO1.ARM7 643 ---NRESISVA-GGVEALVALAQSCSNASTGLQERAAGALWGLSV 683
ARABIDILLO1.ARM8 684 SEANSVAIGRE-GGVPPLIALARSEA---EDVHETAAGALWNLAF 724
ARABIDILLO1.ARM9 726 -PGNALRIVEE-GGVPALVHLCSSSVS--KMARFMAALALAYMFD 766
PHYSCODILLO1A/1B.ARM1 387 AECNAPSLLK--QGIAMMLRLVQSAQ---EDVQERAASALATFVV 426
PHYSCODILLO1A/1B.ARM2 435 DSARAEAVMNG-GGIALLLGLAKSCR---EGVQSEAAKAIANLSV 475
PHYSCODILLO1A/1B.ARM3 476 NTEVAKRVALE-GGISILAGLARSRN---RWVAEEAAGGLWNLSV 516
PHYSCODILLO1A/1B.ARM4 520 ---HKGAIAGA-GAIEALVGLAFKWPAGGEGVLERAAGALANLAA 560
PHYSCODILLO1A/1B.ARM5 562 -DKCSMEVAVA-GGVRALVRLARFCNH--EGVQEQAARALANLAA 602
PHYSCODILLO1A/1B.ARM6 607 -NGNNAAVGREEGALEALVQLTCSNH---EGVRQEAAGALWNLSF 647
PHYSCODILLO1A/1B.ARM7 651 ---NREAIAAA-GGVEALVALAQGCSSGSQGLQERAAGALWGLSV 691
PHYSCODILLO1A/1B.ARM8 692 SEANSIAIGRE-GGVAPLITLAHSDF---EDVHETAVGALWNLVF 732
PHYSCODILLO1A/1B.ARM9 734 PGNALRMVEE--EGVPALVHLCSSSRS--KMARFMAALALAYMFD 774
PHYSCODILLO2.ARM1 386 NAPSLDSFWLK-QGTAMMLRLVQSAQ---EDVQERAAAALAVFVL 426
PHYSCODILLO2.ARM2 435 DSARAEAVMNG-GGIALLLGLAKSCG---EGVQSEAAKAIANLSV 475
PHYSCODILLO2.ARM3 476 NTEVAKRVALE-GGISILAALARSPN---RWVAEEAAGGLWNLSV 516
PHYSCODILLO2.ARM4 520 ---HKGAIAEA-GAIEALVDLAFKWPAGGEGVLERAAGALANLAA 560
PHYSCODILLO2.ARM5 562 -DKCSMEVAVA-GGVRALVRLAQFCH---EGVQEQAARALANLAT 602
PHYSCODILLO2.ARM6 607 -NGNNAAVGREAGALEALVRLTGSNH---EGVRQEAAGALWNLSF 647
PHYSCODILLO2.ARM7 651 ---NREAIAAA-GGVEALVALAQDCSSGSQGLQERAAGALWGLSV 691
PHYSCODILLO2.ARM8 692 SEANSIAIGRE-GGVAPLITLAHSNS---EDVHETAVGALWNLAF 732
PHYSCODILLO2.ARM9 733 --NPGNALRMAEEGVPALVHLCSSSRS--KMARFMAALALAYMFD 773
SELAGIDILLO.ARM1 ??? NNTQGLDAFWLKQGTSVMLRLIKSMQED---VQERAATALATFVV ???
SELAGIDILLO.ARM2 412 DSSRAEAVMHG-GGIRSLLDLARSSREG---VQSEAAKAIANLSV 452
SELAGIDILLO.ARM3 453 NAEVAKAVATE-GGINILAGLARSPNRW---VAEEAAGGLWNLSV 485
SELAGIDILLO.ARM4 497 ---HKGAIADA-GAIEALVDLALKWPAGGEGVLERAAGALANLAA 537
SELAGIDILLO.ARM5 539 -DKCSMKVA-NAGGVNALVNLARFCKHEG--VQEQAARALANLAA 579
SELAGIDILLO.ARM6 584 -NGNNAAVGREAGAIEALVKLTCSNHEG---VRQEAAGALWNLSF 624
SELAGIDILLO.ARM7 628 ---NREAIAAA-GGVEALVALAQGCSNGSQGLQERAAGALWGLSV 657
SELAGIDILLO.ARM8 658 SEENSIAIGRE-GGVAPLVALARSDAED---VHETAAGALWNLAF 709
SELAGIDILLO.ARM9 711 -PGNALRIVEED-GVSALVRLCSSSRSK--MARFMAALALAYMFD 751

3.9.3 Plant Armadillos have highly conserved domain architectures

ARABIDILLO homologues are found throughout all land plant genomes sequenced so far and they all have highly conserved protein sequences. Each possesses an NLS, an F-box domain, a linker region composed of LRRs as defined by Nibau *et al*., (2011) and an ARM repeat domain towards the C-terminus of the protein. Full-length ARABIDILLO1 homologues have already been identified in the genomes of *Physcomitrella patens* (PHYSCODILLO, moss; chapter three), *Selaginella moellendorffii* (SELAGIDILLO, spike moss; chapter three), *Oryza sativa* (ORYZADILLO, rice), *Brachypodium distachyon* (BRACHYDILLO, purple false brome), *Zea mays* (ZEADILLO, maize), *Sorghum bicolor* (SORGHODILLO, Sorghum), *Vitis vinifera* (VITIDILLO1 and -2, Grape Vine), *Populus trichocarpa* (POPLARDILLO1, -2 and -3, Poplar), *Ricinis communis* (RICINODILLO1 and -2, castor oil plant), *Manihot esculenta* (MANIHODILLO1 and -2, cassava), *Mimulus guttatus* (MIMULODILLO1 and -2, monkey-flower), *Arabidopsis lyrata* (A.LYRATA1 and -2, northern rock cress), *Arabidopsis thaliana* (thale cress), *Glycine max* (GLYCINODILLO, soybean) and *Cucumis sativus* (CUCUMIDILLO, cucumber; Nibau *et al*., 2011).

A BLAST search using the ARABIDILLO1 protein sequence was carried out at the Phytozome genome browser, in order to identify additional homologues in land plants. In addition to those obtained by Nibau *et al*., homologues were identified in *Hordeum vulgare* (HORDIDILLO, barley), *Aquilegia coerulea* (AQUILIDILLO1, Colorado blue columbine), *Eucalyptus grandis* (EUCADILLO, *Eucalyptus*), *Citrus sinensis* (CITRUS.SINIDILLO, orange),

Citrus clementina (CITRUS.CLEMIDILLO, clementine) and *Prunus persica* (PRUNIDILLO, peach tree).

Only partial sequences were identified for *Setaria italica* (SETADILLO, foxtail millet) and *Carica papaya* (CARICADILLO, papaya) so these were omitted from further analyses. SETADILLO appeared to have a 50 residue truncation at its N-terminus and therefore had no NLS sequence. This is likely to have been due to an incorrectly defined start codon and consequently the protein was shorter than anticipated (873 residues). CARICADILLO appeared to be heavily truncated within the ARM repeat region at its C-terminus and consequently was only 617 residues long.

Second putative ARABIDILLO homologues were identified in the genomes of *Zea mays* (ZEADILLO2) and *Aquilegia coerulea* (AQUILIDILLO2). However, both proteins were heavily truncated at their C-termini and were only 888 and 809 residues long respectively. ZEADILLO2 and AQUILIDILLO2 were eliminated from phylogenetic analyses since full-length sequences were not available (**Table 3.1**). A full alignment of all of the ARABIDILLO homologues from land plants can be found in **Appendix 8.6**.

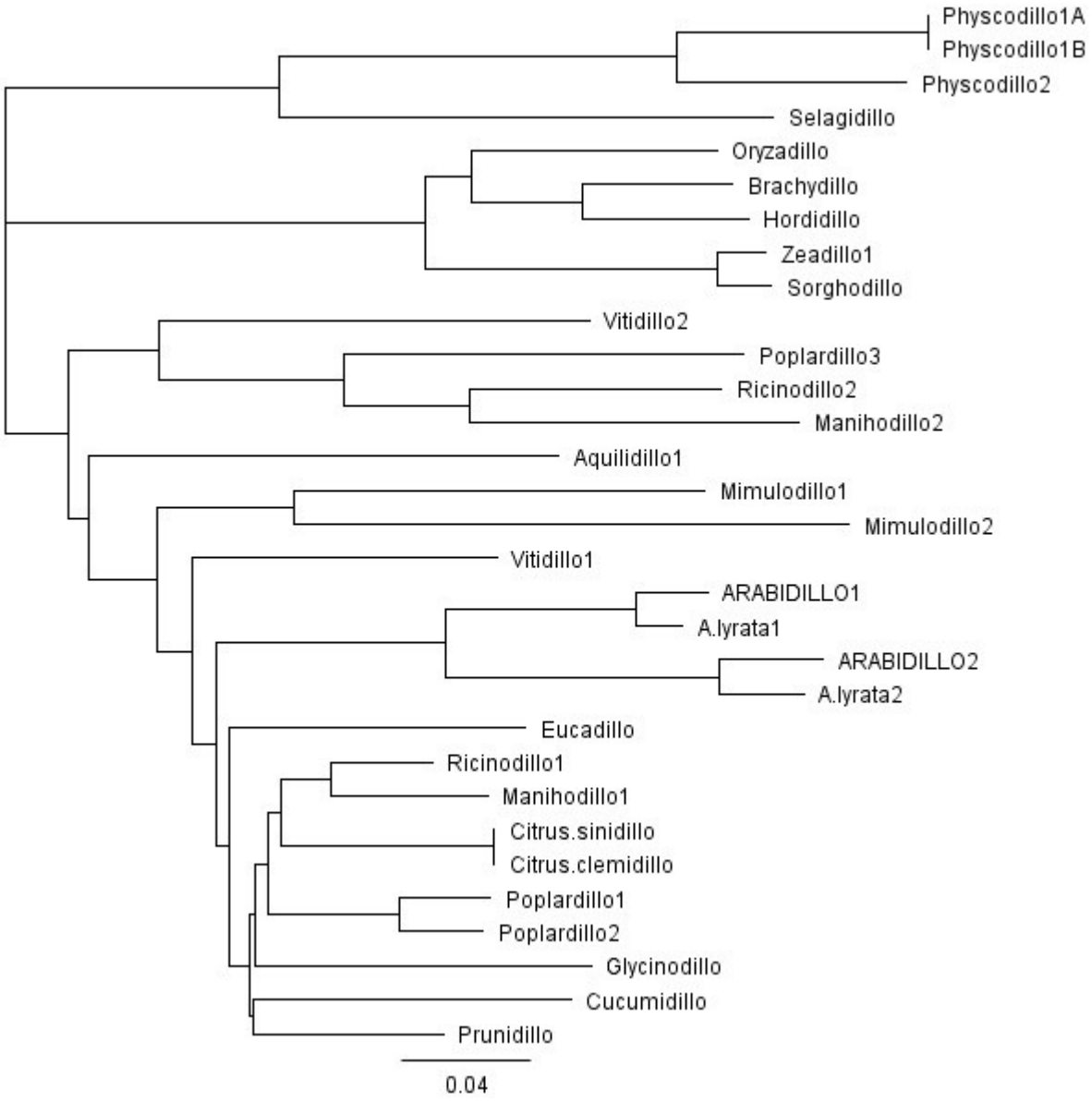
Phylogenetic analyses show that PHYSCODILLO1A/1B are the closest relatives of PHYSCODILLO2 and are likely to have arisen from a whole genome duplication event. The bryophyte representatives are most closely related to SELAGIDILLO (lycophyte) and this is perhaps unsurprising since they are evolutionarily less distant to one another than they both are to the angiosperms. Bryophyte and lycophyte representatives are more closely related to the monocot representatives than the dicot representatives; to support this theory, their protein

sequences have higher homology to monocot ARABIDILLOs than dicot ARABIDILLOs. ARABIDILLO1 and A.LYRATA1 are very closely related, as are ARABIDILLO2 and A.LYRATA2. These in turn are more closely related to ARABIDILLO representatives from the woody perennial species, Poplar and *Eucalyptus* than they are to basal land plant species (**Figure 3.19**).

Gene	Length genomic DNA	Length cDNA	No. exons	No. intron	Average length of introns	protein
PHYSCODILLO1A	6251	2823	12	11	311.6	941
PHYSCODILLO1B	6251	2823	12	11	311.6	941
PHYSCODILLO2	5860	2823	12	11	276.1	941
SELAGIDILLO	3379	2736	12	11	58.5	912
Oryzadillo	8158	2826	12	11	484.7	942
Brachydillo	8270	2817	12	11	495.7	939
Zeadillo1	13819	2769	12	11	1004.5	923
Sorghodillo	13725	2766	12	11	996.3	922
Hordidillo	unknown	2817	unknown	unknown	unknown	939
ARABIDILLO1	4044	2793	11	10	125.1	931
ARABIDILLO2	3791	2787	9	8	125.5	929
A.LYRATA1	4049	2790	11	10	125.9	930
A.LYRATA2	3803	2787	9	8	127.0	929
Glycinodillo	10991	2766	12	11	747.7	922
Ricinodillo1	7321	2763	12	11	414.4	921
Ricinodillo2	7376	2757	12	11	419.9	919
Manihodillo1	8191	2781	12	11	491.8	927
Manihodillo2	6582	2769	12	11	346.6	923
Cucumidillo	12269	2775	12	11	863.1	925
Mimulodillo1	5486	2673	12	11	255.7	891
Mimulodillo2	6481	2529	12	11	359.3	843
Aquilidillo1	8660	2724	12	11	539.6	908
Poplardillo1	7891	2757	12	11	466.7	919
Poplardillo2	8648	2757	12	11	535.5	919
Poplardillo3	9027	2766	12	11	569.2	922
Prunidillo	9879	2757	12	11	647.5	919
Citrus.sinidillo	6442	2760	12	11	334.7	920
Citrus.clemidillo	6442	2760	12	11	334.7	920
Eucadillo	11597	2760	12	11	803.4	920
Vitidillo1	52647	2673	13	12	4164.5	891

Table 3.1 ARABIDILLO gene homologues in land plants.

Figure 3.19 Phylogenetic tree of the land plant ARABIDILLOs. Full-length protein sequences were aligned using ClustalW2 or Seaview and trees subsequently constructed in either Treeview or Seaview respectively. The tree featured is the consensus of two independent outputs (neighbor joining). The tree features homologues from *Physcomitrella patens* (three PHYSCODILLOs), *Selaginella moellendorffii* (SELAGIDILLO), *Oryza sativa* (ORYZADILLO), *Brachypodium distachyon* (BRACHYDILLO), *Hordeum vulgare* (HORDIDILLO), *Zea mays* (ZEADILLO), *Sorghum bicolor* (SORGHODILLO), *Vitis vinifera* (two VITIDILLOs), *Populus trichocarpa* (three POPLARDILLOs), *Ricinis communis* (two RICINODILLOs), *Manihot esculenta* (two MANIHODILLOs), *Aquilegia coerulea* (AQUILIDILLO), *Mimulus guttatus* (two MIMULODILLOs), *Arabidopsis lyrata* (two A.LYRATAs), *Arabidopsis thaliana* (two ARABIDILLOs), *Eucalyptus grandis* (EUCADILLO), *Citrus sinensis* (CITRUS.SINIDILLO), *Citrus clementina* (CITRUS.CLEMIDILLO), *Glycine max* (GLYCINODILLO), *Cucumis sativus* (CUCUMIDILLO) and *Prunus persica* (PRUNIDILLO).


Physcodillo1A
Physcodillo1B
Physcodillo2
Selagidillo
Oryzadillo
Brachydillo
Hordidillo
Zeadillo1
Sorghodillo
Vitidillo2
Poplardillo3
Ricinodillo2
Manihodillo2
Aquilidillo1
Mimulodillo1
Mimulodillo2
Vitidillo1
ARABIDILLO1
A.lyrata1
ARABIDILLO2
A.lyrata2
Eucadillo
Ricinodillo1
Manihodillo1
Citrus.sinidillo
Citrus.clemidillo
Poplardillo1
Poplardillo2
Glycinodillo
Cucumidillo
Prunidillo
0.04

3.9.4 Plant Armadillos have diverse genetic structures

The size of ARABIDILLO gene homologous varies enormously (**Table 3.1**). The smallest representatives are from *Selaginella* (3379bp), *Arabidopsis thaliana* (4044bp and 3791bp) and *Arabidopsis lyrata* (4049bp and 3803bp) since these have very small introns. Lengths of ARABIDILLO homologues in the other dicotyledonous species vary considerably from 5486bp (*MIMULODILLO1*) to 12269bp (*CUCUMIDILLO*). Others may be misannotated since they are extremely large and this includes *VITIDILLO1*, which is annotated as a 52647bp gene. The monocot representatives, *ORYZADILLO*, *BRACHYDILLO*, *SORGHODILLO*, *ZEADILLO* and *HORDIDILLO* are generally very large (all over 8kb) and accordingly, these have large introns. The sizes of monocot genomes varies enormously; rice and *Brachypodium* genomes are approximately the same size as *Physcomitrella*, the *Sorghum* genome is approximately 800Mb, the maize genome is approximately the same size as the human genome (2500Mb) and the barley genome is extremely large (5500Mb).

The majority of ARABIDILLO homologues have 12 exons and 11 introns and these range from the earliest evolving bryophytes and lycophytes to the modern day angiosperms. The *Arabidopsis* representative genes contain fewer exons; *ARABIDILLO1* and *A.LYRATA1* have 11 exons and 10 introns and *ARABIDILLO2* and *A.LYRATA2* have 9 exons and 8 introns. Some representatives have more exons; *VITIDILLO1* has 13 exons and 12 introns. It is therefore likely that the ancestral land plant ARABIDILLO homologue contained 12 exons and 11 introns. Throughout evolution, genes then gained or lost intron sequences; *Arabidopsis* genes lost one or two introns whereas grapevine genes acquired an intron.

3.10 Discussion

3.10.1 *Physcomitrella* ARABIDILLO homologues

ARABIDILLO1 and ARABIDILLO2 promote root branching in the model angiosperm species, *Arabidopsis thaliana* (Coates *et al*., 2006). Homologues of these proteins exist in bryophytes, the most basal land plant lineage that lacks specialised rooting structures and instead produces filamentous rhizoids for anchorage and nutrient acquisition. Filamentous rhizoids are equivalent to the root hairs of higher plants and are thus formed by mechanisms unlike those that form multicellular lateral roots (Menand *et al*., 2007). ARABIDILLO homologues must therefore function differently in land plants that predate the origins of 'true roots'.

Three ARABIDILLO homologues were identified in the draft version of the *Physcomitrella* genome; *PHYSCODILLO1A*, *PHYSCODILLO1B* and *PHYSCODILLO2*. The draft genome suggested that full-length copies of *PHYSCODILLO1A* and *PHYSCODILLO2* were present in the genome but only a truncated version of *PHYSCODILLO1B* existed. A sequencing approach was used to confirm that full-length versions of all three genes were present and that *PHYSCODILLO1A* and *PHYSCODILLO1B* existed as an identical inverted repeat on scaffold 91 in a tail-to-tail orientation, which included their promoter regions. There were two intergenic regions separating the two genes and these also existed as an identical inverted repeat, with the exception of a unique stretch of 111 nucleotides and 7 additional nucleotide differences. The strong identity between *PHYSCODILLO1A* and *PHYSCODILLO1B*, even within the intergenic regions, suggested that a very recent gene duplication event had taken place.

Southern blotting was carried out to confirm the number of *ARABIDILLO* homologues in *Physcomitrella*. The presence of full-length *PHYSCODILLO2* in addition to *PHYSCODILLO1A* and/or *PHYSCODILLO1B* genes was confirmed by using probes specific to each gene. In order to confirm the presence of both *PHYSCODILLO1A* and *PHYSCODILLO1B*, promoter probes were used and a number of different restriction enzymes that were able to distinguish between the two genes according to size. Partial success was achieved using this approach. The presence of both *PHYSCODILLO1A* and *-1B* genes was confirmed using *EcoRI* and *HindIII*. However, an additional band was observed in the *HindIII* lane, a result of the probe hybridising to a DNA fragment of approximately 3kb. On closer inspection, the locus sequence contained an additional *HindIII* site 3.2kb downstream of the *HindIII* site within the *PHYSCODILLO1A* promoter sequence. The probe may have recognised this fragment as a consequence of an incomplete genomic DNA digest. Peculiarly, a similar thing occurred in the *EcoRV* lane. A band confirming the presence of *PHYSCODILLO1B* was present but the band corresponding to the *PHYSCODILLO1A* gene was absent. A smaller than expected band of approximately 3kb was detected instead, but this could be due to non-specific probe hybridisation. Non-specific binding could be remedied by increasing the stringency of the washes, by increasing the temperature. Southern blotting attempts have been repeated but this has not been successful. However, even in the absence of clear and concise Southern blotting data, it is certain that all three genes exist since both *PHYSCODILLO2* and the genomic locus containing both *PHYSCODILLO1A* and *PHYSCODILLO1B* genes have been sequenced in their entirety.

There have been few reports of tandem duplications of gene families in plants. Bizarrely, despite having a larger genome size, the genome of *Physcomitrella* contains only 1% of tandemly

arrayed genes (TAGs) in contrast to 16% in *Arabidopsis*, 14% in rice and 11% in *Poplar* genomes. TAGs are defined as pairs of paralogous genes (closest orthologue within the same species) that reside adjacent to one another on the same region of the chromosome either in head-to-head, tail-to-tail or head-to-tail orientations (Rensing *et al*., 2008; Lang *et al*., 2008). Paralogous pairs are usually situated on opposite strands to one another with the majority in a head-to-head orientation. Theories suggest that homologous recombination between paralogous pairs on opposite strands have reduced sequence divergence by exploiting host DNA repair mechanisms (Rensing *et al*., 2008).

The most similar example to the *PHYSCODILLO1A/1B* gene duplication in recent literature is the Chalcone Synthase (CHS) gene family with 17 members. Among the 17 genes, there are three identical copies of CHS3 and two identical copies of CHS5. CHS3.1 and CHS3.2 exist as a head-to-head duplication with 8.8kb separating their start codons. A third identical copy of the CHS3 gene was located on a different scaffold (Koduri *et al*., 2009). On closer inspection of the scaffold sequence, both 5' and 3' UTR sequences for the three CHS3 genes were also identical. CHS5.1 and CHS5.2 are also 100% identical to each other and exist as a head-to-head duplication with 18.9kb separating their start codons. In contrast to the intergenic regions separating both *PHYSCODILLO1A* and *PHYSCODILLO1B*, there was no significant identity between the 8.8kb and 18.9kb intergenic regions separating CHS3.1/CHS3.2 and CHS5.1/CHS5.2 genes respectively. The CHS3 and CHS5 genes were very similar to each other; there were only 6 nucleotide differences between transcripts of the three CHS3 and two CHS5 genes. The result of this was to create only three amino acid mismatches between CHS3 and CHS5 genes. *PHYSCODILLO1A* and *PHYSCODILLO1B* sit in a tail-to-tail orientation on

scaffold 91. Although not identical like the *PHYSCODILLO* gene pair, CHS8 and CHS9 exist in a tail-to-tail orientation with only 1.4kb separating their two stop codons on scaffold 25 (Koduri *et al*., 2009).

Like CHS3.1/CHS3.2 and CHS5.1/CHS5.2, *PHYSCODILLO1A* and *PHYSCODILLO1B* are tandemly duplicated on opposite strands. However, none of the CHS examples appear to have such an extensive duplication of the intergenic sequences between the two paralogous genes. *PHYSCODILLO1A* and *PHYSCODILLO1B* exist in a tail-to-tail and not a head-to-head orientation like many of the CHS gene pairs.

It is not known whether *PHYSCODILLO1A*, *-1B* and *-2* reside on the same or different chromosomes since genes have not yet been mapped to chromosomes. However, great efforts are being made to fulfil this requirement with the ultimate goal being to obtain 27 linkage groups that correspond to the 27 haploid chromosomes of *Physcomitrella* (Kamisugi *et al*., 2008; Rensing *et al*., 2008).

3.10.2 *PHYSCODILLO* gene expression

In order to determine whether the *PHYSCODILLO* genes exhibited tissue-specific expression patterns, RT-PCR and *PHYSCODILLO*-promoter::GUS analyses were carried out.

Due to the high similarity between *PHYSCODILLO1A/1B* and *PHYSCODILLO2* transcripts, designing primers that would distinguish between them was difficult. There were only stretches of up to 10 base pairs within their cDNA sequences in which to design gene-

specific primers. Consequently, when attempting to amplify *PHYSCODILLO2* cDNA, *PHYSCODILLO1A/1B* was often amplified instead. Although cloning the two full-length cDNA sequences in themselves was not problematic, RT-PCR studies were extremely complex, since it was not clear which of the transcripts were produced in a given tissue. It is possible that both *PHYSCODILLO1A* and *-1B* genes are expressed and that the higher abundance of these identical transcripts than *PHYSCODILLO2* means that *PHYSCODILLO1A/1B* could be preferentially amplified over the *PHYSCODILLO2* transcript.

Riggleman *et al*. (1989) determined that the *Drosophila Armadillo* gene produced two different transcripts, but these encoded the same Armadillo protein. They hypothesised that the use of two transcription initiation sites ensured that enough of the *Armadillo* transcript was produced, to avoid any shortfall in protein supply (Riggleman *et al*., 1989). There were no variations in the *PHYSCODILLO1A/1B* transcript sequence in a number of independently sequenced clones. However, primers used for cloning were predominantly situated within coding regions of the transcript. 3' and 5' RACE (rapid amplification of cDNA ends) would establish whether there are any differences between the sequence lengths of the untranslated regions (UTR) of the two transcripts.

Distinguishing between the transcripts of *PHYSCODILLO1A* and *PHYSCODILLO1B* was impossible since they had identical promoter sequences and UTR regions. It was certain that at least one of the genes was expressed since *PHYSCODILLO1A/1B* cDNA was cloned with relative ease. However, it was only possible to drive the expression of a GUS reporter gene from the full-length *PHYSCODILLO1A* gene promoter. Neither the full-length *PHYSCODILLO1B* promoter

nor the shorter 1.6kb promoter sequence shared by both *PHYSCODILLO1A* and *-1B* genes were able to drive GUS expression. There is an intriguing possibility that promoter elements upstream of the identical region shared by both genes are sufficient to drive *PHYSCODILLO1A* expression and that these are absent from the equivalent region of the *PHYSCODILLO1B* gene promoter. There are several additional putative TATA boxes and also a single ABA-responsive element (ABRE) within this upstream region. It is also possible that the *PHYSCODILLO1B* gene promoter is only active under certain conditions and this should also be investigated in the future.

Stable *PHYSCODILLO1A*-promoter::GUS and *PHYSCODILLO2*-promoter::GUS lines were generated and these exhibited identical expression patterns throughout both the haploid and diploid phases of the life cycle. Strong GUS expression in both protonemata and gametophores was observed after just 1 hour. Only weak GUS expression was observed in rhizoids, even after overnight incubation with X-gluc. This suggests that promoters are less active in rhizoids than in protonemata and the leaf-like structures of gametophores. Interestingly, GUS signals were detected throughout male antheridia but only in the upper cells of female archegonia containing unfertilised egg cells. After fertilisation, this localised expression was not observed, suggesting that the *PHYSCODILLOs* may have involvement in providing an open channel through which sperm can gain access to egg cells. GUS expression was observed at the base of the spore capsule but not elsewhere in the diploid sporophyte generation. The *PHYSCODILLOs* may therefore be important during fertilisation and for the dispersal of spores. Interestingly, *ARABIDILLO1*-promoter::GUS expression was observed specifically in stigma (female reproductive structures of flowers in angiosperms) and also to a lesser extent in anthers (male reproductive structures; Juliet Coates, unpublished data).

3.10.3 SELAGIDILLO

The *Selaginella moellendorffi* genome encodes a single ARABIDILLO homologue that has been named *SELAGIDILLO* (*Selaginella* ARABIDILLO). Like *Physcomitrella*, there were imperfections in the *Selaginella* genome assembly. The *SELAGIDILLO* genomic DNA that was cloned and sequenced was the same length as that annotated in the genome but there were a large number of single nucleotide differences. *SELAGIDILLO* genomic DNA and cDNA were cloned from a different haplotype to that sequenced by the JGI and therefore single nucleotide substitutions are likely to be due to haplotype variability. Despite this variability, the cDNA sequence was significantly shorter and contained a shorter version of exon 10 than that annotated in the genome assembly. *SELAGIDILLO* was the smallest of the *ARABIDILLO* gene homologues identified and this was due to short intron lengths. This is likely to be due to the fact that the *Selaginella moellendorffii* genome is the smallest land plant genome sequenced to date and is approximately two thirds of the size of the *Arabidopsis* genome (Banks *et al*., 2011).

3.10.4 ARABIDILLO homologues are highly conserved across the land plants

ARABIDILLO homologues have been identified in all land plant genomes sequenced to date and they all share the same structural characteristics. All possess a nuclear localisation signal (NLS), an N-terminal F-box domain, a linker region composed of leucine-rich repeats (LRR) and an Armadillo (ARM)-repeat domain and the sequence identity is enhanced within these conserved regions. No full-length homologues have been identified in the sequenced genomes of algal species. However, partial sequences have been identified in Charophyte algae

but it is not known yet whether they assemble to encode an F-box protein with LRRs and an Arm repeat domain (Juliet Coates, personal communication). Their presence in all land plant genomes suggests that the ARABIDILLOs have been conserved at least since plants transitioned from water to land approximately 500 million years ago.

PHYSCODILLO1A/1B and PHYSCODILLO2 are 85% identical to one another and approximately 60% identical to the ARABIDILLO proteins. From approximately 220 F-box proteins in *Physcomitrella*, the PHYSCODILLOs are the only proteins that contain downstream LRRs and ARM repeats. SELAGIDILLO was 63% identical to the ARABIDILLOs and 69% identical to the PHYSCODILLOs. *Selaginella* has over 300 putative F-box proteins and similarly to the PHYSCODILLO proteins, SELAGIDILLO is the only F-box protein encoded by the *Selaginella* genome that contains downstream LRRs and ARM repeats. The PHYSCODILLOs and SELAGIDILLO are more related to each other than they are to the ARABIDILLOs. This is perhaps unsurprising since they are evolutionarily less distant to one another than they are to the ARABIDILLOs that reside within the angiosperm lineage.

Gibbs *et al*. determined that ARABIDILLO1 was a *bona fide* F-box protein since it was able to interact with ASK1 and ASK2, *Arabidopsis* homologues of SKP1 from *Saccharomyces cerevisiae* (Daniel Gibbs, PhD thesis; Nibau *et al*., 2011). SKP1 interacts with Cullin1 and an F-box protein to form SCF complexes that consequently function as multisubunit E3 ubiquitin-protein ligases, which target proteins for proteasome-mediated degradation (Bai *et al*., 2006; Skowyra *et al*., 2007). F-box proteins appear to confer substrate specificity to these complexes by binding to different protein targets (Lechner *et al*., 2006). The PHYSCODILLOs and

SELAGIDILLO contain many of the key residues that are required for ARABIDILLO F-box function in *Arabidopsis*. Furthermore, a number of ASK and Cullin proteins are encoded by both *Physcomitrella* and *Selaginella* genomes, suggesting that both PHYSCODILLO and SELAGIDILLO proteins could function as *bona fide* F-box proteins.

Ongoing and unpublished work in the Coates lab examines the interaction partners of the ARABIDILLO1 ARM repeat region. Using the ARM repeat region as bait in a yeast 2-hybrid screen of an *Arabidopsis* root library, the Coates lab have shown that the ARABIDILLO1 ARM repeat region interacts with the R2R3 MYB transcription factors; AtMYB92, AtMYB93 and AtMYB53. They hypothesised that the MYB proteins were inhibitors of lateral root development in *Arabidopsis* and that they were specifically targeted for degradation by the ARABIDILLOs (Daniel Gibbs, PhD thesis). Although many MYB transcription factors are encoded by both *Physcomitrella* and *Selaginella* genomes, there are no clear homologues of AtMYB92, AtMYB93 or AtMYB53 and these appear to be restricted to higher plant genomes. It is possible that these particular MYB transcription factors were acquired during evolution to fulfil a rooting function and this correlates well with their absence in bryophytes and lycophytes, two lineages that lack true rooting structures. The interacting partners of the ARM repeats of PHYSCODILLOs in early-evolving land plants is currently being investigated in the Coates lab.

3.10.5 Summary

The work in this chapter has determined that there are three ARABIDILLO gene homologues in *Physcomitrella*; these encode proteins with high conservation throughout the land plants. The aim of the next chapter is to attempt to elucidate roles of *Physcomitrella* ARABIDILLOs.

Chapter IV:

Phenotypic characterisation of *physcodillo* deletion mutants and novel responses to ABA

4.1 Introduction

ARABIDILLO1 and ARABIDILLO2 are 80% identical to each other and function redundantly to promote root branching in *Arabidopsis*. Single *arabidillo1* or *arabidillo2* knockout mutants yield no detectable mutant phenotypes. However, double *arabidillo1/arabidillo2* knockout mutants have fewer lateral roots compared to wild type plants (Coates *et al*., 2006). Furthermore, *arabidillo1/2* mutants exhibit ABA insensitive responses during seed germination (Gibbs and Coates, unpublished). Interestingly, roots of both wild type and *arabidillo1/2* mutant plants exhibit normal responses to ABA, which plays a role in the repression of lateral root formation (Coates *et al*., 2006; Nibau *et al* 2011; de Smet *et al* 2003). Since PHYSCODILLO1A/1B and PHYSCODILLO2 are 85% identical to each other, it is likely that they also function redundantly. It would therefore be necessary to disrupt all three of the *PHYSCODILLO* genes in order to detect mutant phenotypes.

Conveniently, *Physcomitrella* is the only land plant in which targeted transgene integration by homologous recombination occurs at a similar efficiency to that of the yeast, *Saccharomyces cerevisiae* (Schaefer and Zryd, 1997). Targeted gene replacement by homologous recombination has therefore become an extremely useful reverse genetics tool in which to target constructs to specific pre-determined locations within the *Physcomitrella* genome.

The work in this chapter describes the generation of *PHYSCODILLO* deletion mutants by sequential gene targeting and their subsequent phenotypic characterisation.

4.2 Generation of stable *physcodillo* deletion mutants

4.2.1 Generation of stable *physcodillo2* deletion mutants

A *PHYSCODILLO2* single deletion construct was generated by cloning two flanking sequences homologous to *PHYSCODILLO2* genomic DNA and inserting them into the pAHG1 vector (Knight *et al*., 2002; Machuka *et al*., 2005) either side of a Hygromycin resistance cassette (**Figure 4.1A**). *Physcomitrella* protoplasts were transformed using the *PHYSCODILLO2* deletion construct and after two rounds of Hygromycin selection, a number of putative transformants were obtained. An initial PCR screen of all 68 putative *physcodillo2* deletion mutant lines was carried out using a pair of *PHYSCODILLO2* gene-specific primers. The forward primer was designed to anneal to a sequence located within a region of the *PHYSCODILLO2* gene that would no longer be present if successful transgene integration had taken place. The reverse primer was designed to anneal to a sequence located at the 3' end of the *PHYSCODILLO2* gene. The forward primer would be unable to anneal if the central portion of the gene had been successfully replaced by the Hygromycin resistance cassette following protoplast transformation (**Figure 4.1A**).

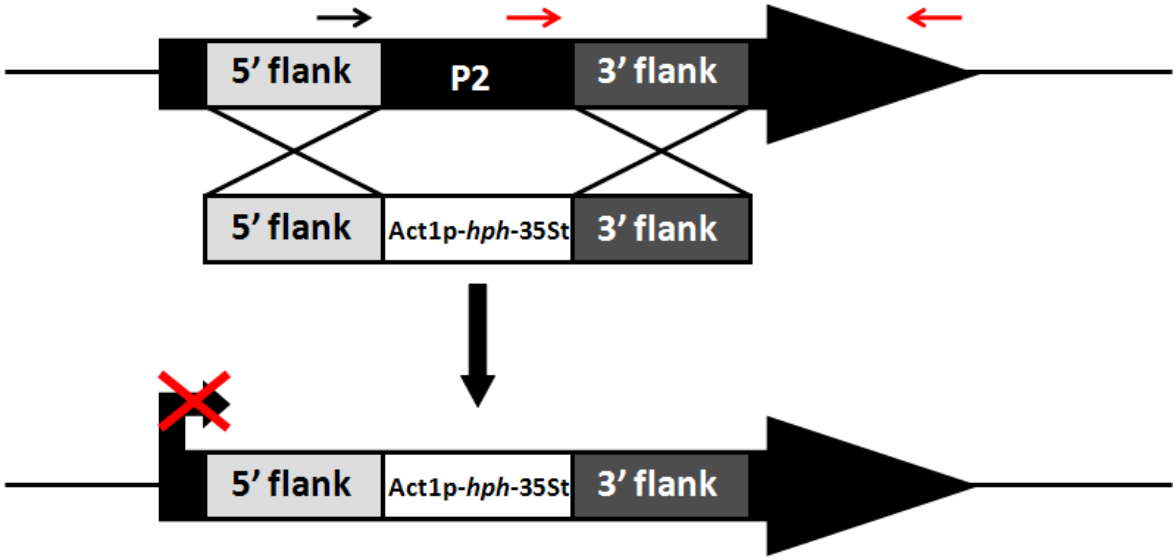
41 out of 68 lines produced PCR products consistent with the presence of an intact *PHYSCODILLO2* gene; disruption of the PHYSCODILLO2 open-reading frame by gene targeting had therefore been unsuccessful. The Hygromycin resistance cassette had probably integrated elsewhere in the genome. The remaining lines produced no bands at all, which suggested that they could all be positive for the *physcodillo2* deletion (**Figure 4.1B**).

Genomic DNA was isolated from five of the remaining lines: 101, 105, 107, 109 and 110. Full-length *PHYSCODILLO2* gene products were then amplified from these templates using P2-5'-1 and P2-3'-1 primers and sequenced using a primer designed to anneal within the 5' flanking sequence of the deletion construct (**Figure 4.1B**). There were two possible

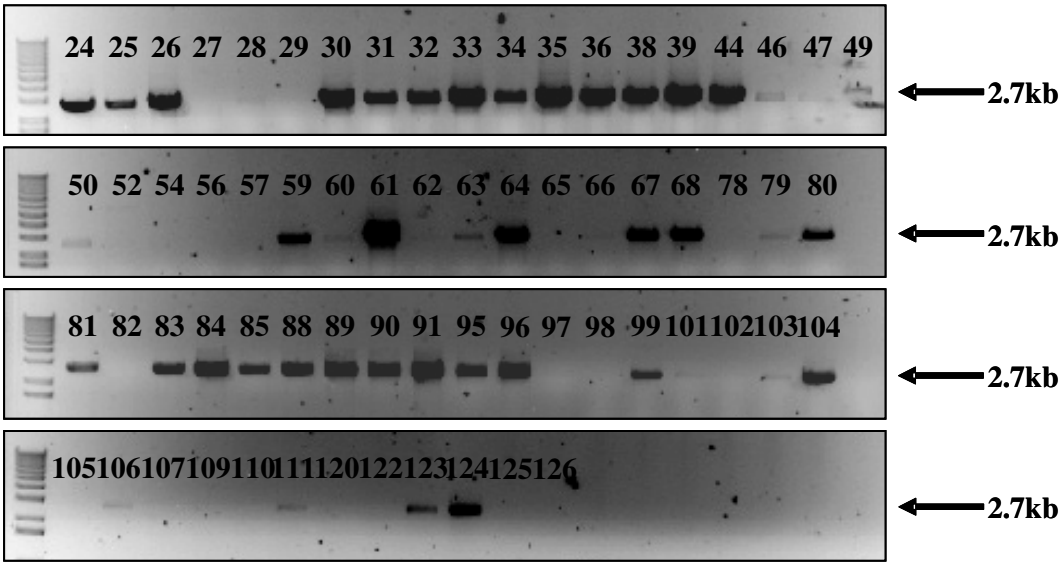
sequencing outcomes: *PHYSCODILLO2* sequences obtained (negative for the *physcodillo2* deletion) or Hygromycin cassette sequences obtained (positive for *physcodillo2* deletion).

Three of the five lines (101, 107 and 109) were negative for the *physcodillo2* deletion and all possessed an intact *PHYSCODILLO2* gene. The Hygromycin cassette had probably integrated randomly elsewhere in the genome sequence. Two of the five lines (105 and 110) were positive for the *physcodillo* deletion. Not only had gene-targeting been successful, but only a single copy of the construct had been incorporated at the *PHYSCODILLO2* gene locus (**Figure 4.1C**). *physcodillo2.105* (*ppdillo2.105*) and *physcodillo2.110* (*ppdillo2.110*) single deletion mutants had no obvious morphological phenotypes. Furthermore, they were able to complete their life cycles within three months and produce viable spores.

A)


5' flank
P2
3' flank
5' flank
Act1p-*hph*-35St
3' flank
5' flank
Act1p-*hph*-35St
3' flank

B)


24 25 26 27 28 29 30 31 32 33 34 35 36 38 39 44 46 47 49
2.7kb
50 52 54 56 57 59 60 61 62 63 64 65 66 67 68 78 79 80
2.7kb
81 82 83 84 85 88 89 90 91 95 96 97 98 99 101102103104
2.7kb
105106107109110111201221231241251261
2.7kb

C)

```
P2KO.101              GTAGGGTTGAATGGCAACTGGTTCACGATGATGGCGGTTTTTCGAAGTTGACCCATGTTC 60
P2.gDNA               GTAGGGTTGAATGGCAACTGGTTCACGATGATGGCGGTTTTTCGAAGTTGACCCATGTTC 1800
Hygromycin.cassette   ------------------------------------------------------------
P2KO.101              AATTAAACCTCCCTCGAGTGTAGGCGTGGTTTGCGTGGGTTACTCAGTATAGT--GAATT 118
P2.gDNA               AATTAAACCTCCCTCGAGTGTAGGCGTGGTTTGCGTGGGTTACTCAGTATAGT--GAATT 1858
Hygromycin.cassette   ------------CTCGAGGTCATTCATATGCTTGAGAAGAGAGTCGGGATAGTCCAAAAT 48
-------------------------------------------------------------------------------
P2KO.105              GTAGGGTTGAATGGCAACTGGTTCACGATGATGGCGGTTTTTCGAAGTTGACCCATGTTC 60
P2.gDNA               GTAGGGTTGAATGGCAACTGGTTCACGATGATGGCGGTTTTTCGAAGTTGACCCATGTTC 1800
Hygromycin.cassette   ------------------------------------------------------------
P2KO.105              AATTAAACCTCCCTCGAGGTCATTCATATGCTTGAGAAGAGAGTCGGGATAGTCCAAAAT 120
P2.gDNA               AATTAAACCTCCCTCGAGTGTAGGCGTGGTTTGCGTGGGTTACTCAGTATAGT-GAATT  1858
Hygromycin.cassette   ------------CTCGAGGTCATTCATATGCTTGAGAAGAGAGTCGGGATAGTCCAAAAT 48
-------------------------------------------------------------------------------
P2KO.107              GTAGGGTTGAATGGCAACTGGTTCACGATGATGGCGGTTTTTCGAAGTTGACCCATGTTC 60
P2.gDNA               GTAGGGTTGAATGGCAACTGGTTCACGATGATGGCGGTTTTTCGAAGTTGACCCATGTTC 1800
Hygromycin.cassette   ------------------------------------------------------------
P2KO.107              AATTAAACCTCCCTCGAGTGTAGGCGTGGTTTGCGTGGGTTACTCAGTATAGT--GAATT 118
P2.gDNA               AATTAAACCTCCCTCGAGTGTAGGCGTGGTTTGCGTGGGTTACTCAGTATAGT--GAATT 1858
Hygromycin.cassette   ------------CTCGAGGTCATTCATATGCTTGAGAAGAGAGTCGGGATAGTCCAAAAT 48
-------------------------------------------------------------------------------
P2KO.109              GTAGGGTTGAATGGCAACTGGTTCACGATGATGGCGGTTTTTCGAAGTTGACCCATGTTC 60
P2.gDNA               GTAGGGTTGAATGGCAACTGGTTCACGATGATGGCGGTTTTTCGAAGTTGACCCATGTTC 1800
Hygromycin.cassette   ------------------------------------------------------------
P2KO.109              AATTAAACCTCCCTCGAGTGTAGGCGTGGTTTGCGTGGGTTACTCAGTATAGT--GAATT 118
P2.gDNA               AATTAAACCTCCCTCGAGTGTAGGCGTGGTTTGCGTGGGTTACTCAGTATAGT--GAATT 1858
Hygromycin.cassette   ------------CTCGAGGTCATTCATATGCTTGAGAAGAGAGTCGGGATAGTCCAAAAT 48
-------------------------------------------------------------------------------
P2KO.110              GTAGGGTTGAATGGCAACTGGTTCACGATGATGGCGGTTTTTCGAAGTTGACCCATGTTC 60
P2.gDNA               GTAGGGTTGAATGGCAACTGGTTCACGATGATGGCGGTTTTTCGAAGTTGACCCATGTTC 1800
Hygromycin.cassette   ------------------------------------------------------------
P2KO.110              AATTAAACCTCCCTCGAGGTCATTCATATGCTTGAGAAGAGAGTCGGGATAGTCCAAAAT 120
P2.gDNA               AATTAAACCTCCCTCGAGTGTAGGCGTGGTTTGCGTGGGTTACTCAGTATAGT-GAATT  1858
Hygromycin.cassette   ------------CTCGAGGTCATTCATATGCTTGAGAAGAGAGTCGGGATAGTCCAAAAT 48
```

Figure 4.1 Generation of *physcodillo2* deletion mutants.

A) A *PHYSCODILLO2* deletion construct was generated by amplifying two *PHYSCODILLO2* homologous sequences and inserting them either side of a Hygromycin resistance cassette in pAHG1, driven by the rice actin promoter. *physcodillo2* deletion mutants were generated by inserting the Hygromycin resistance cassette into the open reading frame of the *PHYSCODILLO2* gene by homologous recombination. 5' (nucleotides 98-1628) and 3' (nucleotides 4103-5900) *PHYSCODILLO2* homologous flanking sequences are denoted by light and dark grey boxes respectively. The Hygromycin cassette is denoted by a white box and the *PHYSCODILLO2* gene denoted by a solid black arrow. Recombination events have been indicated using an 'X'. Primers used to amplify PCR products during screening are indicated by a pair of red arrows. The primer used to sequence full-length *PHYSCODILLO2*gene products is indicated by a black arrow.

B) An initial PCR screen of 68 putative *physcodillo2* deletion mutants was carried out using *PHYSCODILLO2* gene-specific primers (red arrows in (A)). A 2.7kb fragment was expected for wild-type plants and this has been indicated on the gel image. 41/68 lines were negative for the *physcodillo2* deletion and carried a wild-type version of the *PHYSCODILLO2* gene. 27/68 lines produced no PCR products, so remained putative *physcodillo2* deletion mutants.

C) Full-length *PHYSCODILLO2* genomic DNA products were amplified from 5/27 of the remaining lines. Products were then sequenced using a primer designed to detect an integration event (black arrow in (A)). Sequence alignments show that two lines named P2KO.105 and P2KO.110 had been successfully transformed using the construct and that a single insert event had taken place.

4.2.2 Generation of stable *physcodillo* triple deletion mutants

As detailed in chapter three, *PHYSCODILLO1A* and *PHYSCODILLO1B* exist as an identical tail-to-tail repeat within a large genetic locus located on scaffold 91 of the *Physcomitrella* genome assembly. The intergenic regions that separate the two identical genes and their promoter regions are also identical. These factors meant that it would not be possible to delete both *PHYSCODILLO1A* and *PHYSCODILLO1B* individually using a conventional gene replacement strategy. A novel approach was used to remove both genes simultaneously and replace the entire 23kb *PHYSCODILLO1A/1B* locus with a G418 resistance cassette. Although large deletions via gene targeting have been described in other organisms (e.g. mouse, *Dictyostelium*, *Plasmodium*), it has not yet been attempted in *Physcomitrella* (D. Schaefer and J.C. Coates, personal communication).

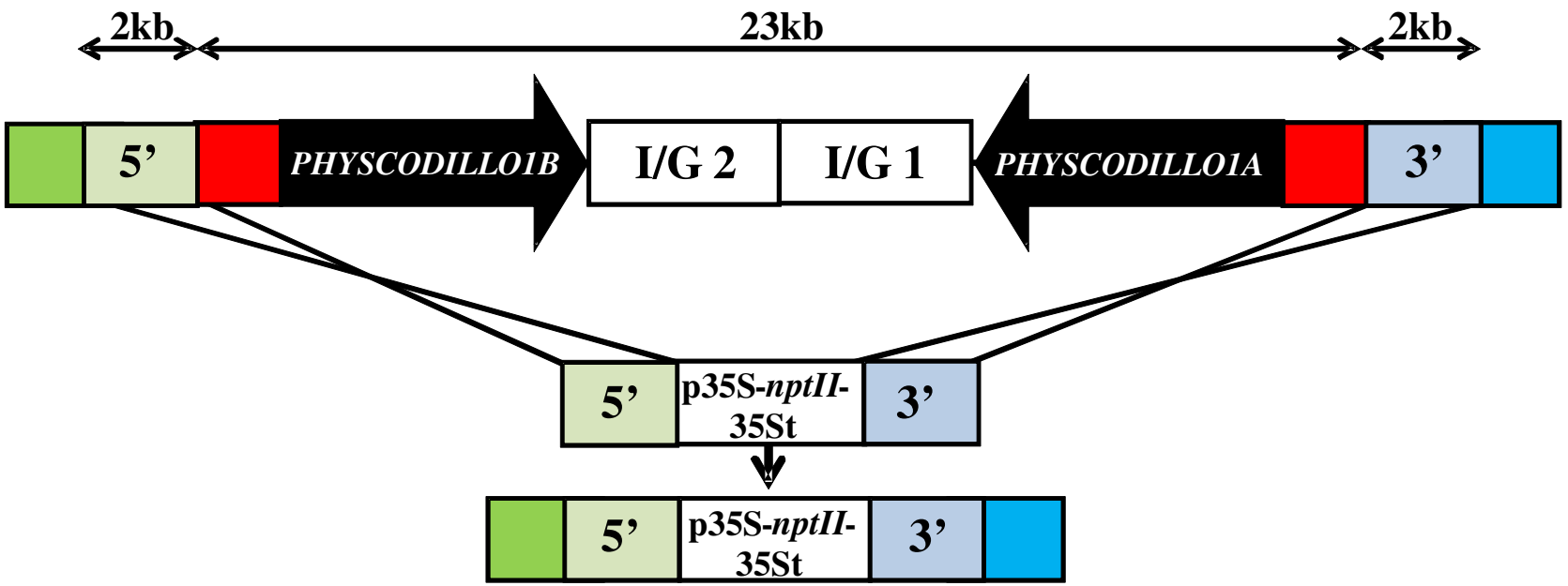
A *PHYSCODILLO1A/1B* double deletion construct was generated by cloning two flanking sequences homologous to regions outside of the *PHYSCODILLO1A/1B* locus (i.e. which did not share homology to either *PHYSCODILLO1A* or *PHYSCODILLO1B* genes or their promoters) and inserting them into the pMBL10a vector (Knight *et al*., 2002; Machuka *et al*., 2005), either side of the G418 resistance cassette (**Figure 4.3**). In order to generate a *physcodillo1a/physcodillo1b/physcodillo2* triple deletion mutant (*ppdillo1a/1b/2*), the *PHYSCODILLO1A/1B* double deletion construct was transformed into protoplasts isolated from *ppdillo2.105* mutant protonemata.

Transforming a Hygromycin resistant *ppdillo2.105* single deletion mutant with a G418 resistant *PHYSCODILLO1A/1B* double deletion construct enabled straightforward screening of putative transformants using antibiotics. After two rounds of antibiotic selection using both Hygromycin and G418, 15 putative triple deletion mutants were obtained. Genomic DNA was isolated from the 15 putative deletion mutants. PCR was then carried out to detect

homologous recombination and correct integration at the endogenous *PHYSCODILLO1A/1B* locus.

Correct 3' integration was checked using a forward primer designed to anneal to the G418 resistance cassette and a reverse primer designed to anneal to the endogenous locus, in sequence upstream of the 3' flanking sequence used to prepare the deletion construct. Seven of these 15 lines (16, 23, 24, 25, 33, 34 and 39) generated bands corresponding to correct 3' integration (**Figure 4.4A**). Correct 5' integration was checked using a reverse primer designed to anneal to the G418 resistance cassette and a forward primer designed to anneal to the endogenous locus, in sequence upstream of the 5' flanking sequence used to prepare the deletion construct. Seven of these lines (1, 9, 12, 16, 18, 33 and 39) produced correctly sized products (**Figure 4.4B**). Three of the lines (16, 33 and -39) showed correct integration at both 5' and 3' sites either side of the *PHYSCODILLO1A/1B* locus. *physcodillo1a/1b/2-16* (*ppdillo1a/1b/2-16*), *physcodillo1a/1b/2-33* (*ppdillo1a/1b/2-33*) and *physcodillo1a/1b/2-39* (*ppdillo1a/1b/2-39*) were therefore confirmed as *physcodillo* triple deletion mutants. Two of the three triple deletion lines, *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* were selected for phenotypic analyses.

Figure 4.3 Generation of *ppdillo1a/1b/2* triple deletion mutants. A *PHYSCODILLO1A/1B* double deletion construct was generated by amplifying two sequences homologous to unique sequences either side of the *PHYSCODILLO1A/1B* locus (that shared no homology to either gene) and inserting them either side of a G418 resistance cassette. *PHYSCODILLO1A/1B* double deletion mutants were generated by replacing the entire *PHYSCODILLO1A/1B* locus with the G418 resistance cassette via homologous recombination. Full-length *PHYSCODILLO1A* and *PHYSCODILLO1B* genes are indicated by black arrows and their promoters are denoted by red boxes. Intergenic regions of both genes have been indicated by white boxes. 5' and 3' homologous sequences are indicated by light green and light blue boxes respectively. The G418 cassette is denoted by a white box and recombination events marked by an 'X'. The resulting product is indicated at the bottom of the diagram. Green and blue boxes represent unique sequences either side of the *PHYSCODILLO1A/1B* locus. In order to generate triple *physcodillo* deletion mutants, the construct was transformed into protoplasts isolated from *physcodillo2.105* (*ppdillo2.105*) single deletion mutants.


2kb
23kb
2kb
5'
PHYSCODILLO1B
I/G 2
I/G 1
PHYSCODILLO1A
3'
5'
p35S-nptII-35St
3'
5'
p35S-nptII-35St
3'

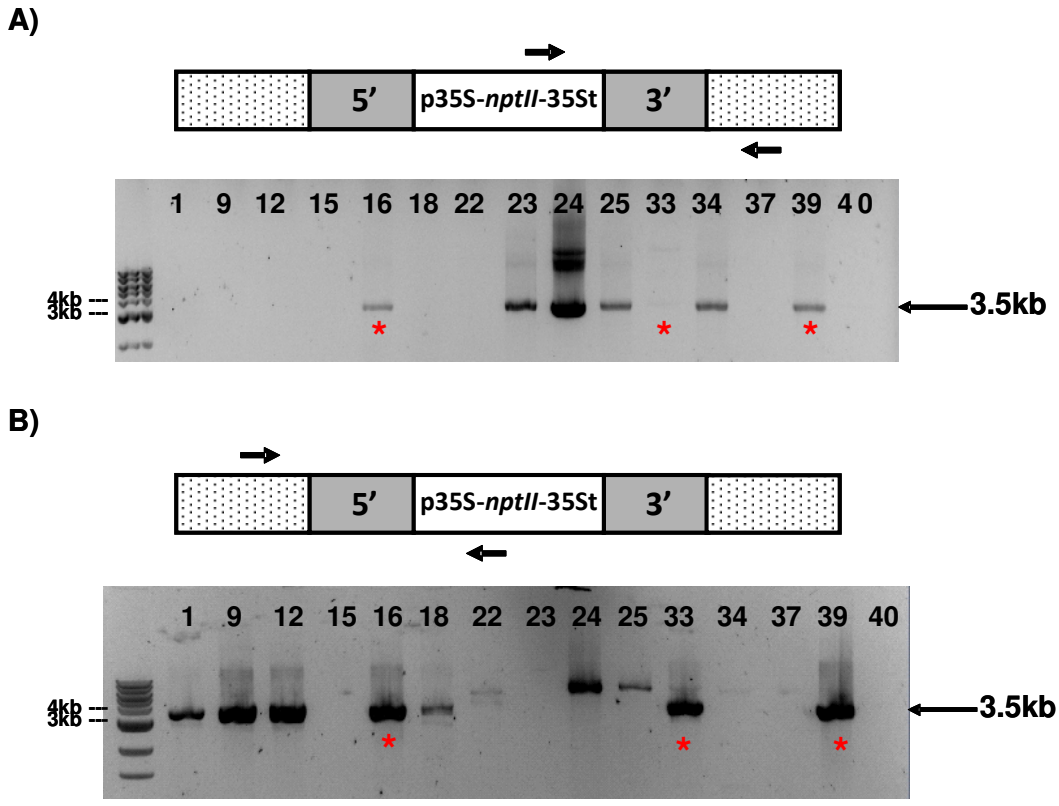


Figure 4.4 Genotyping to determine the presence of a *physcodillo1a/1b* deletion in the *physcodillo2.105* deletion background. 15 putative triple *physcodillo1a/physcodillo1b/physcodillo2* (*ppdillo1a/1b/*2) deletion mutants were obtained following two rounds of G418 selection.

A) 3' integration was examined using a forward primer designed to anneal to the G418 resistance cassette and a reverse primer designed to anneal to endogenous sequences beyond the targeting site (indicated by arrows on diagram). 7/15 lines tested showed correct 3' integration of the deletion construct.

B) 5' integration was examined using a forward primer designed to anneal to endogenous sequences beyond the targeting site and a reverse primer designed to anneal to the G418 resistance cassette (indicated by arrows on diagram). 7/15 of lines tested showed correct 5' integration of the deletion construct. 3/15 lines (indicated by red asterisks (*)) showed both 3' and 5' integration at the *PHYSCODILLO1A/1B* locus (*ppdillo11a/1b/2-16*, *ppdillo1a/1b/2-33* and *ppdillo1a/1b/2-39*).

4.3 Phenotypic characterisation of *physcodillo* triple deletion mutants

The *physcodillo1a/1b/2* triple deletion mutants *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* triple deletion mutants had no obvious morphological phenotypes. They exhibited normal responses to both light and gravity (data not shown). In addition, there were no significant differences in the sizes or morphology of female archegonia or male antheridia compared to wild type (**Figure 4.6**). Furthermore, triple deletion mutants produced sporophytes and these contained fully viable spores. The deletion mutants were therefore able to successfully complete the *Physcomitrella* life cycle.

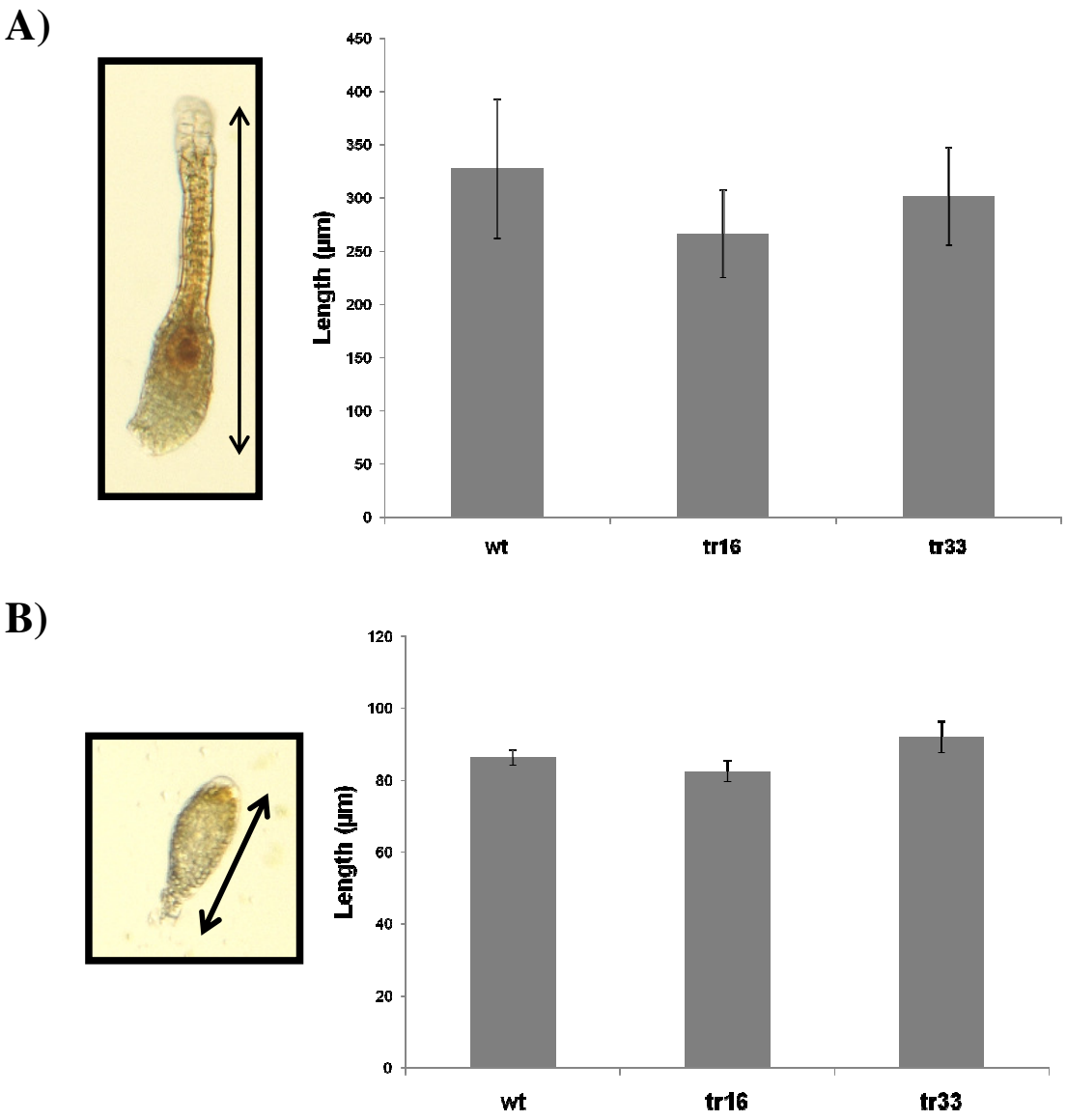


Figure 4.6 Antheridia and archegonia of wild type and *ppdillo1a/1b/2* mutants.

A) Female archegonia were isolated from mature gametophores of wild type and *ppdillo1a/1b/2-16* (tr16) and *ppdillo1a/1b/2-33* (tr33) mutants. Lengths of archegonia were then measured; there were no differences between wild type and mutant archegonia.

B) Male antheridia were isolated from mature gametophores of wild type and *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* mutants. Lengths of antheridia were then measured; there were no differences between wild type and mutant antheridia.

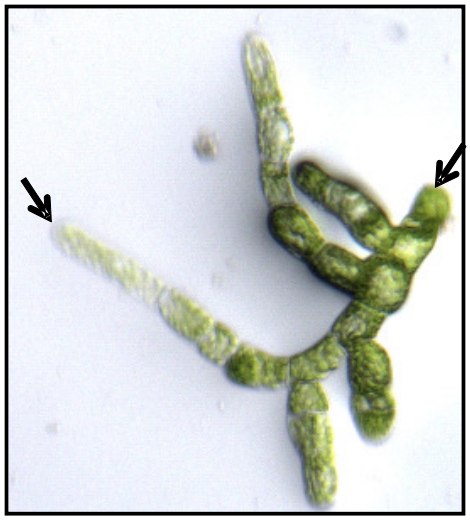
4.3.1 Growth of filaments from regenerating protoplasts

In order to determine whether *ppdillo1a/1b/2* triple deletion mutants exhibit defects during early developmental processes, their ability to regenerate from protoplasts was investigated. Protoplasts were isolated from wild type (control), *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* protonemata and allowed to regenerate. Lengths of the longest filaments within each regenerating colony were measured after a two week period of growth. In a preliminary experiment, filaments regenerating from triple *physcodillo* deletion mutant protoplasts were shorter than those regenerating from wild type protoplasts (**Figure 4.7**). The average length of 13 wild type filaments was 791.4µm ± 83.9µm (standard error of the mean, SEM) after two weeks whereas those of *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* deletion mutants were 584.2µm ± 40.94µm and 572.82µm ± 79.24µm respectively (*t* test, *, $P < 0.05$). Preliminary analyses of branch number suggested that neither *ppdillo1a/1b/2-16* nor *ppdillo1a/1b/2-33* exhibited defects in protonemal filament branching.

Sizes of *Physcomitrella* colonies formed from regenerating protoplasts were then examined in order to determine whether *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* colonies were smaller than a wild type control. Protoplasts were isolated from wild type, *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* protonemata and allowed to regenerate. The surface areas of colonies that formed from regenerating protoplasts were then measured after approximately one week of growth. In two independent protoplast regeneration experiments, *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* colonies were significantly smaller than those regenerating from wild type protoplasts (**Figure 4.8**). In the first experiment; the average size of 26 colonies regenerating from wild type protoplasts was 32517µm^2 ± 1761µm^2 (SEM) compared to 15478µm^2 ± 1815µm^2 (n=26; t test, **, $P < 0.01$) and 17746µm^2 ± 2388µm^2 (n=26; t test, **, $P < 0.01$) for *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* respectively. In experiment two; the average size of 35 colonies regenerating from wild type protoplasts was

18012 μm^2 ± 1587μm^2 compared to 10942μm^2 ± 1672μm^2 (n=35; t test, ***, $P < 0.001$) and 11719μm^2 ± 1118μm^2 (n=35; t test, ***, $P < 0.001$) for *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* respectively. Colonies that regenerated from protoplasts were generally larger in experiment one than in experiment two, due to a marginally longer growth period. It was not possible to repeat the protoplast regeneration experiment for a third time due to time constraints. Nevertheless, it appears that *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* mutants exhibit growth defects during early stages of development.

A)



B)

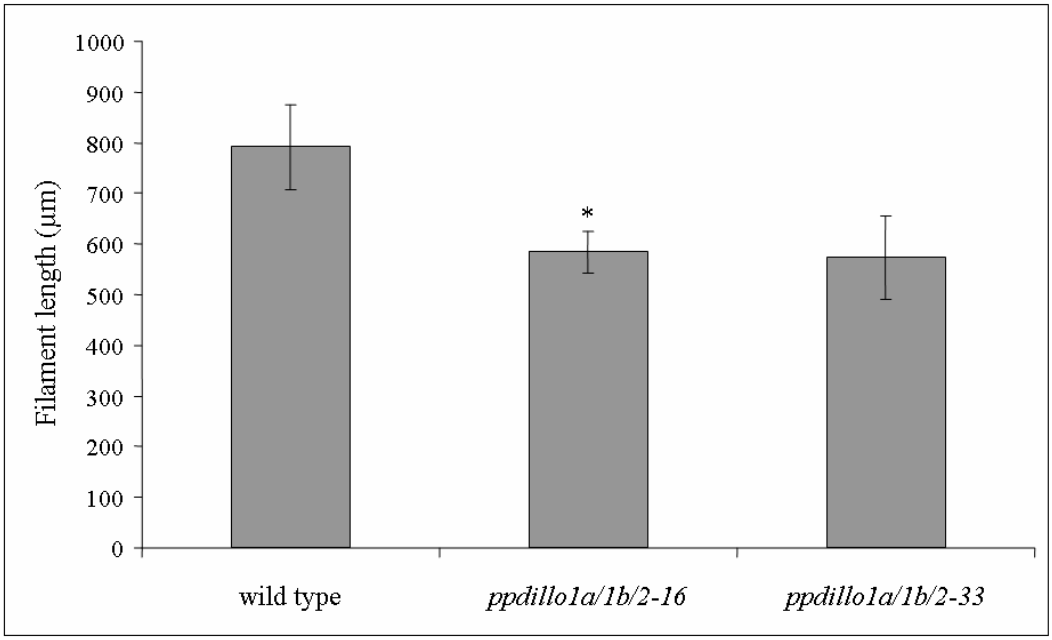
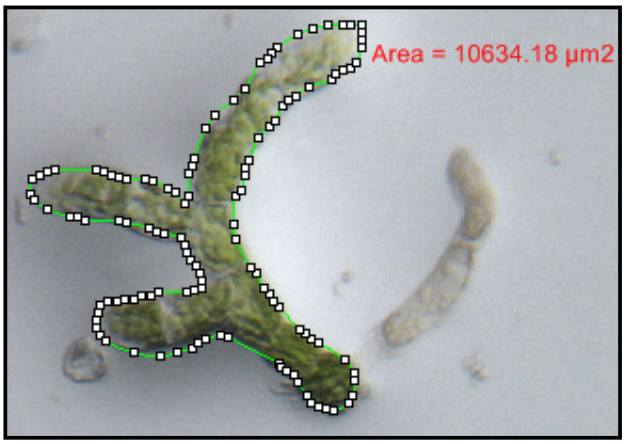
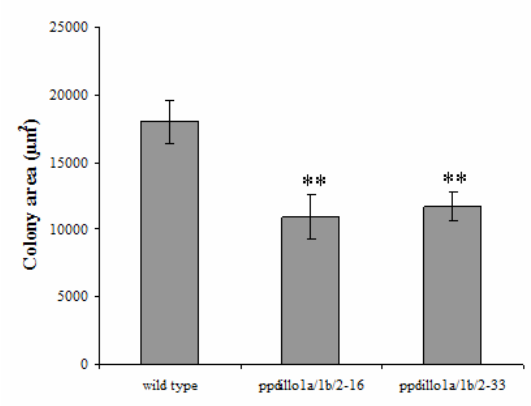


Figure 4.7 Average lengths of filaments emerging from wild type and *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* mutants. A) Protoplast after 14 days of growth. Arrows indicated an example of a filament measured. **B)** Wild type colonies regenerating from protoplasts had longer filaments than *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* deletion mutants. Error bars indicate ± standard error of the mean (SEM). Mean filament lengths (n=13): Wild type, 791.4µm ± 83.9µm; *ppdillo1a/1b/2-16*, 584.2µm ± 40.9µm (t test, *, $P < 0.05$); *ppdillo1a/1b/2-33*, 572.82µm ± 79.2µm.

A)



B)



C)

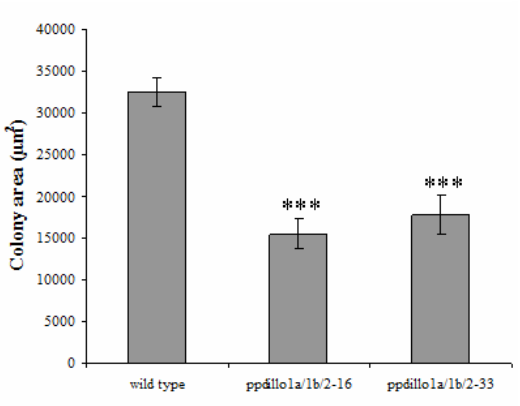


Figure 4.8 Mean colony sizes formed from regenerating protoplasts. Protoplasts were isolated from wild type and *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* triple deletion mutants and grown for one week. **A)** Image shows an example of a protoplast after one week of growth that was measured using Nikon imaging software. **B)** Wild type colonies were significantly larger than *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* colonies in experiment one. Error bars indicate ± standard error of the mean (SEM). Mean colony areas: Wild type 32517µm^2 ± 1761µm^2; *ppdillo1a/1b/2-16* 15478µm^2 ± 1815µm^2 (t test, ***, $P < 0.001$); *ppdillo1a/1b/2-33* 17746µm^2 ± 2388µm^2 (t test, ***, $P < 0.001$). **C)** Wild type colonies were significantly larger than *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* colonies in experiment two. Mean colony areas: Wild type 18012µm^2 ± 1587µm^2; *ppdillo1a/1b/2-16* 10942µm^2 ± 1672µm^2 (t test, **, $P < 0.01$); *ppdillo1a/1b/2-33* 11719µm^2 ± 1118µm^2 (t test, **, $P < 0.01$).

4.3.2 Effect of ABA on *Physcomitrella* spore germination

ABA responses are evolutionarily conserved and are believed to have been essential for plants to successfully colonise dry land (Tougane *et al*., 2010). The role of ABA in desiccation and freezing tolerance in bryophytes has been well documented. In evolutionarily more recent plants, ABA plays a key role in ensuring correct timing of seed germination, by maintaining seeds in dormancy until environmental conditions are suitable (Holdsworth *et al*., 2008).

arabidillo1/2 double knockout mutants have impaired germination capabilities; they are able to germinate in the presence of higher concentrations of ABA than wild type (Gibbs and Coates, unpublished). In order to determine whether PHYSCODILLO and ARABIDILLO proteins share conserved functions during germination, it was first necessary to determine whether *Physcomitrella* spore germination was also inhibited by ABA. Wild type sporophytes were therefore harvested and spores immediately germinated on minimal medium containing different concentrations of ABA. ABA concentrations were selected based on independent trials within the Coates lab and also on work published by other groups. *Physcomitrella* can tolerate much higher concentrations of ABA than *Arabidopsis*; *Arabidopsis* seed germination is fully inhibited by 10µm ABA and 1µm ABA is sufficient to inhibit lateral root initiation (Finkelstein and Somerville, 1990; deSmet *et al*., 2003; Khandelwal *et al*., 2010).

In an initial experiment, spores were treated with 0µm (control), 1µm, 10µm and 100µm ABA. After a six day growth period, the percentage of germinating spores was calculated. Spores from which at least a single filament emerged were counted as 'germinated' and those that did not produce a filament were counted as 'ungerminated'. All of the ABA concentrations tested inhibited spore germination: 33.6% spores were unable to germinate when treated with 1µm ABA, 62.9% spores were unable to germinate when treated

with 10µm ABA and 100% spores were unable to germinate when treated with 100µm ABA. In the untreated control, all of the spores were able to germinate (**Figure 4.9A**).

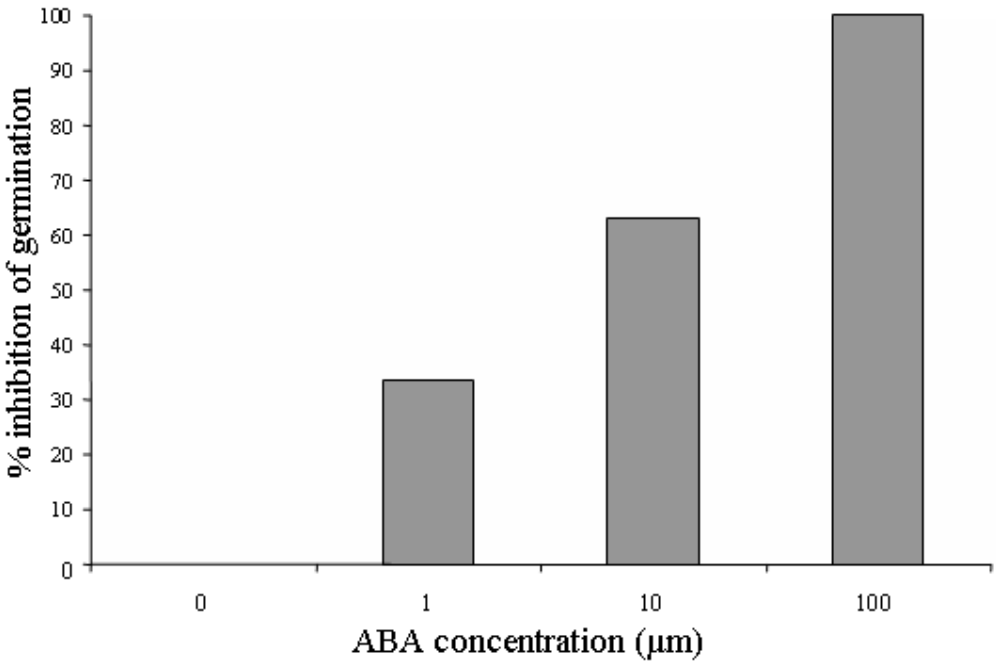
Since there were large differences between the inhibitory effects of spore germination at 10µm and 100µm ABA, spores were treated with a wider range of concentrations in a second experiment (0µm, 1µm, 10µm, 25µm, 50µm and 100µm ABA). After a six day growth period, the percentage of germinating spores was calculated. The inhibitory effects of ABA on spore germination observed during the second experiment were less severe than the preliminary experiment; 5.4% spores were unable to germinate when treated with 1µm ABA, 12.9% spores were unable to germinate when treated with 10µm ABA, 77% spores were unable to germinate when treated with 25µm ABA, 97.8% spores were unable to germinate when treated with 50µm ABA and 100% spores were unable to germinate when treated with 100µm ABA. In the untreated control, all of the spores were able to germinate, confirming that the inhibition of spore germination is due to the presence of ABA (**Figure 4.9B**). ABA inhibits *Physcomitrella* spore germination in a dose-dependent manner.

Figure 4.9 Relative inhibition of newly harvested spore germination with ABA. Wild type sporophytes were harvested and spores immediately germinated in the presence of different ABA concentrations. After six days, the percentage of germinating spores was calculated.

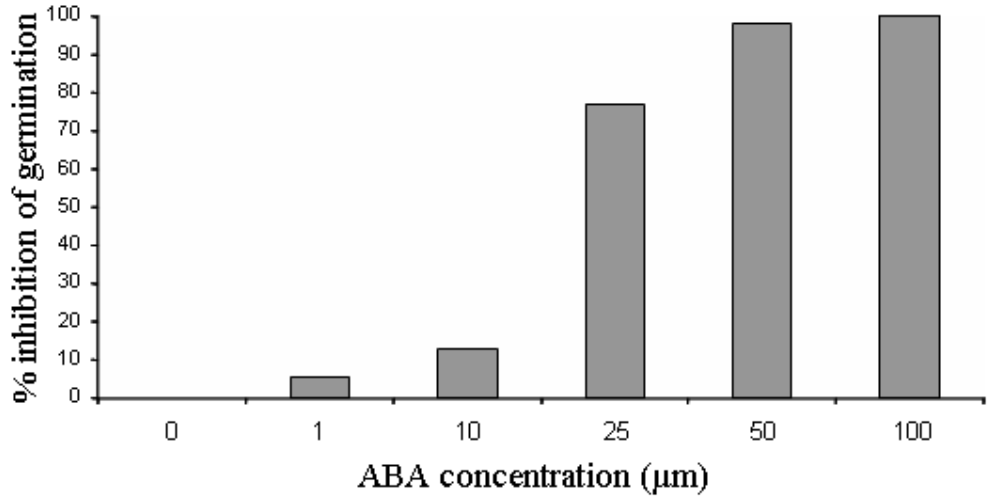
A) Spores were germinated in the presence of 0 (control), 1, 10 and 100µm ABA. As ABA concentration increases, the percentage inhibition of germination also increased. In the untreated control, all spores germinated.

B) Spores were germinated in the presence of 0, 1, 10, 25, 50 and 100µm ABA. As ABA concentration increased, the inhibition of germination also increased. In the untreated control, all spores germinated. ABA inhibits spore germination in a dose-dependent manner.

A)


% inhibition of germination
100
90
80
70
60
50
40
30
20
10
0
0
1
10
100
ABA concentration (µm)

B)


% inhibition of germination
100
90
80
70
60
50
40
30
20
10
0
0
1
10
25
50
100
ABA concentration (µm)

4.3.3 Effect of ABA on 'after-ripened' *Physcomitrella* spore germination

Angiosperm seeds undergo a specialised form of after-ripening before they are able to germinate and this process is regulated by ABA. In dormant barley grains and *Arabidopsis* seeds, endogenous levels of ABA are high and this prevents them from germinating during unfavourable environmental conditions. In after-ripened seeds, the ABA content of declines rapidly after a period of imbibition; these are consequently able to germinate (Jacobsen *et al*., 2002; Ali-Rachedi *et al*., 2004). To determine whether there are any 'after-ripening' effects on *Physcomitrella* spore germination, wild type sporophytes were harvested and stored at room temperature for a period of three months. Spores were then germinated in the presence of 0µm (control), 10µm, 25µm, 50µm and 100µm ABA. A methanol solvent control was also included to ensure that any effects observed were due to ABA and not the ABA solvent. After a six day growth period, the percentage of germinating spores was calculated. In two independent experiments, the inhibitory effects of ABA on 'after-ripened' spores were less severe than the inhibitory effects on newly harvested spores; all of the spores treated with 10µm ABA were able to germinate, 11.3% ± 3.67% spores were unable to germinate when treated with 25µm ABA, 22.8% ± 0.67% spores were unable to germinate when treated with 50µm ABA and 45.8% ± 9.73% spores were unable to germinate when treated with 100µm ABA. In the untreated control, all of the spores were able to germinate. In the methanol solvent control, all of the spores were able to germinate, confirming that ABA inhibits spore germination (**Figure 4.10**).

None of the freshly harvested spores germinated in the presence of 100µm ABA whereas over half of the 'after-ripened' spores germinated. Furthermore, the germination of freshly harvested spores was severely impaired at 10µm ABA whereas 'after-ripened' spores treated with 10µm ABA were all able to germinate. 'After-ripened' spores were more ABA tolerant than freshly harvested spores. This suggested that the mechanism controlling ABA-

mediated inhibition of both seed and spore germination may be ancient and evolutionarily conserved. It also demonstrates an example of co-option of function into diploid seeds that make the diploid sporophyte in later-evolving land plants and into haploid spores that make the haploid gametophyte in early-evolving land plants.

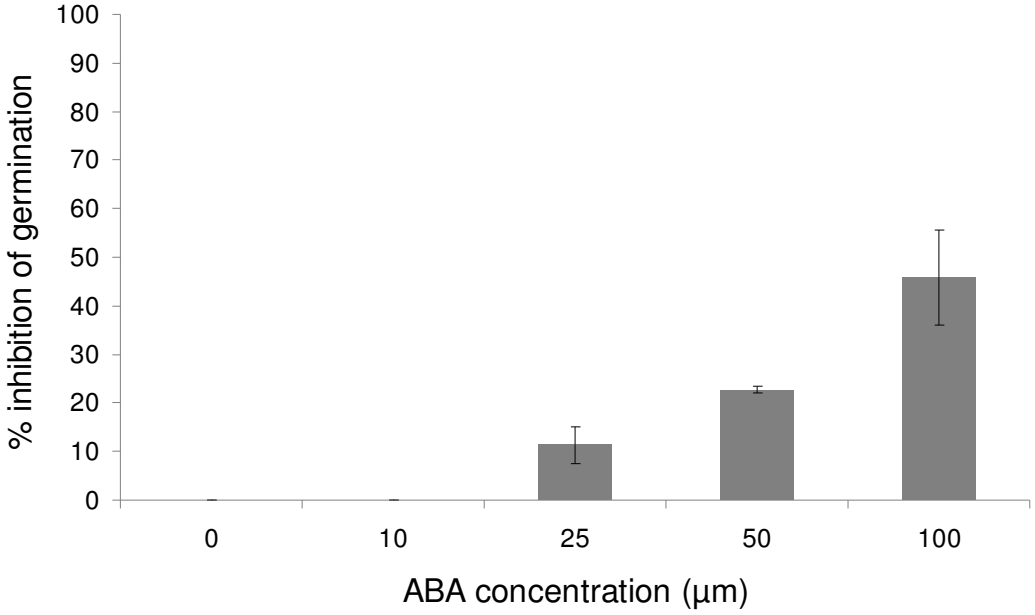


Figure 4.10 Relative inhibition of 'after-ripened' spore germination with ABA. Wild type sporophytes were harvested and stored for three months. Spores were then germinated in the presence of 10, 25, 50 and 100µm ABA. Spores were also germinated in the absence of ABA (untreated control and methanol solvent control). As ABA concentration increased, the percentage inhibition of germination increased. All spores were able to germinate in the untreated control, methanol solvent control and in the presence of 10µm ABA. 'After-ripened' spores were more ABA tolerant than freshly harvested spores. Error bars indicate ± standard error of the mean (SEM). Mean percentage inhibition of spore germination: Untreated control, 0% ± 0%; methanol solvent control, 0% ± 0%; 10µm ABA, 0% ± 0%; 25µm ABA, 11.3% ± 3.67%; 50µm ABA, 22.8% ± 0.67%; 100µm, 45.8% ± 9.73%.

4.3.3.1 Effect of ABA on *Physcomitrella* colony area

As an effect of ABA on spore germination has not been documented previously, the effects of ABA on filament growth were also tested for the first time. To examine the effects of ABA on early vegetative growth of *Physcomitrella*, areas of colonies regenerating from both germinating spores and protoplasts were examined.

4.3.3.1.1 Effect of ABA on colonies formed from germinating spores

Wild type spores were germinated in the presence of different concentrations of ABA (0, 10, 50 and 100µm ABA) and allowed to grow for a six day growth period. Areas of colonies formed from germinating spores were then measured (n=20). The mean size of colonies regenerating from untreated spores was 13797.7µm^2 ± 1218.4µm^2; the mean colony size after 10µm ABA treatment was 11187.2µm^2 ± 767.9µm^2; the mean colony size after 50µm ABA treatment was 10089.5µm^2 ± 832.5µm^2 (t test, *, $P < 0.05$); the mean colony size after 100µm ABA treatment was 9114.6µm^2 ± 1369.6µm^2 (t test, *, $P < 0.05$). As ABA concentration increased, colony area decreased in a dose-dependent manner (**Figure 4.11**).

4.3.3.1.2 Effect of ABA on colonies formed from regenerating protoplasts

Wild type protoplasts were prepared and regenerated in the presence of different concentrations of ABA (0, 10, 25, 50 and 100µm). After a growth period of eight days, areas of colonies formed from wild type protoplasts were then measured (n=25). The mean size of colonies regenerating from untreated protoplasts was 13253µm^2 ± 1112.3µm^2; the mean size of colonies regenerating after 10µm ABA treatment was 10577µm^2 ± 613.6µm^2 (t test, *, $P < 0.05$); the mean size of colonies regenerating after 25µm ABA treatment was 5819µm^2 ± 314.6µm^2 (t test, ***, $P < 0.001$); the mean size of colonies regenerating after 50µm ABA treatment was 4195µm^2 ± 355.9µm^2 (t test, ***, $P < 0.001$); the mean size of colonies

regenerating after 100µm ABA treatment was 4360µm^2 ± 326.4µm^2 (t test, ***, P < 0.001). Once again, as ABA concentration increased, colony area decreased in a dose-dependent manner (**Figure 4.12**).

4.3.3.1.3 Effect of ABA on cell expansion

In order to determine whether ABA induced changes in cell division or cell elongation or both, protoplasts were isolated from a transgenic line constitutively expressing nuclear GFP (NLS4, Bezanilla *et al*., 2003). NLS4 protoplasts were regenerated in the presence of different concentrations of ABA (0, 10 and 25µm). After a growth period of seven days, the mean cell length was calculated by measuring the length of a filament and dividing it by the number of fluorescent nuclei in the filament (n=10 colonies). In a preliminary experiment, the mean number of cell nuclei of untreated colonies was 10; the mean number of cell nuclei of colonies treated with 10µm ABA was 9.7; the mean number of cell nuclei of colonies treated with 25µm ABA was 6.6. Consequently, the mean cell length of untreated colonies was 30.55µm ± 1.82µm; the mean cell length of colonies treated with 10µm ABA was 29.16µm ± 2.11µm; the mean cell length of colonies treated with 25µm ABA was 17.59µm ± 1.58µm (t test, ***, P < 0.001). ABA therefore appears to affect both cell division and cell elongation in *Physcomitrella* (**Figure 4.13**).

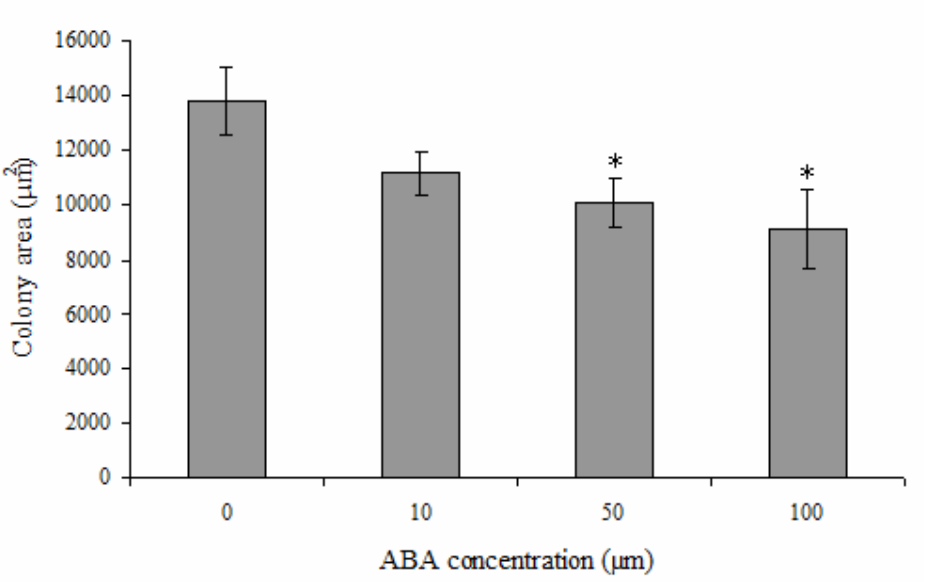


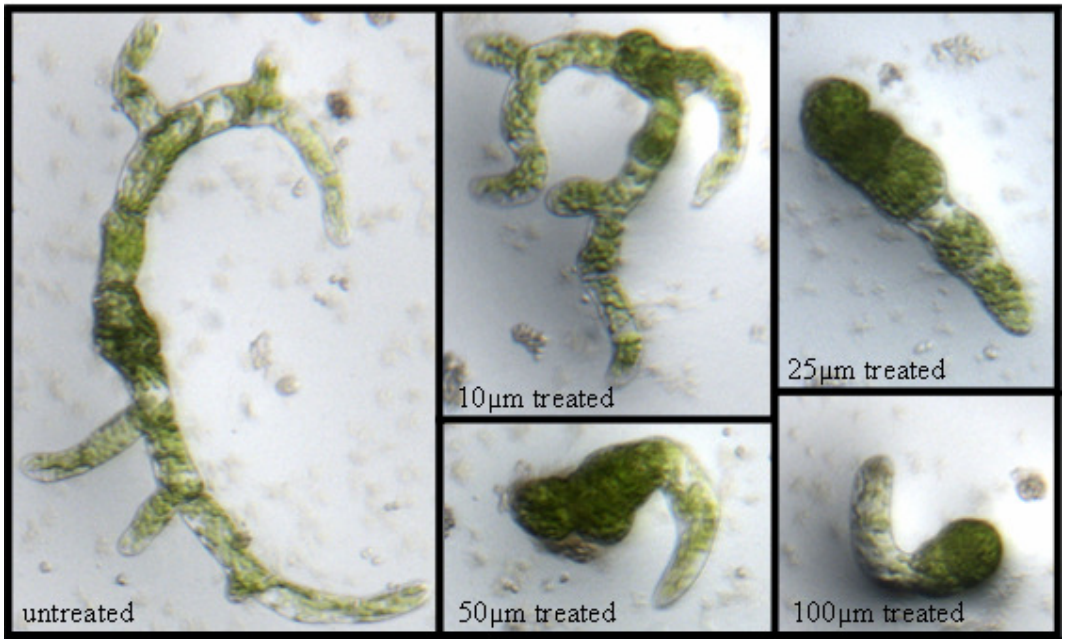
Figure 4.11 Mean colony sizes formed from germinating spores. Wild type spores were germinated in the presence of 0, 10, 50 and 100µm ABA. After a six day growth period, areas of colonies formed from germinating spores were then measured (n=20 for each treatment). As ABA concentration increased, colony area decreased in a dose-dependent manner. Error bars indicate ± standard error of the mean (SEM). Mean colony area: Untreated spores, 13797.7µm^2 ± 1218.4µm^2; 10µm ABA, 11187.2µm^2 ± 767.9µm^2; 50µm ABA, 10089.5µm^2 ± 832.5µm^2 (t test, *, $P < 0.05$); 100µm ABA, 9114.6µm^2 ± 1369.6µm^2 (t test, *, $P < 0.05$).

Figure 4.12 Mean colony sizes formed from regenerating protoplasts with ABA. Protoplasts were isolated from wild type protonemata, regenerated in the presence of different ABA concentrations and grown for eight days before being measured (n=25).

A) Representative images of wild type filaments regenerating on different concentrations of ABA. The corresponding treatment has been indicated at the bottom of each image.

B) Mean colony sizes formed from regenerating protoplasts treated with different ABA concentrations. As ABA concentration increases, colony area decreased in a dose-dependent manner. Error bars indicate ± standard error of the mean (SEM). Mean colony area: Untreated protoplasts, 13253μm^2 ± 1112.3μm^2; 10µm ABA treatment, 10577μm^2 ± 613.6μm^2 (t test, *, $P < 0.05$); 25µm ABA treatment, 5819μm^2 ± 314.6μm^2 (t test, ***, $P < 0.001$); 50µm ABA treatment, 4195μm^2 ± 355.9μm^2 (t test, ***, $P < 0.001$); 100µm ABA treatment, 4360μm^2 ± 326.4μm^2 (t test, ***, $P < 0.001$).

A)


untreated
10µm treated
25µm treated
50µm treated
100µm treated

B)

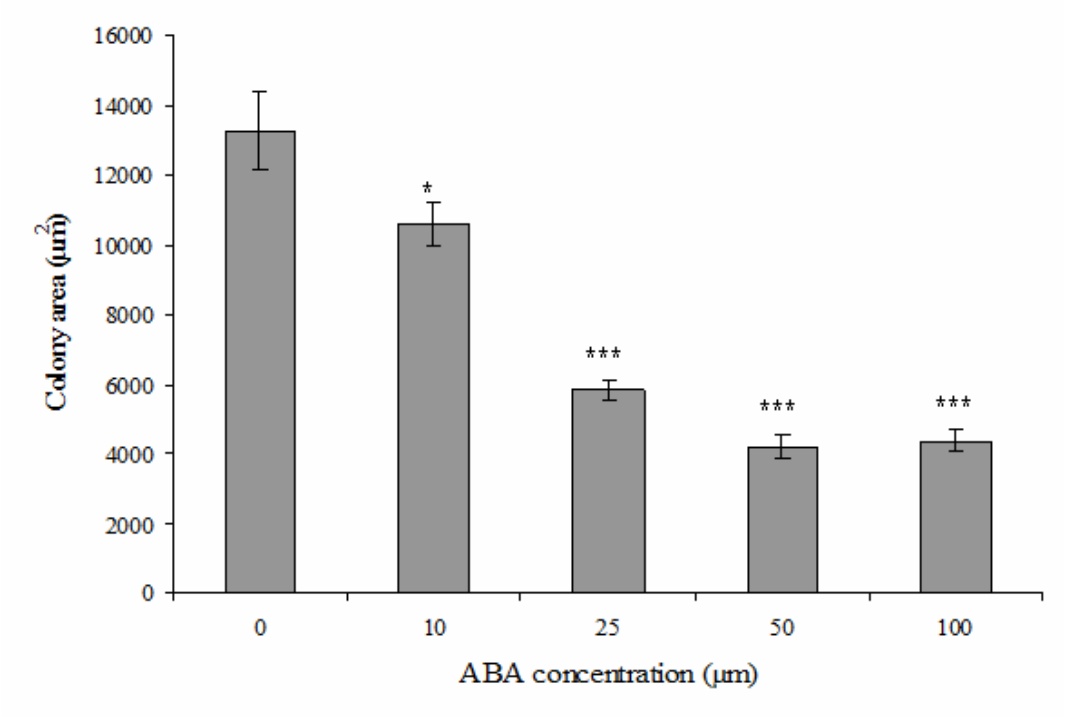
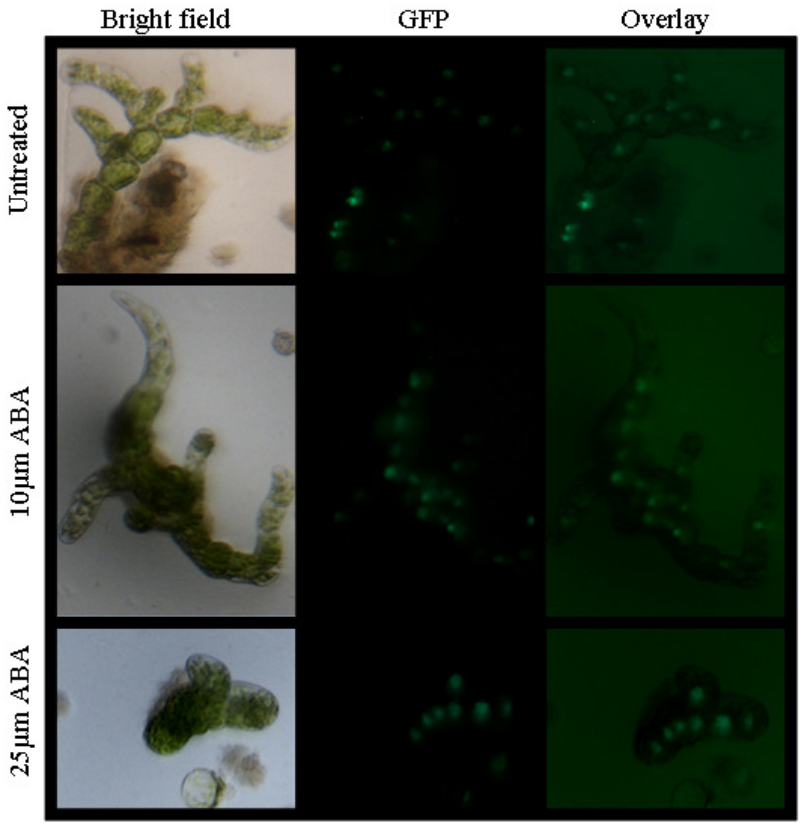
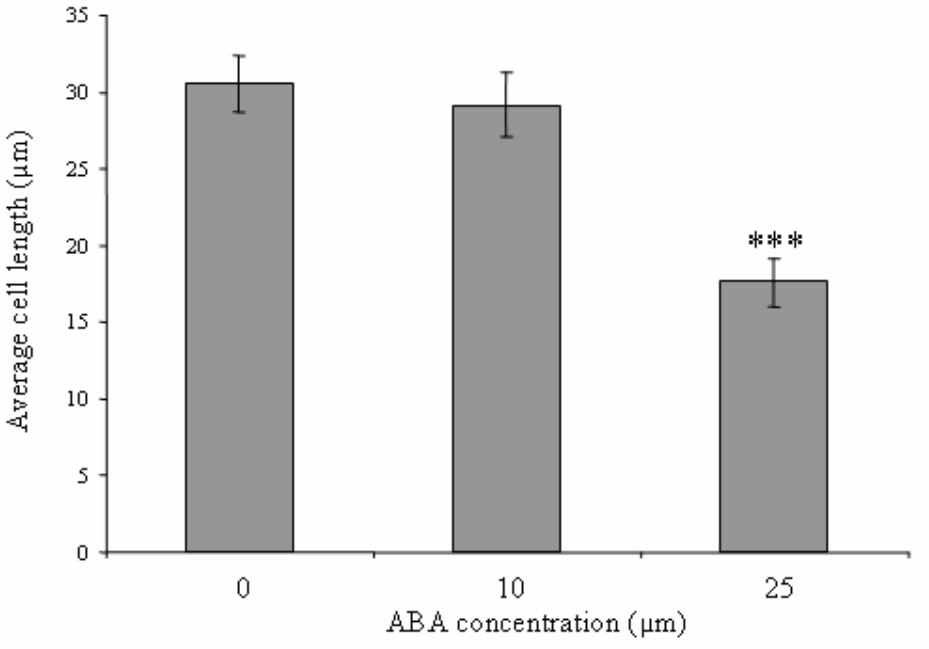

Colony area (µm²)
16000
14000
12000
10000
8000
6000
4000
2000
0
*
***
***
***
0
10
25
50
100
ABA concentration (µm)

Figure 4.13 Mean cell length in regenerating protoplasts treated with ABA. Protoplasts were isolated from the NLS4 line, which constitutively overexpresses nuclear GFP. Protoplasts were then regenerated in the presence of different ABA concentrations and grown for one week. A) Bright field and fluorescent images of untreated control and 10µm and 25µm ABA treated protoplasts. B) Mean cell length of filaments regenerating from ABA treated protoplasts. As ABA concentration increases, the mean cell length decreases. Error bars indicate ± standard error of the mean (SEM). Mean cell length: Untreated, 30.55µm ± 1.82µm; 10µm ABA treated, 29.16µm ± 2.11µm; 25µm ABA treated, 17.59µm ± 1.58µm (t test, ***, $P < 0.001$).

A)


Bright field
GFP
Overlay
Untreated
10µm ABA
25µm ABA

B)


Average cell length (µm)
35
30
25
20
15
10
5
0
***
0
10
25
ABA concentration (µm)

4.3.4 *physcodillo* triple deletion mutants: Response to ABA

arabidillo1/2 double knockout mutants exhibit insensitivity to ABA-mediated inhibition of seed germination. Conversely, ARABIDILLO1 overexpressing mutants exhibit hypersensitivity to ABA (Gibbs and Coates, unpublished data).

Work in this chapter has concluded that *Physcomitrella* spore germination is inhibited by exogenous ABA and that this inhibition takes place in a dose-dependent manner (section **4.3.2** and **4.3.3**). In order to determine whether ARABIDILLO proteins share an evolutionarily conserved role in seed/spore germination, spore germination assays were carried out using both newly harvested and three-month old spores isolated from wild type and *ppdillo1a/1b/2* triple deletion mutants. Spores were germinated in the presence of different concentrations of ABA and the percentage of germinating spores was calculated after a six day growth period.

Newly harvested spores from wild type and *ppdillo1a/1b/2-16* were all able to germinate in the absence of ABA (0µm control). Treatment of newly harvested spores with increasing ABA concentrations (1, 10, 25, 50 and 100µm) resulted in a dramatic decrease in the percentage of germinating spores. When treated with 10µm ABA, 12.9% wild type spores and 21.1% *ppdillo1a/1b/2-16* spores were unable to germinate; when treated with 25µm ABA, 77% wild type and 61% *ppdillo1a/1b/2-16* spores were unable to germinate and when treated with 50µm ABA, 97.8% wild type and 76.7% *ppdillo1a/1b/2-16* spores were unable to germinate. Interestingly, when treated with 100µm ABA, no wild type spores were able to germinate; *ppdillo1a/1b/2-16* spore germination was only inhibited by 94.1% (**Figure 4.14**). This suggested that *ppdillo1a/1b/2* triple deletion mutants exhibited insensitivity to ABA-mediated inhibition of spore germination.

'After-ripened' spores from wild type and both triple deletion lines were all able to germinate in the absence of ABA (0µm control). Similarly to newly harvested spores,

treatment of after-ripened spores with increasing concentrations of ABA (10, 25, 50 and 100µm), the inhibition of wild type spore germination was more severe than in both *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* deletion mutants. However, after-ripened spores were able to germinate in the presence of 100µm ABA whereas freshly harvested spores could not; only 36.1% wild type, 22.6% *ppdillo1a/1b/2-16* and 23.6% *ppdillo1a/1b/2-33* were unable to germinate when treated with 100µm ABA. The effects of germination in the presence of either 25µm or 50µm were also less severe in the mutants compared to wild type (**Figure 4.15**). *ppdillo1a/1b/2* triple deletion mutants therefore exhibit insensitivity to ABA-mediated inhibition of spore germination even after a period of 'after-ripening'. Although the effects are less severe, 'after-ripened' *arabidillo1/2* mutant seed also exhibits ABA insensitivity compared to wild type while ARABIDILLO1-overexpressing seeds are hypersensitive to ABA (Gibbs and Coates unpublished).

4.3.5 *physcodillo* triple deletion mutants: desiccation and freezing tolerance

Khandelwal *et al.* (2010) showed that when *Physcomitrella* protonemata are able to survive a period of desiccation when pretreated with ABA. They demonstrated that *ppabi3* triple deletion mutants were unresponsive to ABA and therefore did not survive after a period of desiccation (Khandelwal *et al*., 2010).

To determine whether the *ppdillo1a/1b/2* triple deletion mutants exhibited defects in ABA responses during both desiccation and freezing tolerance, a similar experiment was carried out. Wild type and *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* protonemata were grown for a period of one week. Tissues were then treated overnight with 0 (untreated control), 10 and 100µm ABA before being subjected to either desiccation or freezing. Tissues were either desiccated by transferring cellophanes to an empty petri dish and incubating in the growth

room for one week, or frozen at -20^{o}C for one week. After treatment, cellophanes with tissues were returned to normal growth conditions and grown for one week.

After one week, both wild type and *ppdillo1a/1b/2* tissues not treated with ABA died after desiccation and freezing whereas wild type and *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* tissues treated with ABA survived (data not shown). These results suggest that the pathway leading ABA to confer desiccation and freezing tolerance is distinct from ABA-mediated inhibition of seed/spore germination.

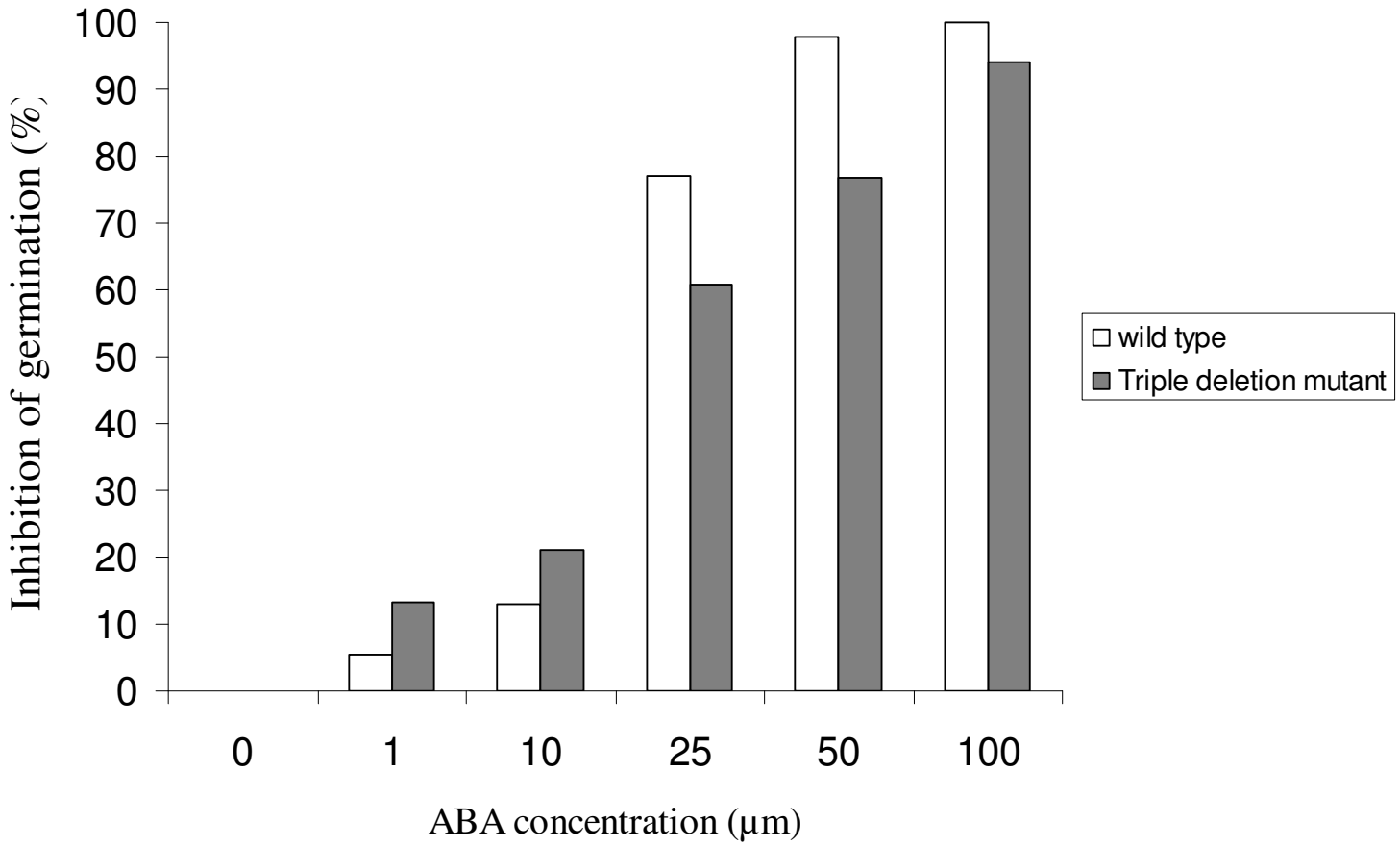


Figure 4.14 Relative levels of germination inhibition of newly harvested spores treated with ABA.
Sporophytes were harvested from wild type and the *ppdillo1a/1b/2-16* deletion mutant. Spores were then germinated immediately in the presence of different ABA concentrations. After a growth period of six days, the percentage inhibition of germination was calculated. As ABA concentration increased, the inhibition of germination increased. *ppdillo1a/1b/2-16* mutant spore germination was less inhibited than wild type.

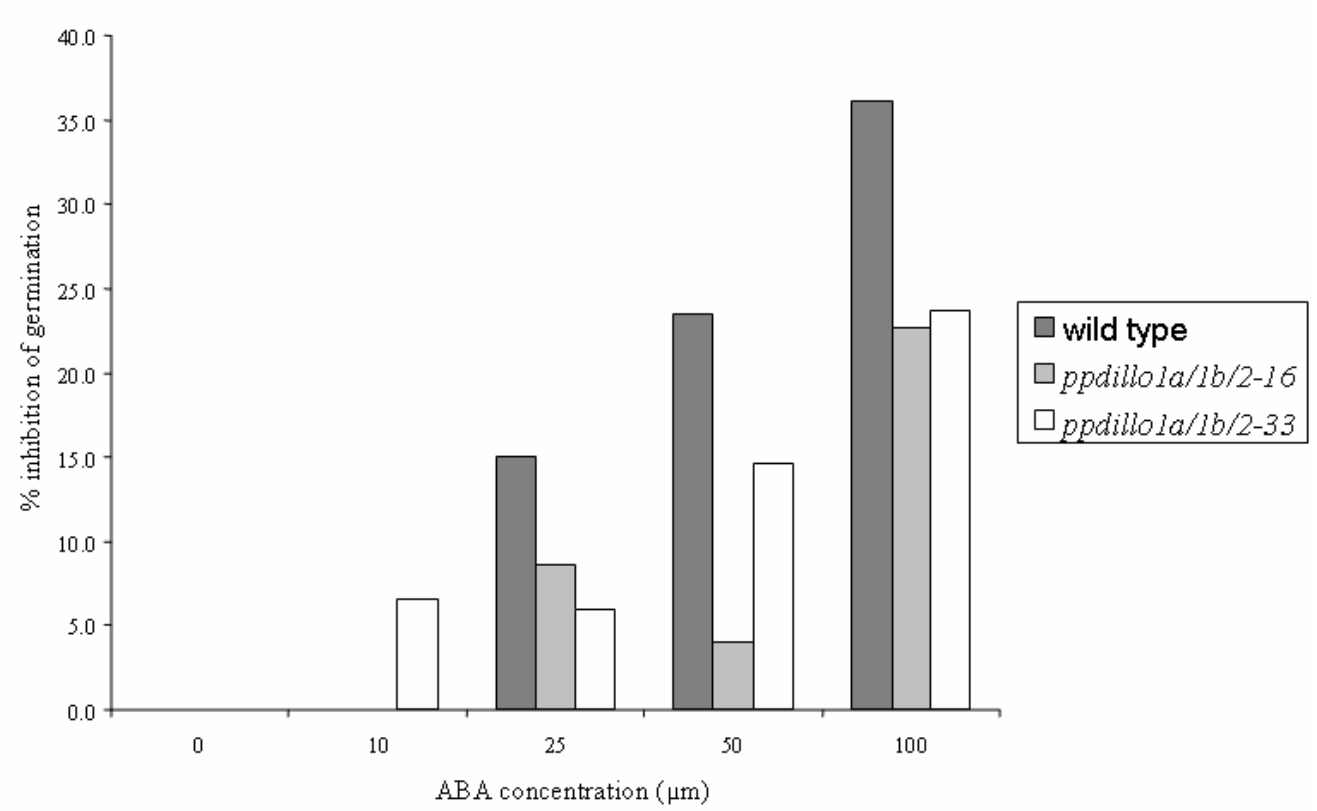


Figure 4.15 Relative levels of germination inhibition of 'after-ripened' spores treated with ABA.
Sporophytes were harvested from wild type and *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* deletion mutants and stored for three months. Spores were then germinated in the presence of different ABA concentrations. After a growth period of six days, the percentage inhibition of germination was calculated. As ABA concentration increased, the inhibition of germination increased. *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* mutant spore germination was less inhibited than wild type.

4.4 Generation of complementation lines of the *ppdillo1a/1b/2* mutants

It was important to generate a number of complementation lines in order to confirm that *ppdillo1a/1b/2* mutant phenotypes were caused by a loss of PHYSCODILLO proteins and not by incorrect integration of deletion constructs elsewhere in the *Physcomitrella* genome. Two different *ppdillo1a/1b/2* rescue constructs were prepared; rice actin promoter driven *PHYSCODILLO1A*-GFP and *PHYSCODILLO2*-promoter driven *PHYSCODILLO1A*-GFP (**Figure 4.16**). Constructs were linearised and transformed into *ppdillo1a/1b/2-16* mutant protoplasts. Since *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* mutants already contained Hygromycin and G418 resistance cassettes, the rescue construct contained a Zeocin resistance cassette. This enabled screening of putative transformants using Hygromycin, G418 and Zeocin antibiotics. After two rounds of selection, a number of putative rescue lines were obtained. Due to time constraints, it was not possible to screen these lines for phenotypes. However, they are currently being propagated in the Coates lab and phenotypes will be analysed at a later date. Since ARABIDILLO-YFP protein fusions are able to complement the *arabidillo1/2* mutant phenotype, it is likely that C-terminally fused GFP will not pose any threat to PHYSCODILLO protein function.

A)

B)

Figure 4.16 Constructs for *ppdillo1a/1b/2* triple deletion mutant complementation.

PHYSCODILLO1A cDNA was cloned and inserted in-frame with the GFP coding sequence in the pHSP-GUS-108-35SNPT vector. The vector contains two sequences homologous to the BS213 targeting sequence; integration sites within these positions of the genome cause no detrimental phenotypes. The expression of the *PHYSCODILLO1A*-GFP would be driven by either the rice actin promoter (A) or the *PHYSCODILLO2* gene promoter (B).

4.5 Discussion

4.5.1 Generation of stable *physcodillo* deletion mutants

Amongst the land plants, *Physcomitrella* has the unique ability to carry out gene targeting by homologous recombination at a similar efficiency to *Saccharomyces cerevisiae*. This technique was therefore an extremely useful means of disrupting the *PHYSCODILLO* genes using a gene replacement strategy in order to investigate their function.

Chapter three described the cloning of three full-length *PHYSCODILLO* genes that were subsequently named *PHYSCODILLO1A*, *PHYSCODILLO1B* and *PHYSCODILLO2*. *PHYSCODILLO2* was located as an individual gene on scaffold 13 whereas *PHYSCODILLO1A* and *PHYSCODILLO1B* genes were identified within the same genetic locus as an identical inverted repeat, in a tail-to-tail orientation.

Sequencing revealed that two of the putative *ppdillo2* single deletion lines were genuine *ppdillo2* deletion mutants; these were *ppdillo2.105* and *ppdillo2.110*. The endogenous *PHYSCODILLO2* genes of both *ppdillo2.105* and *ppdillo2.110* mutants contained a single copy of the antibiotic resistance cassette, confirming that single-copy allele replacement had taken place. Double homologous recombination-mediated allele replacement is often accompanied by ectopic insertions within alternative sites of the genome sequence (Kamisugi *et al*., 2005). Southern blotting will need to be carried out to determine whether this is the case. Following transformation using the *PHYSCODILLO2* deletion construct, relatively low gene targeting efficiencies were observed compared to transformations described throughout the remainder of the thesis. A critical factor that may have contributed to the low targeting efficiency of the *PHYSCODILLO2* gene was the concentration of transforming DNA. Only 10µg transforming DNA was used to generate

ppdillo2 deletion mutants, whereas 20µg transforming DNA was used in subsequent transformations and this greatly enhanced gene targeting efficiencies.

RT-PCR was carried out in order to confirm that the disrupted *PHYSCODILLO2* gene was unable to produce a functional mRNA transcript. Unfortunately, *PHYSCODILLO2* gene-specific primers amplified *PHYSCODILLO1A/1B* in all cases because *PHYSCODILLO2* and *PHYSCODILLO1A/1B* cDNA sequences were too similar (or shared too high a % identity) to one another. This was confirmed by interpreting the digest profiles of PCR products obtained. An attempt was made to use primers specific to sequences within 5' and 3' untranslated regions (UTRs) of the *PHYSCODILLO2* gene but this was also unsuccessful. Although no *PHYSCODILLO2* transcript was detected using *PHYSCODILLO2* UTR-specific primers, it was also not possible to detect the tubulin control transcript. It was therefore not possible to confirm that loss of *PHYSCODILLO2* gene transcript was due to *PHYSCODILLO2* gene disruption rather than cDNA that was not amplifiable by PCR. RT-PCR analyses are in the process of being repeated. Nevertheless, since sequencing of both *ppdillo2.105* and *ppdillo2.110* mutants confirmed the presence of the Hygromycin cassette within the *PHYSCODILLO2* gene, it was generally agreed that these were *bona fide ppdillo2* single deletion mutants.

Since functional redundancy operates within this gene family in *Arabidopsis*, it was necessary to generate triple deletion lines in which all three *PHYSCODILLO* genes had been disrupted. This proved far less straightforward than generating *ppdillo2* single deletion lines. A single deletion construct could have been used to target either *PHYSCODILLO1A* or *PHYSCODILLO1B* or both genes simultaneously. However, there were several reasons why this was not attempted. Screening of putative deletion lines would have been problematic and highly dependent on the ability to generate large 11kb PCR products. Although this was

carried out successfully when cloning and sequencing the *PHYSCODILLO1A/1B* locus (chapter three), it was not always easy. Moreover, if only one of two of the genes were disrupted successfully, targeting the second gene in a subsequent transformation would have been difficult, especially if using an identical deletion construct that differed only in the antibiotic resistance cassette that it contained. This may have carried out homologous recombination with the original *PHYSCODILLO1A/PHYSCODILLO1B* single deletion construct (within either *PHYSCODILLO1A* or *PHYSCODILLO1B* genes) instead of targeting the remaining non-targeted gene.

With these factors in consideration, it was decided that a novel strategy would be adopted to delete both *PHYSCODILLO1A* and *PHYSCODILLO1B* genes simultaneously using a single double deletion construct. This would involve removing a 23kb portion of the *PHYSCODILLO1A/1B* locus, the largest deletion attempted to date in *Physcomitrella*. Taking advantage of unique sequences either side of the *PHYSCODILLO1A/1B* locus that shared no homology to either gene; it was possible to replace the entire locus with a G418 resistance cassette. In order to generate triple deletion mutants, the *PHYSCODILLO1A/1B* double deletion construct was introduced into protoplasts isolated from the *ppdillo2.105* single deletion mutant. Following two rounds of selection using Hygromycin and G418 antibiotics, 15 putative triple deletion mutants were identified. These were subsequently screened for correct 5' and 3' integration using PCR. Genotyping PCR using DNA isolated from putative triple deletion mutant #24 appeared to have very large bands; perhaps this was due to the presence or more than one insert at the *PHYSCODILLO1A/1B* locus or simply non-specific annealing of primer sequences. Three of the lines showed both correct 5' and 3' integration and were therefore confirmed as the triple deletion mutants *ppdillo1a/1b/2-16*,

ppdillo1a/1b/2-33 and *ppdillo1a/1b/2-39*. *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* were selected for phenotypic analysis.

RT-PCR was carried out in order to confirm that *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* mutants did not produce functional *PHYSCODILLO* gene transcripts. Unfortunately, attempts were largely unsuccessful and often resulted in a large excess of bands. Tubulin control transcripts were detected in all cases, suggesting that cDNA was of sufficient quality to yield a PCR amplicon. However, products corresponding to the presence of *PHYSCODILLO* genes were amplified in template-free controls, suggesting that PCR reagents were contaminated with wild type *Physcomitrella* cDNA. This is in the process of being repeated with alternative reagents. Nevertheless, since the genotyping PCR yielded positive results, I am confident that *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* were authentic triple *physcodillo* deletion mutants. In order to verify that these mutants no longer have an intact *PHYSCODILLO1A/1B* locus, genotyping PCR products will be sequenced. If repeated RT-PCR attempts are unsuccessful, northern blotting could be carried out: this would be less sensitive and thus should rule out cross-contamination between the 3 genes. Southern blotting will determine whether there are any ectopic insertions of the *PHYSCODILLO1A/1B* construct elsewhere in the genome.

There are few published examples of large deletions in recent literature, especially deletions greater than 20kb. The earliest report of a large locus deletion using gene targeting was carried out in mouse embryonic stem cells. This involved removing a 15kb portion of the T-cell antigen receptor β-subunit locus and inserting a neomycin phosphotransferase gene by homologous recombination (Mombaerts *et al*., 1991). In *Plasmodium*, it is possible to delete sequences as large as 23kb-50kb using gene replacement by targeted double cross-over

recombination (Rita Tewari, personal communication). The 23kb *PHYSCODILLO1A/1B* deletion appears to be the largest known targeted deletion in *Physcomitrella* so far.

4.5.2 *physcodillo* triple deletion mutants exhibit defective protoplast regeneration

ppdillo1a/1b/2-16 and *ppdillo1a/1b/2-33* had no obvious morphological phenotypes; they exhibited normal responses to light and gravity and produced morphologically normal gametangia. They were able to successfully complete the life cycle within three months, from spore germination to sporophyte production.

In order to determine whether *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* mutants exhibited defects during early developmental processes, their ability to regenerate from protoplasts was investigated. In a preliminary experiment, protoplasts were isolated from wild type and the two triple deletion mutants and grown for a two week period. Filaments regenerating from *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* protoplasts were shorter than filaments regenerating from wild type protoplasts.

To investigate this phenotype further, sizes of colonies regenerating from wild type and *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* protoplasts were examined. After one week of regeneration, sizes of colonies regenerating from *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* mutant protoplasts were significantly smaller than those regenerating from wild type protoplasts and this was observed in two independent experiments. Preliminary analyses of protonemal branching suggested that the differences observed between wild type and *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* mutants were due to defects in cell elongation rather than cell division.

There are several published examples that describe *Physcomitrella* mutants with defects in actin-dependent processes, which have some similarities to the *ppdillo1a/1b/2* triple

deletion mutant phenotype. ARPC1 is an important component of the Arp2/3 complex. Unusually, *arpc1* RNAi lines do not have any caulonemal tissues and consequently do not form buds. *arpc1* RNAi lines possess chloronemata containing irregularly shaped cells, which have abnormal cell division patterns. ARPC1 therefore appears to play a pivotal role in polarised outgrowth of filaments (Harries *et al*., 2005). *profilin* RNAi lines have inhibited tip growth and consequently have fewer cells that are much smaller and rounder than wild type. Profilin interacts with actin and is essential for tip growth of protonemata (Vidali *et al*., 2007). Δ*brk1* mutants produce small filamentous colonies that contain short cells with misplaced cross walls (Perroud and Quatrano, 2008). Plants lacking class II formin genes exhibit defects in polarised tip growth. Growth of filaments is severely stunted and these contain spherical cells that have defects in actin organisation (Vidali *et al*., 2009).

The phenotype of *ppdillo1a/1b/2* triple deletion mutants is much less severe than any of the actin-associated phenotypes described. Preliminary work in the Coates group suggests that PHYSCODILLO proteins are unlikely to interact with the cytoskeleton directly, despite the cytoskeletal functions of animal and *Dictyostelium* Armadillo proteins. To examine this further, transient transformations using an actin marker, such as talin-GFP would determine whether there are any differences between the actin cytoskeletons of wild type plants and *ppdillo1a/1b/2* mutants..

To investigate whether *ppdillo1a/1b/2* triple deletion mutants exhibited defects in cell division, cell walls of protonemata could be stained with either propidium iodide or Calcofluor-white. The mean number of cells within colonies regenerating from both wild type and mutant protoplasts could then be determined more easily. If mutants were smaller due to a delay in cell division, colonies of protonemata regenerating from mutant protoplasts would contain fewer cells than those of wild type. By dividing filament lengths by total

numbers of cells, it would also be possible to determine whether the *ppdillo1a/1b/2* mutants exhibit defects in cell elongation. Alternatively, it would be possible to transiently transform both wild type and *ppdillo1a/1b/2* mutant protoplasts using the NLS4 construct and monitor regeneration over a two week period. Taking advantage of the presence of nuclear GFP, it would be possible to determine the number of cells by counting the number of fluorescent nuclei (Bezanilla *et al*., 2005).

4.5.3 Examining effects of ABA on *Physcomitrella* spore germination

arabidillo1/2 knockout mutants exhibit insensitivity to ABA-mediated inhibition of seed germination. Conversely, ARABIDILLO1 overexpressing mutants exhibit hypersensitive ABA responses (Gibbs and Coates, unpublished data). It has been known for some time that ABA maintains seed dormancy to prevent seeds from germinating in unfavourable environmental conditions (Finkelstein *et al*., 2008).

The fundamental aims of the ABA experiments described in this chapter were to determine whether ARABIDILLO proteins from both *Arabidopsis* and *Physcomitrella* shared a conserved function during seed/spore germination. Unlike angiosperms, there have been no reports indicating that ABA regulates germination of spores from early-evolving land plants. In studies of the fern *Mohria caffrorum*, ABA was shown to have no effect on spore germination. However, they only examined relatively low concentrations of ABA (Chia and Raghavan, 1982). It was therefore important to determine whether ABA had an involvement in spore germination before being able to make any comparisons between the effects of ABA on wild type and *physcodillo* triple deletion mutant spores.

4.5.3.1 Effects of ABA on newly harvested spore germination

Wild type sporophytes were harvested and spores germinated immediately in the presence of different ABA concentrations. A preliminary experiment revealed that as ABA concentration increased (0, 1, 10, 100µm ABA), the percentage of germinating spores decreased dramatically; no spores were able to germinate in the presence of 100µm ABA. A wider range of ABA concentrations were then examined in a repeat experiment and the effects of ABA were less severe. Although they fit the same trend, the large differences in germination inhibition between experiments one and two suggested that the effects of ABA could be dependent on the age of the spores. Sporophytes used in the second experiment may have been more mature than those used in the first experiment, even though they were harvested at the same time, immediately prior to germination. Endogenous levels of ABA within spores may decrease throughout spore maturation in a similar mechanism to those of seeds, therefore the effects of exogenously applied ABA would have less inhibitory effects on spore germination..

4.5.3.2 Effect of ABA on 'after-ripened' spore germination

'After-ripening' has been extensively described in angiosperms; it is the process by which endogenous levels of ABA (normally high in dormant spores) are broken down in a time-dependent manner in order to permit germination. The effects of ABA are less severe on 'after-ripened' seed germination than newly harvested seed germination (Jacobsen *et al*., 2002; Ali-Rachedi *et al*., 2004).

To determine whether spore age affects the severity of ABA-mediated inhibition of spore germination, wild type sporophytes were harvested and stored for a three month period. Similarly to 'after-ripened' seeds, the ABA-mediated inhibition of 'after-ripened' spore

germination was less severe. All of the 'after-ripened' spores were able to germinate in the presence of 10µm ABA and only mild inhibition was observed in the presence of 25µm and 50µm ABA. Over half of the 'after-ripened' spores could germinate in the presence of 100µm ABA whereas none of the freshly harvested spores were able to germinate at all. All of the spores germinated in the methanol solvent control, confirming that spore germination was inhibited by ABA itself and not the solvent.

These findings demonstrate that angiosperm seed germination and *Physcomitrella* spore germination are both regulated by ABA. Furthermore, it appears that the effects of ABA on spore germination are less severe after a period of 'after-ripening'. It therefore appears that the mechanism underlying the ABA-mediated control of seed/spore germination may be evolutionarily conserved from the bryophytes to the angiosperms.

Gibberellins promote seed germination and cell elongation in angiosperms. However, bioactive gibberellins do not appear to be present in bryophytes. Interestingly, precursors of gibberellins appear to be functional in *Physcomitrella* and these regulate spore germination. Spores germinating in the presence of the ent-kaurene (gibberellin precursor) inhibitor AMO-1618 mimic the effects of ABA (Anterola *et al*., 2009). Thus, there appears to be a conserved regulatory mechanism throughout land plants controlling the seed/spore dormancy/germination status by balancing gibberellin and ABA. If this hypothesis is correct, then treating *Physcomitrella* spores with different ratios of ent-kaurene and ABA will alter the proportions of germinating spores. For example, the inhibition of spores germinating in the presence of both 25µm ABA and ent-kaurene would be less severe than those germinating in the presence of ABA alone.

4.5.3.3 The effects of ABA on cell division and cell elongation

In order to determine the effects of ABA on *Physcomitrella* development, sizes of colonies produced from both germinating spores and regenerating protoplasts were examined.

4.5.3.3.1 The effects of ABA on filament growth from spores

To examine the effects of ABA on filamentous growth from spores, wild type sporophytes were harvested and spores germinated immediately in the presence of different ABA concentrations (0, 10, 50 and 100µm). As ABA concentration increased, sizes of filamentous colonies decreased in a dose-dependent manner.

4.5.3.3.2 Effects of ABA on filament growth from protoplasts

To examine the effects of ABA on filamentous growth from protoplasts, wild type protoplasts were isolated and regenerated in the presence of different ABA concentrations (0, 10, 25, 50 and 100µm ABA). Similarly to germinating spores, sizes of filamentous colonies regenerating from protoplasts decreased as ABA concentration increased.

To determine whether the effects of ABA were due to inhibition of cell elongation or cell division, protoplasts were isolated from a line constitutively expressing nuclear GFP (NLS4; Bezanilla *et al*., 2005). They were then regenerated in the presence of different ABA concentrations (0, 10, 25µm ABA). After one week, the average length of cells within each of the ABA treated filaments was determined by counting numbers of fluorescent nuclei within a given length of filament. As ABA concentration increased, the mean cell length decreased. Cells of filaments regenerating from 25µm ABA treated protoplasts were almost half the size of those of untreated filaments. They also contained fewer cells than untreated

filaments. ABA therefore appeared to inhibit elongation and cell division in protonemata. This data is only preliminary and will need to be repeated at a later date.

This apparent role of ABA in *Physcomitrella* protonemata cell division and cell elongation appears to be consistent with other plants. ABA inhibits cell division within the quiescent centre of the *Arabidopsis* root meristem (Han *et al*., 2010). ABA also prevents the progression of cell division within the *Arabidopsis* embryo and consequently prevents seed germination, a process dependent on SMALL ORGAN 2 (SMO2; Hu *et al*., 2010). The role of ABA in cell expansion has also been described in the literature. In coffee, ABA inhibits cell expansion within the embryo and consequently inhibits seed germination. There are also contradictory publications showing that ABA both inhibits and promotes root cell elongation (Pilet and Saugy, 1987).

4.5.4 ARABIDILLO proteins: evolutionarily conserved roles in germination

Work in this chapter has shown that ABA inhibits *Physcomitrella* spore germination and that the severity of this inhibition decreases as spore age increases. *Physcomitrella* spores therefore appear to undergo a specialised 'after-ripening' process comparable to that of *Arabidopsis* and barley seeds, whereby levels of endogenous ABA appear to decrease in a time-dependent manner to allow dormancy release in favourable conditions (Jacobsen *et al*., 2002; Ali-Rachedi *et al*., 2004).

To test whether roles of ARABIDILLO proteins in seed/spore germination were evolutionarily conserved, sporophytes were harvested from wild type and *ppdillo1a/1b/2* triple deletion mutants. Some of the spores were germinated immediately in the presence of

different concentrations of ABA whereas others were germinated after a three month period of 'after-ripening'.

In a preliminary experiment, freshly harvested wild type spores were unable to germinate in the presence of 100µm ABA whereas 5.9% *ppdillo1a/1b/2-16* spores germinated. Similarly, freshly harvested wild type spore germination was inhibited by 97.8% in the presence of 50µm ABA compared to just 76.7% *ppdillo1a/1b/2-16* spores. Germination was also more severely impaired in wild type spores than *ppdillo1a/1b/2-16* spores in the presence of 25µm ABA.

The effects of ABA on spore germination of both wild type and *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* mutants were less severe after a three month period of 'after-ripening'. Nevertheless, mutant spores exhibited insensitivity to ABA-mediated inhibition of spore germination when they were freshly harvested and after a period of 'after-ripening'.

Complementation lines have been generated whereby PHYSCODILLO1A has been re-introduced into *ppdillo1a/1b/2* mutants. Phenotypic analyses of these lines will need to be carried out. This will confirm that mutant phenotypes are due to the loss of *PHYSCODILLO* genes and not the ectopic insertion of the *PHYSCODILLO1A/1B* double deletion construct at an alternative location within the genome sequence.

Although the effects of ABA were rather subtle, these findings have been independently confirmed by other members the Coates lab. Spores of *PHYSCODILLO2* overexpressing mutants exhibit hypersensitive responses to inhibition of germination and protonemal growth by ABA and this corresponds nicely with the ARABIDILLO1 overexpression phenotype in *Arabidopsis*, which also exhibits hypersensitivity to exogenous ABA (Younousse Saidi, unpublished data; Gibbs and Coates, unpublished data). The role of ARABIDILLOs in diploid seed germination to generate the sporophyte may therefore have

been co-opted from an ancestral function, which was to control germination of haploid spores to generate the gametophyte.

4.5.5 Future work

Several lines of evidence are outstanding. For example, the loss of *PHYSCODILLO* transcripts from *ppdillo1a/1b/2* triple deletion mutants has not been confirmed. In order to demonstrate that targeting of the *PHYSCODILLO* deletion constructs has been successful, a combination of Southern and northern blotting could be carried out. Southern blotting will demonstrate that the large *PHYSCODILLO1A/1B* locus was deleted and subsequently replaced with a G418 resistance cassette. As an alternative to the highly sensitive technique PCR, northern blotting can be used to show that no *PHYSCODILLO* transcript is present in the *ppdillo1a/1b/2* triple deletion mutants. A number of *ppdillo1a/1b/2* complementation lines have been generated. Analyses of these lines will determine whether the *ppdillo1a/1b/2* phenotype occurred as a result of the loss of PHYSCODILLO proteins or mistargeting of the *PHYSCODILLO1A/1B* double deletion construct. Since two independent *ppdillo1a/1b/2* triple deletion mutants behave similarly, it is unlikely that the construct would have been mistargeted in two independently transformed protoplasts.

The work in this chapter has shown that PHYSCODILLO proteins play roles in the regulation of spore germination in *Physcomitrella*. It is also known that PHYSCODILLO proteins do not appear to play roles in desiccation or freezing tolerance, other pathways mediated by ABA (Khandelwal *et al*., 2010). It would therefore be interesting to investigate other ABA-dependent processes, such as stomatal opening in *ppdillo1a/1b/2* mutants, compared to wild type. Work in the Coates group is currently investigating whether

arabidillo1/2 double knockout mutants exhibit any impairment in their ability to regulate stomatal aperture compared to wild type *Arabidopsis* plants.

There are still unanswered questions concerning the roles of ABA in other early-evolving land plants. It would therefore be interesting to determine whether ABA inhibits seed/spore germination in the lycophytes and ferns, which evolved after the bryophytes and before the angiosperms, in addition to evolutionarily ancient algae.

Chapter V:

ARABIDILLO cross-species complementation

5.1 Introduction

The bryophytes were the earliest plants to successfully transition from water to land approximately 500 million years ago. In order to colonise progressively drier habitats, plants subsequently evolved a number of adaptations that enabled them to survive and reproduce on land, such as vascular tissues, roots, seeds and flowers. Throughout evolution, plants have acquired diverse morphological characteristics. Most notably, mosses and lycophytes lack sophisticated rooting systems that are present in vascular plants, such as rice and *Arabidopsis*. They also reproduce by dispersing haploid spores and not by producing seeds. The presence of highly conserved ARABIDILLO homologues in all of the diverse land plant lineages is therefore intriguing in land plant lineages that lack both roots and seeds.

The purpose of the work in this chapter was to determine whether ARABIDILLO homologues from evolutionarily ancient plants that lack both roots and seeds are functionally equivalent to ARABIDILLO proteins.

5.2 Generation of Physcodillo1A-GFP and Selagidillo-GFP constructs

In conjunction with loss-of-function studies, overexpression analyses are a useful way of investigating protein function. It is also possible to visualise subcellular localisations of specific proteins of interest by generating green fluorescent protein (GFP) fusion constructs. ARABIDILLO-GFP fusion proteins driven by the 35S promoter are detected exclusively in the nuclei of root cells. Furthermore, ARABIDILLO1 and ARABIDILLO2 overexpressing mutants have complementary phenotypes to that of *arabidillo1/2* double knockout mutants; overexpressing mutants have more lateral roots than wild type (Coates *et al.*, 2006).

In order to determine the subcellular localisation of Physcodillo1A and Selagidillo proteins, Physcodillo1A and Selagidillo cDNA sequences were cloned (as detailed in Chapter

three) and inserted in-frame with the green fluorescent protein (GFP) coding sequence in pGreenII 0029. The pGreenII 0029 vector also contained a constitutive Cauliflower mosaic virus (CaMV) 35S promoter sequence in order to drive the expression of the resulting fusion transcripts (Hellens *et al*., 2000; Coates *et al*., 2006). Physcodillo1A-GFP and Selagidillo-GFP fusion constructs can be seen in **Figure 5.1A** and **5.2B** respectively.

5.3 Subcellular localisation of Physcodillo1A-GFP and Selagidillo-GFP

Kalderon et al. discovered the first nuclear localisation signal (NLS) in the SV40 large T-antigen (Kalderon et al., 1984). A consensus sequence for monopartite NLSs was later defined as K-(K/R)-X-(K/R) where 'X' denotes any amino acid (Dingwall et al., 1988).

ARABIDILLO1 and ARABIDILLO2 proteins are expressed exclusively in the nuclei of root cells. Their NLSs therefore appear to be functional, especially as mutating residues within the NLS abolishes nuclear targeting of ARABIDILLO proteins (Coates *et al*., 2006; Nibau *et al*., 2011). In addition, Physcodillo1A, Physcodillo1B, Physcodillo2 and Selagidillo all appear to possess an NLS and these fit the consensus sequence defined by Dingwall et al. It is therefore likely that they are also targeted to cell nuclei (**Figure 5.1**).

In order to determine the subcellular localisation of Physcodillo1A and Selagidillo proteins and elucidate where they might function in cells, Physcodillo1A-GFP and Selagidillo-GFP constructs were transiently transformed into *Physcomitrella* protoplasts. The localisation of the GFP fusion proteins was visualised using confocal microscopy two days after transformation, with the assistance of Younousse Saidi. Similarly to ARABIDILLO1 and -2 in *Arabidopsis*, Physcodillo1A-GFP was detected exclusively in the nucleus of *Physcomitrella* protoplasts, suggesting that the Physcodillo proteins have a nuclear function

(**Figure 5.1B**). Interestingly, Selagidillo-GFP was detected predominantly in the nucleus but, to varying degrees, also as punctuate foci within the cytosol (**Figure 5.2).**

Figure 5.1 Subcellular localisation of Physcodillo1A.

A) The Physcodillo1A-GFP construct was prepared by inserting *Physcodillo1A* cDNA in-frame with the GFP coding sequence. The expression of the Physcodillo1A-GFP fusion protein was driven by the constitutive CaMV35S promoter.

B) Subcellular localisation of the Physcodillo1A-GFP fusion protein in two independently transformed *Physcomitrella* protoplasts (as indicated). Images on the left show Physcodillo1A-GFP green fluorescence, images in the middle show red chlorophyll autofluoresence and images on the right show merged images of both GFP and chlorophyll autofluorescence.

A)


35S
Physcodillo1A
GFP

B)

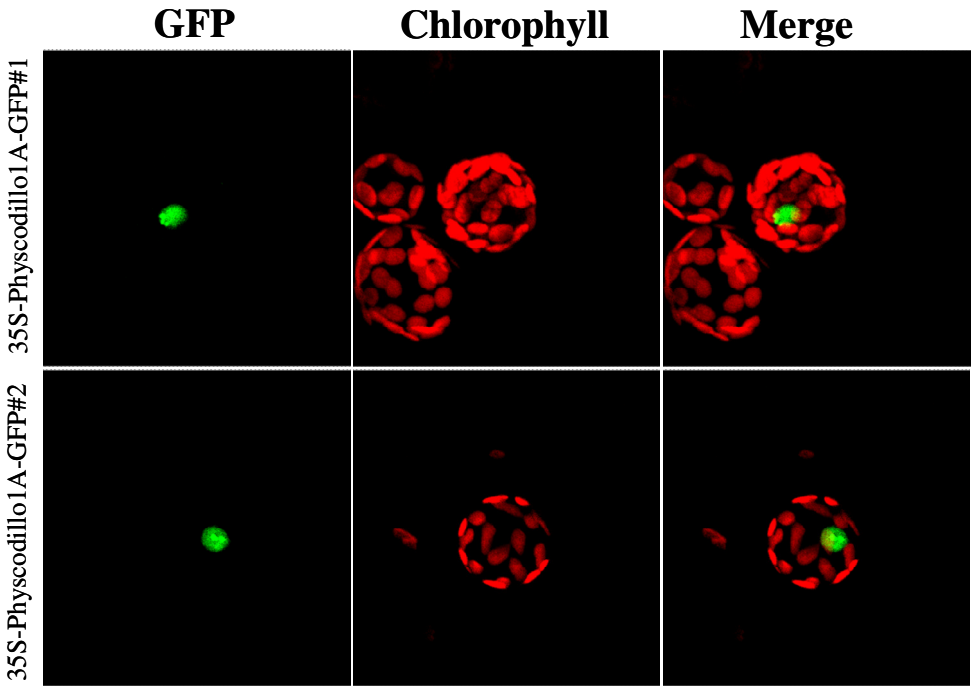
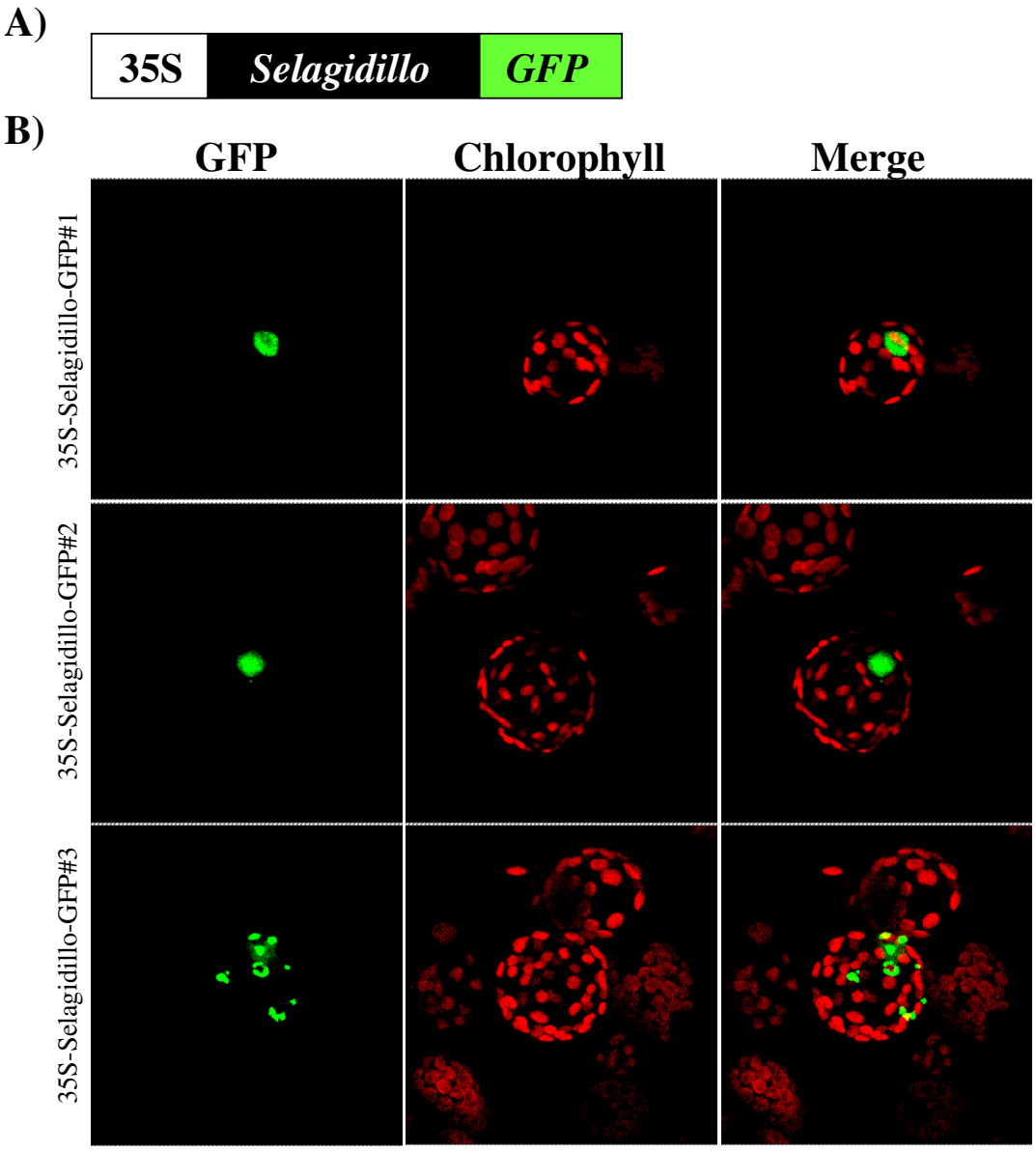

GFP
Chlorophyll
Merge
35S-Physcodillo1A-GFP#1
35S-Physcodillo1A-GFP#2

Figure 5.2 Subcellular localisation of Selagidillo.

A) The Selagidillo-GFP construct was prepared by inserting *Selagidillo* cDNA in-frame with the GFP coding sequence. The expression of the Selagidillo-GFP fusion protein was driven by the constitutive CaMV35S promoter.

B) Subcellular localisation of the Selagidillo-GFP fusion protein in three independently transformed *Physcomitrella* protoplasts (as indicated). Images on the left show Selagidillo-GFP green fluorescence, images in the middle show red chlorophyll autofluoresence and images on the right show merged images of both GFP and chlorophyll autofluorescence.


A)
35S
Selagidillo
GFP
B)
GFP
Chlorophyll
Merge
35S-Selagidillo-GFP#1
35S-Selagidillo-GFP#2
35S-Selagidillo-GFP#3

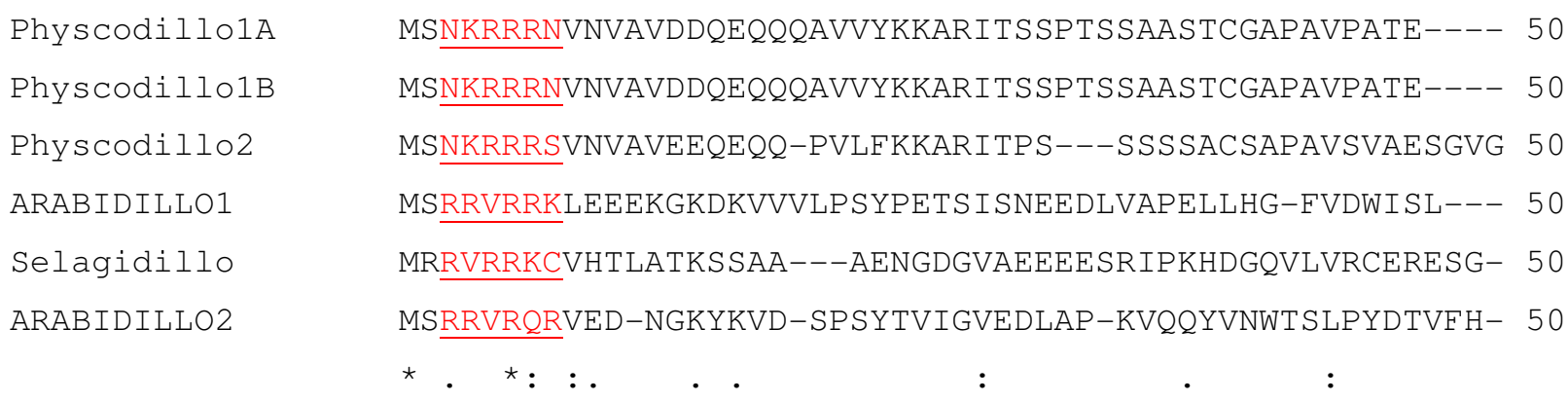


Figure 5.3 Putative nuclear localisation signals (NLS) in the N-terminal portions of ARABIDILLO homologues.

The N-terminal regions of Physcodillo1A, Physcodillo1B, Physcodillo2, Selagidillo, ARABIDILLO1 and ARABIDILLO2 were aligned using ClustalW2. Putative NLSs are indicated in red type and have been underlined. Numbers to the right of the alignment indicate the relative position of amino acids from the start residue methionine. Identical residues are denoted by asterisks (*), conserved residues denoted by a colon (:) and semi-conserved residues by a period (.).

5.4 Generation of *Arabidopsis* plants expressing Physcodillo1A or Selagidillo

To determine whether *Physcomitrella* or *Selaginella* ARABIDILLOs could function in *Arabidopsis*, transgenic plants expressing either Physcodillo1A or Selagidillo were generated. Expressing either of these proteins in wild type *Arabidopsis* plants would determine whether there were any phenotypes as a result of their overexpression. Expressing either of these proteins in *arabidillo1/arabidillo2* double knockout mutants would determine whether they are capable of rescuing both lateral root and seed germination phenotypes.

5.4.1 Generation of Physcodillo1A-GFP homozygous overexpression lines

Physcodillo1A-GFP was transformed into both wild type *Arabidopsis* and *arabidillo1/2* mutant plants using the floral dip method and then allowed to self-pollinate and set seed (T1). T1 seeds were subsequently screened using kanamycin since the pGreen plasmid contains a kanamycin resistance gene. Seven wild type (WT) seedlings and five *arabidillo1/2* (ara1/2) mutant seedlings survived selection and were therefore likely to contain the Physcodillo1A-GFP insertion. These heterozygous seedlings were then transferred to soil, allowed to self-pollinate and set seed. T2 seeds were then sown on medium containing 50µg/ml kanamycin for segregation analysis. Ten days after germination, the numbers of kanamycin resistant (Kan^R) and kanamycin sensitive (Kan^S) seedlings were determined. Seven out of 12 T2 lines from heterozygous T1 plants segregated at a ratio of approximately 3:1 (resistant to susceptible to kanamycin), suggesting that these lines contained a single insert of the Physcodillo1A-GFP construct (**Table 5.1**). The following lines were grown up in order to obtain homozygous T3 seed: WT.35S-Physcodillo1A-GFP.2, WT.35S-Physcodillo1A-GFP.3, WT.35S-Physcodillo1A-GFP.5 and WT.35S-

Physcodillo1A-GFP.6 and ara1/2.35S-Physcodillo1A-GFP.3 and ara1/2.35S-Physcodillo1A-GFP.5. Due to time constraints, it was not possible to carry out phenotypic analysis on these lines. However, they will be examined for both lateral root and seed germination phenotypes at a later date to determine whether ARABIDILLO proteins have evolutionarily conserved functions. The remaining lines did not segregate at a 3:1 ratio and therefore will be omitted from further analyses.

5.4.2 Generation of Selagidillo-GFP homozygous overexpression lines

Selagidillo-GFP was transformed into both wild type *Arabidopsis* and *arabidillo1/2* mutant plants using the floral dip method and then allowed to set T1 seed, which were subsequently screened using kanamycin. Six wild type (WT) seedlings and five *arabidillo1/2* (ara1/2) mutant seedlings survived selection and were therefore likely to contain the Selagidillo-GFP insertion. These heterozygous seedlings were transferred to soil, allowed to self-pollinate and set seed. T2 seeds were then sown on medium containing 50µg/ml kanamycin for segregation analysis. Ten days after germination, the numbers of kanamycin resistant (Kan^R) and kanamycin sensitive (Kan^S) seedlings were determined. 10 out of the 11 T2 lines from heterozygous T1 plants segregated at a ratio of approximately 3:1 (resistant to susceptible to kanamycin), suggesting that these lines contained a single insert of the Selagidillo-GFP construct (**Table 5.2**). The following lines were grown up in order to obtain homozygous T3 seed: WT.35S-Selagidillo-GFP.1, WT.35S-Selagidillo-GFP.3, WT.35S-Selagidillo-GFP.5, WT.35S-Selagidillo-GFP.6, ara1/2.35S-Selagidillo-GFP.2, ara1/2.35S-Selagidillo-GFP.3, ara1/2.35S-Selagidillo-GFP.4 and ara1/2.35S-Selagidillo-GFP.5. Again, due to time constraints, it was not possible to carry out phenotypic analysis on these lines but they will be examined for lateral root

and seed germination phenotypes at a later date. ara1/2.35S-Selagidillo-GFP.1 segregated at a ratio of 9:1, indicating that it contained more than one insert. This was therefore omitted from further analyses. WT.35S-Selagidillo-GFP.2 and WT.35S-Selagidillo-GFP.4 segregated close to a 3:1 ratio and were stored for use if required.

Transformant	Number of plants			% Kan^R	Segregation ratio
	Kan^R	Kan^S	Total		
WT.35S-Physcodillo1A-GFP.1	40	8	48	83.3	5:1
WT.35S-Physcodillo1A-GFP.2	64	25	89	71.9	2.56:1 *
WT.35S-Physcodillo1A-GFP.3	60	18	78	76.9	3.3:1 *
WT.35S-Physcodillo1A-GFP.4	55	13	68	80.9	4.23:1
WT.35S-Physcodillo1A-GFP.5	52	18	70	74.3	2.89:1 *
WT.35S-Physcodillo1A-GFP.6	70	26	96	72.9	2.69:1 *
WT.35S-Physcodillo1A-GFP.7	120	49	169	71	2.45:1 *
ara1/2.35S-Physcodillo1A-GFP.1	58	6	64	90.6	9.67:1
ara1/2.35S-Physcodillo1A-GFP.2	66	2	68	97.1	33:1
ara1/2.35S-Physcodillo1A-GFP.3	55	15	70	78.6	3.67:1 *
ara1/2.35S-Physcodillo1A-GFP.4	56	2	58	96.6	28:1
ara1/2.35S-Physcodillo1A-GFP.5	58	19	77	75.3	3.05:1 *

Table 5.1 Segregation of T2 progeny of 35S-Physcodillo1A-GFP transformants. T2 seeds were harvested and germinated on medium containing 50µg/ml kanamycin. Ten days after germination, the number of Kan^R and Kan^S seedlings was counted and the segregation ratio of resistant to susceptible to kanamycin determined. Transgenic lines that segregated at a 3:1 ratio were self-pollinated in order to obtain homozygous T3 seed. These lines have been indicated by asterisks (*).

Transformant	Number of plants			% Kan^R	Segregation ratio
	Kan^R	Kan^S	Total		
WT.35S-Selagidillo-GFP.1	56	22	78	71.8	2.54:1 *
WT.35S-Selagidillo-GFP.2	60	26	86	69.8	2.31:1
WT.35S-Selagidillo-GFP.3	101	43	144	70.1	2.35:1 *
WT.35S-Selagidillo-GFP.4	47	13	60	78.3	3.62:1
WT.35S-Selagidillo-GFP.5	43	14	57	75.4	3.07:1 *
WT.35S-Selagidillo-GFP.6	42	15	57	73.7	2.8:1 *
ara1/2.35S-Selagidillo-GFP.1	90	10	100	90	9:1
ara1/2.35S-Selagidillo-GFP.2	40	11	51	78.4	3.64:1 *
ara1/2.35S-Selagidillo-GFP.3	90	30	120	75	3:1 *
ara1/2.35S-Selagidillo-GFP.4	68	23	91	74.7	2.96:1 *
ara1/2.35S-Selagidillo-GFP.5	73	19	92	79.3	3.84:1 *

Table 5.2 Segregation of T2 progeny of 35S-Selagidillo-GFP transformants. T2 seeds were harvested and germinated on medium containing 50µg/ml kanamycin. Ten days after germination, the number of Kan^R and Kan^S seedlings was counted and the segregation ratio of resistant to susceptible to kanamycin determined. Transgenic lines that segregated at a 3:1 ratio were self-pollinated in order to obtain homozygous T3 seed. These lines have been indicated by asterisks (*).

5.5 Discussion

5.5.1 Localisation of Physcodillo1A and Selagidillo-GFP fusion proteins

ARABIDILLO genes are expressed in a number of tissues but the proteins they encode are found exclusively in the nuclei of root cells (Coates et al., 2006). ARABIDILLO proteins are highly unstable and are targeted for ubiquitination and degradation via the proteasome, as the proteasomal degradation inhibitor MG132 stabilises ARABIDILLO1 and -2 (Nibau et al., 2011).

Like the ARABIDILLO proteins, Selagidillo and the Physcodillo proteins possess an NLS at their N-termini and these fit the consensus sequence defined by Dingwall et al (1988). Given the presence of an NLS in the protein sequences of all ARABIDILLO homologues, Physcodillo1A and Selagidillo were expected to localise to cell nuclei. Physcodillo1A was found exclusively in nuclei of protoplasts isolated from *Physcomitrella* protonemata. However, although Selagidillo was predominantly detected in nuclei, GFP signal was often observed throughout the cytosol as large punctuate foci. This could be due to some of the 35S-Selagidillo-GFP transformed protoplasts having high copy numbers of plasmids, which boost protein levels to excessive levels. This could interfere with normal protein expression patterns and lead to aberrant localisation. The presence of these foci may be due to expression of Selagidillo in a heterologous host system. However, this seems unlikely since ARABIDILLO1 exhibits nuclear localisation in *Physcomitrella* protoplasts and vice versa (Younousse Saidi, unpublished data). Similar punctuate foci are observed when truncated versions of Physcodillo and ARABIDILLO1, lacking NLSs, are expressed in *Physcomitrella* and *Arabidopsis* protoplasts respectively (Nibau et al., 2011; Younousse Saidi, unpublished data).

Preliminary results in the Coates lab suggest that Physcodillo proteins are more stable than the ARABIDILLOs but that they are also turned over by the proteasome. Stable lines expressing soybean heat shock promoter driven Physcodillo1A/1B and Physcodillo2 localise to nuclei of all tissues examined; protonemata, leafy gametophores, rhizoids, archegonia and antheridia (Younousse Saidi, unpublished data). However, the subcellular localisation of Physcodillo proteins driven by their own promoter is unknown. A number of complementation lines have been generated and these have been discussed in chapter four of this thesis. The purpose of the complementation lines was to confirm that *physcodillo1a/1b/2* triple mutant phenotypes were a result of a loss of *Physcodillo* gene expression and not mistargeting by the double knockout construct to an alternative location within the *Physcomitrella* genome. Due to time constraints however, it was not possible to analyse these lines for tissue-specific expression patterns of Physcodillo proteins. In the future, the localisation of Physcodillo proteins driven by their endogenous promoters will need to be investigated. This will determine where they are likely to function in the plant. Physcodillo-promoter::GUS activity was detected throughout *Physcomitrella* tissues with the exception of archegonia and the diploid sporophyte. It is therefore likely that Physcodillo proteins will not be found in these tissues either. Investigating where Physcodillo proteins are expressed will determine whether they are turned over more rapidly in specific tissues and reveal more about their possible functions.

5.5.2 Generation of transgenic plants expressing Physcodillo1A or Selagidillo

In order to determine whether ARABIDILLO proteins have functions that are conserved across the land plants, transgenic *Arabidopsis* lines expressing either Physcodillo1A or Selagidillo were generated using the floral dip method. To establish

whether bryophyte and/or lycophyte representatives could rescue *arabidillo1/2* lateral root and seed germination phenotypes, both Physcodillo1A and Selagidillo were expressed in *arabidillo1/2* mutant backgrounds. To establish whether bryophyte and/or lycophyte representatives could induce phenotypic changes in *Arabidopsis* plants, Physcodillo1A and Selagidillo were expressed in wild type backgrounds.

A number of heterozygous T2 individuals were obtained that segregated at a ratio of 3:1 (resistant to susceptible to kanamycin). This indicated that a single copy insert had integrated into the *Arabidopsis* genome. These were subsequently allowed to self-pollinate in order to obtain homozygous T3 individuals. The majority of the remaining transformants displayed irregular non-Mendelian segregation ratios (e.g. 28:1 and 33:1) with a very high proportion of resistant plants among the progeny. This suggested that more than one copy of the construct had integrated at a number of independent sites within the genome.

Due to time constraints, it was not possible to test whether ARABIDILLOs were capable of cross-species complementation between *Physcomitrella*, *Selaginella* and *Arabidopsis*. The generation of a number of T3 homozygous *Arabidopsis* lines expressing either Physcodillo1A-GFP or Selagidillo-GFP fusion proteins is underway. Once homozygous lines have been identified, they will be subjected to phenotypic analyses.

5.5.3 Future work: Cross-species complementation experiments

A number of research groups have demonstrated that orthologous proteins from land plants are capable of fulfilling functions across species. Consequently, there are numerous publications demonstrating conserved functions across 400 million years of separation, from basal land plants to modern day angiosperms. For example, OPEN STOMATA 1 (OST1) homologues from *Physcomitrella* and *Selaginella* were recently shown to be capable of

controlling ABA-mediated control of stomatal aperture in *Arabidopsis* and fully complementing *Atost1* mutants (Chater et al., 2011; Ruszala et al., 2011). This is particularly interesting since stomata are found exclusively in the diploid sporophyte of *Physcomitrella*.

Conversely, there are a number of examples showing that *Physcomitrella* proteins can only partially complement *Arabidopsis* mutant phenotypes and vice versa. Golden2-like (GLK) transcription factors regulate chloroplast development in maize, *Arabidopsis* and *Physcomitrella*. Loss-of-function mutant phenotypes include impaired thylakoid formation and reduced chlorophyll biosynthesis in both *Arabidopsis* and *Physcomitrella*. *Physcomitrella GLK* (*PpGLK1*) can partially rescue the *Arabidopsis glk1/glk2* double knockout mutant phenotype whereas *Arabidopsis* GLK (*GLK1*) cannot complement *Ppglk1/glk2* double deletion mutants at all. GLK genes therefore exhibit a uni-directional complementation between *Arabidopsis* and *Physcomitrella* (Yasumura et al., 2005; Bravo-Garcia et al., 2009). Another notable example is ABSCISIC ACID INSENSITIVE 3 (ABI3), which confers desiccation tolerance to both *Physcomitrella* vegetative tissues and *Arabidopsis* seeds. *Arabidopsis abi3* mutants therefore have green, desiccation intolerant seeds that are able to germinate in the presence of high concentrations of ABA. *PpABI3A*, a *Physcomitrella* homologue of *ABI3*, can partially rescue *abi3* mutant phenotypes; seeds are no longer green but still germinate in the presence of high concentrations of exogenous ABA. It is unknown whether *Arabidopsis* ABI3 can rescue *ppabi3* mutant phenotypes as this does not appear to have been examined (Marella et al., 2006; Khandelwal et al., 2010).

5.5.3.1 Cross-species complementation of the *arabidillo1/2* lateral root phenotype

Arabidopsis plants have complex rooting structures that are heavily branched to permit more efficient water and nutrient uptake. *arabidillo1/2* double knockout mutants have

fewer lateral roots than wild type and are therefore likely to be compromised in their ability to locate vital water and nutrients (Coates et al., 2006). It is believed that ARABIDILLOs promote cell division in the pericycle in order to initiate lateral root growth (Coates et al., 2006). Unlike angiosperms, *Physcomitrella* and *Selaginella* lack sophisticated rooting structures; *Physcomitrella* possesses filamentous rhizoids that are equivalent to *Arabidopsis* root hairs and *Selaginella* produces rhizophores that differentiate into simple bifurcating rooting structures. The mechanism by which *Selaginella* rooting structures bifurcate is largely unknown. It would therefore be interesting to establish whether ARABIDILLO homologues from organisms that lack roots are able to promote root branching in *Arabidopsis*. Wild type and *arabidillo1/2* mutant plants expressing either Physcodillo1A-GFP or Selagidillo-GFP have been generated. These will be grown on vertical plates and their lateral root densities determined. If Physcodillo1A and Selagidillo are capable of carrying out a rooting function in *Arabidopsis*, an increase in lateral root density will be observed. This would also demonstrate that early-evolving ARABIDILLO homologues are able to interact with angiosperm-specific transcription factors, especially since Physcodillos and Selagidillo are likely to have a nuclear function.

5.5.3.2 Cross-species complementation of the arabidillo1/2 seed germination phenotype

arabidillo1/2 mutant seeds exhibit insensitivity to ABA-mediated inhibition of germination (Gibbs and Coates, unpublished data). Unlike angiosperms, *Physcomitrella* and *Selaginella* do not produce seeds but instead reproduce by generating and dispersing spores by aerial means. It would therefore be interesting to examine whether Physcodillo1A and Selagidillo are able to rescue seed germination phenotypes of *arabidillo1/2* mutants. Since ppdillo1a/1b/2 triple deletion mutants exhibit insensitivity to ABA-mediated inhibition of

germination, it is not unreasonable to hypothesise that Physcodillo1A and Selagidillo rescue lines will exhibit hypersensitive responses to exogenous ABA.

These cross-species complementation experiments will determine whether ARABIDILLO proteins have evolutionarily conserved roles in seed/spore germination.

5.5.4 Future work: Cross-species complementation experiments in *Physcomitrella*

Work in this thesis has shown that *Physcomitrella* spore germination is controlled by ABA, suggesting that ABA-mediated control of germination is evolutionarily conserved from the early-evolving bryophytes to the angiosperms. Moreover, similarly to *arabidillo1/2* mutants, *ppdillo1a/1b/2* triple deletion mutants exhibit insensitivity to ABA-mediated inhibition of spore germination. It would therefore be interesting to determine whether ARABIDILLO1 and Selagidillo can rescue the spore germination phenotypes of *ppdillo1a/1b/2* mutants.

Work in this thesis also showed that filaments regenerating from protoplasts prepared from *ppdillo1a/1b/2* mutant tissues produced smaller colonies than those of wild type. To determine whether ARABIDILLO1 and Selagidillo are able to rescue protoplast regeneration phenotypes, they will be introduced into both wild type and mutant lines.

Chapter VI:

General Discussion

6.1 Introduction

The field of plant evolutionary developmental biology (evo-devo) research has become very popular in recent years. Combined with modern molecular techniques, the ability to transform a variety of plants and access genome sequences of a number of evolutionarily divergent species, it is possible to answer fundamental questions about plant evolution at the molecular level. Understanding more about how individual gene families have evolved will highlight potential gene targets in order to ultimately improve yields of important crops for use as both food and fuel.

Land plants evolved from single-celled aquatic algal ancestors and therefore had to develop a number of responses that enabled them to survive stresses that accompanied the movement to land, such as gravity, UV, salinity and desiccation. The earliest plants to successfully transition from water to land approximately 500 million years ago were the bryophytes: the liverworts, mosses and hornworts (Kenrick and Crane, 1997). Plants subsequently evolved a number of morphological adaptations that enabled them to survive and reproduce on land, such as vascular tissues, roots, seeds and flowers.

In the angiosperm *Arabidopsis*, ARABIDILLO1 and ARABIDILLO2 function redundantly to promote root branching. They also regulate seed germination in an ABA-dependent manner and their gene expression is upregulated by ABA (Coates *et al*., 2006; Gibbs and Coates, unpublished data).

ARABIDILLO homologues are found throughout the land plants and this includes the early-evolving bryophytes and lycophytes. ARABIDILLO homologues have highly conserved protein structures; all have a nuclear localisation signal (NLS), an N-terminal F-box domain, a series of leucine-rich repeats (LRR) and an Armadillo (ARM) repeat domain (Nibau *et al*., 2011).

The aim of the work in this thesis was to determine the ancestral roles of ARABIDILLO-like proteins, by characterising their function in a moss species that lacks both seeds and multicellular roots. The general results of this work and ideas for future experimentation are discussed below.

6.2 The Physcodillos

Cloning and sequencing revealed that there were three ARABIDILLO homologues in the *Physcomitrella* genome: *Physcodillo1A*, *Physcodillo1B* and *Physcodillo2*. *Physcodillo1A* and *Physcodillo1B* are 100% identical to each other and exist in a tail-to-tail orientation, suggesting that a recent gene duplication event had taken place; *Physcodillo2* appears to be located elsewhere in the genome.

It remains unclear whether *Physcodillo1A* and/or *Physcodillo1B* are expressed *in planta* since their transcript sequences are indistinguishable. However, Physcodillo proteins share a striking resemblance to the ARABIDILLO proteins. Like the ARABIDILLOs, Physcodillo proteins are found exclusively in the nucleus, suggesting that they may also have a nuclear function. Stable lines expressing either Physcodillo1A/1B:GFP or Physcodillo2:GFP driven from a soybean heat-shock promoter have been generated in the Coates lab. Physcodillo:GFP fusion proteins are found exclusively in the nucleus of all *Physcomitrella* tissues. Furthermore, Physcodillo:GFP fusion proteins are stabilised by the proteasome inhibitor MG132, demonstrating that Physcodillo proteins are turned over by the proteasome in a ubiquitin-dependent manner (Younousse Saidi, unpublished data), similarly to ARABIDILLOs (Nibau *et al*., 2011). Since the heat-shock promoter drives extremely high levels of gene expression (Saidi *et al*., 2005) results may not be a true reflection of endogenous *Physcodillo* protein expression. By generating transgenic lines in which

functional Physcodillo:GFP fusions are knocked into endogenous *Physcodillo* gene loci and driven by *Physcodillo* gene promoters, it would be possible to determine whether Physcodillo protein expression is restricted to specific tissue types. However, since Physcodillo proteins are relatively unstable, it may not be possible to detect a Physcodillo:GFP signal using this approach without the proteasome inhibitor MG132.

ARABIDILLO proteins have *bona fide* F-box domains that are able to interact with *Arabidopsis* SKP1-like (ASK) proteins in a yeast two-hybrid system (Nibau *et al*., 2011). Homologues of ASK and Cullin proteins have been identified in all land plant genomes, so it seems reasonable to assume that ARABIDILLO homologues may also function as *bona fide* F-box proteins. Using the Physcodillo F-box domain as 'bait' and *Physcomitrella* homologues of ASK proteins as 'prey' in a yeast 2-hybrid assay would determine whether Physcodillo proteins are able to form SCF E3 ubiquitin ligases in *Physcomitrella*.

The ARABIDILLO1 ARM repeat domain interacts with a group of R2R3 MYB transcription factors, which were consequently named the ARABIDILLO-interacting MYBs (AIMs; Daniel Gibbs, PhD thesis). *aim* knockout mutants have more lateral roots than wild type whereas AIM overexpressing mutants have fewer lateral roots than wild type (Gibbs *et al*., unpublished data). ARABIDILLO proteins are therefore likely to target inhibitors of lateral root initiation (AIMs) for ubiquitination and degradation via the proteasome.

Both *Physcomitrella* and *Selaginella* genomes encode numerous MYB transcription factors but there are no obvious AIM homologues. Furthermore, neither *Physcomitrella* nor *Selaginella* has 'true roots'. *Physcomitrella* has filamentous rhizoids for nutrient acquisition and anchorage to substrata and these are equivalent to root hairs of *Arabidopsis* (Menand *et al*., 2007). *Selaginella* has rhizophores that differentiate into multicellular rooting structures. However, these roots branch dichotomously at the tip and have no obvious lateral root

counterparts (Banks, 2008). AIMs may therefore have been acquired by later-evolving plants to carry out a rooting function, such as the ferns which were among the earliest plants to evolve 'true' multicellular roots. Determining which proteins interact with Physcodillo proteins will help to elucidate the ancestral roles of the ARABIDILLOs and this work is ongoing in the Coates group.

Since ARABIDILLO proteins interact with MYB transcription factors, they may accordingly activate or repress the transcription of a number of target genes in cell nuclei. Physcodillo proteins may also interact with transcription factors in cell nuclei that collectively activate transcription of a distinct or overlapping group of genes. Similarly, although homologues of the AIM transcription factors do not exist in the *Physcomitrella* genome, they may interact with alternative transcription factors in the nucleus. Research in the Coates group will compare differences between proteomes of wild type and *ppdillo1a/1b/2* triple deletion mutants using stable isotope labelling with amino acids in cell culture (SILAC), a popular method used to quantify proteins.

6.3 The role of Physcodillos during early development

Work in this thesis described the generation of a number of *physcodillo* (*ppdillo*) deletion mutants. *ppdillo2* single deletion mutants had no obvious morphological phenotypes, which is perhaps unsurprising given the redundant nature of ARABIDILLO genes. Two independent triple deletion mutants, *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* exhibited growth defects during early developmental processes. Colonies regenerating from *ppdillo1a/1b/2-16* and *ppdillo1a/1b/2-33* protoplasts were significantly smaller than wild type. However, it is unclear whether mutants exhibit a delay in cell division (fewer cells) or slower tip growth (shorter cells). It would be relatively straightforward to determine numbers

and lengths of their cells by using propidium iodide to stain cell walls of regenerating protoplasts. *ppdillo1a/1b/2* mutants with fewer cells than wild type would therefore have delayed cell division; *ppdillo1a/1b/2* mutants with shorter cells than wild type but equivalent cell numbers would have defects in cell elongation; *ppdillo1a/1b/2* mutants with fewer and shorter cells would have defects in both cell division and cell elongation. ARABIDILLO proteins appear to regulate cell division events leading to the emergence of lateral root primordia (Coates *et al*., 2006).

6.4 Physcodillos and ABA

ARABIDILLO proteins clearly show some involvement in ABA-dependent processes in plants. In the presence of ABA, seed germination is impaired in the *arabidillo1/2* mutant. Furthermore, ARABIDILLO genes are significantly upregulated by ABA. However, *arabidillo1/2* mutants exhibit normal responses to ABA during lateral root development, suggesting that the ABA-mediated control of seed germination and lateral root elongation involve two discrete signalling pathways (Coates *et al*., 2006; Daniel Gibbs, PhD thesis).

To determine whether *Physcodillo* genes are upregulated by ABA, a number of investigations can be carried out. Transforming both wild type and *ppdillo1a/1b/2* triple deletion mutants with the ABA-inducible Em-GUS gene would establish whether mutants exhibit insensitivity to ABA during later developmental processes. In addition, Real-time RT-PCR will quantify whether *Physcodillo* genes are upregulated by ABA and will be a more sensitive technique.

6.5 ABA and spore germination

ABA plays important and diverse roles throughout the plant kingdom. ABA controls stomatal opening to prevent water loss, confers desiccation and freezing tolerance to plant tissues and regulates seed germination. Most notably, ABA responses are detected in early-evolving bryophytes, including *Marchantia* and *Physcomitrella*, believed to be among the earliest plants to transition from water to land approximately 500 million years ago (Knight *et al*., 1995; Tougane *et al*., 2010).

ABA confers desiccation and freezing tolerance to *Physcomitrella* vegetative tissues (Minami *et al*., 2003; Khandelwal *et al*., 2010). ABA also confers desiccation tolerance to the later-evolving lycophyte and resurrection plant, *Selaginella tamariscina*, which evolved after the bryophytes but before the angiosperms (Liu *et al*., 2008). Desiccation tolerance appears to be restricted to angiosperm seeds and has been lost from vegetative tissues throughout evolution (Khandelwal *et al*., 2010).

Most relevant to the work in this thesis is the role of ABA in maintaining seed dormancy until environmental conditions are appropriate for germination (Gubler *et al*., 2005). A number of proteins play important roles in the ABA-dependent regulation of seed germination. SMALL ORGAN2 is required for correct cell division during *Arabidopsis* seed germination; *smo2* mutants consequently have delayed germination (Hu *et al*., 2010). EID1-like protein 3 (EDL3) positively regulates ABA-dependent seed germination in *Arabidopsis* (Koops *et al*., 2011). Very little is known about ABA responses during germination of early-evolving spores.

Work in this thesis implies that ABA also controls the germination of spores from early-evolving land plants and that endogenous ABA levels within spores may decrease in a time-dependent manner. These findings support the hypothesis that mechanisms underlying

dormancy release in seed/spores are evolutionarily conserved from bryophytes to angiosperms. Genes controlling these processes are likely to have been co-opted from the haploid gametophyte (haploid spores) to the diploid sporophyte (diploid seeds) after the divergence of the bryophytes from other plant species. To build an evolutionary history of the roles of ABA in spore/seed germination, ABA responses of a number of additional plants can be investigated. *Marchantia polymorpha* is a liverwort and therefore belongs to the most basal embryophyte lineage. *Marchantia* is ABA responsive and is therefore able to drive the expression of the Em-GUS gene (Tougane *et al*., 2010). *Marchantia* would therefore be an ideal choice of organism in which to determine whether ABA controls spore germination in an evolutionarily more ancient plant. *Selaginella kraussiana* and *Selaginella moellendorffii* would be ideal organisms in which to determine whether ABA regulates spore germination in a later-evolving and seedless plant. The Coates group is currently investigating whether algal spore germination is regulated by ABA by studying ABA responses of the zoospores of *Ulva linza*, a green marine alga that diverged from the land plant lineage before plants made the transition from water to land.

6.6 The effect of ABA on cell division and cell elongation

Protoplast regeneration experiments demonstrated that filament colony size decreases as the concentration of ABA increases. Further studies using protoplasts isolated from the NLS4 line hinted that ABA appears to inhibit both cell division and cell elongation in moss protonemata. ABA controls cell elongation and cell division in a number of additional plants. In other bryophytes, ABA induces brood cell formation; these are short spherical protonemal cells normally produced in response to desiccation (Goode *et al*., 1993).

ABA appears to play key roles during cell division. In maize, ABA maintains cells in the G1 stage of the cell cycle and consequently delays cell division (Sánchez *et al*., 2005); ABA inhibits early cell division events during coffee seed germination (Da Silva *et al*., 2008); ABA inhibits cell division within the quiescent centre of the *Arabidopsis* root meristem (Han *et al*., 2010).

ABA appears to both promote and inhibit root elongation in plants. ABA inhibits lateral root elongation in *Arabidopsis* but also appears to rescue arrested lateral root elongation in *Medicago latd* mutants (deSmet *et al*., 2003; Liang *et al*., 2007).

6.7 ARABIDILLO homologues have conserved functions

arabidillo1/2 mutants are insensitive to ABA-mediated inhibition of seed germination whereas ARABIDILLO overexpressing mutants are hypersensitive (Gibbs and Coates, unpublished data). Work in this thesis confirmed that Physcodillos also play roles in spore germination. *ppdillo1a/1b/2* triple deletion mutants are insensitive to ABA-mediated inhibition of spore germination. Work in the Coates lab has also confirmed that Physcodillo1A/1B and Physcodillo2 overexpressing mutants are hypersensitive to ABA-mediated inhibition of spore germination (Younousse Saidi, unpublished data).

Similarly to *Arabidopsis*, the inhibitory effects of ABA are less severe on 'after-ripened' spores. However, the trend is still the same (i.e. *ppdillo1a/1b/2* and Physcodillo overexpressing mutants remain insensitive and hypersensitive respectively).

6.8 Cross-species complementation

In order to determine whether ARABIDILLOs are functionally equivalent, transgenic *Arabidopsis* lines expressing either Physcodillo or Selagidillo have been generated. They will

be analysed for lateral root and seed germination phenotypes at a later date. If Physcodillo and Selagidillo proteins are capable of rescuing lateral root phenotypes of *arabidillo1/2* knockout mutants, this would suggest that they are functionally equivalent to ARABIDILLOs and therefore can interact with AIMs *in vivo*. Yeast 2-hybrid assays could be carried out to determine whether Physcodillo and Selagidillo proteins are able to interact with AIMs, by using Physcodillo and Selagidillo ARM repeat domains as 'bait' and AIM proteins as 'prey' or vice versa. Alternatively, bimolecular fluorescence complementation (BiFC) could be used to investigate this interaction.

In order to determine whether ARABIDILLOs and Selagidillo can function in *Physcomitrella*, transgenic lines expressing either ARABIDILLO2 or Selagidillo are in the process of being generated. These lines will be analysed for protoplast regenerating and seed germination phenotypes at a later date. This will determine whether ARABIDILLO proteins have cross-species functions.

6.9 Final words

The work in this thesis has begun to answer questions about the ancestral roles of ARABIDILLO proteins in land plants. *ppdillo1a/1b/2* triple deletion mutants exhibit defects during early developmental processes and like *arabidillo1/2* knockout mutants, exhibit insensitivity to ABA-mediated inhibition of spore germination. This suggests that the mechanism of dormancy release in seeds/spores is evolutionarily conserved from early-evolving spore-producing bryophytes to seed-producing angiosperms.

Understanding more about how this seemingly important gene family has evolved will reveal whether ARABIDILLO ancestors from part of an evolutionarily conserved signalling pathways.

Chapter VII:

References

Abedin, M. and King, N. (2010) Diverse evolutionary paths to cell adhesion. Trends Cell Biol 20(12):734-42.

Acharya, B.R. and Assmann, S.M. (2009). Hormone interactions in stomatal function. Plant Mol Biol 69(4):451-62.

Ali-Rachedi, S. et al. (2004) Changes in endogenous abscisic acid levels during dormancy release and maintenance of mature seeds: studies with the Cape Verde Islands ecotype, the dormant model of Arabidopsis thaliana. Planta. 219(3):479-88.

Alonso, J.M. et al. (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. Science 301(5633):653-7.

Amador, V. et al. (2001) Gibberellins signal nuclear import of PHOR1, a photoperiod-responsive protein with homology to Drosophila armadillo. Cell 106(3):343-54.

Andrade, M.A. et al. (2001) Protein repeats: structures, functions, and evolution. J Struct Biol. 134(2-3):117-31.

Anterola, A. et al. (2009) Gibberellin precursor is involved in spore germination in the moss Physcomitrella patens. Planta. 229(4):1003-7.

Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 14;408(6814):796-815.

Ariizumi, T. et al. (2011) The role of two f-box proteins, SLEEPY1 and SNEEZY, in Arabidopsis gibberellin signaling. Plant Physiol. 155(2):765-75.

Ashton, N.W. et al. (1979). Analysis of gametophytic development in the moss, *Physcomitrella patens*, using auxin and cytokinin resistant mutants. *Planta* 144, 427–435.

Bai, M.Y. et al. (2007) Functions of OsBZR1 and 14-3-3 proteins in brassinosteroid signaling in rice. Proc Natl Acad Sci U S A. 104(34):13839-44.

Banks, J.A. et al. (2011) The Selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science 332(6032):960-3.

Banks, J.A. (2009) Selaginella and 400 million years of separation. Annu Rev Plant Biol. **60**:223-38.

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

Bennici, A. (2008). Origin and early evolution of land plants: Problems and considerations. Commun Integr Biol. 1(2):212-8.

Bezanilla, M. et al. (2003) RNA interference in the moss Physcomitrella patens. Plant Physiol. 133(2):470-4.

Bezanilla, M. et al. (2005) An RNAi system in Physcomitrella patens with an internal marker for silencing allows for rapid identification of loss of function phenotypes. Plant Biol (Stuttg). 7(3):251-7.

Bonner, J.T. (1999) The origins of multicellularity. Integrative Biology: Issues, News, and Reviews Volume 1(1):27–36.

Bowman, J.L. et al. (2007) Green genes-comparative genomics of the green branch of life. Cell 129(2):229-34.

Brodribb, T.J. et al. (2009) Evolution of stomatal responsiveness to CO(2) and optimization of water-use efficiency among land plants. New Phytol 183 (3):839-47.

Carol, R.J. and Dolan, L. Building a hair: tip growth in Arabidopsis thaliana root hairs. Philos Trans R Soc Lond B Biol Sci. 357(1422):815-21.

Chaloner, B. (1998). The greening of the land. Trends Ecol Evol. 13(6):251.

Chater, C. et al. (2011) Regulatory mechanism controlling stomatal behavior conserved across 400 million years of land plant evolution. Curr Biol 21(12):1025-9.

Clough, S.J. and Bent, A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J. 16(6):735-43.

Clouse, S.D. (1996) Molecular genetic studies confirm the role of brassinosteroids in plant growth and development. Plant J. 10(1):1-8.

Coates, J.C. et al. (2002). Loss of the beta-catenin homologue aardvark causes ectopic stalk formation in Dictyostelium. Mech Dev 116(1-2):117-27.

Coates, J.C. et al. (2006). Armadillo-related proteins promote lateral root development in Arabidopsis. Proc Natl Acad Sci U S A 31;103(5):1621-6.

Cooke, T.J. et al. (2002) Evolutionary patterns in auxin action. Plant Mol Biol. 49(3-4):319-38.

Cove, D. et al. (2006) Mosses as model systems for the study of metabolism and development. Annu Rev Plant Biol. 57:497-520.

Cuming, A.C. (1984) Developmental regulation of gene expression in wheat embryos. Molecular cloning of a DNA sequence encoding the early-methionine-labelled (Em) polypeptide. Eur J Biochem. 145(2):351-7.

Cuming, A.C. et al. (2007). Microarray analysis of transcriptional responses to abscisic acid and osmotic, salt, and drought stress in the moss, Physcomitrella patens. New Phytol 176(2):275-87.

Da Silva, E.A. et al. (2008) ABA inhibits embryo cell expansion and early cell division events during coffee (Coffea arabica 'Rubi') seed germination. Ann Bot. 102(3):425-33.

Depuydt, S. **and Hardtke, C.S**. (2011) Hormone signalling crosstalk in plant growth regulation. Curr Biol. 21(9):R365-73.

deSmet, I. et al. (2010) Embryogenesis - the humble beginnings of plant life. The Plant Journal, 61(6), 959-70.

Dharmasiri, N. et al. (2005) The F-box protein TIR1 is an auxin receptor. Nature 435(7041):441-5.

Dickinson, D.J. et al. (2011) A polarized epithelium organized by beta- and alpha-catenin predates cadherin and metazoan origins. Science 331(6022):1336-9.

Dolan, L. (2009). Body building on land: morphological evolution of land plants. Curr Opin Plant Biol. 12(1):4-8.

Eklund, D.M. et al. (2010) Homologues of the Arabidopsis thaliana SHI/STY/LRP1 genes control auxin biosynthesis and affect growth and development in the moss Physcomitrella patens. Development 137(8):1275-84.

Farrant, J.M. and Moore, J.P. (2011). Programming desiccation-tolerance: from plants to seeds to resurrection plants. Curr Opin Plant Biol. 14(3):340-5. **Finkelstein, R.R. and Somerville, C.R.** (1990) Three Classes of Abscisic Acid (ABA)-Insensitive Mutations of Arabidopsis Define Genes that Control Overlapping Subsets of ABA Responses. Plant Physiol. 94(3):1172-9.

Finkelstein, R. et al. (2008) Molecular aspects of seed dormancy. Annu Rev Plant Biol. 59:387-415.

Fonseca, S. et al. (2009). The jasmonate pathway: the ligand, the receptor and the core signalling module. Curr Opin Plant Biol. 12(5):539-47.

Frank, W. et al. (2005) Physcomitrella patens is highly tolerant against drought, salt and osmotic stress. Planta 220(3):384-94.

Fujita, T. et al. (2008) Convergent evolution of shoots in land plants: lack of auxin polar transport in moss shoots. Evol Dev. 10(2):176-86.

Fujii, H. et al. (2009) In vitro reconstitution of an abscisic acid signalling pathway. Nature 462(7273):660-4.

Gao, Y. et al. (2007) Genetic characterization reveals no role for the reported ABA receptor, GCR2, in ABA control of seed germination and early seedling development in Arabidopsis. Plant J. 52(6):1001-13.

Gerrienne, P. and Gonez, P. (2010): Early evolution of life cycles in embryophytes: A focus on the fossil evidence of gametophyte/sporophyte size and morphological complexity. Journal of Systematics and Evolution, 49: 1-16.

Giraudat, J. et al. (1992) Isolation of the Arabidopsis ABI3 gene by positional cloning. Plant Cell 4(10):1251-61.

Goff, S.A. et al. (2002) A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296(5565):92-100.

Gomez-Roldan, V. et al. (2008) Strigolactone inhibition of shoot branching. Nature 455(7210):189-94.

Grimson, M.J. et al. (2000) Adherens junctions and beta-catenin-mediated cell signalling in a non-metazoan organism. Nature 408(6813):727-31.

Gubler, F. et al. (2005) Dormancy release, ABA and pre-harvest sprouting. Curr Opin Plant Biol. 8(2):183-7.

Guilfoyle, T.J. and Hagen, G. (2007) Auxin response factors. Curr Opin Plant Biol. 10(5):453-60.

Han, W. et al. (2010) Fluridone affects quiescent centre division in the Arabidopsis thaliana root stem cell niche. BMB Rep. 43(12):813-7.

Hanada, K. et al. (2011) Origin and evolution of genes related to ABA metabolism and its signaling pathways. J Plant Res. 124(4):455-65.

Harries, P.A. et al. (2005) Actin-related protein2/3 complex component ARPC1 is required for proper cell morphogenesis and polarized cell growth in Physcomitrella patens. Plant Cell.17(8):2327-39.

Harwood, A.J. (2008) Dictyostelium development: a prototypic Wnt pathway? Methods Mol Biol. 469:21-32.

Hauser, F. (2011) Evolution of abscisic acid synthesis and signaling mechanisms. Curr Biol. 21(9):R346-55.

Hayashi, K. et al. (2008) Small-molecule agonists and antagonists of F-box protein-substrate interactions in auxin perception and signaling. Proc Natl Acad Sci U S A. 105(14):5632-7.

Hayashi, K. et al. (2010) Endogenous diterpenes derived from ent-kaurene, a common gibberellin precursor, regulate protonema differentiation of the moss Physcomitrella patens. Plant Physiol. 153(3):1085-97.

He, Z. et al. (2000) Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. Science 288(5475):2360-3.

Hedges, S.B. (2002). The origin and evolution of model organisms. Nat Rev Genet.3(11):838-49.

Hellwege, E.M. et al. (1996) Abscisic acid causes changes in gene expression involved in the induction of the landform of the liverwort Riccia fluitans L. Planta. 198(3):423-32.

Hirano, K. (2007) The GID1-mediated gibberellin perception mechanism is conserved in the Lycophyte Selaginella moellendorffii but not in the Bryophyte Physcomitrella patens. Plant Cell 19(10):3058-79.

Ho, M.S. et al. (2008) The utility F-box for protein destruction. Cell Mol Life Sci. 65(13):1977-2000.

Hu, Z. et al. (2010) Arabidopsis SMO2 regulates seed germination and ABA response. Plant Signal Behav. 5(3):325-7.

Huber, A.H. et al. (1997) Three-dimensional structure of the armadillo repeat region of beta-catenin. Cell. 90(5):871-82.

Jacobsen, J.V. et al. (2002) Abscisic acid, phaseic acid and gibberellin contents associated with dormancy and germination in barley. Physiol Plant. 115(3):428-441.

Jang, G. et al. (2011). RSL genes are sufficient for rhizoid system development in early diverging land plants. Development 138(11):2273-81.

Jang, G. and Dolan, L. (2011). Auxin promotes the transition from chloronema to caulonema in moss protonema by positively regulating *PpRSL1* and *PpRSL2* in *Physcomitrella patens*. New Phytologist 192 (2): 319-327.

Kamisugi, Y. and Cuming, A.C. (2005) The evolution of the abscisic acid-response in land plants: comparative analysis of group 1 LEA gene expression in moss and cereals. Plant Mol Biol. 59(5):723-37.

Kamisugi, Y. et al. (2008) A sequence-anchored genetic linkage map for the moss, Physcomitrella patens. Plant J. 56(5):855-66.

Kenrick, P. and Crane, P.R. (1997) The origin and early evolution of plants on land. Nature **389** (6646): 33–39

Kepinski, S. and Leyser, O. (2005) The Arabidopsis F-box protein TIR1 is an auxin receptor. Nature 435(7041):446-51.

Khandelwal, A. et al. (2010) Role of ABA and ABI3 in desiccation tolerance. Science 327(5965):546.

Khraiwesh, B. (2008) Specific gene silencing by artificial MicroRNAs in Physcomitrella patens: an alternative to targeted gene knockouts. Plant Physiol. 148(2):684-93.

Klingensmith, J. et al. (1989) The segment polarity phenotype of Drosophila involves differential tendencies toward transformation and cell death. Dev Biol. 134(1):130-45.

Knight, C.D. et al. (1995) Molecular Responses to Abscisic Acid and Stress Are Conserved between Moss and Cereals. Plant Cell 7(5):499-506.

Koduri, P.K. et al. (2010) Genome-wide analysis of the chalcone synthase superfamily genes of Physcomitrella patens. Plant Mol Biol.72(3):247-63.

Koltai, H. (2011) Strigolactones are regulators of root development. New Phytol. 190(3):545-9.

Komatsu, K. et al. (2009) Functional analyses of the ABI1-related protein phosphatase type 2C reveal evolutionarily conserved regulation of abscisic acid signaling between Arabidopsis and the moss Physcomitrella patens. Plant Physiol 152(3):1529-43.

Koops, P. (2011) EDL3 is an F-box protein involved in the regulation of abscisic acid signalling in Arabidopsis thaliana. J Exp Bot.

Krysan, P.J. (1999) T-DNA as an insertional mutagen in Arabidopsis. Plant Cell. 11(12):2283-90.

Kuromori, T. et al. (2010) ABC transporter AtABCG25 is involved in abscisic acid transport and responses. Proc Natl Acad Sci U S A. 107(5):2361-6.

Lang, D. et al. (2008) Exploring plant biodiversity: the Physcomitrella genome and beyond. Trends Plant Sci. 13(10):542-9.

Langdale, J.A. (2008) Evolution of developmental mechanisms in plants. Curr Opin Genet Dev. 18(4):368-73.

Lechner, E. et al. (2006) F-box proteins everywhere. Curr Opin Plant Biol. 9(6):631-8.

Lee, S.J. et al. (2010) AtNEK6 interacts with ARIA and is involved in ABA response during seed germination. Mol Cells. 29(6):559-66.

Leyser, O. (2005) The fall and rise of apical dominance. Curr Opin Genet Dev. 15(4):468-71.

Li, J. and Chory, J. (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. Cell 90(5):929-38.

Liang, Y. et al. (2007) Abscisic acid rescues the root meristem defects of the Medicago truncatula latd mutant. Dev Biol. 304(1):297-307.

Liu, X. et al. (2007) A G protein-coupled receptor is a plasma membrane receptor for the plant hormone abscisic acid. Science 315(5819):1712-6.

Liu, M.S. et al. (2008) Constitutive components and induced gene expression are involved in the desiccation tolerance of Selaginella tamariscina. Plant Cell Physiol. 49(4):653-63.

MacMillan, J. (2001) Occurrence of Gibberellins in Vascular Plants, Fungi, and Bacteria. J Plant Growth Regul. 20(4):387-442.

Ma, Y. et al. (2009) Regulators of PP2C phosphatase activity function as abscisic acid sensors. Science 324(5930):1064-8.

Marella, H.H. et al. (2006) Characterization and functional analysis of ABSCISIC ACID INSENSITIVE3-like genes from Physcomitrella patens. Plant J. 46(6):1032-44.

Masucci, J.D. and Schiefelbein, J.W. (1994) The rhd6 Mutation of Arabidopsis thaliana Alters Root-Hair Initiation through an Auxin- and Ethylene-Associated Process. Plant Physiol. 106(4):1335-1346.

Meinke, D.W. et al. (1998) Arabidopsis thaliana: a model plant for genome analysis. Science 282(5389):662, 679-82.

Melcher, K. (2009) A gate-latch-lock mechanism for hormone signalling by abscisic acid receptors. Nature 462(7273):602-8.

Menand, B. et al. (2007). An ancient mechanism controls the development of cells with a rooting function in land plants. Science 8;316(5830):1477-80.

Meyerowitz, E.M. (1987) Arabidopsis thaliana. Annu Rev Genet. 21:93-111.

Meyerowitz, E.M. (1999) Plants, animals and the logic of development. Trends Cell Biol 9(12):M65-8.

Minami, A. et al. (2005). Cold acclimation in bryophytes: low-temperature-induced freezing tolerance in Physcomitrella patens is associated with increases in expression levels of stress-related genes but not with increase in level of endogenous abscisic acid. Planta 220(3):414-23.

Miyazono, K. et al. (2009) Structural basis of abscisic acid signalling. Nature 462(7273):609-14.

Mombaerts, P. (1991) Creation of a large genomic deletion at the T-cell antigen receptor beta-subunit locus in mouse embryonic stem cells by gene targeting. Proc Natl Acad Sci U S A. 88(8):3084-7.

Muschietti, J. and McCormick, S. (2010) Abscisic acid (ABA) receptors: light at the end of the tunnel. F1000 Biol Rep. 24;2. pii: 15.

Müller, A.H. and Hansson, M. (2009) The barley magnesium chelatase 150-kd subunit is not an abscisic acid receptor. Plant Physiol. 150(1):157-66.

Nagao, M. et al. (2006) Accumulation of theanderose in association with development of freezing tolerance in the moss Physcomitrella patens. Phytochemistry 67(7):702-9.

Nelson, D.C. (2011) F-box protein MAX2 has dual roles in karrikin and strigolactone signaling in Arabidopsis thaliana. Proc Natl Acad Sci U S A. 108(21):8897-902.

Nibau, C. et al. (2011) ARABIDILLO proteins have a novel and conserved domain structure important for the regulation of their stability. Plant Mol Biol. 75(1-2):77-92.

Nibau, C. et al. (2008) Branching out in new directions: the control of root architecture by lateral root formation. New Phytol. 179(3):595-614.

Niklas, K.J. and Kutschera, U. (2010) The evolution of the land plant life cycle. New Phytol. 185(1):27-41.

Nishimura, N. et al. (2009) Structural mechanism of abscisic acid binding and signaling by dimeric PYR1. Science 326(5958):1373-9.

Oda, H. et al. (1993) Identification of a Drosophila homologue of alpha-catenin and its association with the armadillo protein. J Cell Biol. 121(5):1133-40.

Oda, H. et al. (1997) Phenotypic analysis of null mutants for DE-cadherin and Armadillo in Drosophila ovaries reveals distinct aspects of their functions in cell adhesion and cytoskeletal organization. Genes Cells. 2(1):29-40.

Oldenhof, H. et al. (2006) Freezing and desiccation tolerance in the moss Physcomitrella patens: an in situ Fourier transform infrared spectroscopic study. Biochim Biophys Acta.1760(8):1226-34.

Pandey, S. et al. (2009) Two novel GPCR-type G proteins are abscisic acid receptors in Arabidopsis. Cell 136(1):136-48.

Parcy, F. et al. (1994) Regulation of gene expression programs during Arabidopsis seed development: roles of the ABI3 locus and of endogenous abscisic acid. Plant Cell 6(11):1567-82.

Park, S.Y. et al. (2009) Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. Science 324(5930):1068-71.

Peifer, M. et al. (1991) The segment polarity gene armadillo interacts with the wingless signaling pathway in both embryonic and adult pattern formation. Development 111(4):1029-43.

Peifer, M. et al. (1993) A role for the Drosophila segment polarity gene armadillo in cell adhesion and cytoskeletal integrity during oogenesis. Development 118(4):1191-207.

Peifer, M. et al. (1994) A repeating amino acid motif shared by proteins with diverse cellular roles. Cell 76(5):789-91.

Pence, V. C. (1998). Cryopreservation of bryophytes: the effects of abscisic acid and encapsulation dehydration. Bryologist 101: 278-281.

Perroud, P.F. and Quatrano, R.S. (2008) BRICK1 is required for apical cell growth in filaments of the moss Physcomitrella patens but not for gametophore morphology. Plant Cell. 20(2):411-22.

Petroski, M.D. and Deshaies, R.J. (2005) Function and regulation of cullin-RING ubiquitin ligases. Nat Rev Mol Cell Biol. 6(1):9-20.

Pilet, P.E. and Saugy, M. (1987) Effect on Root Growth of Endogenous and Applied IAA and ABA: A Critical Reexamination. Plant Physiol. 83(1):33-8.

Pils, B. and Heyl, A. (2009) Unraveling the evolution of cytokinin signaling. Plant Physiol. 151(2):782-91.

Prigge, M.J. and Bezanilla, M. (2010) Evolutionary crossroads in developmental biology: Physcomitrella patens. Development. 137(21):3535-43.

Proust, H. (2011) Strigolactones regulate protonema branching and act as a quorum sensing-like signal in the moss Physcomitrella patens. Development 138(8):1531-9.

Raghavendra, A.S. et al. (2010) ABA perception and signalling. Trends Plant Sci. 15(7):395-401.

Razem, F.A. et al. (2006) The RNA-binding protein FCA is an abscisic acid receptor. Nature 439(7074):290-4.

Rensing, S.A. et al. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 319(5859):64-9.

Reski, R. and Frank, W. (2005) Moss (Physcomitrella patens) functional genomics--Gene discovery and tool development, with implications for crop plants and human health. Brief Funct Genomic Proteomic 4(1):48-57.

Richardt, S. et al. (2010) Microarray analysis of the moss Physcomitrella patens reveals evolutionarily conserved transcriptional regulation of salt stress and abscisic acid signalling. Plant Mol Biol. 72(1-2):27-45.

Riggleman, B. et al. (1989) Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene. Genes Dev. 3(1):96-113.

Risk, J.M. (2008) FCA does not bind abscisic acid. Nature 11;456(7223):E5-6.

Rodríguez-Gacio, M.C. et al. (2009) Seed dormancy and ABA signaling: the breakthrough goes on. Plant Signal Behav. 4(11):1035 - 49.

Ruszala, E.M. et al. (2011) Land plants acquired active stomatal control early in their evolutionary history. Curr Biol 21(12):1030-5.

Sakakibara, K. et al. (2003) Involvement of auxin and a homeodomain-leucine zipper I gene in rhizoid development of the moss Physcomitrella patens. Development 130(20):4835-46.

Saidi, Y. et al. (2005) Controlled expression of recombinant proteins in Physcomitrella patens by a conditional heat-shock promoter: a tool for plant research and biotechnology. Plant Mol Biol. 59(5):697-711.

Sakata, Y. et al. (2010) Regulation of the ABA-responsive Em promoter by ABI3 in the moss Physcomitrella patens: role of the ABA response element and the RY element. Plant Signal Behav. 5(9):1061-6.

Santiago, J. et al. (2009) Modulation of drought resistance by the abscisic acid receptor PYL5 through inhibition of clade A PP2Cs. Plant J. 60(4):575-88.

Santiago, J. et al. (2009) The abscisic acid receptor PYR1 in complex with abscisic acid. Nature 462(7273):665-8.

Savage, N.S. and Schmidt, W. (2008) From priming to plasticity: the changing fate of rhizodermic cells. Bioessays 30(1):75-81.

Schaap, P. (2011) Evolutionary crossroads in developmental biology: Dictyostelium discoideum. Development 138(3):387-96.

Schaefer, D.G. and Zryd, J.P (2001). The moss Physcomitrella patens, now and then. Plant Physiol 127 (4), 1430-8.

Sethi, J.K. and Vidal-Puig, A. (2010) Wnt signalling and the control of cellular metabolism. Biochem J. 427(1):1-17.

Shapiro, L. (1997) The multi-talented beta-catenin makes its first appearance. Structure 5(10):1265-8.

Shen, Y.Y. et al. (2006) The Mg-chelatase H subunit is an abscisic acid receptor. Nature 443(7113):823-6.

Schumaker, K.S. and Dietrich, M.A. (1998) HORMONE-INDUCED SIGNALING DURING MOSS DEVELOPMENT. Annu Rev Plant Physiol Plant Mol Biol. 49:501-523.

Sieberer, B.J. (2005) Microtubules guide root hair tip growth. New Phytol. 167(3):711-9.

Southern, E. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98: 503-508.

Southworth, D. and Cresti, M. (1997) Comparison of flagellated and nonflagellated sperm in plants. Am J Bot 84(9):1301.

Stirk, W.A. et al. (2003) Cytokinins in macroalgae. Plant Growth Regul 41: 13–24

Stirnberg, P. et al. (2007) MAX2 participates in an SCF complex which acts locally at the node to suppress shoot branching. Plant J. 50(1):80-94.

Spíchal, L. et al. (2004) Two cytokinin receptors of Arabidopsis thaliana, CRE1/AHK4 and AHK3, differ in their ligand specificity in a bacterial assay. Plant Cell Physiol. 45(9):1299-305.

Strader, L.C. et al. (2004) Recessive-interfering mutations in the gibberellin signaling gene SLEEPY1 are rescued by overexpression of its homologue, SNEEZY. Proc Natl Acad Sci U S A. 101(34):12771-6.

Tewari, R. et al. (2010) Armadillo-repeat protein functions: questions for little creatures. Trends Cell Biol. 20(8):470-81.

Thomas, R.J. et al. (1983) Endogenous auxin and ethylene in pellia (bryophyta). Plant Physiol. 73(2):395-7.

Tougane, K. et al. (2010) Evolutionarily conserved regulatory mechanisms of abscisic acid signaling in land plants: characterization of ABSCISIC ACID INSENSITIVE1-like type 2C protein phosphatase in the liverwort Marchantia polymorpha. Plant Physiol. 152(3):1529-43.

Ueguchi-Tanaka, M. et al. (2005) GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin. Nature 29;437(7059):693-8.

Umehara, M. et al. (2008) Inhibition of shoot branching by new terpenoid plant hormones. Nature 455(7210):195-200.

Umezawa, T. et al. (2009) Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in Arabidopsis. Proc Natl Acad Sci U S A. 106(41):17588-93.

Vandenbussche, F. et al. (2007) Evolutionary conservation of plant gibberellin signalling pathway components. BMC Plant Biol. 29;7:65.

van Noort, M. et al. (2002) Wnt signaling controls the phosphorylation status of beta-catenin. J Biol Chem. 277(20):17901-5.

Vidali, L. et al. (2009) Rapid formin-mediated actin-filament elongation is essential for polarized plant cell growth. Proc Natl Acad Sci U S A. 106(32):13341-6.

Vierstra, R.D. (2003) The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. Trends Plant Sci. 8(3):135-42.

Wang, X. et al. (2009). Proteomic analysis of the cold stress response in the moss, Physcomitrella patens. Proteomics. 9(19):4529-38.

Wang, Z. et al. (2009) LFR, which encodes a novel nuclear-localized Armadillo-repeat protein, affects multiple developmental processes in the aerial organs in Arabidopsis. Plant Mol Biol. 69(1-2):121-31.

Wang, X. et al. (2010) Conservation between higher plants and the moss Physcomitrella patens in response to the phytohormone abscisic acid: a proteomics analysis. BMC Plant Biol. 27;10:192.

Xu, G. et al. (2009) Evolution of F-box genes in plants: different modes of sequence divergence and their relationships with functional diversification. Proc Natl Acad Sci U S A. 106(3):835-40.

Yang, Y. et al. (2010) Silencing Sl-EBF1 and Sl-EBF2 expression causes constitutive ethylene response phenotype, accelerated plant senescence, and fruit ripening in tomato. J Exp Bot. 61(3):697-708.

Yasumura, Y. et al. (2007) Step-by-step acquisition of the gibberellin-DELLA growth-regulatory mechanism during land-plant evolution. Curr Biol. 17(14):1225-30.

Ye, H. et al. (2011) Recent advances in the regulation of brassinosteroid signaling and biosynthesis pathways(f). J Integr Plant Biol. 53(6):455-68.

Zhu, Z. and Guo, H. (2008) Genetic Basis of Ethylene Perception and Signal Transduction in *Arabidopsis.* Journal of Integrative Plant Biology 50, 808–815.

Chapter VIII:

Appendix

8.1 Alignment of *PHYSCODILLO1A* with pseudogene

```
pseudogene            ATATCCAACAAGCGGCGGCGCGAAGTCAATGTGGCGGTGGACGATCAGGAGCAGCAGCAG 60
PHYSCODILLO1A/1B      ATGTCCAACAAGCGGCGGCGCAATGTCAATGTGGCGGTGGACGATCAGGAGCAGCAGCAG 60
                      ** ****************** * ************************************

pseudogene            ACTGTTGTCTATAAGAAAGCCAGGATAACATCGTCGCCGAC------------GTCCGCT 108
PHYSCODILLO1A/1B      GCTGTTGTATATAAGAAAGCCAGGATAACATCGTCGCCGACATCGTCGGCGGCGTCCACT 120
                       ******* ********************************            **** **

pseudogene            TGTGGTGCATCGGCGGTGCCCCCTATAGAATCAGGTGTTAAAATTGGCGCGGGTCGCGAT 168
PHYSCODILLO1A/1B      TGTGGTGCACCGGCGGTGCCGGCTACAGAATCAGGTGTTGAAAGTGACGCGGGTCGCGAT 180
                      ********* **********  *** ************* *** ** **************

pseudogene            GCCCATTGGACAAGTCTGCCAGATGAAACCGTGCTTGGGTTGTTTAATCTGCTGAACCAC 228
PHYSCODILLO1A/1B      GCGCATTGGACAAGTCTTCCAGATGAAACCGTGCTTGGGTTGTTTAATCTGCTGAACCAC 240
                      ** ************** ******************************************

pseudogene            AGGGACATGGCGAGCCTTGCGTCCGTGTGAAAGGGATGGCAGGTGCTGGGGAGCTCGCAG 288
PHYSCODILLO1A/1B      AGGGACAGGGCGAGCCTTGCGTCCGTGTGCAAGGGATGGCAGGTGCTGGGGAGCTCGCCA 300
                      ******* ********************** ****************************

pseudogene            TCGCTGTGGAATTCGTCGGATCTTCGGTTGCACACGTTGAATTCGGAGATGGTGTCTGCG 348
PHYSCODILLO1A/1B      TCGCTGTGGAATTCGCTGGATCTTCGGTCGCACTCGTTGAATTCGGAGATGGTGTCTGCG 360
                      ***************  ************ **** *************************

pseudogene            CTGGCAGGACGATGCTCGAATTTGGAGGCGCTGAAGTTCCGCAGAGGTGCCTTTGCGAGC 408
PHYSCODILLO1A/1B      CTGGCAGGACGATGCTCGAATTTGGAGGCGCTGAAGTTCCGCAGAGGTGCCTCTGCGAGC 420
                      ************************************************** *******

pseudogene            TCAATTGTGGGGTTGCAGGCGAAGGGGCTTAGGGAAATGAGCGGCGATTGCTGCAGCCAG 468
PHYSCODILLO1A/1B      TCAATTGTGGGGTTGCAGGCGAAGGGGCTTAGGGAGCTGAGCGGCGATTGCTGCAGCCAG 480
                      ********************************** ************************

pseudogene            CTGTCGGATGCAACCCTGTCCATGGTGGTCGCCCGGCATGCGAACCTGGAGAGCA----- 523
PHYSCODILLO1A/1B      CTGTCGGATGCAACCCTGTCCATGGTGGTCGCCCGGCATGCGAACCTGGAGAGCTTGCTG 540
                      ************************************************

pseudogene            -------------------------------AAGGCCTT--------TTG----------- 534
PHYSCODILLO1A/1B      CTGGGTTCGGATTGCGAGAGAGTGACGAGCGAGGCATTGAAGGTGATTGCAGTGTGTTGT 600
                                                    **** **        ***
```

8.2 *PHYSCODILLO1A*, *PHYSCODILLO1B* and *PHYSCODILLO2* genomic DNA alignment

```
PpDILLO1A.gDNA      ATGTCCAACAAGCGGCGGCGCAATGTCAATGTGGCGGTGGACGATCAGGAGCAGCAGCAG 60
PpDILLO1B.gDNA      ATGTCCAACAAGCGGCGGCGCAATGTCAATGTGGCGGTGGACGATCAGGAGCAGCAGCAG 60
PpDILLO2.gDNA       ATGTCCAACAAGCGGCGGCGGAGCGTCAACGTGGCGGTGGAGGAGCAGGAGCAGCAGCCG 60
                    ******************** *  ***** *********** ** ************ *

PpDILLO1A.gDNA      GCTGTTGTATATAAGAAAGCCAGGATAACATCGTCGCCGACATCGTCGGCGGCGTCCACT 120
PpDILLO1B.gDNA      GCTGTTGTATATAAGAAAGCCAGGATAACATCGTCGCCGACATCGTCGGCGGCGTCCACT 120
PpDILLO2.gDNA       ---GTTTTGTTCAAGAAGGCCAGGATAACACCGTCGTCG---TCGTCG------TCCGCT 108
                       *** * *  ***** ************ ***** **   ******      *** **

PpDILLO1A.gDNA      TGTGGTGCACCGGCGGTGCCGGCTACAGAATCAGGTGTTGAAAGTGACGCGGGTCGCGAT 180
PpDILLO1B.gDNA      TGTGGTGCACCGGCGGTGCCGGCTACAGAATCAGGTGTTGAAAGTGACGCGGGTCGCGAT 180
PpDILLO2.gDNA       TGTAGTGCCCCGGCAGTGTCTGTTGCGGAGTCGGGCGTGGGAATTGCTGGGGATCGCGAT 168
                    *** **** ***** *** * * * * ** ** ** ** * ** **  * ** *******

PpDILLO1A.gDNA      GCGCATTGGACAAGTCTTCCAGATGAAACCGTGCTTGGGTTGTTTAATCTGCTGAACCAC 240
PpDILLO1B.gDNA      GCGCATTGGACAAGTCTTCCAGATGAAACCGTGCTTGGGTTGTTTAATCTGCTGAACCAC 240
PpDILLO2.gDNA       GCGCATTGGACGGAGCTGCCGGATGATACTGTGTTCGGGTTGTTCAATTTGCTGAACTAT 228
                    ***********    ** ** ***** ** *** * ******** *** ******** *

PpDILLO1A.gDNA      AGGGACAGGGCGAGCCTTGCGTCCGTGTGCAAGGGATGGCAGGTGCTGGGGAGCTCGCCA 300
PpDILLO1B.gDNA      AGGGACAGGGCGAGCCTTGCGTCCGTGTGCAAGGGATGGCAGGTGCTGGGGAGCTCGCCA 300
PpDILLO2.gDNA       AGGGATAGGGCGAGTCTCGCGTCCGTGTGCAGGGCATGGCGCGGGTTGGGGAGCTCGACG 288
                    ***** ******** ** ************* ** *****  * * *********** *

PpDILLO1A.gDNA      TCGCTGTGGAATTCGCTGGATCTTCGGTCGCACTCGTTGAATTCGGAGATGGTGTCTGCG 360
PpDILLO1B.gDNA      TCGCTGTGGAATTCGCTGGATCTTCGGTCGCACTCGTTGAATTCGGAGATGGTGTCTGCG 360
PpDILLO2.gDNA       TCGTTGTGGACGTCGTTGGATCTTCGCTTGCACAGGTTGGATTCGGAGGTCGTGTCGGCG 348
                    *** ******  *** ********** * ****  **** ******** * ***** ***

PpDILLO1A.gDNA      CTGGCAGGACGATGCTCGAATTTGGAGGCGCTGAAGTTCCGCAGAGGTGCCTCTGCGAGC 420
PpDILLO1B.gDNA      CTGGCAGGACGATGCTCGAATTTGGAGGCGCTGAAGTTCCGCAGAGGTGCCTCTGCGAGC 420
PpDILLO2.gDNA       CTTGCGGGCCGGTGCTCGAATTTGCAGTGTCTGAAATTTCGCGGAGGCGCGTTCGCGAAC 408
                    ** ** ** ** ************ **   ***** ** *** **** ** *  **** *

PpDILLO1A.gDNA      TCAATTGTGGGGTTGCAGGCGAAGGGGCTTAGGGAGCTGAGCGGCGATTGCTGCAGCCAG 480
PpDILLO1B.gDNA      TCAATTGTGGGGTTGCAGGCGAAGGGGCTTAGGGAGCTGAGCGGCGATTGCTGCAGCCAG 480
PpDILLO2.gDNA       TCAATTGTGGGGTTTGCATGCGAGGGAGCTGAGGGAGCTGAGCGGTGATTGGTGCAGCCTT 468
                    *********** ***** **** ** *** *************** ***** *******

PpDILLO1A.gDNA      CTGTCGGATGCAACCCTGTCCATGGTGGTCGCCCGGCATGCGAACCTGGAGAGCTTGCTG 540
PpDILLO1B.gDNA      CTGTCGGATGCAACCCTGTCCATGGTGGTCGCCCGGCATGCGAACCTGGAGAGCTTGCTG 540
PpDILLO2.gDNA       CTGTCGGATGCAACCCTGTCCATGGTGGTTGCTCGGCATGGGAACCTGGAGAGCTTGCAG 528
                    ***************************** ** ******* ***************** *

PpDILLO1A.gDNA      CTGGGTTCGGATTGCGAGAGAGTGACGAGCGAGGCATTGAAGGTGATTGCAGTGTGTTGT 600
PpDILLO1B.gDNA      CTGGGTTCGGATTGCGAGAGAGTGACGAGCGAGGCATTGAAGGTGATTGCAGTGTGTTGT 600
PpDILLO2.gDNA       CTGGGTTCGGATTGCGAGAGGGTGACGAGCGAGGCGCTGAAAGTGGTAGCAGTTTGTTGT 588
                    ******************** **************  **** *** * ***** ******

PpDILLO1A.gDNA      CCGAAATTGCGACGGCTGTGCGTGTCGGGCGTCCTAAAAGTGGAGAGAGACGCCATCCAG 660
PpDILLO1B.gDNA      CCGAAATTGCGACGGCTGTGCGTGTCGGGCGTCCTAAAAGTGGAGAGAGACGCCATCCAG 660
PpDILLO2.gDNA       CCTAAATTGCGACGGTTGTGCATCTCAGGCATTCGGGATGTGGACCGAGACGCCATTCAG 648
                    ** ************ ***** * ** *** * *   * *****  ********** ***

PpDILLO1A.gDNA      GCGTTGTTCCAGCATTGCAAGGGTCTGACGGAGCTGGGGTTCCTGGACAGTCACACCATC 720
PpDILLO1B.gDNA      GCGTTGTTCCAGCATTGCAAGGGTCTGACGGAGCTGGGGTTCCTGGACAGTCACACCATC 720
PpDILLO2.gDNA       GCAATGTTCCAGCATTGCCAGGGTCTGACAGAGCTGGGGTTTCTGGACAGTCACACCATC 708
                    **  ************** ********** *********** ******************

PpDILLO1A.gDNA      GACGAAGGGGCGTTCGGGGGTGCGAGCAGCCTGCGATTTTTGTCGGTAGCGGGATGCAGG 780
PpDILLO1B.gDNA      GACGAAGGGGCGTTCGGGGGTGCGAGCAGCCTGCGATTTTTGTCGGTAGCGGGATGCAGG 780
PpDILLO2.gDNA       GACGAGGGAGCTTTCGGGGCTGCGAGGAACCTACGATTTTTGTCCGTGGCGGGATGCAGG 768
                    ***** ** ** ******* ****** * *** *********** ** ************

PpDILLO1A.gDNA      TGTATAGTGTGGAGCACGGCGGCGCATTGGTGGAGCAAATTGCCGAACCTGGCAGGGCTG 840
PpDILLO1B.gDNA      TGTATAGTGTGGAGCACGGCGGCGCATTGGTGGAGCAAATTGCCGAACCTGGCAGGGCTG 840
PpDILLO2.gDNA       TGCATAGCGTGGAGCACGGCAGCGCAGTGCTGGAGCAAATTGCCCAACTTGAGTGGGCTG 828
                    ** **** ************ ***** ** ************** *** **   ******
```

```
PpDILLO1A.gDNA      GATGTTTCGCGGACAGATATCACTCCGACTGCGCTGATGCAGGTCCTGGCAGGGCCCGAG 900
PpDILLO1B.gDNA      GATGTTTCGCGGACAGATATCACTCCGACTGCGCTGATGCAGGTCCTGGCAGGGCCCGAG 900
PpDILLO2.gDNA       GATGTTTCTCGGACGGATATAACTCCGAGCACACTGGCGCAGGTGCTTGCATGTCCCGAA 888
                    ******** ***** ***** *******   * ***  ****** ** *** * *****

PpDILLO1A.gDNA      CTGAGGGTGGTTTGCGCTCTGAACTGCCCTGTTCTGGAGGAGGGCAGCAACCCCGTGACA 960
PpDILLO1B.gDNA      CTGAGGGTGGTTTGCGCTCTGAACTGCCCTGTTCTGGAGGAGGGCAGCAACCCCGTGACA 960
PpDILLO2.gDNA       CTGAAGGTGGTGTGCGGCCTGAACTGCCCCCTTCTGGAGGAGGGTAGCAACCCGATCCCG 948
                    **** ****** ****  ***********  ************* ********  *  *

PpDILLO1A.gDNA      CTGCCGTCGTCGAAGAAGACGGTGGTGCTGGCTCGGTTCACGGATGTGATGGAAGGGTTG 1020
PpDILLO1B.gDNA      CTGCCGTCGTCGAAGAAGACGGTGGTGCTGGCTCGGTTCACGGATGTGATGGAAGGGTTG 1020
PpDILLO2.gDNA       TCGCCATCGTCAAAGACGGCGGTGTTGCTGGCACGGTTCACAGATCTTATGGAAGGATTG 1008
                      *** ***** **** * ***** ******* ******** *** * ******** ***

PpDILLO1A.gDNA      GATGCTTTGCTGAGCCCGAGCAATTGGAAGGAGGAGGCAGGCTCGTGCGCGAGGTCGAGG 1080
PpDILLO1B.gDNA      GATGCTTTGCTGAGCCCGAGCAATTGGAAGGAGGAGGCAGGCTCGTGCGCGAGGTCGAGG 1080
PpDILLO2.gDNA       GAGGTTTTGCTGAACCCGAGCAATTCGAAGGAGGAGGCGTGTTCGTCTGGACGATCCAGG 1068
                    ** * ******** *********** ************  * ****  *   * ** ***

PpDILLO1A.gDNA      TGCGGCGGGCGCGCGCGCGCGGATTCGGAAGTGGCGAAGTGGACGGAGTGGATGTTGTCG 1140
PpDILLO1B.gDNA      TGCGGCGGGCGCGCGCGCGCGGATTCGGAAGTGGCGAAGTGGACGGAGTGGATGTTGTCG 1140
PpDILLO2.gDNA       TATGGCAGGCGCATGCGTGTGGACCCGGAAGTGGCGAAGTGGACGGAGAGAATGTTGTCA 1128
                    *  *** *****  *** * ***  ********************** * ********

PpDILLO1A.gDNA      CATGCGGTTTTGAAGATCGCGGAGTGCAATGCACCGAGTCTG------------CTGAAG 1188
PpDILLO1B.gDNA      CATGCGGTTTTGAAGATCGCGGAGTGCAATGCACCGAGTCTG------------CTGAAG 1188
PpDILLO2.gDNA       CATGCACTTTTGAAGATTGCGGAGAGCAATGCGCCCAGCCTGGATAGCTTCTGGCTGAAG 1188
                    *****  ********** ****** ******* ** ** ***            ******

PpDILLO1A.gDNA      CAGGGGATCGCGATGATGCTTCGACTGGTTCAAAGCGCGCAGGAGGATGTACAGGAGAGG 1248
PpDILLO1B.gDNA      CAGGGGATCGCGATGATGCTTCGACTGGTTCAAAGCGCGCAGGAGGATGTACAGGAGAGG 1248
PpDILLO2.gDNA       CAGGGAACCGCCATGATGCTTCGACTGGTCCAAAGCGCACAGGAGGATGTACAGGAAAGG 1248
                    ***** * *** ***************** ******** ***************** ***

PpDILLO1A.gDNA      GCTGCGTCGGCTTTGGCGACGTTTGTAGTGGTGGACGACGAGAACGCGACTGTAGATTCT 1308
PpDILLO1B.gDNA      GCTGCGTCGGCTTTGGCGACGTTTGTAGTGGTGGACGACGAGAACGCGACTGTAGATTCT 1308
PpDILLO2.gDNA       GCTGCGGCTGCTTTGGCGGTGTTTGTTTTGGTGGATGATGAGAATGCGACTGTGGATTCT 1308
                    ****** * *********  ******  ******* ** ***** ******** ******

PpDILLO1A.gDNA      GCGCGGGCGGAAGCTGTGATGAACGGGGGCGGGATCGCGCTGCTGTTGGGGCTGGCGAAG 1368
PpDILLO1B.gDNA      GCGCGGGCGGAAGCTGTGATGAACGGGGGCGGGATCGCGCTGCTGTTGGGGCTGGCGAAG 1368
PpDILLO2.gDNA       GCACGGGCGGAAGCTGTGATGAACGGTGGCGGCATTGCGTTGCTGCTAGGTCTGGCAAAG 1368
                    ** ********************** ***** ** *** ***** * ** ***** ***

PpDILLO1A.gDNA      TCTTGCCGCGAGGGGGTGCAGTCGGAAGCTGCGAAGGTAGGGG-GGTTGGTGGGGAATGT 1427
PpDILLO1B.gDNA      TCTTGCCGCGAGGGGGTGCAGTCGGAAGCTGCGAAGGTAGGGG-GGTTGGTGGGGAATGT 1427
PpDILLO2.gDNA       TCTTGTGGCGAGGGAGTGCAATCCGAAGCCGCCAAGGTAGTTCTAGTTTCTGAGGAACGA 1428
                    *****  ******* ***** ** ***** ** *******     ***  ** **** *

PpDILLO1A.gDNA      AGTGGAGTGTGAGTGTGCACAAGGCAGGCGTGTGCGCAGGAAGGATTGTA---------- 1477
PpDILLO1B.gDNA      AGTGGAGTGTGAGTGTGCACAAGGCAGGCGTGTGCGCAGGAAGGATTGTA---------- 1477
PpDILLO2.gDNA       GTCGGAGTGCGAGTGTGCACGGGACAG-TTTTTGTGGAGAGAGGGTAGTGTTTACTAGGT 1487
                       ****** **********  * ***   * ** * **  *** * **

PpDILLO1A.gDNA      -----------GGGTAG-CGGGGGTATTGCGTGAGTTGTGGAGCGTGGGGTGGAGCAGGG 1525
PpDILLO1B.gDNA      -----------GGGTAG-CGGGGGTATTGCGTGAGTTGTGGAGCGTGGGGTGGAGCAGGG 1525
PpDILLO2.gDNA       GGTTGGCAGGCGGGCAGGCGGAAGCATTGCATTTCTTTCCG-GCGTTAGCTGAAGTAGGG 1546
                              *** ** ***  * ***** *   **   * ****  * ** ** ****

PpDILLO1A.gDNA      TCGGGGGAGAGCAGGTGGACGACGATGGTGGTATTTGCGGAATGGGTTGATGTTTGTGGG 1585
PpDILLO1B.gDNA      TCGGGGGAGAGCAGGTGGACGACGATGGTGGTATTTGCGGAATGGGTTGATGTTTGTGGG 1585
PpDILLO2.gDNA       TTGAATGGCAACTGGTTCACGATGATGGCGGT-TTTTCGAAGTTGACCCATGTTCAATTA 1605
                    * *   *  * * ***  **** ***** *** *** ** * * *    *****

PpDILLO1A.gDNA      AGTGTGGGTCGCATGCGGGCACAGTGCGTGTGGGGTGCAGAAGCTGTGGGGAGTTG-GAC 1644
PpDILLO1B.gDNA      AGTGTGGGTCGCATGCGGGCACAGTGCGTGTGGGGTGCAGAAGCTGTGGGGAGTTG-GAC 1644
PpDILLO2.gDNA       AACCTCCCTCGAGTGTAGGCGTGGTTTGCGTGGGTTACTCAG--TATAGTGAATTACGAC 1663
                    *   *   ***  **  ***   **  * ***** * *  *   * * * ** **  ***

PpDILLO1A.gDNA      TGACTGGTAGCTGTTGGGT-TGGTTGGTGCAGGCCATCGCAAATTTGTCCGTGAACACGG 1703
PpDILLO1B.gDNA      TGACTGGTAGCTGTTGGGT-TGGTTGGTGCAGGCCATCGCAAATTTGTCCGTGAACACGG 1703
PpDILLO2.gDNA       TAA-TGGCACACATTCCGTGTGGCTGTTGCAGGCCATTGCAAATTTATCTGTAAATACAG 1722
                    * * *** *    **  ** *** ** ********** ******** ** ** ** ** *
```

```
PpDILLO1A.gDNA      AGGTTGCGAAAAGGGTTGCTTTGGAAGGGGGCATCAGTATCCTGGCTGGTTTGGCGAGGT 1763
PpDILLO1B.gDNA      AGGTTGCGAAAAGGGTTGCTTTGGAAGGGGGCATCAGTATCCTGGCTGGTTTGGCGAGGT 1763
PpDILLO2.gDNA       AAGTTGCGAAGAGGGTTGCTTTGGAAGGGGGAATCAGTATCCTGGCTGCGTTAGCGAGGT 1782
                    * ******** ******************** ****************  ** *******

PpDILLO1A.gDNA      CCCGGAACCGGTGGGTTGCTGAGGAGGCGGCGGGGGGTTTATGGAATCTGTCGGTGGGAG 1823
PpDILLO1B.gDNA      CCCGGAACCGGTGGGTTGCTGAGGAGGCGGCGGGGGGTTTATGGAATCTGTCGGTGGGAG 1823
PpDILLO2.gDNA       CTCCGAATCGGTGGGTTGCCGAGGAGGCAGCGGGGGGTTTATGGAACCTGTCGGTCGGGG 1842
                    * * *** ************ ******** ***************** ******** ** *

PpDILLO1A.gDNA      AGGAGCACAAGGTGAGGCGCTGTGGTTGGCAATTGCGATGAAGTGTTTT---GCTGCGAA 1880
PpDILLO1B.gDNA      AGGAGCACAAGGTGAGGCGCTGTGGTTGGCAATTGCGATGAAGTGTTTT---GCTGCGAA 1880
PpDILLO2.gDNA       AAGAGCATAAGGTGGGGTGCCATGCTTTCCAACAATGATGAGGCGTTTGCGAATTGCGGA 1902
                    * ***** ****** ** **  ** **  ***    ***** * ****      **** *

PpDILLO1A.gDNA      TTGGTGGTGTGAGCGCGTCGGTGTG-----------GACATTGTAGA----GCGAGGGG- 1924
PpDILLO1B.gDNA      TTGGTGGTGTGAGCGCGTCGGTGTG-----------GACATTGTAGA----GCGAGGGG- 1924
PpDILLO2.gDNA       TT--TTGTGTGTGGGGGTCGGGATGCTGTTGCTCAAGGAACTGCACATTCGATGATGGAA 1960
                    **  * ***** * * *****  **           *  * ** * *      ** **

PpDILLO1A.gDNA      ---TGAGGAAGGCAGGTGGTGTGGGAG--GTCGGGGTTGTGGAAGGGACAGGGCG--CGG 1977
PpDILLO1B.gDNA      ---TGAGGAAGGCAGGTGGTGTGGGAG--GTCGGGGTTGTGGAAGGGACAGGGCG--CGG 1977
PpDILLO2.gDNA       TTTTTACGGAGTCTCGTCATACGGATTCTGTTGTGATAGTCAATGTTACCGAAAGACCTG 2020
                       * * * ** *  **  *  **     ** * * * **  * *  ** *   *  * *

PpDILLO1A.gDNA      CACGGATGTTGAGGTTGGGAACAGTGTAGAGTTGGAAAAGAGGAACGTGT---TTGTGCG 2034
PpDILLO1B.gDNA      CACGGATGTTGAGGTTGGGAACAGTGTAGAGTTGGAAAAGAGGAACGTGT---TTGTGCG 2034
PpDILLO2.gDNA       TGCAGAC----AGATTGTGG-TGGTGCA-AGTTACAACAGTGCATCTTGTGAACTGTTTG 2074
                      * **     ** *** *    *** * ****  ** ** * * * ***    ***  *

PpDILLO1A.gDNA      -AGTCGGTGAGGAAAGCGGAGGGTGGAGAGTGTAGTGAGGAGCCAGGGGTTAGGG----G 2089
PpDILLO1B.gDNA      -AGTCGGTGAGGAAAGCGGAGGGTGGAGAGTGTAGTGAGGAGCCAGGGGTTAGGG----G 2089
PpDILLO2.gDNA       CAGTTACTGTGCA---CGATAGGTTGT---TTTTGTGATTGGCCGGGAGTTAAGATCATA 2128
                     ***   ** * *   **   *** *    * * ****   *** ** **** *

PpDILLO1A.gDNA      GTGGGAGATGG-------ATGGTCGGGTTGGTTGTCATGGTAATGGTCATGGTAGCGCAC 2142
PpDILLO1B.gDNA      GTGGGAGATGG-------ATGGTCGGGTTGGTTGTCATGGTAATGGTCATGGTAGCGCAC 2142
PpDILLO2.gDNA       GTGAGC-ATGGTTTTGAAATAGTTGG-TTGGTTGGCT--GTCTTGGTCACATTTTCGCAC 2184
                    *** *  ****       ** ** ** ******* *   **  ******   *  *****

PpDILLO1A.gDNA      TGACAAATGAGGGCGTGAATGCAAACAGGGTGCGATAGCTGGCGCAGGCGCAATCGAAGC 2202
PpDILLO1B.gDNA      TGACAAATGAGGGCGTGAATGCAAACAGGGTGCGATAGCTGGCGCAGGCGCAATCGAAGC 2202
PpDILLO2.gDNA       TTACAAATACGACCGTTGATGCTAATAGGGTGCTATAGCCGAGGCAGGCGCAATTGAAGC 2244
                    * ******  *  ***  **** ** ******* ***** *  ************ *****

PpDILLO1A.gDNA      ACTGGTGGGATTGGCATTCAAGTGGCCCGCCGGAGGGGAGGGTGTCCTGGTGAGCACGTG 2262
PpDILLO1B.gDNA      ACTGGTGGGATTGGCATTCAAGTGGCCCGCCGGAGGGGAGGGTGTCCTGGTGAGCACGTG 2262
PpDILLO2.gDNA       ACTGGTGGATCTGGCATTTAAGTGGCCTGCTGGAGGGGAGGGTGTTCTGGTGAGCACGTA 2304
                    ********   ******* ******** ** ************** *************

PpDILLO1A.gDNA      CATCGCCCAGTAGCTTTAGTGTTGTGGCATGATCCG---CACAGCTAGGTTTGACGTTGG 2319
PpDILLO1B.gDNA      CATCGCCCAGTAGCTTTAGTGTTGTGGCATGATCCG---CACAGCTAGGTTTGACGTTGG 2319
PpDILLO2.gDNA       CATTCCTCACCAG-TTTGGGATTGT--TATGAGCCGAGCTATAGTTGTTTTTAAATCTGA 2361
                    ***  * **  ** *** *  ****   **** ***    * ** *   *** *   **

PpDILLO1A.gDNA      GAGTGGTTGACGGGTGGTTGCTGTGATGG-TTGGCATT-TGGGTGTTGGGATCTTGTTGG 2377
PpDILLO1B.gDNA      GAGTGGTTGACGGGTGGTTGCTGTGATGG-TTGGCATT-TGGGTGTTGGGATCTTGTTGG 2377
PpDILLO2.gDNA       CAT--GTTGACGA-TGTTTGCATTTCCGAATTGGTGTTGCGGGTTCTCA-ATCT-GTTGC 2416
                     *   *******  ** ****  *   *  ****  **  ****  *   **** ****

PpDILLO1A.gDNA      AAATGAATGGTGTGGTGCGGCGAGAGGAAGAAGGGATGCGTG----AGGCTACTACGTTT 2433
PpDILLO1B.gDNA      AAATGAATGGTGTGGTGCGGCGAGAGGAAGAAGGGATGCGTG----AGGCTACTACGTTT 2433
PpDILLO2.gDNA       GA-TGATTGATTTTGTCGGATGTGATGAAGAAGTGACGCTTGTTTAGGACCCTTACCCTT 2475
                     * *** ** * * **  *  * ** ******* ** ** **     * *   ***  **

PpDILLO1A.gDNA      GGTGTGTGGCATGGGGTGGTGCGAGGTGTGTTGGAAAGGCTGGTGGTGGGCATTGGTGTT 2493
PpDILLO1B.gDNA      GGTGTGTGGCATGGGGTGGTGCGAGGTGTGTTGGAAAGGCTGGTGGTGGGCATTGGTGTT 2493
PpDILLO2.gDNA       G--------CACAG---------ATGTGTGT-----AGGCTGT------GCAC-GGTGTA 2506
                    *        **  *         * ******     ******       ***  *****

PpDILLO1A.gDNA      GGTTTGAGGGTGGGTTGTGATGAGCATGTAGATGAGAGAGTGCAGCGGCGAAAGCGGAAG 2553
PpDILLO1B.gDNA      GGTTTGAGGGTGGGTTGTGATGAGCATGTAGATGAGAGAGTGCAGCGGCGAAAGCGGAAG 2553
PpDILLO2.gDNA       GTTTTGA--------TCAGAAGCGCATGGGATTGCTTGGATCCTTCG-----------AA 2547
                    * *****        *  ** * *****    **   *  * *  **           *
```

```
PpDILLO1A.gDNA      CGGGGTTGGGATGCGGTTGGGATTGACGTTGAGCATTTTGTTTCCGGATAGCAGGAGAGA 2613
PpDILLO1B.gDNA      CGGGGTTGGGATGCGGTTGGGATTGACGTTGAGCATTTTGTTTCCGGATAGCAGGAGAGA 2613
PpDILLO2.gDNA       CAGGGTTCAAGAACAGTGGACACTAACGCTGAGCACTTTGTAT--GGATGGTAGGAGCGA 2605
                    * *****      * ** *  * * *** ****** ***** *  **** * ***** **

PpDILLO1A.gDNA      GCAGCAGGTGCGTTGGCCAACCTAGCAGCCGATGACAAGTGCAGCATGGAAGTCGCAGTG 2673
PpDILLO1B.gDNA      GCAGCAGGTGCGTTGGCCAACCTAGCAGCCGATGACAAGTGCAGCATGGAAGTCGCAGTG 2673
PpDILLO2.gDNA       GCAGCAGGTGCTCTGGCTAACCTAGCAGCTGATGACAAGTGCAGCATGGAAGTTGCTGTG 2665
                    ***********  **** *********** ********************** ** ***

PpDILLO1A.gDNA      GCGGGGGGTGTGCGGGCTCTGGTGAGGCTGGCGCGGTTTTGTAACCACGAAGGCGTGCAA 2733
PpDILLO1B.gDNA      GCGGGGGGTGTGCGGGCTCTGGTGAGGCTGGCGCGGTTTTGTAACCACGAAGGCGTGCAA 2733
PpDILLO2.gDNA       GCTGGCGGTGTACGGGCCCTAGTTAGACTTGCACAGTTTTGTAATCACGAAGGGGTGCAA 2725
                    ** ** ***** ***** ** ** ** ** ** * ********* ******** ******

PpDILLO1A.gDNA      GAGCAGGTACGAGGCAACGGGAAGCTGAGAGCAGCGGTTGGTAAAGGGGGATGGGGTTTG 2793
PpDILLO1B.gDNA      GAGCAGGTACGAGGCAACGGGAAGCTGAGAGCAGCGGTTGGTAAAGGGGGATGGGGTTTG 2793
PpDILLO2.gDNA       GAGCAGGTGCGGGGCAATCAGACTCTGAG-------------------------------- 2754
                    ******** ** *****   **  *****

PpDILLO1A.gDNA      GAAGGCTGAGGTTGGGTCGTGATGAGGGGGTAGCGGAGTGAGAAACACACACATGGGGTT 2853
PpDILLO1B.gDNA      GAAGGCTGAGGTTGGGTCGTGATGAGGGGGTAGCGGAGTGAGAAACACACACATGGGGTT 2853
PpDILLO2.gDNA       ---------------GTCGCGGTG-----------------------------------T 2764
                                   **** * **                                    *

PpDILLO1A.gDNA      TGGAATGAGTGGCAGGGGCGCAGAGTGGAAGGAGGGGTTGGAGGTTTGCGAGGTTAAGTG 2913
PpDILLO1B.gDNA      TGGAATGAGTGGCAGGGGCGCAGAGTGGAAGGAGGGGTTGGAGGTTTGCGAGGTTAAGTG 2913
PpDILLO2.gDNA       TGGAAT--------------CGGAGT-----------TT---GGTTTTCGTGGATGAG-- 2794
                    ******              * ****           **   ***** ** ** * **

PpDILLO1A.gDNA      GTGCAGGACATGGTTGAGTTGGCAGAGAAGAAGTGATTGAGACAGTGGCGTTGCGGTGAA 2973
PpDILLO1B.gDNA      GTGCAGGACATGGTTGAGTTGGCAGAGAAGAAGTGATTGAGACAGTGGCGTTGCGGTGAA 2973
PpDILLO2.gDNA       GTGCACGTC-------------CAGAG--------------------------------- 2808
                    ***** * *             *****

PpDILLO1A.gDNA      GGTGTTGCGTGGGGATGGATTTGGTTAGCGTGATTAGCATTGAGGATTAATGAGGGCGTG 3033
PpDILLO1B.gDNA      GGTGTTGCGTGGGGATGGATTTGGTTAGCGTGATTAGCATTGAGGATTAATGAGGGCGTG 3033
PpDILLO2.gDNA       ---GTTGTATGGGC---------------------------------------------- 2819
                       ****  ****

PpDILLO1A.gDNA      ACGATAGTGAGGGTTGCGCGTAAGTGGAGAGTGCATTGTGTGGGTCCTGGAGGCGTTTGG 3093
PpDILLO1B.gDNA      ACGATAGTGAGGGTTGCGCGTAAGTGGAGAGTGCATTGTGTGGGTCCTGGAGGCGTTTGG 3093
PpDILLO2.gDNA       ---------------------------------------TGTAGTTTTTGGGAGAGCTTGT 2841
                                                           *** * *  ***  * * ***

PpDILLO1A.gDNA      GGCGTGAGTGGAGTGTTTGCACGATCGGTGGAGCCGGTTTTAGGTGCAGCGTTGGCGAAG 3153
PpDILLO1B.gDNA      GGCGTGAGTGGAGTGTTTGCACGATCGGTGGAGCCGGTTTTAGGTGCAGCGTTGGCGAAG 3153
PpDILLO2.gDNA       GGTG----------GTTC------TCAGTACAG---GTCTT-------------------- 2863
                    ** *          ***       ** **  **   ** **

PpDILLO1A.gDNA      AGCATGCGGAATGGGAGAGGAGGTGACGAGTTGCGAGGTGTGACTGGTGTTTTGGAACTG 3213
PpDILLO1B.gDNA      AGCATGCGGAATGGGAGAGGAGGTGACGAGTTGCGAGGTGTGACTGGTGTTTTGGAACTG 3213
PpDILLO2.gDNA       -------GAAAT---------------------------CGTGATTG----TTTG------ 2880
                           * ***                             **** **    ****

PpDILLO1A.gDNA      TGTTTGATAGGCAGCACGCGCGTTGGCGAATCTTGCGGCTCATGGAGACAGCAACGGTAA 3273
PpDILLO1B.gDNA      TGTTTGATAGGCAGCACGCGCGTTGGCGAATCTTGCGGCTCATGGAGACAGCAACGGTAA 3273
PpDILLO2.gDNA       -------CAGGCAGCGCGCGCCTTGGCCAATCTCGCCACTCATGGAGACAGCAACGGTAA 2933
                            ******* ***** ***** ***** **  **********************

PpDILLO1A.gDNA      CAACGCTGCTGTAGGGCGGGAGGAAGGCGCGCTGGAAGCTCTGGTGCAGCTGACTTGCTC 3333
PpDILLO1B.gDNA      CAACGCTGCTGTAGGGCGGGAGGAAGGCGCGCTGGAAGCTCTGGTGCAGCTGACTTGCTC 3333
PpDILLO2.gDNA       TAATGCTGCTGTGGGGCGGGAGGCAGGCGCACTTGAAGCTCTCGTGCGCCTGACTGGGTC 2993
                     ** ******** ********** ****** ** ******** ****  ****** * **

PpDILLO1A.gDNA      AAATCATGAAGGTGTCAGGTATGCG-GATTTGTTGGGGTG-GTGGACAGCAGGGAGGTGT 3391
PpDILLO1B.gDNA      AAATCATGAAGGTGTCAGGTATGCG-GATTTGTTGGGGTG-GTGGACAGCAGGGAGGTGT 3391
PpDILLO2.gDNA       TAACCATGAAGGCGTTAGGTACGTGTGATCT-CTGAGGCGAGTGTGCGGCATTGAAA--- 3049
                     ** ******** ** ***** * * *** *  ** ** * ***  * ***  **

PpDILLO1A.gDNA      TTTTGTTGTGCAGGGGTGATGTGAAGGGTTGAGGGTAGAGATGGGCCGGTGTTCTAAGAG 3451
PpDILLO1B.gDNA      TTTTGTTGTGCAGGGGTGATGTGAAGGGTTGAGGGTAGAGATGGGCCGGTGTTCTAAGAG 3451
PpDILLO2.gDNA       -TTGGTT------GGTCAATGAGATTGGATAAGGTCCAAACTCGGTGACTTACCTGACAG 3102
                     ** ***      **   *** **  ** * ***    *  * **    *   ** * **
```

```
PpDILLO1A.gDNA      TGTTTTGGGCGTCCCCTTGGACGGCGCTTGCAGACAAGAAGCAGCGGGAGCATTGTGGAA 3511
PpDILLO1B.gDNA      TGTTTTGGGCGTCCCCTTGGACGGCGCTTGCAGACAAGAAGCAGCGGGAGCATTGTGGAA 3511
PpDILLO2.gDNA       TATTTTGG-TATCGGCTTTCA---CATGTGCAGACAAGAAGCAGCTGGAGCACTGTGGAA 3158
                    * ******   **  ***  *   *   ***************** ****** *******

PpDILLO1A.gDNA      TCTGTCGTTCGACGACAGAAACCGAGAAGCTATTGCAGCGGCTGGTGGCGTGGAGGCACT 3571
PpDILLO1B.gDNA      TCTGTCGTTCGACGACAGAAACCGAGAAGCTATTGCAGCGGCTGGTGGCGTGGAGGCACT 3571
PpDILLO2.gDNA       TCTGTCTTTTGATGATAGAAACCGAGAAGCTATTGCAGCAGCTGGTGGCGTGGAGGCACT 3218
                    ****** ** ** ** ********************** ********************

PpDILLO1A.gDNA      GGTGAGTGAGAGTGGACACGAACCGTGGGCGTAGGAAGAGATACGAAAGGGGCGTTCTGG 3631
PpDILLO1B.gDNA      GGTGAGTGAGAGTGGACACGAACCGTGGGCGTAGGAAGAGATACGAAAGGGGCGTTCTGG 3631
PpDILLO2.gDNA       GGTGAGTCAGAGTTGACACGAGTCTTGACTGCAGCAAGAGAT--GGAGGAGTCAGTGTTG 3276
                    ******* ***** *******  * **   * ** *******  * * * * *  * * *

PpDILLO1A.gDNA      GGGCAGAGGTGAATGG-GGGTGATTCGAGAGCGTGGCCGGGGGGTTGAGCAGAGCTGGAG 3690
PpDILLO1B.gDNA      GGGCAGAGGTGAATGG-GGGTGATTCGAGAGCGTGGCCGGGGGGTTGAGCAGAGCTGGAG 3690
PpDILLO2.gDNA       TCGTTCGGTTCCGTGGTGGGTCACGTGACAGTGCAGTACAAGGGTGCAGTGGAACAGGAG 3336
                      *    * *   *** **** *   ** ** *  *     ****  **  ** * ****

PpDILLO1A.gDNA      GTGAAGATGAAGGAT-GCAGGCTTAGTGCGGACATGTAGAGTGATGGGGG-GCTGCATGA 3748
PpDILLO1B.gDNA      GTGAAGATGAAGGAT-GCAGGCTTAGTGCGGACATGTAGAGTGATGGGGG-GCTGCATGA 3748
PpDILLO2.gDNA       TTGAAAATGAAAAATTGCTCCCGTAATGCAGACGC-TAAATTAGTTGTGGTGGTGAATGA 3395
                     **** *****  ** **   * ** *** ***   ** * *  * * ** * ** ****

PpDILLO1A.gDNA      CAGTTGAGCGTGGGGTGGTTGTGGAGGGTAGGATCGGGGTGCGTTGATGTTGAGATCGGG 3808
PpDILLO1B.gDNA      CAGTTGAGCGTGGGGTGGTTGTGGAGGGTAGGATCGGGGTGCGTTGATGTTGAGATCGGG 3808
PpDILLO2.gDNA       CCCCTGGAAATCGTCTGGTTGTGATGTGTACCATTGGGATGTGGTCATGATTATATCAGG 3455
                    *   **    * *  ********  * ***  ** *** ** * * *** * * *** **

PpDILLO1A.gDNA      TTGGGG--GGCAGTCGTGTGGC----GAGTTGTGCGAGGTTTTGCGAGGCGTGAACAGGC 3862
PpDILLO1B.gDNA      TTGGGG--GGCAGTCGTGTGGC----GAGTTGTGCGAGGTTTTGCGAGGCGTGAACAGGC 3862
PpDILLO2.gDNA       TTGAGGACGACATGGGGATGACTTTCACCTTTTTTCAGGTTGGGAGTAAATTGCACATGT 3515
                    *** **  * **   *  ** *       ** *   *****  * *     ** *** *

PpDILLO1A.gDNA      TGT-GTT-GAGGCAAGG-GCGCACGGTGGTGTGGTG-TGGAGCAGAGTGGGGGTGGCATT 3918
PpDILLO1B.gDNA      TGT-GTT-GAGGCAAGG-GCGCACGGTGGTGTGGTG-TGGAGCAGAGTGGGGGTGGCATT 3918
PpDILLO2.gDNA       CGTTGTCAGAGGTAGACTACCCACAATACTGTGGGGATCCAGCAGAGGAATGTTGGCACA 3575
                     ** **  **** *     * ***  *  ***** * *  *******    * *****

PpDILLO1A.gDNA      GGGTGCATGACGATGGTGTGGTGGTTGTGCAGGTGGCTTTGGCTCAAGGGTGTAGCAGCG 3978
PpDILLO1B.gDNA      GGGTGCATGACGATGGTGTGGTGGTTGTGCAGGTGGCTTTGGCTCAAGGGTGTAGCAGCG 3978
PpDILLO2.gDNA       GAATGCATGACGAACCTGTGGTGGTTGTGTAGGTTGCTTTAGCTCAAGATTGTAGCAGCG 3635
                    *  **********   ************** **** ***** *******  **********

PpDILLO1A.gDNA      GATCTCAAGGTCTGCAAGAGAGGGCTGCGGGTGCATTATGGGGGCTGTCTGTGTCGGAGG 4038
PpDILLO1B.gDNA      GATCTCAAGGTCTGCAAGAGAGGGCTGCGGGTGCATTATGGGGGCTGTCTGTGTCGGAGG 4038
PpDILLO2.gDNA       GATCTCAAGGTCTGCAGGAGAGGGCTGCTGGTGCATTGTGGGGTCTATCTGTTTCTGAAG 3695
                    *************** *********** ******** ***** ** ***** ** ** *

PpDILLO1A.gDNA      CGAACAGGTGAGGGCATTGAGCGCGATGGGAGGTGGTGTTGTTGGCGGAAGCGTTGGGCG 4098
PpDILLO1B.gDNA      CGAACAGGTGAGGGCATTGAGCGCGATGGGAGGTGGTGTTGTTGGCGGAAGCGTTGGGCG 4098
PpDILLO2.gDNA       CCAACAGGTGAGAACGTTTTTCGTGATGGGAAGCTGTGTTCTTCGCGGAAGCCTTGGGTG 3755
                    * **********  * **   ** ******* *  ***** ** ******** ***** *

PpDILLO1A.gDNA      TGGGTTGAAGGTCGCAGGGGTAGAGTCGTCGGGGATGGAGCGGATTGTGTGAAGTATTGC 4158
PpDILLO1B.gDNA      TGGGTTGAAGGTCGCAGGGGTAGAGTCGTCGGGGATGGAGCGGATTGTGTGAAGTATTGC 4158
PpDILLO2.gDNA       TTTGTTGAAGGGTGCAGGTCTTCATCTGTGGGTGCTGAAGTCGATGATATGTAC--TTGT 3813
                    *  ********  *****  *  *   ** ** * ** **  ***  * ** *   ***

PpDILLO1A.gDNA      GCGGAGTGAGAGTCGATGGGTTGGATTTGGCGATTATGGGGATTTTGGTTGCGTAGCAGA 4218
PpDILLO1B.gDNA      GCGGAGTGAGAGTCGATGGGTTGGATTTGGCGATTATGGGGATTTTGGTTGCGTAGCAGA 4218
PpDILLO2.gDNA       GTGGAG-------------GTTGGGGTT-----TTATGGA----TCGGTT-TGCAGC--- 3847
                    * ****             *****  **     ******     * ****  * ***

PpDILLO1A.gDNA      GGTTGGAGGTGGGTGACGGGGTGTGGTGCGATGACCGTGTGTCTGACGTGTGTGTGTGTG 4278
PpDILLO1B.gDNA      GGTTGGAGGTGGGTGACGGGGTGTGGTGCGATGACCGTGTGTCTGACGTGTGTGTGTGTG 4278
PpDILLO2.gDNA       --CTG-----------CGGA---------GATGA---TATCACTGAAGTGCGTGC---TG 3879
                       **           ***          *****   * *  **** *** ***    **

PpDILLO1A.gDNA      TGTGTGTGTGTGTTGGGAACTTGCGTGGTGGCGCAGCATTGCTATCGGGCGCGAAGGCGG 4338
PpDILLO1B.gDNA      TGTGTGTGTGTGTTGGGAACTTGCGTGGTGGCGCAGCATTGCTATCGGGCGCGAAGGCGG 4338
PpDILLO2.gDNA       TGAATGTGTGTTTTGG---------------CTCAGCATTGCTATCGGTCGTGAAGGTGG 3924
                    **  ******* ****               * *************** ** ***** **
```

```
PpDILLO1A.gDNA      AGTGGCTCCACTGATCACCCTGGCGCACTCTGATTTTGAGGTGAGTTGTGGGATGAGTTG 4398
PpDILLO1B.gDNA      AGTGGCTCCACTGATCACCCTGGCGCACTCTGATTTTGAGGTGAGTTGTGGGATGAGTTG 4398
PpDILLO2.gDNA       AGTTGCCCCGCTGATCACTCTGGCACACTCGAACTCTGAGGTGAGGTGCTGGATTTGGCG 3984
                    *** ** ** ******** ***** *****  * * ********* **  ****  *  *

PpDILLO1A.gDNA      CAGGGTTTTGGGGGACGAGGTTTTGGGT---TTGGTTAGCGCAGCGTGCATGTCA---TG 4452
PpDILLO1B.gDNA      CAGGGTTTTGGGGGACGAGGTTTTGGGT---TTGGTTAGCGCAGCGTGCATGTCA---TG 4452
PpDILLO2.gDNA       CATGGTTTTGCAGCATGAGGCATGTGATATTCTGGTCAGCGGAGTATATATGGTAGCTTG 4044
                    ** *******  * * ****  *  * *    **** **** **  *  ***  *   **

PpDILLO1A.gDNA      TGAGTGACA-CGGTTAGACGCGAGTGGTTGGGG-----TGTGATGGTG--GAGGCTGGGT 4504
PpDILLO1B.gDNA      TGAGTGACA-CGGTTAGACGCGAGTGGTTGGGG-----TGTGATGGTG--GAGGCTGGGT 4504
PpDILLO2.gDNA       TAAGTCTCAGCAGTTGGAGGTT-GTGGATGAAACTGATTACGAATATGTCGAGCTCAAGG 4103
                    * ***  ** * *** ** *   **** **        *  **   **  ***     *

PpDILLO1A.gDNA      GGTGGTGGCATACAAAG-TGCAA---AGTGCAGGGAGGCGAGAGC-TGGATGCATGGGTG 4559
PpDILLO1B.gDNA      GGTGGTGGCATACAAAG-TGCAA---AGTGCAGGGAGGCGAGAGC-TGGATGCATGGGTG 4559
PpDILLO2.gDNA       GCTGGTGCTTCGTTGGGATGCTTTCTAGTATGCGGAGGTCAGAACATACATTAAACAGTA 4163
                    * *****         * ***     ***    *****  *** * *  **  *   **

PpDILLO1A.gDNA      AGTGGGAGGAATGGGGTGTAGTGGAAATGGAAAGGGGGCGTG-GTGTGGAGCCATGGTTG 4618
PpDILLO1B.gDNA      AGTGGGAGGAATGGGGTGTAGTGGAAATGGAAAGGGGGCGTG-GTGTGGAGCCATGGTTG 4618
PpDILLO2.gDNA       TGTAGGGCAGATGGTGTGCTGTTGAGGTGTGAGGAATAGCTGTGCATGGAGC--TGTTCG 4221
                     ** **    **** ***  ** **  **  * *      ** *  ******  ** * *

PpDILLO1A.gDNA      GGAGATGGTTGTACGTGGC--GGGGCATTGGTGGTGGTATCGGTGTGGTGGGACGATGAT 4676
PpDILLO1B.gDNA      GGAGATGGTTGTACGTGGC--GGGGCATTGGTGGTGGTATCGGTGTGGTGGGACGATGAT 4676
PpDILLO2.gDNA       TGAAATAGCTTTTTGTACCATGGTGCAGTCTGGACGGTGTTGGCCCGTTGGCTCAGCAAT 4281
                     ** ** * * *  **  *  ** *** *   *  *** * **   * ***  *    **

PpDILLO1A.gDNA      GG--GCGTTTAGGAGGTGGGTTTC---TGAAGTTTCCGGCGATGGGAGTGCAGGACGTAC 4731
PpDILLO1B.gDNA      GG--GCGTTTAGGAGGTGGGTTTC---TGAAGTTTCCGGCGATGGGAGTGCAGGACGTAC 4731
PpDILLO2.gDNA       GAATGCGTATTGAGTGTTGGTTTGGAATGGGATT------GACGCAATTGCAGGACGTAC 4335
                    *   **** * *   ** *****    **   **      ** *  * ************

PpDILLO1A.gDNA      ATGAAACAGCGGTGGGAGCGTTGTGGAACCTGGTTTTCAATCCTGGCAACGCACTGCGGA 4791
PpDILLO1B.gDNA      ATGAAACAGCGGTGGGAGCGTTGTGGAACCTGGTTTTCAATCCTGGCAACGCACTGCGGA 4791
PpDILLO2.gDNA       ATGAGACAGCGGTGGGCGCGTTGTGGAATCTGGCTTTCAATCCCGGTAACGCGTTGCGCA 4395
                    **** *********** *********** **** ********* ** *****  **** *

PpDILLO1A.gDNA      TGGTGGAAGAGGAGGGAGTCCCGGCGCTGGTGCATTTGTGCTCGTCGTCGAGATCTAAGA 4851
PpDILLO1B.gDNA      TGGTGGAAGAGGAGGGAGTCCCGGCGCTGGTGCATTTGTGCTCGTCGTCGAGATCTAAGA 4851
PpDILLO2.gDNA       TGGCGGAAGAAG---GAGTCCCTGCGCTTGTACATTTGTGCTCGTCGTCAAGATCTAAAA 4452
                    *** ****** *   ******* ***** ** ***************** ******** *

PpDILLO1A.gDNA      TGGCGCGGTTTATGGCAGCTCTGGCTCTGGCGTACATGTTCGATGGCAGGTAAAGCGGTG 4911
PpDILLO1B.gDNA      TGGCGCGGTTTATGGCAGCTCTGGCTCTGGCGTACATGTTCGATGGCAGGTAAAGCGGTG 4911
PpDILLO2.gDNA       TGGCACGTTTTATGGCAGCTCTGGCCCTCGCGTACATGTTCGATGGCAGGTATAGCACTC 4512
                    **** ** ***************** ** ********************** ***  *

PpDILLO1A.gDNA      GCGTGTGGGTGTGTGAGGGTGCGGTCGTAGGACGTGGGGGTGGTGAGATGACGTGAGACG 4971
PpDILLO1B.gDNA      GCGTGTGGGTGTGTGAGGGTGCGGTCGTAGGACGTGGGGGTGGTGAGATGACGTGAGACG 4971
PpDILLO2.gDNA       ATATCTGCCTGTATGAGCTAGATGTTGCCGAACTTGG------TGGAATTGTGTGGAATG 4566
                       * **  *** ****   *  ** *  * ** ***      **  **   ***  * *

PpDILLO1A.gDNA      GGGAGGAGGAAGTTCGTAGGCGTTGGATAGGTGCGGTGAGGGAGAGTCGGGTGAAGTGGA 5031
PpDILLO1B.gDNA      GGGAGGAGGAAGTTCGTAGGCGTTGGATAGGTGCGGTGAGGGAGAGTCGGGTGAAGTGGA 5031
PpDILLO2.gDNA       GGTGGGTGGAAGTCCGCGGAAAGCGTCTTGATGTGGTCAGGGTGTGTGCGGTGTGGTGGA 4626
                    **  ** ****** **  *     *  * * ** *** **** * **  ****  *****

PpDILLO1A.gDNA      GCAGTGGGGAGTGGGTGACGTGGTTTTGGAGA--ACCGGA---AGATATGGAAGAAGGGC 5086
PpDILLO1B.gDNA      GCAGTGGGGAGTGGGTGACGTGGTTTTGGAGA--ACCGGA---AGATATGGAAGAAGGGC 5086
PpDILLO2.gDNA       GCAGTGTAGT-CGAGTGTTTTCGTTTGGGAAATAACTAGGTCCGGTGATGGTGACGGTGC 4685
                    ******  *   * ***   * **** *** *  **  *     *  ****     * **

PpDILLO1A.gDNA      GCGGTTAGGAATGAGGAGATTGCTGTTGCGTTGTTGGGACGAGGTTGGGGCGGGGGTGAG 5146
PpDILLO1B.gDNA      GCGGTTAGGAATGAGGAGATTGCTGTTGCGTTGTTGGGACGAGGTTGGGGCGGGGGTGAG 5146
PpDILLO2.gDNA       CCAGCAATGGACGATGAAGTT-TTGTTGCATAACAGGAGCTCA-TTCGGATATGTTTCGA 4743
                     * *  * * * ** **  **  ****** *    **  *    ** **    *  *

PpDILLO1A.gDNA      AAGTGGAGGGGGTTGA---GGTGCGTAGGAGCGCGCGTTTGTGTGCATGATTTG-----A 5198
PpDILLO1B.gDNA      AAGTGGAGGGGGTTGA---GGTGCGTAGGAGCGCGCGTTTGTGTGCATGATTTG-----A 5198
PpDILLO2.gDNA       TAGTTGAGGTAGTTAATTAGGCGTTTTGTGCAGCG-------GTGCACGACTTGTTCGAA 4796
                     *** ****  *** *   ** *  * *    ***       ***** ** ***     *
```

```
PpDILLO1A.gDNA      AGTGGTTGGGCGGCGTTTGGAGCGGTTGAGTGGGTGTGGCAAAGCAGTGAGTGGGAGTGA 5258
PpDILLO1B.gDNA      AGTGGTTGGGCGGCGTTTGGAGCGGTTGAGTGGGTGTGGCAAAGCAGTGAGTGGGAGTGA 5258
PpDILLO2.gDNA       AGTCCTGAGGCGGCCTGTGTAGCGATAGAGTTG-TCCGGCAGAGGAGTGTGGGTAAATGC 4855
                    ***  *  ****** * ** **** * **** * *  **** ** **** * *  * **

PpDILLO1A.gDNA      TGGAGAGTGAACGTGATCGGAGTGGTTGACGTCATAGTTGGGCTCCTCTGACGCAGGATG 5318
PpDILLO1B.gDNA      TGGAGAGTGAACGTGATCGGAGTGGTTGACGTCATAGTTGGGCTCCTCTGACGCAGGATG 5318
PpDILLO2.gDNA       TGCGAGATAGACGTG-TTGTAG--ACTAGCGTCACATTTGGGCTCCGCTGGTGCAGGATG 4912
                    **     *  ***** * * **    *  ***** * ********* ***  ********

PpDILLO1A.gDNA      GATGAGGTTGCAGTTGGATTGTCGAGTGTGGAGAACAACGGGAGGACAGTGAGCTTGGAG 5378
PpDILLO1B.gDNA      GATGAGGTTGCAGTTGGATTGTCGAGTGTGGAGAACAACGGGAGGACAGTGAGCTTGGAG 5378
PpDILLO2.gDNA       GATGAGGTTGCTGTCAGAGTATCGAGCGGAGAGAACCATGGCAGGACAGTGAACCTGGAA 4972
                    *********** **  ** * ***** *  ****** * ** ********** * ****

PpDILLO1A.gDNA      GCAGCAAGGAAACTGGCGTTGCGAAACATCGATGCGTTCGTGCAAACGTTTTTTGATCCG 5438
PpDILLO1B.gDNA      GCAGCAAGGAAACTGGCGTTGCGAAACATCGATGCGTTCGTGCAAACGTTTTTTGATCCG 5438
PpDILLO2.gDNA       GCGATTCGGAAATTGGCGTTGCGAAGCATTGATGCGTTTGTGCTAACTTTCTGTGATCAA 5032
                    **     ***** ************ *** ******** **** *** ** * *****

PpDILLO1A.gDNA      CAGTCTTTGACAGCTGCTGCGTCATCGTGGGCCGGTGCGACGTTGAACCAGGTCGCGGAG 5498
PpDILLO1B.gDNA      CAGTCTTTGACAGCTGCTGCGTCATCGTGGGCCGGTGCGACGTTGAACCAGGTCGCGGAG 5498
PpDILLO2.gDNA       CAAGCTCTGACAGCTGCTGCGTCATCGTGGGCTCCTGCCACCCTGAACCAAGTTGCGGAG 5092
                    **  ** ************************   *** **  ******* ** ******

PpDILLO1A.gDNA      ACCGCCACAATCCAAGAAGCCGGGCATCTGAGATGCAGGTAGGCGATGCGTACGTAGGG- 5557
PpDILLO1B.gDNA      ACCGCCACAATCCAAGAAGCCGGGCATCTGAGATGCAGGTAGGCGATGCGTACGTAGGG- 5557
PpDILLO2.gDNA       ACCGCCAGGATCCAAGAAGCTGGACTTCTCAGATGCAGGTAAGGGGTGTGTGTGTGGGGA 5152
                    *******  *********** ** * *** *********** * * ** **  ** ***

PpDILLO1A.gDNA      CGTGAATGTGTGGATGGTACAGAATTAGATGCAACGGATTTGGATGC-ATGG-------- 5608
PpDILLO1B.gDNA      CGTGAATGTGTGGATGGTACAGAATTAGATGCAACGGATTTGGATGC-ATGG-------- 5608
PpDILLO2.gDNA       TGGGAAGATGCACATC-CATTGAAGTCGGAGCAATGAGATTGAGTGGGATAGCTGTTTTC 5211
                     * ***  **   **   *  *** * *  **** *   ***  **  ** *

PpDILLO1A.gDNA      ----GC--AGCGGAAGCAACGTGGGCGCGTG--GTAAGGAATGGACGCGAGTGGGTGT-G 5659
PpDILLO1B.gDNA      ----GC--AGCGGAAGCAACGTGGGCGCGTG--GTAAGGAATGGACGCGAGTGGGTGT-G 5659
PpDILLO2.gDNA       TTTTGCTGAGTGGAAGCAATTTAAGCCCGTCTTGTATGACATGGGCATTGGCAAGTTTTG 5271
                        **  ** ********  *  ** ***   *** *  **** *    *   ** * *

PpDILLO1A.gDNA      GGCATGACAGGCGGTCGGGGTGATGGGACAGTGGCGGTGATGAGGTAGCGGATAGTGAAG 5719
PpDILLO1B.gDNA      GGCATGACAGGCGGTCGGGGTGATGGGACAGTGGCGGTGATGAGGTAGCGGATAGTGAAG 5719
PpDILLO2.gDNA       AGTATTG-GGTTCATGTAGTTGATAGGGAAGTTGAGGC--TTGGACAGTGGGTAACGACA 5328
                     * **    *    *   * **** **  *** * **   *  *  ** ** **  **

PpDILLO1A.gDNA      G--GCGGTGTGGAGTGCGGTGCGTGGAGTGTTGTTTGTATGGACGAGGCTAAGAGCGTGC 5777
PpDILLO1B.gDNA      G--GCGGTGTGGAGTGCGGTGCGTGGAGTGTTGTTTGTATGGACGAGGCTAAGAGCGTGC 5777
PpDILLO2.gDNA       GCAGTGGTTCCCAATGCAAT---TGGA----------CGTGGACAAAATTT-GAGAGTTC 5374
                    *  * ***    * ***  *   ****           ***** *   *  *** ** *

PpDILLO1A.gDNA      GGGTGGGTTGGTTGCAGTGTAGCCGAGCTCGGGCGGTTTGTGGCCATGTTGAGGAACGGA 5837
PpDILLO1B.gDNA      GGGTGGGTTGGTTGCAGTGTAGCCGAGCTCGGGCGGTTTGTGGCCATGTTGAGGAACGGA 5837
PpDILLO2.gDNA       GGCGTGGTTGGTTGCAGTGGAGCGGAGCTTGGACGGTTTGTGGCCATGTTGAGGAATGGT 5434
                    **   ************** *** ***** ** ********************** **

PpDILLO1A.gDNA      CATGCAGTTTTGCGATCGTGTGCGGCATTCGCTCTTCTGCAGGTAAGCAGTGTGAGTGAG 5897
PpDILLO1B.gDNA      CATGCAGTTTTGCGATCGTGTGCGGCATTCGCTCTTCTGCAGGTAAGCAGTGTGAGTGAG 5897
PpDILLO2.gDNA       TCTGCAGTTCTCCGAACATGTGCTGCGTTTGCTCTTCTGCAGGTAATCACTG---GTGCT 5491
                      ******* * *** * ***** ** ** ***************** ** **   ***

PpDILLO1A.gDNA      CATGAAGAGGACAAAGCGGTGGTGAGGGACGCGAGGTGGGTGTGATGGTAGGGTTGTTA- 5956
PpDILLO1B.gDNA      CATGAAGAGGACAAAGCGGTGGTGAGGGACGCGAGGTGGGTGTGATGGTAGGGTTGTTA- 5956
PpDILLO2.gDNA       GATCGACAGGACAGAGT-----TGACGGA-GCGA--TGGG-GTGGTTGTATCATTTTTAG 5542
                     **  * ****** **      *** *** ****  **** *** * ***   ** ***

PpDILLO1A.gDNA      --TTGATTCGCCG-GTGGAGTGATGTGCGTGAGAGA-GAGGAAGTGTTGGAGCGAGGTT- 6011
PpDILLO1B.gDNA      --TTGATTCGCCG-GTGGAGTGATGTGCGTGAGAGA-GAGGAAGTGTTGGAGCGAGGTT- 6011
PpDILLO2.gDNA       GTTCGTTTTGTTACGGGGAGTGTTGGAC-TGAGGTTTAAGGTACCTTTCGACTTAGGTTT 5601
                      * * ** *    * ****** **  * ****     *** *   ** **   *****

PpDILLO1A.gDNA      -----------GGAGGTTGGTTT-TGGCGTGATGGATTG-------GGATGTGGTTGCGT 6052
PpDILLO1B.gDNA      -----------GGAGGTTGGTTT-TGGCGTGATGGATTG-------GGATGTGGTTGCGT 6052
PpDILLO2.gDNA       ACTGACCCGTAGGGTGTGGGTGTGTGGAACGGTTTACTGACCCGTAGGGTGTGGGTGTGT 5661
                              **  ** *** * ***   * *  * **       ** ***** ** **
```

```
PpDILLO1A.gDNA      GTCACAGTTCACGATGCCTGGAGGCAGGCATGCGCTGCACCACGCAGTTCTGCTGCAGAA 6112
PpDILLO1B.gDNA      GTCACAGTTCACGATGCCTGGAGGCAGGCATGCGCTGCACCACGCAGTTCTGCTGCAGAA 6112
PpDILLO2.gDNA       GGAACAGTTTACTATGCCTGGAGGTAGGCATGCGATGCACCATGCAGATCTTCTGCAGAA 5721
                    *  ****** ** *********** ********* ******* **** *** ********

PpDILLO1A.gDNA      GGTTGGTGCGAGTCGGGTTCTGCGGACTGTGGCTGCAGCAGCGACGGCAGCAATGCAGGC 6172
PpDILLO1B.gDNA      GGTTGGTGCGAGTCGGGTTCTGCGGACTGTGGCTGCAGCAGCGACGGCAGCAATGCAGGC 6172
PpDILLO2.gDNA       GACCGGTGCAAGTCGAGTGCTGCGGACTGTGGCTGCAGCAGCAACTGCGGCAATGCAGGC 5781
                    *   ***** ***** ** ********************** ** ** ***********

PpDILLO1A.gDNA      CAAAGTGTTCGCCCGAGTGGTTCTTCGAAACCTGGAGCATCATCAGGCGGAGAGCAACAC 6232
PpDILLO1B.gDNA      CAAAGTGTTCGCCCGAGTGGTTCTTCGAAACCTGGAGCATCATCAGGCGGAGAGCAACAC 6232
PpDILLO2.gDNA       TAAAGTGTTTGCCCGAATCGTTCTTCGTAACCTCGAGCATCATCAAGCGGAGAGCAACGC 5841
                     ******** ****** * ******** ***** ********** ************ *

PpDILLO1A.gDNA      GGTGGTGGCACTGTCGTAA 6251
PpDILLO1B.gDNA      GGTGGTGGCACTGTCGTAA 6251
PpDILLO2.gDNA       AGTGGTGGCACTATTATAA 5860
                     ********** *  ***
```

8.3 *PHYSCODILLO* genomic DNA aligned with cDNA

```
PpDILLO1A.cDNA      ATGTCCAACAAGCGGCGGCGCAATGTCAATGTGGCGGTGGACGATCAGGAGCAGCAGCAG 60
PpDILLO1A.gDNA      ATGTCCAACAAGCGGCGGCGCAATGTCAATGTGGCGGTGGACGATCAGGAGCAGCAGCAG 60
                    ************************************************************

PpDILLO1A.cDNA      GCTGTTGTATATAAGAAAGCCAGGATAACATCGTCGCCGACATCGTCGGCGGCGTCCACT 120
PpDILLO1A.gDNA      GCTGTTGTATATAAGAAAGCCAGGATAACATCGTCGCCGACATCGTCGGCGGCGTCCACT 120
                    ************************************************************

PpDILLO1A.cDNA      TGTGGTGCACCGGCGGTGCCGGCTACAGAATCAGGTGTTGAAAGTGACGCGGGTCGCGAT 180
PpDILLO1A.gDNA      TGTGGTGCACCGGCGGTGCCGGCTACAGAATCAGGTGTTGAAAGTGACGCGGGTCGCGAT 180
                    ************************************************************

PpDILLO1A.cDNA      GCGCATTGGACAAGTCTTCCAGATGAAACCGTGCTTGGGTTGTTTAATCTGCTGAACCAC 240
PpDILLO1A.gDNA      GCGCATTGGACAAGTCTTCCAGATGAAACCGTGCTTGGGTTGTTTAATCTGCTGAACCAC 240
                    ************************************************************

PpDILLO1A.cDNA      AGGGACAGGGCGAGCCTTGCGTCCGTGTGCAAGGGATGGCAGGTGCTGGGGAGCTCGCCA 300
PpDILLO1A.gDNA      AGGGACAGGGCGAGCCTTGCGTCCGTGTGCAAGGGATGGCAGGTGCTGGGGAGCTCGCCA 300
                    ************************************************************

PpDILLO1A.cDNA      TCGCTGTGGAATTCGCTGGATCTTCGGTCGCACTCGTTGAATTCGGAGATGGTGTCTGCG 360
PpDILLO1A.gDNA      TCGCTGTGGAATTCGCTGGATCTTCGGTCGCACTCGTTGAATTCGGAGATGGTGTCTGCG 360
                    ************************************************************

PpDILLO1A.cDNA      CTGGCAGGACGATGCTCGAATTTGGAGGCGCTGAAGTTCCGCAGAGGTGCCTCTGCGAGC 420
PpDILLO1A.gDNA      CTGGCAGGACGATGCTCGAATTTGGAGGCGCTGAAGTTCCGCAGAGGTGCCTCTGCGAGC 420
                    ************************************************************

PpDILLO1A.cDNA      TCAATTGTGGGGTTGCAGGCGAAGGGGCTTAGGGAGCTGAGCGGCGATTGCTGCAGCCAG 480
PpDILLO1A.gDNA      TCAATTGTGGGGTTGCAGGCGAAGGGGCTTAGGGAGCTGAGCGGCGATTGCTGCAGCCAG 480
                    ************************************************************

PpDILLO1A.cDNA      CTGTCGGATGCAACCCTGTCCATGGTGGTCGCCCGGCATGCGAACCTGGAGAGCTTGCTG 540
PpDILLO1A.gDNA      CTGTCGGATGCAACCCTGTCCATGGTGGTCGCCCGGCATGCGAACCTGGAGAGCTTGCTG 540
                    ************************************************************

PpDILLO1A.cDNA      CTGGGTTCGGATTGCGAGAGAGTGACGAGCGAGGCATTGAAGGTGATTGCAGTGTGTTGT 600
PpDILLO1A.gDNA      CTGGGTTCGGATTGCGAGAGAGTGACGAGCGAGGCATTGAAGGTGATTGCAGTGTGTTGT 600
                    ************************************************************

PpDILLO1A.cDNA      CCGAAATTGCGACGGCTGTGCGTGTCGGGCGTCCTAAAAGTGGAGAGAGACGCCATCCAG 660
PpDILLO1A.gDNA      CCGAAATTGCGACGGCTGTGCGTGTCGGGCGTCCTAAAAGTGGAGAGAGACGCCATCCAG 660
                    ************************************************************

PpDILLO1A.cDNA      GCGTTGTTCCAGCATTGCAAGGGTCTGACGGAGCTGGGGTTCCTGGACAGTCACACCATC 720
PpDILLO1A.gDNA      GCGTTGTTCCAGCATTGCAAGGGTCTGACGGAGCTGGGGTTCCTGGACAGTCACACCATC 720
                    ************************************************************

PpDILLO1A.cDNA      GACGAAGGGGCGTTCGGGGGTGCGAGCAGCCTGCGATTTTTGTCGGTAGCGGGATGCAGG 780
PpDILLO1A.gDNA      GACGAAGGGGCGTTCGGGGGTGCGAGCAGCCTGCGATTTTTGTCGGTAGCGGGATGCAGG 780
                    ************************************************************

PpDILLO1A.cDNA      TGTATAGTGTGGAGCACGGCGGCGCATTGGTGGAGCAAATTGCCGAACCTGGCAGGGCTG 840
PpDILLO1A.gDNA      TGTATAGTGTGGAGCACGGCGGCGCATTGGTGGAGCAAATTGCCGAACCTGGCAGGGCTG 840
                    ************************************************************

PpDILLO1A.cDNA      GATGTTTCGCGGACAGATATCACTCCGACTGCGCTGATGCAGGTCCTGGCAGGGCCCGAG 900
PpDILLO1A.gDNA      GATGTTTCGCGGACAGATATCACTCCGACTGCGCTGATGCAGGTCCTGGCAGGGCCCGAG 900
                    ************************************************************

PpDILLO1A.cDNA      CTGAGGGTGGTTTGCGCTCTGAACTGCCCTGTTCTGGAGGAGGGCAGCAACCCCGTGACA 960
PpDILLO1A.gDNA      CTGAGGGTGGTTTGCGCTCTGAACTGCCCTGTTCTGGAGGAGGGCAGCAACCCCGTGACA 960
                    ************************************************************

PpDILLO1A.cDNA      CTGCCGTCGTCGAAGAAGACGGTGGTGCTGGCTCGGTTCACGGATGTGATGGAAGGGTTG 1020
PpDILLO1A.gDNA      CTGCCGTCGTCGAAGAAGACGGTGGTGCTGGCTCGGTTCACGGATGTGATGGAAGGGTTG 1020
                    ************************************************************

PpDILLO1A.cDNA      GATGCTTTGCTGAGCCCGAGCAATTGGAAGGAGGAGGCAGGCTCGTGCGCGAGGTCGAGG 1080
PpDILLO1A.gDNA      GATGCTTTGCTGAGCCCGAGCAATTGGAAGGAGGAGGCAGGCTCGTGCGCGAGGTCGAGG 1080
                    ************************************************************
```

```
PpDILLO1A.cDNA      TGCGGCGGGCGCGCGCGCGCGGATTCGGAAGTGGCGAAGTGGACGGAGTGGATGTTGTCG 1140
PpDILLO1A.gDNA      TGCGGCGGGCGCGCGCGCGCGGATTCGGAAGTGGCGAAGTGGACGGAGTGGATGTTGTCG 1140
                    ************************************************************

PpDILLO1A.cDNA      CATGCGGTTTTGAAGATCGCGGAGTGCAATGCACCGAGTCTGCTGAAGCAGGGGATCGCG 1200
PpDILLO1A.gDNA      CATGCGGTTTTGAAGATCGCGGAGTGCAATGCACCGAGTCTGCTGAAGCAGGGGATCGCG 1200
                    ************************************************************

PpDILLO1A.cDNA      ATGATGCTTCGACTGGTTCAAAGCGCGCAGGAGGATGTACAGGAGAGGGCTGCGTCGGCT 1260
PpDILLO1A.gDNA      ATGATGCTTCGACTGGTTCAAAGCGCGCAGGAGGATGTACAGGAGAGGGCTGCGTCGGCT 1260
                    ************************************************************

PpDILLO1A.cDNA      TTGGCGACGTTTGTAGTGGTGGACGACGAGAACGCGACTGTAGATTCTGCGCGGGCGGAA 1320
PpDILLO1A.gDNA      TTGGCGACGTTTGTAGTGGTGGACGACGAGAACGCGACTGTAGATTCTGCGCGGGCGGAA 1320
                    ************************************************************

PpDILLO1A.cDNA      GCTGTGATGAACGGGGGCGGGATCGCGCTGCTGTTGGGGCTGGCGAAGTCTTGCCGCGAG 1380
PpDILLO1A.gDNA      GCTGTGATGAACGGGGGCGGGATCGCGCTGCTGTTGGGGCTGGCGAAGTCTTGCCGCGAG 1380
                    ************************************************************

PpDILLO1A.cDNA      GGGGTGCAGTCGGAAGCTGCGAAGG---------------------------------- 1405
PpDILLO1A.gDNA      GGGGTGCAGTCGGAAGCTGCGAAGGTAGGGGGGTTGGTGGGGAATGTAGTGGAGTGTGAG 1440
                    *************************

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      TGTGCACAAGGCAGGCGTGTGCGCAGGAAGGATTGTAGGGTAGCGGGGGTATTGCGTGAG 1500


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      TTGTGGAGCGTGGGGTGGAGCAGGGTCGGGGGAGAGCAGGTGGACGACGATGGTGGTATT 1560


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      TGCGGAATGGGTTGATGTTTGTGGGAGTGTGGGTCGCATGCGGGCACAGTGCGTGTGGGG 1620


PpDILLO1A.cDNA      ------------------------------------------------------CCAT 1409
PpDILLO1A.gDNA      TGCAGAAGCTGTGGGGAGTTGGACTGACTGGTAGCTGTTGGGTTGGTTGGTGCAGGCCAT 1680
                                                                         ****

PpDILLO1A.cDNA      CGCAAATTTGTCCGTGAACACGGAGGTTGCGAAAAGGGTTGCTTTGGAAGGGGGCATCAG 1469
PpDILLO1A.gDNA      CGCAAATTTGTCCGTGAACACGGAGGTTGCGAAAAGGGTTGCTTTGGAAGGGGGCATCAG 1740
                    ************************************************************

PpDILLO1A.cDNA      TATCCTGGCTGGTTTGGCGAGGTCCCGGAACCGGTGGGTTGCTGAGGAGGCGGCGGGGGG 1529
PpDILLO1A.gDNA      TATCCTGGCTGGTTTGGCGAGGTCCCGGAACCGGTGGGTTGCTGAGGAGGCGGCGGGGGG 1800
                    ************************************************************

PpDILLO1A.cDNA      TTTATGGAATCTGTCGGTGGGAGAGGAGCACAAGG-------------------------- 1564
PpDILLO1A.gDNA      TTTATGGAATCTGTCGGTGGGAGAGGAGCACAAGGTGAGGCGCTGTGGTTGGCAATTGCG 1860
                    **********************************

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      ATGAAGTGTTTTGCTGCGAATTGGTGGTGTGAGCGCGTCGGTGTGGACATTGTAGAGCGA 1920


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GGGGTGAGGAAGGCAGGTGGTGTGGGAGGTCGGGGTTGTGGAAGGGACAGGGCGCGGCAC 1980


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GGATGTTGAGGTTGGGAACAGTGTAGAGTTGGAAAAGAGGAACGTGTTTGTGCGAGTCGG 2040


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      TGAGGAAAGCGGAGGGTGGAGAGTGTAGTGAGGAGCCAGGGGTTAGGGGGTGGGAGATGG 2100


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      ATGGTCGGGTTGGTTGTCATGGTAATGGTCATGGTAGCGCACTGACAAATGAGGGCGTGA 2160
```

```
PpDILLO1A.cDNA      -----------GTGCGATAGCTGGCGCAGGCGCAATCGAAGCACTGGTGGGATTGGCATT 1613
PpDILLO1A.gDNA      ATGCAAACAGGGTGCGATAGCTGGCGCAGGCGCAATCGAAGCACTGGTGGGATTGGCATT 2220
                              **********************************************

PpDILLO1A.cDNA      CAAGTGGCCCGCCGGAGGGGAGGGTGTCCTGG---------------------------- 1645
PpDILLO1A.gDNA      CAAGTGGCCCGCCGGAGGGGAGGGTGTCCTGGTGAGCACGTGCATCGCCCAGTAGCTTTA 2280
                    ********************************

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GTGTTGTGGCATGATCCGCACAGCTAGGTTTGACGTTGGGAGTGGTTGACGGGTGGTTGC 2340


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      TGTGATGGTTGGCATTTGGGTGTTGGGATCTTGTTGGAAATGAATGGTGTGGTGCGGCGA 2400


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GAGGAAGAAGGGATGCGTGAGGCTACTACGTTTGGTGTGTGGCATGGGGTGGTGCGAGGT 2460


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GTGTTGGAAAGGCTGGTGGTGGGCATTGGTGTTGGTTTGAGGGTGGGTTGTGATGAGCAT 2520


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GTAGATGAGAGAGTGCAGCGGCGAAAGCGGAAGCGGGGTTGGGATGCGGTTGGGATTGAC 2580


PpDILLO1A.cDNA      ----------------------------AGAGAGCAGCAGGTGCGTTGGCCAACCTAGCA 1677
PpDILLO1A.gDNA      GTTGAGCATTTTGTTTCCGGATAGCAGGAGAGAGCAGCAGGTGCGTTGGCCAACCTAGCA 2640
                                               ********************************

PpDILLO1A.cDNA      GCCGATGACAAGTGCAGCATGGAAGTCGCAGTGGCGGGGGGGTGTGCGGGCTCTGGTGAGG 1737
PpDILLO1A.gDNA      GCCGATGACAAGTGCAGCATGGAAGTCGCAGTGGCGGGGGGGTGTGCGGGCTCTGGTGAGG 2700
                    ************************************************************

PpDILLO1A.cDNA      CTGGCGCGGTTTTGTAACCACGAAGGCGTGCAAGAGCAGG---------------------- 1777
PpDILLO1A.gDNA      CTGGCGCGGTTTTGTAACCACGAAGGCGTGCAAGAGCAGGTACGAGGCAACGGGAAGCTG 2760
                    ****************************************

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      AGAGCAGCGGTTGGTAAAGGGGGATGGGGTTTGGAAGGCTGAGGTTGGGTCGTGATGAGG 2820


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GGGTAGCGGAGTGAGAAACACACACATGGGGTTTGGAATGAGTGGCAGGGGCGCAGAGTG 2880


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GAAGGAGGGGTTGGAGGTTTGCGAGGTTAAGTGGTGCAGGACATGGTTGAGTTGGCAGAG 2940


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      AAGAAGTGATTGAGACAGTGGCGTTGCGGTGAAGGTGTTGCGTGGGGATGGATTTGGTTA 3000


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GCGTGATTAGCATTGAGGATTAATGAGGGCGTGACGATAGTGAGGGTTGCGCGTAAGTGG 3060


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      AGAGTGCATTGTGTGGGTCCTGGAGGCGTTTGGGGCGTGAGTGGAGTGTTTGCACGATCG 3120


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GTGGAGCCGGTTTTAGGTGCAGCGTTGGCGAAGAGCATGCGGAATGGGAGAGGAGGTGAC 3180


PpDILLO1A.cDNA      ---------------------------------------------CAGCACGCGCGTTGGC 1793
PpDILLO1A.gDNA      GAGTTGCGAGGTGTGACTGGTGTTTTGGAACTGTGTTTGATAGGCAGCACGCGCGTTGGC 3240
                                                                 ***************
```

```
PpDILLO1A.cDNA      GAATCTTGCGGCTCATGGAGACAGCAACGGTAACAACGCTGCTGTAGGGCGGGAGGAAGG 1853
PpDILLO1A.gDNA      GAATCTTGCGGCTCATGGAGACAGCAACGGTAACAACGCTGCTGTAGGGCGGGAGGAAGG 3300
                    ************************************************************

PpDILLO1A.cDNA      CGCGCTGGAAGCTCTGGTGCAGCTGACTTGCTCAAATCATGAAGGTGTCAG--------- 1904
PpDILLO1A.gDNA      CGCGCTGGAAGCTCTGGTGCAGCTGACTTGCTCAAATCATGAAGGTGTCAGGTATGCGGA 3360
                    *********************************************

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      TTTGTTGGGGTGGTGGACAGCAGGGAGGTGTTTTTGTTGTGCAGGGGTGATGTGAAGGGT 3420


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      TGAGGGTAGAGATGGGCCGGTGTTCTAAGAGTGTTTTGGGCGTCCCCTTGGACGGCGCTT 3480


PpDILLO1A.cDNA      ----ACAAGAAGCAGCGGGAGCATTGTGGAATCTGTCGTTCGACGACAGAAACCGAGAAG 1960
PpDILLO1A.gDNA      GCAGACAAGAAGCAGCGGGAGCATTGTGGAATCTGTCGTTCGACGACAGAAACCGAGAAG 3540
                        ********************************************************

PpDILLO1A.cDNA      CTATTGCAGCGGCTGGTGGCGTGGAGGCACTGGTG-------------------------- 1995
PpDILLO1A.gDNA      CTATTGCAGCGGCTGGTGGCGTGGAGGCACTGGTGAGTGAGAGTGGACACGAACCGTGGG 3600
                    **********************************

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      CGTAGGAAGAGATACGAAAGGGGCGTTCTGGGGGCAGAGGTGAATGGGGGTGATTCGAGA 3660


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GCGTGGCCGGGGGGTTGAGCAGAGCTGGAGGTGAAGATGAAGGATGCAGGCTTAGTGCGG 3720


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      ACATGTAGAGTGATGGGGGGCTGCATGACAGTTGAGCGTGGGGTGGTTGTGGAGGGTAGG 3780


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      ATCGGGGTGCGTTGATGTTGAGATCGGGTTGGGGGGCAGTCGTGTGGCGAGTTGTGCGAG 3840


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GTTTTGCGAGGCGTGAACAGGCTGTGTTGAGGCAAGGGCGCACGGTGGTGTGGTGTGGAG 3900


PpDILLO1A.cDNA      ------------------------------------------------------GCTTTGG 2002
PpDILLO1A.gDNA      CAGAGTGGGGGTGGCATTGGGTGCATGACGATGGTGTGGTGGTTGTGCAGGTGGCTTTGG 3960
                                                                           *******

PpDILLO1A.cDNA      CTCAAGGGTGTAGCAGCGGATCTCAAGGTCTGCAAGAGAGGGCTGCGGGTGCATTATGGG 2062
PpDILLO1A.gDNA      CTCAAGGGTGTAGCAGCGGATCTCAAGGTCTGCAAGAGAGGGCTGCGGGTGCATTATGGG 4020
                    ************************************************************

PpDILLO1A.cDNA      GGCTGTCTGTGTCGGAGGCGAACAG----------------------------------- 2087
PpDILLO1A.gDNA      GGCTGTCTGTGTCGGAGGCGAACAGGTGAGGGCATTGAGCGCGATGGGAGGTGGTGTTGT 4080
                    *************************

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      TGGCGGAAGCGTTGGGCGTGGGTTGAAGGTCGCAGGGGTAGAGTCGTCGGGGATGGAGCG 4140


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GATTGTGTGAAGTATTGCGCGGAGTGAGAGTCGATGGGTTGGATTTGGCGATTATGGGGA 4200


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      TTTTGGTTGCGTAGCAGAGGTTGGAGGTGGGTGACGGGGTGTGGTGCGATGACCGTGTGT 4260


PpDILLO1A.cDNA      ----------------------------------------------------------CATTGC 2093
PpDILLO1A.gDNA      CTGACGTGTGTGTGTGTGTGTGTGTGTGTTGGGAACTTGCGTGGTGGCGCAGCATTGC 4320
                                                                             ******
```

```
PpDILLO1A.cDNA      TATCGGGCGCGAAGGCGGAGTGGCTCCACTGATCACCCTGGCGCACTCTGATTTTGAGG- 2152
PpDILLO1A.gDNA      TATCGGGCGCGAAGGCGGAGTGGCTCCACTGATCACCCTGGCGCACTCTGATTTTGAGGT 4380
                    ************************************************************

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GAGTTGTGGGATGAGTTGCAGGGTTTTGGGGGACGAGGTTTTGGGTTTGGTTAGCGCAGC 4440

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GTGCATGTCATGTGAGTGACACGGTTAGACGCGAGTGGTTGGGGTGTGATGGTGGAGGCT 4500

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GGGTGGTGGTGGCATACAAAGTGCAAAGTGCAGGGAGGCGAGAGCTGGATGCATGGGTGA 4560

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GTGGGAGGAATGGGGTGTAGTGGAAATGGAAAGGGGGCGTGGTGTGGAGCCATGGTTGGG 4620

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      AGATGGTTGTACGTGGCGGGGCATTGGTGGTGGTATCGGTGTGGTGGGACGATGATGGGC 4680

PpDILLO1A.cDNA      ---------------------------------------------ACGTACATGAAACAG 2167
PpDILLO1A.gDNA      GTTTAGGAGGTGGGTTTCTGAAGTTTCCGGCGATGGGAGTGCAGGACGTACATGAAACAG 4740
                                                                  ***************

PpDILLO1A.cDNA      CGGTGGGAGCGTTGTGGAACCTGGTTTTCAATCCTGGCAACGCACTGCGGATGGTGGAAG 2227
PpDILLO1A.gDNA      CGGTGGGAGCGTTGTGGAACCTGGTTTTCAATCCTGGCAACGCACTGCGGATGGTGGAAG 4800
                    ************************************************************

PpDILLO1A.cDNA      AGGAGGGAGTCCCGGCGCTGGTGCATTTGTGCTCGTCGTCGAGATCTAAGATGGCGCGGT 2287
PpDILLO1A.gDNA      AGGAGGGAGTCCCGGCGCTGGTGCATTTGTGCTCGTCGTCGAGATCTAAGATGGCGCGGT 4860
                    ************************************************************

PpDILLO1A.cDNA      TTATGGCAGCTCTGGCTCTGGCGTACATGTTCGATGGCAGG---------------------- 2328
PpDILLO1A.gDNA      TTATGGCAGCTCTGGCTCTGGCGTACATGTTCGATGGCAGGTAAAGCGGTGGCGTGTGGG 4920
                    ****************************************

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      TGTGTGAGGGTGCGGTCGTAGGACGTGGGGGTGGTGAGATGACGTGAGACGGGGAGGAGG 4980

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      AAGTTCGTAGGCGTTGGATAGGTGCGGTGAGGGAGAGTCGGGTGAAGTGGAGCAGTGGGG 5040

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      AGTGGGTGACGTGGTTTTGGAGAACCGGAAGATATGGAAGAAGGGCGCGGTTAGGAATGA 5100

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GGAGATTGCTGTTGCGTTGTTGGGACGAGGTTGGGGCGGGGGTGAGAAGTGGAGGGGGTT 5160

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GAGGTGCGTAGGAGCGCGCGTTTGTGTGCATGATTTGAAGTGGTTGGGCGGCGTTTGGAG 5220

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      CGGTTGAGTGGGTGTGGCAAAGCAGTGAGTGGGAGTGATGGAGAGTGAACGTGATCGGAG 5280

PpDILLO1A.cDNA      ----------------------------------ATGGATGAGGTTGCAGTTGGATTGT 2353
PpDILLO1A.gDNA      TGGTTGACGTCATAGTTGGGCTCCTCTGACGCAGGATGGATGAGGTTGCAGTTGGATTGT 5340
                                                     *************************

PpDILLO1A.cDNA      CGAGTGTGGAGAACAACGGGAGGACAGTGAGCTTGGAGGCAGCAAGGAAACTGGCGTTGC 2413
PpDILLO1A.gDNA      CGAGTGTGGAGAACAACGGGAGGACAGTGAGCTTGGAGGCAGCAAGGAAACTGGCGTTGC 5400
                    ************************************************************
```

```
PpDILLO1A.cDNA      GAAACATCGATGCGTTCGTGCAAACGTTTTTTGATCCGCAGTCTTTGACAGCTGCTGCGT 2473
PpDILLO1A.gDNA      GAAACATCGATGCGTTCGTGCAAACGTTTTTTGATCCGCAGTCTTTGACAGCTGCTGCGT 5460
                    ************************************************************

PpDILLO1A.cDNA      CATCGTGGGCCGGTGCGACGTTGAACCAGGTCGCGGAGACCGCCACAATCCAAGAAGCCG 2533
PpDILLO1A.gDNA      CATCGTGGGCCGGTGCGACGTTGAACCAGGTCGCGGAGACCGCCACAATCCAAGAAGCCG 5520
                    ************************************************************

PpDILLO1A.cDNA      GGCATCTGAGATGCAG---------------------------------------------- 2549
PpDILLO1A.gDNA      GGCATCTGAGATGCAGGTAGGCGATGCGTACGTAGGGCGTGAATGTGTGGATGGTACAGA 5580
                    ****************

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      ATTAGATGCAACGGATTTGGATGCATGGGCAGCGGAAGCAACGTGGGCGCGTGGTAAGGA 5640


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      ATGGACGCGAGTGGGTGTGGGCATGACAGGCGGTCGGGGTGATGGGACAGTGGCGGTGAT 5700


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GAGGTAGCGGATAGTGAAGGGCGGTGTGGAGTGCGGTGCGTGGAGTGTTGTTTGTATGGA 5760


PpDILLO1A.cDNA      ----------------------------------TGTAGCCGAGCTCGGGCGGTTTGTGG 2575
PpDILLO1A.gDNA      CGAGGCTAAGAGCGTGCGGGTGGGTTGGTTGCAGTGTAGCCGAGCTCGGGCGGTTTGTGG 5820
                                                     **************************

PpDILLO1A.cDNA      CCATGTTGAGGAACGGACATGCAGTTTTGCGATCGTGTGCGGCATTCGCTCTTCTGCAG- 2634
PpDILLO1A.gDNA      CCATGTTGAGGAACGGACATGCAGTTTTGCGATCGTGTGCGGCATTCGCTCTTCTGCAGG 5880
                    ************************************************************

PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      TAAGCAGTGTGAGTGAGCATGAAGAGGACAAAGCGGTGGTGAGGGACGCGAGGTGGGTGT 5940


PpDILLO1A.cDNA      ------------------------------------------------------------
PpDILLO1A.gDNA      GATGGTAGGGTTGTTATTGATTCGCCGGTGGAGTGATGTGCGTGAGAGAGAGGAAGTGTT 6000


PpDILLO1A.cDNA      ------------------------------------------------------------T 2635
PpDILLO1A.gDNA      GGAGCGAGGTTGGAGGTTGGTTTTGGCGTGATGGATTGGGATGTGGTTGCGTGTCACAGT 6060
                                                                               *

PpDILLO1A.cDNA      TCACGATGCCTGGAGGCAGGCATGCGCTGCACCACGCAGTTCTGCTGCAGAAGGTTGGTG 2695
PpDILLO1A.gDNA      TCACGATGCCTGGAGGCAGGCATGCGCTGCACCACGCAGTTCTGCTGCAGAAGGTTGGTG 6120
                    ************************************************************

PpDILLO1A.cDNA      CGAGTCGGGTTCTGCGGACTGTGGCTGCAGCAGCGACGGCAGCAATGCAGGCCAAAGTGT 2755
PpDILLO1A.gDNA      CGAGTCGGGTTCTGCGGACTGTGGCTGCAGCAGCGACGGCAGCAATGCAGGCCAAAGTGT 6180
                    ************************************************************

PpDILLO1A.cDNA      TCGCCCGAGTGGTTCTTCGAAACCTGGAGCATCATCAGGCGGAGAGCAACACGGTGGTGG 2815
PpDILLO1A.gDNA      TCGCCCGAGTGGTTCTTCGAAACCTGGAGCATCATCAGGCGGAGAGCAACACGGTGGTGG 6240
                    ************************************************************

PpDILLO1A.cDNA      CACTGTCGTAA 2826
PpDILLO1A.gDNA      CACTGTCGTAA 6251
                    ***********
```

```
PpDILLO2.gDNA      ATGTCCAACAAGCGGCGGCGGAGCGTCAACGTGGCGGTGGAGGAGCAGGAGCAGCAGCCG 60
PpDILLO2.cDNA      ATGTCCAACAAGCGGCGGCGGAGCGTCAACGTGGCGGTGGAGGAGCAGGAGCAGCAGCCG 60
                   ************************************************************

PpDILLO2.gDNA      GTTTTGTTCAAGAAGGCCAGGATAACACCGTCGTCGTCGTCGTCCGCTTGTAGTGCCCCG 120
PpDILLO2.cDNA      GTTTTGTTCAAGAAGGCCAGGATAACACCGTCGTCGTCGTCGTCCGCTTGTAGTGCCCCG 120
                   ************************************************************

PpDILLO2.gDNA      GCAGTGTCTGTTGCGGAGTCGGGCGTGGGAATTGCTGGGGATCGCGATGCGCATTGGACG 180
PpDILLO2.cDNA      GCAGTGTCTGTTGCGGAGTCGGGCGTGGGAATTGCTGGGGATCGCGATGCGCATTGGACG 180
                   ************************************************************

PpDILLO2.gDNA      GAGCTGCCGGATGATACTGTGTTCGGGTTGTTCAATTTGCTGAACTATAGGGATAGGGCG 240
PpDILLO2.cDNA      GAGCTGCCGGATGATACTGTGTTCGGGTTGTTCAATTTGCTGAACTATAGGGATAGGGCG 240
                   ************************************************************

PpDILLO2.gDNA      AGTCTCGCGTCCGTGTGCAGGGCATGGCGCGGGTTGGGGAGCTCGACGTCGTTGTGGACG 300
PpDILLO2.cDNA      AGTCTCGCGTCCGTGTGCAGGGCATGGCGCGGGTTGGGGAGCTCGACGTCGTTGTGGACG 300
                   ************************************************************

PpDILLO2.gDNA      TCGTTGGATCTTCGCTTGCACAGGTTGGATTCGGAGGTCGTGTCGGCGCTTGCGGGCCGG 360
PpDILLO2.cDNA      TCGTTGGATCTTCGCTTGCACAGGTTGGATTCGGAGGTCGTGTCGGCGCTTGCGGGCCGG 360
                   ************************************************************

PpDILLO2.gDNA      TGCTCGAATTTGCAGTGTCTGAAATTTCGCGGAGGCGCGTTCGCGAACTCAATTGTGGGT 420
PpDILLO2.cDNA      TGCTCGAATTTGCAGTGTCTGAAATTTCGCGGAGGCGCGTTCGCGAACTCAATTGTGGGT 420
                   ************************************************************

PpDILLO2.gDNA      TTGCATGCGAGGGAGCTGAGGGAGCTGAGCGGTGATTGGTGCAGCCTTCTGTCGGATGCA 480
PpDILLO2.cDNA      TTGCATGCGAGGGAGCTGAGGGAGCTGAGCGGTGATTGGTGCAGCCTTCTGTCGGATGCA 480
                   ************************************************************

PpDILLO2.gDNA      ACCCTGTCCATGGTGGTTGCTCGGCATGGGAACCTGGAGAGCTTGCAGCTGGGTTCGGAT 540
PpDILLO2.cDNA      ACCCTGTCCATGGTGGTTGCTCGGCATGGGAACCTGGAGAGCTTGCAGCTGGGTTCGGAT 540
                   ************************************************************

PpDILLO2.gDNA      TGCGAGAGGGTGACGAGCGAGGCGCTGAAAGTGGTAGCAGTTTGTTGTCCTAAATTGCGA 600
PpDILLO2.cDNA      TGCGAGAGGGTGACGAGCGAGGCGCTGAAAGTGGTAGCAGTTTGTTGTCCTAAATTGCGA 600
                   ************************************************************

PpDILLO2.gDNA      CGGTTGTGCATCTCAGGCATTCGGGATGTGGACCGAGACGCCATTCAGGCAATGTTCCAG 660
PpDILLO2.cDNA      CGGTTGTGCATCTCAGGCATTCGGGATGTGGACCGAGACGCCATTCAGGCAATGTTCCAG 660
                   ************************************************************

PpDILLO2.gDNA      CATTGCCAGGGTCTGACAGAGCTGGGGTTTCTGGACAGTCACACCATCGACGAGGGAGCT 720
PpDILLO2.cDNA      CATTGCCAGGGTCTGACAGAGCTGGGGTTTCTGGACAGTCACACCATCGACGAGGGAGCT 720
                   ************************************************************

PpDILLO2.gDNA      TTCGGGGCTGCGAGGAACCTACGATTTTTGTCCGTGGCGGGATGCAGGTGCATAGCGTGG 780
PpDILLO2.cDNA      TTCGGGGCTGCGAGGAACCTACGATTTTTGTCCGTGGCGGGATGCAGGTGCATAGCGTGG 780
                   ************************************************************

PpDILLO2.gDNA      AGCACGGCAGCGCAGTGCTGGAGCAAATTGCCCAACTTGAGTGGGCTGGATGTTTCTCGG 840
PpDILLO2.cDNA      AGCACGGCAGCGCAGTGCTGGAGCAAATTGCCCAACTTGAGTGGGCTGGATGTTTCTCGG 840
                   ************************************************************

PpDILLO2.gDNA      ACGGATATAACTCCGAGCACACTGGCGCAGGTGCTTGCATGTCCCGAACTGAAGGTGGTG 900
PpDILLO2.cDNA      ACGGATATAACTCCGAGCACACTGGCGCAGGTGCTTGCATGTCCCGAACTGAAGGTGGTG 900
                   ************************************************************

PpDILLO2.gDNA      TGCGGCCTGAACTGCCCCCTTCTGGAGGAGGGTAGCAACCCGATCCCGTCGCCATCGTCA 960
PpDILLO2.cDNA      TGCGGCCTGAACTGCCCCCTTCTGGAGGAGGGTAGCAACCCGATCCCGTCGCCATCGTCA 960
                   ************************************************************

PpDILLO2.gDNA      AAGACGGCGGTGTTGCTGGCACGGTTCACAGATCTTATGGAAGGATTGGAGGTTTTGCTG 1020
PpDILLO2.cDNA      AAGACGGCGGTGTTGCTGGCACGGTTCACAGATCTTATGGAAGGATTGGAGGTTTTGCTG 1020
                   ************************************************************

PpDILLO2.gDNA      AACCCGAGCAATTCGAAGGAGGAGGCGTGTTCGTCTGGACGATCCAGGTATGGCAGGCGC 1080
PpDILLO2.cDNA      AACCCGAGCAATTCGAAGGAGGAGGCGTGTTCGTCTGGACGATCCAGGTATGGCAGGCGC 1080
                   ************************************************************
```

```
PpDILLO2.gDNA      ATGCGTGTGGACCCGGAAGTGGCGAAGTGGACGGAGAGAATGTTGTCACATGCACTTTTG 1140
PpDILLO2.cDNA      ATGCGTGTGGACCCGGAAGTGGCGAAGTGGACGGAGAGAATGTTGTCACATGCACTTTTG 1140
                   ************************************************************

PpDILLO2.gDNA      AAGATTGCGGAGAGCAATGCGCCCAGCCTGGATAGCTTCTGGCTGAAGCAGGGAACCGCC 1200
PpDILLO2.cDNA      AAGATTGCGGAGAGCAATGCGCCCAGCCTGGATAGCTTCTGGCTGAAGCAGGGAACCGCC 1200
                   ************************************************************

PpDILLO2.gDNA      ATGATGCTTCGACTGGTCCAAAGCGCACAGGAGGATGTACAGGAAAGGGCTGCGGCTGCT 1260
PpDILLO2.cDNA      ATGATGCTTCGACTGGTCCAAAGCGCACAGGAGGATGTACAGGAAAGGGCTGCGGCTGCT 1260
                   ************************************************************

PpDILLO2.gDNA      TTGGCGGTGTTTGTTTTGGTGGATGATGAGAATGCGACTGTGGATTCTGCACGGGCGGAA 1320
PpDILLO2.cDNA      TTGGCGGTGTTTGTTTTGGTGGATGATGAGAATGCGACTGTGGATTCTGCACGGGCGGAA 1320
                   ************************************************************

PpDILLO2.gDNA      GCTGTGATGAACGGTGGCGGCATTGCGTTGCTGCTAGGTCTGGCAAAGTCTTGTGGCGAG 1380
PpDILLO2.cDNA      GCTGTGATGAACGGTGGCGGCATTGCGTTGCTGCTAGGTCTGGCAAAGTCTTGTGGCGAG 1380
                   ************************************************************

PpDILLO2.gDNA      GGAGTGCAATCCGAAGCCGCCAAGGTAGTTCTAGTTTCTGAGGAACGAGTCGGAGTGCGA 1440
PpDILLO2.cDNA      GGAGTGCAATCCGAAGCCGCCAAGG---------------------------------- 1405
                   *************************

PpDILLO2.gDNA      GTGTGCACGGGACAGTTTTTGTGGAGAGAGGGTAGTGTTTACTAGGTGGTTGGCAGGCGG 1500
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      GCAGGCGGAAGCATTGCATTTCTTTCCGGCGTTAGCTGAAGTAGGGTTGAATGGCAACTG 1560
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      GTTCACGATGATGGCGGTTTTTCGAAGTTGACCCATGTTCAATTAAACCTCCCTCGAGTG 1620
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      TAGGCGTGGTTTGCGTGGGTTACTCAGTATAGTGAATTACGACTAATGGCACACATTCCG 1680
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      TGTGGCTGTTGCAGGCCATTGCAAATTTATCTGTAAATACAGAAGTTGCGAAGAGGGTTG 1740
PpDILLO2.cDNA      ---------------CCATTGCAAATTTATCTGTAAATACAGAAGTTGCGAAGAGGGTTG 1450
                                  ********************************************


PpDILLO2.gDNA      CTTTGGAAGGGGGAATCAGTATCCTGGCTGCGTTAGCGAGGTCTCCGAATCGGTGGGTTG 1800
PpDILLO2.cDNA      CTTTGGAAGGGGGAATCAGTATCCTGGCTGCGTTAGCGAGGTCTCCGAATCGGTGGGTTG 1510
                   ************************************************************

PpDILLO2.gDNA      CCGAGGAGGCAGCGGGGGGTTTATGGAACCTGTCGGTCGGGGAAGAGCATAAGGTGGGGT 1860
PpDILLO2.cDNA      CCGAGGAGGCAGCGGGGGGTTTATGGAACCTGTCGGTCGGGGAAGAGCATAAGG------ 1564
                   ********************************************************

PpDILLO2.gDNA      GCCATGCTTTCCAACAATGATGAGGCGTTTGCGAATTGCGGATTTTGTGTGTGGGGGTCG 1920
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      GGATGCTGTTGCTCAAGGAACTGCACATTCGATGATGGAATTTTTACGGAGTCTCGTCAT 1980
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      ACGGATTCTGTTGTGATAGTCAATGTTACCGAAAGACCTGTGCAGACAGATTGTGGTGGT 2040
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      GCAAGTTACAACAGTGCATCTTGTGAACTGTTTGCAGTTACTGTGCACGATAGGTTGTTT 2100
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      TTGTGATTGGCCGGGAGTTAAGATCATAGTGAGCATGGTTTTGAAATAGTTGGTTGGTTG 2160
PpDILLO2.cDNA      ------------------------------------------------------------
```

```
PpDILLO2.gDNA      GCTGTCTTGGTCACATTTTCGCACTTACAAATACGACCGTTGATGCTAATAGGGTGCTAT 2220
PpDILLO2.cDNA      ------------------------------------------------------GTGCTAT 1571
                                                                        *******

PpDILLO2.gDNA      AGCCGAGGCAGGCGCAATTGAAGCACTGGTGGATCTGGCATTTAAGTGGCCTGCTGGAGG 2280
PpDILLO2.cDNA      AGCCGAGGCAGGCGCAATTGAAGCACTGGTGGATCTGGCATTTAAGTGGCCTGCTGGAGG 1631
                   ************************************************************


PpDILLO2.gDNA      GGAGGGTGTTCTGGTGAGCACGTACATTCCTCACCAGTTTGGGATTGTTATGAGCCGAGC 2340
PpDILLO2.cDNA      GGAGGGTGTTCTGG-------------------------------------------------- 1645
                   **************


PpDILLO2.gDNA      TATAGTTGTTTTTAAATCTGACATGTTGACGATGTTTGCATTTCCGAATTGGTGTTGCGG 2400
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      GTTCTCAATCTGTTGCGATGATTGATTTTGTCGGATGTGATGAAGAAGTGACGCTTGTTT 2460
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      AGGACCCTTACCCTTGCACAGATGTGTGTAGGCTGTGCACGGTGTAGTTTTGATCAGAAG 2520
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      CGCATGGGATTGCTTGGATCCTTCGAACAGGGTTCAAGAACAGTGGACACTAACGCTGAG 2580
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      CACTTTGTATGGATGGTAGGAGCGAGCAGCAGGTGCTCTGGCTAACCTAGCAGCTGATGA 2640
PpDILLO2.cDNA      --------------------AGCGAGCAGCAGGTGCTCTGGCTAACCTAGCAGCTGATGA 1685
                                       ************************************


PpDILLO2.gDNA      CAAGTGCAGCATGGAAGTTGCTGTGGCTGGCGGTGTACGGGCCCTAGTTAGACTTGCACA 2700
PpDILLO2.cDNA      CAAGTGCAGCATGGAAGTTGCTGTGGCTGGCGGTGTACGGGCCCTAGTTAGACTTGCACA 1745
                   ************************************************************


PpDILLO2.gDNA      GTTTTGTAATCACGAAGGGGTGCAAGAGCAGGTGCGGGGCAATCAGACTCTGAGGTCGCG 2760
PpDILLO2.cDNA      GTTTTGTAATCACGAAGGGGTGCAAGAGCAGG----------------------------- 1777
                   ********************************


PpDILLO2.gDNA      GTGTTGGAATCGGAGTTTGGTTTTCGTGGATGAGGTGCACGTCCAGAGGTTGTATGGGCT 2820
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      GTAGTTTTTGGGAGAGCTTGTGGTGGTTCTCAGTACAGGTCTTGAAATCGTGATTGTTTG 2880
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      CAGGCAGCGCGCGCCTTGGCCAATCTCGCCACTCATGGAGACAGCAACGGTAATAATGCT 2940
PpDILLO2.cDNA      ----CAGCGCGCGCCTTGGCCAATCTCGCCACTCATGGAGACAGCAACGGTAATAATGCT 1833
                       ********************************************************


PpDILLO2.gDNA      GCTGTGGGGCGGGAGGCAGGCGCACTTGAAGCTCTCGTGCGCCTGACTGGGTCTAACCAT 3000
PpDILLO2.cDNA      GCTGTGGGGCGGGAGGCAGGCGCACTTGAAGCTCTCGTGCGCCTGACTGGGTCTAACCAT 1893
                   ************************************************************


PpDILLO2.gDNA      GAAGGCGTTAGGTACGTGTGATCTCTGAGGCGAGTGTGCGGCATTGAAATTGGTTGGTCA 3060
PpDILLO2.cDNA      GAAGGCGTTAG--------------------------------------------------- 1904
                   ***********


PpDILLO2.gDNA      ATGAGATTGGATAAGGTCCAAACTCGGTGACTTACCTGACAGTATTTTGGTATCGGCTTT 3120
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      CACATGTGCAGACAAGAAGCAGCTGGAGCACTGTGGAATCTGTCTTTTGATGATAGAAAC 3180
PpDILLO2.cDNA      -----------ACAAGAAGCAGCTGGAGCACTGTGGAATCTGTCTTTTGATGATAGAAAC 1953
                              ************************************************


PpDILLO2.gDNA      CGAGAAGCTATTGCAGCAGCTGGTGGCGTGGAGGCACTGGTGAGTCAGAGTTGACACGAG 3240
PpDILLO2.cDNA      CGAGAAGCTATTGCAGCAGCTGGTGGCGTGGAGGCACTGGT------------------- 1994
                   ****************************************
```

```
PpDILLO2.gDNA      TCTTGACTGCAGCAAGAGATGGAGGAGTCAGTGTTGTCGTTCGGTTCCGTGGTGGGTCAC 3300
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      GTGACAGTGCAGTACAAGGGTGCAGTGGAACAGGAGTTGAAAATGAAAAATTGCTCCCGT 3360
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      AATGCAGACGCTAAATTAGTTGTGGTGGTGAATGACCCCTGGAAATCGTCTGGTTGTGAT 3420
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      GTGTACCATTGGGATGTGGTCATGATTATATCAGGTTGAGGACGACATGGGGATGACTTT 3480
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      CACCTTTTTTCAGGTTGGGAGTAAATTGCACATGTCGTTGTCAGAGGTAGACTACCCACA 3540
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      ATACTGTGGGGATCCAGCAGAGGAATGTTGGCACAGAATGCATGACGAACCTGTGGTGGT 3600
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      TGTGTAGGTTGCTTTAGCTCAAGATTGTAGCAGCGGATCTCAAGGTCTGCAGGAGAGGGC 3660
PpDILLO2.cDNA      ---------TGCTTTAGCTCAAGATTGTAGCAGCGGATCTCAAGGTCTGCAGGAGAGGGC 2045
                           **************************************************

PpDILLO2.gDNA      TGCTGGTGCATTGTGGGGTCTATCTGTTTCTGAAGCCAACAGGTGAGAACGTTTTTCGTG 3720
PpDILLO2.cDNA      TGCTGGTGCATTGTGGGGTCTATCTGTTTCTGAAGCCAACAG------------------- 2087
                   ****************************************

PpDILLO2.gDNA      ATGGGAAGCTGTGTTCTTCGCGGAAGCCTTGGGTGTTTGTTGAAGGGTGCAGGTCTTCAT 3780
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      CTGTGGGTGCTGAAGTCGATGATATGTACTTGTGTGGAGGTTGGGGTTTTATGGATCGGT 3840
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      TTGCAGCCTGCGGAGATGATATCACTGAAGTGCGTGCTGTGAATGTGTGTTTTGGCTCAG 3900
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      CATTGCTATCGGTCGTGAAGGTGGAGTTGCCCCGCTGATCACTCTGGCACACTCGAACTC 3960
PpDILLO2.cDNA      CATTGCTATCGGTCGTGAAGGTGGAGTTGCCCCGCTGATCACTCTGGCACACTCGAACTC 2147
                   ************************************************************

PpDILLO2.gDNA      TGAGGTGAGGTGCTGGATTTGGCGCATGGTTTTGCAGCATGAGGCATGTGATATTCTGGT 4020
PpDILLO2.cDNA      TGAGG------------------------------------------------------- 2152
                   *****

PpDILLO2.gDNA      CAGCGGAGTATATATGGTAGCTTGTAAGTCTCAGCAGTTGGAGGTTGTGGATGAAACTGA 4080
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      TTACGAATATGTCGAGCTCAAGGGCTGGTGCTTCGTTGGGATGCTTTCTAGTATGCGGAG 4140
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      GTCAGAACATACATTAAACAGTATGTAGGGCAGATGGTGTGCTGTTGAGGTGTGAGGAAT 4200
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      AGCTGTGCATGGAGCTGTTCGTGAAATAGCTTTTTGTACCATGGTGCAGTCTGGACGGTG 4260
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      TTGGCCCGTTGGCTCAGCAATGAATGCGTATTGAGTGTTGGTTTGGAATGGGATTGACGC 4320
PpDILLO2.cDNA      ------------------------------------------------------------
```

```
PpDILLO2.gDNA      AATTGCAGGACGTACATGAGACAGCGGTGGGCGCGTTGTGGAATCTGGCTTTCAATCCCG 4380
PpDILLO2.cDNA      ---------ACGTACATGAGACAGCGGTGGGCGCGTTGTGGAATCTGGCTTTCAATCCCG 2203
                           **************************************************

PpDILLO2.gDNA      GTAACGCGTTGCGCATGGCGGAAGAAGGAGTCCCTGCGCTTGTACATTTGTGCTCGTCGT 4440
PpDILLO2.cDNA      GTAACGCGTTGCGCATGGCGGAAGAAGGAGTCCCTGCGCTTGTACATTTGTGCTCGTCGT 2263
                   ************************************************************

PpDILLO2.gDNA      CAAGATCTAAAATGGCACGTTTTATGGCAGCTCTGGCCCTCGCGTACATGTTCGATGGCA 4500
PpDILLO2.cDNA      CAAGATCTAAAATGGCACGTTTTATGGCAGCTCTGGCCCTCGCGTACATGTTCGATGGCA 2323
                   ************************************************************

PpDILLO2.gDNA      GGTATAGCACTCATATCTGCCTGTATGAGCTAGATGTTGCCGAACTTGGTGGAATTGTGT 4560
PpDILLO2.cDNA      GG------------------------------------------------------------ 2325
                   **

PpDILLO2.gDNA      GGAATGGGTGGGTGGAAGTCCGCGGAAAGCGTCTTGATGTGGTCAGGGTGTGTGCGGTGT 4620
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      GGTGGAGCAGTGTAGTCGAGTGTTTTCGTTTGGGAAATAACTAGGTCCGGTGATGGTGAC 4680
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      GGTGCCCAGCAATGGACGATGAAGTTTTGTTGCATAACAGGAGCTCATTCGGATATGTTT 4740
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      CGATAGTTGAGGTAGTTAATTAGGCGTTTTGTGCAGCGGTGCACGACTTGTTCGAAAGTC 4800
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      CTGAGGCGGCCTGTGTAGCGATAGAGTTGTCCGGCAGAGGAGTGTGGGTAAATGCTGCGA 4860
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      GATAGACGTGTTGTAGACTAGCGTCACATTTGGGCTCCGCTGGTGCAGGATGGATGAGGT 4920
PpDILLO2.cDNA      ---------------------------------------------------ATGGATGAGGT 2336
                                                                     ***********

PpDILLO2.gDNA      TGCTGTCAGAGTATCGAGCGGAGAGAACCATGGCAGGACAGTGAACCTGGAAGCGATTCG 4980
PpDILLO2.cDNA      TGCTGTCAGAGTATCGAGCGGAGAGAACCATGGCAGGACAGTGAACCTGGAAGCGATTCG 2396
                   ************************************************************

PpDILLO2.gDNA      GAAATTGGCGTTGCGAAGCATTGATGCGTTTGTGCTAACTTTCTGTGATCAACAAGCTCT 5040
PpDILLO2.cDNA      GAAATTGGCGTTGCGAAGCATTGATGCGTTTGTGCTAACTTTCTGTGATCAACAAGCTCT 2456
                   ************************************************************

PpDILLO2.gDNA      GACAGCTGCTGCGTCATCGTGGGCTCCTGCCACCCTGAACCAAGTTGCGGAGACCGCCAG 5100
PpDILLO2.cDNA      GACAGCTGCTGCGTCATCGTGGGCTCCTGCCACCCTGAACCAAGTTGCGGAGACCGCCAG 2516
                   ************************************************************

PpDILLO2.gDNA      GATCCAAGAAGCTGGACTTCTCAGATGCAGGTAAGGGGTGTGTGTGTGGGGATGGGAAGA 5160
PpDILLO2.cDNA      GATCCAAGAAGCTGGACTTCTCAGATGCAG------------------------------ 2546
                   ****************************

PpDILLO2.gDNA      TGCACATCCATTGAAGTCGGAGCAATGAGATTGAGTGGGATAGCTGTTTTCTTTTGCTGA 5220
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      GTGGAAGCAATTTAAGCCCGTCTTGTATGACATGGGCATTGGCAAGTTTTGAGTATTGGG 5280
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      TTCATGTAGTTGATAGGGAAGTTGAGGCTTGGACAGTGGGTAACGACAGCAGTGGTTCCC 5340
PpDILLO2.cDNA      ------------------------------------------------------------

PpDILLO2.gDNA      AATGCAATTGGACGTGGACAAAATTTGAGAGTTCGGCGTGGTTGGTTGCAGTGGAGCGGA 5400
PpDILLO2.cDNA      ---------------------------------------------------------TGGAGCGGA 2555
                                                                           *********
```

```
PpDILLO2.gDNA      GCTTGGACGGTTTGTGGCCATGTTGAGGAATGGTTCTGCAGTTCTCCGAACATGTGCTGC 5460
PpDILLO2.cDNA      GCTTGGACGGTTTGTGGCCATGTTGAGGAATGGTTCTGCAGTTCTCCGAACATGTGCTGC 2615
                   ************************************************************

PpDILLO2.gDNA      GTTTGCTCTTCTGCAGGTAATCACTGGTGCTGATCGACAGGACAGAGTTGACGGAGCGAT 5520
PpDILLO2.cDNA      GTTTGCTCTTCTGCAG---------------------------------------------- 2631
                   ***************

PpDILLO2.gDNA      GGGGTGGTTGTATCATTTTTAGGTTCGTTTTGTTACGGGGAGTGTTGGACTGAGGTTTAA 5580
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      GGTACCTTTCGACTTAGGTTTACTGACCCGTAGGGTGTGGGTGTGTGGAACGGTTTACTG 5640
PpDILLO2.cDNA      ------------------------------------------------------------


PpDILLO2.gDNA      ACCCGTAGGGTGTGGGTGTGTGGAACAGTTTACTATGCCTGGAGGTAGGCATGCGATGCA 5700
PpDILLO2.cDNA      ----------------------------TTTACTATGCCTGGAGGTAGGCATGCGATGCA 2663
                                               ********************************

PpDILLO2.gDNA      CCATGCAGATCTTCTGCAGAAGACCGGTGCAAGTCGAGTGCTGCGGACTGTGGCTGCAGC 5760
PpDILLO2.cDNA      CCATGCAGATCTTCTGCAGAAGACCGGTGCAAGTCGAGTGCTGCGGACTGTGGCTGCAGC 2723
                   ************************************************************

PpDILLO2.gDNA      AGCAACTGCGGCAATGCAGGCTAAAGTGTTTGCCCGAATCGTTCTTCGTAACCTCGAGCA 5820
PpDILLO2.cDNA      AGCAACTGCGGCAATGCAGGCTAAAGTGTTTGCCCGAATCGTTCTTCGTAACCTCGAGCA 2783
                   ************************************************************

PpDILLO2.gDNA      TCATCAAGCGGAGAGCAACGCAGTGGTGGCACTATTATAA 5860
PpDILLO2.cDNA      TCATCAAGCGGAGAGCAACGCAGTGGTGGCACTATTATAA 2823
                   ****************************************
```

8.4 *PHYSCODILLO* gene promoter sequence alignment

```
                    PHYSCODILLO1A.promoter.F
P1A_promoter        ACGGTGACAAGTGCCGGACGAAACAAGAGGACGAACACTTTCGATGGTCAAGGAGTAAAC 60
P1B.promoter        ------------------------------------------------------------
P2_promoter         ----------------------------------------TTCACAAGCTGAGATGTTCGGA 22
                                                           PHYSCODILLO2.promoter.F

P1A_promoter        TTCTCAGTTCCTCACCGACCCAATTCTAGCCCCCACTCAACCCCACAGCTTCTCCGGAAT 120
P1B.promoter        ------------------------------------------------------------
P2_promoter         GCAAATGGAGATGAGCTCAACTGATGTGGCCCATTGTGAATGTCAAATCATGTCGAGTAG 82
                                                         PHYSCODILLO1B.promoter.F

P1A_promoter        TCCAGGGACGGAAAAGTCGAGTGGTCAAGATCAGTTGCAGATGCATTTAAAAGCCGCCCA 180
P1B.promoter        ------------------------------TACGGCGAAAAAGTTCCTG------GCGCA 24
P2_promoter         TTACTGGGCACCTATCACTACATCCAAAGCCTCGAAAGGATCACATC--------ATTCA 134
                                                   *                      **

P1A_promoter        AGCAGAGTACCAGTATTTTCAGAAATACGTTCAAGAACATAAGTCCTTTGATAAGTTGGT 240
P1B.promoter        GGTACAACGTGAAAAAGTAGGAGAGTGCAATTCACTACAGAGAACTTTC-------CGAG 77
P2_promoter         GATGTC-TGTCTAAACACATGTAAATGCTTCCGGTGTAATTTGTTCTTTTTTGA---AAA 190
                          *        * * *          *       **

P1A_promoter        GGAAATTCATCACGCAAATGCACAACAATCCGTTATCGACTTCCAATCCATACCTATTCC 300
P1B.promoter        CGAAGTCCATCA---AATTTTGGGTTCTGGCACTGCCTCATCGCGACGAACGACGGT-CC 133
P2_promoter         GAGAATTGATGT---ACCCTCAAACTGGCATAGTACGTTGTTCAAATACAAGCTTGGTCC 247
                       * *  **     *                  *      *    *   *        **

P1A_promoter        ACCTTGACAAAAGAAAAAAAAAAAAA----AAAAAAAAAAAAGATAAAAAAAAAGGAAAA 356
P1B.promoter        AACATGGTCAACGAAACCGAGAGAAC----AACTTTCTTGTTCGTCATCAGAGCTTCATT 189
P2_promoter         AACGCAGTGCAGTGCGACGTACATACTTTTAATGGTTTCTATTATAGAAAAAACGACATT 307
                   * *       *              *     **            *    * *     *

P1A_promoter        AAAAGCACTACACGATGTAAAAGAA--------AAATTAAATAAACAGTAACTCCAA-AT 407
P1B.promoter        TCCGTTCTTACCAGACGCCGTTCAA--------GGGCGAAGTCAACTCCAACAACAACAA 241
P2_promoter         TTGTTGATTTTTTGTGGTGAAAGAAATGGTTTTGGACCTAATCAAGATTACTGGTGAAAA 367
                           *    *  *      **              * * **    *      * *

P1A_promoter        TGGCTCGAAAAATCCGAAACCATCTCGAAACAATCCAGTGAATTTTAAATCAAAAGTAAT 467
P1B.promoter        CAGCTTCCCTGAGATAAATGCCTTCTAAAAGAGAGTACT--ATTACCAGCTTGTGACAGC 299
P2_promoter         GAGCTGGCTTGG---AAGTGTCATTTGAAAGCATGTG----ATTTGAAGCCGAGATTAGT 420
                      ***          *          ***          ***   *        *

P1A_promoter        GCAGG-ATTAACTAAAGAGTTGACTACAGAACAAATAGCACTTAATCCAT--AGAATACA 524
P1B.promoter        CGAGA-AC--ATTAAGAATTATACTAC-------------TTTAATTTCT--CGAATTTA 341
P2_promoter         TTTGACATGTTATTGTCAGTGTATGGATAGCAAAGTGATGTCTATTTCCTGTTAATTCTA 480
                       *  *     *    * *  *                    ** *   *    * *  *

P1A_promoter        CATTTTCATAAAATACAGAACTTG-ATACAAAACTACAATTTAACTTACCATGGACAGAT 583
P1B.promoter        --TCAAAATAAGTCCAACAACATATGCATAAAAATA-AATATAA---ATCATG--CATAT 393
P2_promoter         AACTATGATATATCAACTAATGTTTAC-TATACTAATAGTGTAA--TGGTATG---AAAC 534
                           ***        **  *      * *   * * * ***      ***   * *

                                                        PHYSCODILLO1A/1B.promoter.F
P1A_promoter        GCAGAAAAACGACAAGAGTCATCCCACAGGGAACATGTTAA---ACATCCCTCTCGAGTG 640
P1B.promoter        GCATAAAAATAA--------ATCCCACAGAGAACATGTTAA---ACATCCCTCTCGAGTG 442
P2_promoter         ACTTAGGTAGAGTAA----CTTTTTACTTTTAAATTTTGAACCTACATTGCTTCAAAACG 590
                     *  *   *             *   **    **  * * **   ****  **    *  *

P1A_promoter        TCGCAACTACTGAGAATTACGAACACGACTCC----GTGCATCTCAGCCACGCATGTATT 696
P1B.promoter        TCGCAACTACTGAGAATTACGAACACGACTCC----GTGCATCTCAGCCACGCATGTATT 498
P2_promoter         TTTAAACTTCACGTTATACAGAATATGGTTTCAATTGTAGATGTTATAAATACTTTCACT 650
                   *   **** *     **   *** * *  * *    **  ** * *   *  * *  * *

P1A_promoter        CTTATACAGTTCAGTGCAATTTTCAGTTCCTCCCTGCCCGTACTCTCGGACTTGCACTCG 756
P1B.promoter        CTTATACAGTTCAGTGCAATTTTCAGTTCCTCCCTGCCCGTACTCTCGGACTTGCACTCG 558
P2_promoter         TTAACAGTATACAATTTCAATGTATATTCTCTATGACGGACATCTTTTTATCTGAAAAAA 710
                     * * *   * ** *   * * *   ***       *    *   *   *  ** *

P1A_promoter        TCAAATCCTTTGTTCCCATCAAAAGATACGAACCCATATCGTGGGCACAACGTGCACAAA 816
P1B.promoter        TCAAATCCTTTGTTCCCATCAAAAGATACGAACCCATATCGTGGGCACAACGTGCACAAA 618
P2_promoter         ACAGAATATGGAATGAAAATTTAATAACAAAAACAACACCAAAGACCAAAAATGTACATG 770
                     ** *   *    *   *    ** *    ** * * * *   * *  **  ** ***
```

```
P1A_promoter      ---ACAGAGACAATTTACTGAAAAACCACCGATGTAAACCAAAACGCACA----AAGAGC 869
P1B.promoter      ---ACAGAGACAATTTACTGAAAAACCACCGATGTAAACCAAAACGCACA----AAGAGC 671
P2_promoter       TGTAAATGTATAAATATATATATATATATATATATATATTCAAATAAATTCTTTGAAAGC 830
                     * *   * ** *   *  * *   *   ** ** *   ***   *       * ***

P1A_promoter      AGAGGTG----GAAGGGGGTGAAGCAAACGATTGCAACACACAAACACAGACGAGTTCTG 925
P1B.promoter      AGAGGTG----GAAGGGGGTGAAGCAAACGATTGCAACACACAAACACAGACGAGTTCTG 727
P2_promoter       AAAAGTTTTTTGTGTAGAGAATCATAAAGAGTTATCTCATGACTTAACTTTCAAATAAAA 890
                  * * **     *    * *      ***   **    **       **   * * *

P1A_promoter      ATTGGCTTGTGAGCATGAGGTGGACAAGGCATGCACACAG--ATTTGGAAGGGAAGCAAA 983
P1B.promoter      ATTGGCTTGTGAGCATGAGGTGGACAAGGCATGCACACAG--ATTTGGAAGGGAAGCAAA 785
P2_promoter       ATAATGTCTTTTACATGTGTGTGCTTAAACATATACATTGGTATCTTAAATCCAAAATCA 950
                  **    *  *   **** *   *   *  ***  ***  *  ** *  **   **    *

P1A_promoter      TGGGGTAATCAGCCTTGAGGGGTGGTGGGTGTGTTGAAGTGGGCA-CAGATTCCATATTC 1042
P1B.promoter      TGGGGTAATCAGCCTTGAGGGGTGGTGGGTGTGTTGAAGTGGGCA-CAGATTCCATATTC 844
P2_promoter       TAAAAGTACCGCATATATATATATATATATATATTATCAAACAATCAAACCCATGTTT 1010
                  *      * *     *          *   * * *   *     **    *  **** **

P1A_promoter      CGCGGTGGTGGCATTGATTCCAGATTCCAGACGGGGCA-----AAGAAAGGAAATAAACA 1097
P1B.promoter      CGCGGTGGTGGCATTGATTCCAGATTCCAGACGGGGCA-----AAGAAAGGAAATAAACA 899
P2_promoter       TAAA--AGTTGCACAAACCTCCATACACACACAAACAATTTTTGTCAATAGCTATAACCT 1068
                        ** ***   *   *      ** **     *        **  *  **** *

P1A_promoter      GAAGAAAGGGA-AAGGAAAGAAAAGAAAAGAAAAAGAAGAAGAGAAATATAAAGGGCGCC 1156
P1B.promoter      GAAGAAAGGGA-AAGGAAAGAAAAGAAAAGAAAAAGAAGAAGAGAAATATAAAGGGCGCC 958
P2_promoter       GCATCATGTGACAAAGTGACAATTGAATATTTCAACGCCCAAAGCTGCATACATGAGAAT 1128
                  * *  * * ** ** *  * **  *** *    **     * **    *** * *

P1A_promoter      ------AATTATAGGGTGCCTTTCTGGGCGAGAAGGGCAGGCGGTCAT-CGGTCAACGC- 1208
P1B.promoter      ------AATTATAGGGTGCCTTTCTGGGCGAGAAGGGCAGGCGGTCAT-CGGTCAACGC- 1010
P2_promoter       TCGAAAAAGTGTCCAAAATCAATGTGACCGTAATAAACATGAAATCTTGCCCCCAACGTT 1188
                        ** * *       *  * **  **  *    ** *   ** * *   *****

P1A_promoter      --AGGGGAACACAAAAACAGCGAGCAGCGAGGAAGAGGCGACGAAGGCGACACAGGGATG 1266
P1B.promoter      --AGGGGAACACAAAAACAGCGAGCAGCGAGGAAGAGGCGACGAAGGCGACACAGGGATG 1068
P2_promoter       CAACTGTCTTTCAAAAAGTAAGACAACTTTAAAAATCGCAACCGTGTTACTTGTGAATAG 1248
                    *  *     ******    **  *      **   ** **   *        *    *

P1A_promoter      ACGACGACGACGAGAAGGAGGACGCAGAAGAAG--AAGAGCGGCGGGGGGGGGGTGGTCGT 1324
P1B.promoter      ACGACGACGACGAGAAGGAGGACGCAGAAGAAG--AAGAGCGGCGGGGGGGGGGTGGTCGT 1126
P2_promoter       ACAAGGATGCCGCAGAGTCCGTCATCCATCTATCCAGATCCATCCATCATTACACCTCAT 1308
                  ** * ** * **   **   * *    *   *   *    *  *             ** *

P1A_promoter      GGACTCCCTAAGGAAGGAAGGAAGAGGAAGCTGCCACATGAAAACCAAAGCTTCTTTCCT 1384
P1B.promoter      GGACTCCCTAAGGAAGGAAGGAAGAGGAAGCTGCCACATGAAAACCAAAGCTTCTTTCCT 1186
P2_promoter       TCGTCCATCCAGTAACACCCAAACATCATCCATATATTCGTTCATCACACATCCATAACA 1368
                       *    ** **      ** *  *  *    *   *   * ** *  * * *  *

P1A_promoter      CCT--CCATGAAGAATGGAACAGACTCTCTCTGCAGGAAAAACTCAACCCCACTCGCTCG 1442
P1B.promoter      CCT--CCATGAAGAATGGAACAGACTCTCTCTGCAGGAAAAACTCAACCCCACTCGCTCG 1244
P2_promoter       TGTGTTCATCAAAACTTCATCAATCCCATATTTCATCGAAAGTCTGGTTTCATTCGTGAA 1428
                    *   *** ** * *  * **  * *    * **   ***         ** ***

P1A_promoter      CTCCCTCACTCACTCACTCACTCACTCTGTCTG--CTGTTTCTCATTCGTGTCCCA 1500
P1B.promoter      CTCCCTCACTCACTCACTCACTCACTCTGTCTG--CTGTTTCTCATTCGTGTCCCA 1302
P2_promoter       CAA---AAATTGCCCGGACGGGCAGGGAGACTAACGGAGCTGCCGCAAGGTCCCGGTCAA 1485
                  *      * *  * *   *   **   *  **  * *  ***   *    **  *  * *

P1A_promoter      CGCTCTCTTGCTCTTGTTCTCGCTCTCGCTCTCACTCGTCGTCGCTACGCCTGCCCTTCC 1560
P1B.promoter      CGCTCTCTTGCTCTTGTTCTCGCTCTCGCTCTCACTCGTCGTCGCTACGCCTGCCCTTCC 1362
P2_promoter       CGCTGGAACAAGGCGG--CCCGGGCTGACCCACCCACGCCACGCCTAGTCCAGCCCGGCC 1543
                  ****           *  * **  **  * * * * ** *    ***  ** ****  **

P1A_promoter      CCCTCCCCCTCCCGTGTCTATCCCGCAATCCTTAGCGCTTCCTCCCTTCTCGCCACACCC 1620
P1B.promoter      CCCTCCCCCTCCCGTGTCTATCCCGCAATCCTTAGCGCTTCCTCCCTTCTCGCCACACCC 1422
P2_promoter       CGGCCCGGCCCAGGTGGGT---TGGCACTGCGCTG-GCTGGCTGGCTGCGCAG-GCAGTA 1598
                  *   **  * *  ***  *     *** * *   * ***  **  ** * *    **

P1A_promoter      CCTCTTTTTCTTTCTCTTTCTCTCTCTACTTCCTGCTTCCGCTCTTCCGTTTCTTCTCAT 1680
P1B.promoter      CCTCTTTTTCTTTCTCTTTCTCTCTCTACTTCCTGCTTCCGCTCTTCCGTTTCTTCTCAT 1482
P2_promoter       TGTAGTTCTCTTTCTCCTCTCCACTGGTGCTTCTTCTTCTTCACCTCCTCCTCATATCCT 1658
                    *  ** ******** *   * **     * ** ****  * * ***   ** * ** *
```

```
P1A_promoter      CTGCTCTCTTCCCTCTTCGATGCC-ATGCTG-CTTGCCCGCGTCTTTTCGGTTGGAAGTG 1738
P1B.promoter      CTGCTCTCTTCCCTCTTCGATGCC-ATGCTG-CTTGCCCGCGTCTTTTCGGTTGGAAGTG 1540
P2_promoter       TTGCCTTCCTCT-TGTGAGCTGTTTGTGTAGACAGGGTTGCGATTGCTCTGCTAG--GCG 1715
                   ***  ** **  * *  * **    **  * *  *   ***  *  ** * * *  * *

P1A_promoter      AAGAATGATGCCGCAGTTTGGGGTTTGAGGCTGGTTTCGGTGGTGGCTGCTTAGTATTCA 1798
P1B.promoter      AAGAATGATGCCGCAGTTTGGGGTTTGAGGCTGGTTTCGGTGGTGGCTGCTTAGTATTCA 1600
P2_promoter       GTGTTTGCTGATGCTGTCCGTGA--CGGAATTGGGGTTTGTATGAGCCGTAACGTATACG 1773
                    *  ** **  ** **  * *    *    ***  *  **    ** *    **** *

P1A_promoter      AGGATGATTG--TGGGTGATGC----CGC-CAGGGGATGGTTTCTTT-CGAGGAAGAGGG 1850
P1B.promoter      AGGATGATTG--TGGGTGATGC----CGC-CAGGGGATGGTTTCTTT-CGAGGAAGAGGG 1652
P2_promoter       CTGCCGGTTGCTTGGGTGATGAAAAACGCGCAGAGCCTGGTTACGTTGCGGGGGCGCAGT 1833
                    *  * ***  *********     *** *** *  ***** * ** ** **  *  *

P1A_promoter      G--AGGTGTAGATCGGTGCAGGGTTATGCAGCAAGAATATA--TTCGATTT---GGGCCG 1903
P1B.promoter      G--AGGTGTAGATCGGTGCAGGGTTATGCAGCAAGAATATA--TTCGATTT---GGGCCG 1705
P2_promoter       GGCAGCAGGGGATGCGTAGAGGAGCAGACATTGAGTGCAAGGGTATGGTTCCGAAGGAAG 1893
                 *  **  *  ***  **  ***   *  **   **   *    *  * **     **  *

P1A_promoter      CGTGATGCAATGCTTGAGGATAGGTAGAACACGAAGCGGGGGATGTGACCTTTGATTGCG 1963
P1B.promoter      CGTGATGCAATGCTTGAGGATAGGTAGAACACGAAGCGGGGGATGTGACCTTTGATTGCG 1765
P2_promoter       GACGTTGTTAATTTTGAGGGTGGGAGGGAGATGAGG-GAGAAAGGAGGAGGGGAATTATG 1952
                     * **  *   ****** * **  * * * ** * * *  * * *       ***  *

P1A_promoter      TGCAG-GAGCGGGTGACGTTCTGATGGGCGAGGACGGAGGAATTGCAGAGTGCGCGGAGT 2022
P1B.promoter      TGCAG-GAGCGGGTGACGTTCTGATGGGCGAGGACGGAGGAATTGCAGAGTGCGCGGAGT 1824
P2_promoter       ATTAGCGCATGCGTGGTGTTTCGTGGTGTGGGAATAGCTAA----CAACGT-CGCGGAAT 2007
                     ** *   * ***  ***  *  * * * * *  *   *    **  ** ****** *

P1A_promoter      GGAAGGAATAATCGGCGCGGATATTGGGAGTTGGCAACCCTAAGAGAGGTTGAAGCGGCA 2082
P1B.promoter      GGAAGGAATAATCGGCGCGGATATTGGGAGTTGGCAACCCTAAGAGAGGTTGAAGCGGCA 1884
P2_promoter       GGGATCGA-----GTCCTGGGGAGAGGGAATGGGTGA-----ATAGGTGCAGGAG-GGTG 2056
                 ** *   *     * *  **  *  **** * **  *     * **  *  * ** **

P1A_promoter      ATTTGTGTGGGTGATGACGAAGCAGTGACGGC-ACGAAGATGGTGCATGCGCGGTGCGCA 2141
P1B.promoter      ATTTGTGTGGGTGATGACGAAGCAGTGACGGC-ACGAAGATGGTGCATGCGCGGTGCGCA 1943
P2_promoter       AAGAGTGTGATGGTTGGATAACAAGGGATCGCGACCAAGGTGTTGCGTTCGCGTTGCGCT 2116
                 *   *****   * **   **  ** **  ** ** *** ** *** * **** *****

                                  PHYSCODILLO1A/1B.promoter.R
P1A_promoter      TGCAAATGACACAG---ATCATCGGAGGAGCAGCTGATGACATG 2182
P1B.promoter      TGCAAATGACACAG---ATCATCGGAGGAGCAGCTGATGACATG 1984
P2_promoter       AGATAATGAAGCGGGCGATTGTGGGAGGAGGTGATGATGATATG 2160
                   *  *****  * *   **  * *******  * ****** ***
                                  PHYSCODILLO2.promoter.R
```

8.5 *SELAGIDILLO* genomic DNA: sequenced version vs. theoretical

```
SELg.seq2       ATGCGTCGCGTTCGCCGGAAATGCGTCCACACTCTCGCGACCAAGAGCTCCGCGGCAGCG 60
SELg.seq1       ATGCGTCGCGTTCGCCGGAAATGCGTCCACACTCTCGCGACCAAGAGCTCCGCGGCAGCG 60
SELg.Sc26       ATGCGTCGCGTTCGCCGGAAATGCGTCCACACTCTCGCGACCAAGAGCTCCGCGGCAGCG 60
SELg.Sc24       ATGCGTCGCGTTCGCCGGAAATGCGTCCACACTCTCGCGACTAAGAGCTCCGCGGGAGCG 60
                ************************************** ************ ****

SELg.seq2       GAGAATGGCGATGGAGTCGCCGAGGAGGAGGAGTCGAGAATCCCCAAGCACGACGGCCAG 120
SELg.seq1       GAGAATGGCGATGGAGTCGCCGAGGAGGAGGAGTCGAGAATCCCCAAGCACGACGGCCAG 120
SELg.Sc26       GAGAATGGCGATGGAATCGCCGAGGAGGAGGAGTCGAGAATCCCCAAGCACGACGGCCAG 120
SELg.Sc24       GAGAATGGCGATGGAATCGCCGAGGAGGAGGAGTCGAGAATCCCCAAGCACGACGGCCAG 120
                *************** ********************************************

SELg.seq2       GTACTAGTCCGGTGTGAGAGGGAGAGCGGCGTGGATTGGACGCGCCTGGCCGACGATACG 180
SELg.seq1       GTACTAGTCCGGTGTGAGAGGGAGAGCGGCGTGGATTGGACGCGCCTGGCCGACGATACG 180
SELg.Sc26       GTAATAGTCCGGTGTGAGAGGGAGAGCGGCGTGGATTGGACGCGCCTGGCCGACGATACG 180
SELg.Sc24       GTACTAGTCCGGTGTGAGAGGGAGAGCGGCGTGGATTGGACGCGCCTGGCCGACGATACG 180
                *** ********************************************************

SELg.seq2       CTGCTGGGATTGTTTGCGCTGCTCAATTACCGCGACCGGGCCAGCGTTGGATCCGTGTGC 240
SELg.seq1       CTGCTGGGATTGTTTGCGCTGCTCAATTACCGCGACCGGGCCAGCGTTGGATCCGTGTGC 240
SELg.Sc26       CTGCTGGGATTGTTTGCGCTGCTCAATTACCGCGACCGGGCCAGCGTTGGATCCGTGTGC 240
SELg.Sc24       CTGCTGGGATTGTTTTCGCTGCTCAATTACCGCGACCGGGCCAGCGTTGGATCCGTGTGC 240
                *************** ********************************************

SELg.seq2       CGGGCATGGCACGCGCTCTCCTCCTCGCCGTCACTATGGACCAGCCTGGATTTGCGAGCT 300
SELg.seq1       CGGGCATGGCACGCGCTCTCCTCCTCGCCGTCACTATGGACCAGCCTGGATTTGCGAGCT 300
SELg.Sc26       CGGGCATGGCACGCGCTCTCCTCCTCGCCGTCACTATGGACCAGCCTGGATTTGCGAGCT 300
SELg.Sc24       CGGGCATGGCACGCGCTCTCCTCCTCGCCGTCACTGTGGACCAGCCTGGATTTGCGAGCT 300
                ********************************** ************************

SELg.seq2       CACACTCTCGACTCCAACATGGCGTCGGCGCTTGCTTCCAGATGCGCGAAGCTCTCCAAG 360
SELg.seq1       CACACTCTCGACTCCAACATGGCGTCGGCGCTTGCTTCCAGATGCGCGAAGCTCTCCAAG 360
SELg.Sc26       CACACTCTCGATTCCAACATGGCGTCGTCGCTTGCTTCCAGATGCGCGAAGCTCTCCAAG 360
SELg.Sc24       CACACTCTCGATTCCAACATGGCGTCGTCGCTTGCTTCCAGATGCGCGAAGCTCTCCAAG 360
                *********** *************** ********************************

SELg.seq2       CTAAAGTTCCGTGGGGCCTCAGGTGCTAGCTTGATCATCGACTTGCAAGCCAGGCAGCTC 420
SELg.seq1       CTAAAGTTCCGTGGGGCCTCAGGTGCTAGCTTGATCATCGACTTGCAAGCCAGGCAGCTC 420
SELg.Sc26       CTAAAGTTCCGTGGGGCCTCAGGTGCTAGCTTGATCATCGACTTGCAAGCCAGGCAGCTC 420
SELg.Sc24       CTAAAGTTCCGTGGGGCCTCAGGTGCTAGCTTGATCATCGACTTGCAAGCGAGGCAGCTC 420
                ************************************************* *********

SELg.seq2       AAGGGGCTCATAGGTGACGGGTGCAAGGATCTCACCGACGCGACGCTGTCGATGCTCGTC 480
SELg.seq1       AAGGGGCTCATAGGTGACGGGTGCAAGGATCTCACCGACGCGACGCTGTCGATGCTCGTC 480
SELg.Sc26       AAGGGGCTCATAGGTGACGGGTGCAAGGATCTCACCGACGCGACGCTGTCGATGCTCGTC 480
SELg.Sc24       AAGGGGCTCATAGGTGACGGGTGCAAGGATCTCACCGACGCGACGCTGTCGATGCTCGTC 480
                ************************************************************

SELg.seq2       GCCCGCCACGAGAATCTGGAAAGCCTCCAGCTTGGGCCGGAGCTGGAGAAGATCACGAAC 540
SELg.seq1       GCCCGCCACGAGAATCTGGAAAGCCTCCAGCTTGGGCCGGAGCTGGAGAAGATCACGAAC 540
SELg.Sc26       GCCCGCCACGAGAATCTGGAAAGCCTCCAGCTTGGGCCGGAGCTGGAGAAGATCACGAAC 540
SELg.Sc24       GCCCGCCATGAGAATCTGGAAAGCCTCCAGCTTGGGCCGGAGCTGGAGAAGATCACGAAC 540
                ******** ****************************************************

SELg.seq2       GAGGCGATCAAGGTGGTAGCCGTGTGCTGCCGGAGACTCAAGTGCCTGCGGTTGGCGGGG 600
SELg.seq1       GAGGCGATCAAGGTGGTAGCCGTGTGCTGCCGGAGACTCAAGTGCCTGCGGTTGGCGGGG 600
SELg.Sc26       GAGGCGATCAAGGTGGTAGCCGTGTGCTGCCGGAGACTCAAGTGCCTGCGGCTGGCGGGG 600
SELg.Sc24       GAGGCGATCAAGGTGGTAGCCGTGTGCTGCCGGAGACTCAAGTGCCTGCGGCTGGCGGGG 600
                ************************************************** ********

SELg.seq2       ATCCGGGACGTGGACAGTGAAGCTATCGGGGACTTGGTGAAGCACTGCCCGAGCTTGACG 660
SELg.seq1       ATCCGGGACGTGGACAGTGAAGCTATCGGGGACTTGGTGAAGCACTGCCCGAGCTTGACG 660
SELg.Sc26       ATCCGGGACGTGGACAGTGAAGCTATCGGGGACTTGGTGAAGCACTGCCCGAGCTTGACG 660
SELg.Sc24       ATCCGGGACGTGGACAGTGAAGCTATCGGGGACTTGGTGAAGCACTGCCCGAGCTTGACG 660
                ************************************************************

SELg.seq2       GAGCTGGCTTTGCTCGACTGCGCGGTGGTGGACGAGGCTGCGCTCGGCGAGGCAAAGTCT 720
SELg.seq1       GAGCTGGCTTTGCTCGACTGCGCGGTGGTGGACGAGGCTGCGCTCGGCGAGGCAAAGTCT 720
SELg.Sc26       GAGCTGGCTTTGCTCGACTGCGCGGTGGTGGACGAGGCTGCGCTCGGCGAGGCAAAGTCT 720
SELg.Sc24       GAGCTGGCTTTGCTCGACTGCGCGGTGGTGGACGAGGCTGCGCTCGGCGAGGCAAAGTCT 720
                ************************************************************
```

```
SELg.seq2       CTGAGGTACTTGTCGGTGGCCGGGTCCAGGAACATAATGTGGACACAGGCCATGCAAGCA 780
SELg.seq1       CTGAGGTACTTGTCGGTGGCCGGGTCCAGGAACATAATGTGGACACAGGCCATGCAAGCA 780
SELg.Sc26       CTGAGGTACTTGTCGGTGGCCGGGTCCAGGAACATAATGTGGACACAGGCCATGCAAGCA 780
SELg.Sc24       CTGAGGTACTTGTCGGTGGCCGGGTCCAGGAACATAATGTGGACACAGGCCATGCAAGCA 780
                ************************************************************

SELg.seq2       TGGAGTAAGCTGGAGAACTTGGTAGCACTTGATGTGTCCAGGACTGAGGTAACACCCGCG 840
SELg.seq1       TGGAGTAAGCTGGAGAACTTGGTAGCACTTGATGTGTCCAGGACTGAGGTAACACCCGCG 840
SELg.Sc26       TGGAGTAAGCTGGAGAACTTGGTAGCACTTGATGTGTCCAGGACTGAGGTAACACCCGCG 840
SELg.Sc24       TGGAGTAAGCTGGAGAACTTGGTAGCACTTGATGTGTCCAGGACTGAGGTAACACCCGCG 840
                ************************************************************

SELg.seq2       GCTGTCATGTCTTTTCTTTCTGCTCCTCGTTTAAGGGTTCTTTGTGCGCTCAGCTGCTCG 900
SELg.seq1       GCTGTCATGTCTTTTCTTTCTGCTCCTCGTTTAAGGGTTCTTTGTGCGCTCAGCTGCTCG 900
SELg.Sc26       GCTGTCATGTCTTTTCTTTCTGCTCCTCGTTTAAGGGTTCTTTGTGCGCTCAGCTGCTCG 900
SELg.Sc24       GCTGTCATGTCTTTTCTTTCTGCTCCTCGTTTAAGGGTTCTTTGTGCGCTCAGCTGCTCG 900
                ************************************************************

SELg.seq2       GCCCTGGAAGATGGAAGCAACTCTGTGAGCTACGTTAGTAAGGACAGGGTTTTGCTGGCT 960
SELg.seq1       GCCCTGGAAGATGGAAGCAACTCTGTGAGCTACGTTAGTAAGGACAGGGTTTTGCTGGCT 960
SELg.Sc26       GCCCTGGAAGATGGAAGCAACTCTGTGAGCTACGTTAGTAAGGACAGGGTTTTGCTGGCT 960
SELg.Sc24       GCCCTGGAAGATGGAAGCAACTCTGTGAGCTACGTTAGTAAGGACAGGGTTTTGCTGGCT 960
                ************************************************************

SELg.seq2       CGGTTCACAGAGTTGATGAACGGGCTGGCGTGCATTTCGAGTTTAGAGCAGCAGGACGAG 1020
SELg.seq1       CGGTTCACAGAGTTGATGAACGGGCTGGCGTGCATTTCGAGTTTAGAGCAGCAGGACGAG 1020
SELg.Sc26       CGCTTCACAGAGTTGATGAACGGGCTGGCGTGCATTTCGAGCGTAGAGCAGCAGGACGAG 1020
SELg.Sc24       CGGTTCACAGAGTTGATGAACGGGCTGGCATGCATTTCGAGCTTAGAGCAGCAGGACGAG 1020
                ** ************************** ************  ******************

SELg.seq2       AGCAGGGTGCTTGTTTGCTGGACGGAATGGGTGCTCTCCCACGCACTGCTGAGGATAGCG 1080
SELg.seq1       AGCAGGGTGCTTGTTTGCTGGACGGAATGGGTGCTCTCCCACGCACTGCTGAGGATAGCG 1080
SELg.Sc26       AGCAGGGTGCTTGTTTGCTGGACGGAATGGGTGCTCTCGCACGCACTGCTGAGGATAGCG 1080
SELg.Sc24       AGCAGGGTGCTTGTTTGCTGGACGGAATGGGTGCTCTCCCACGCACTGCTGAGGATAGCA 1080
                ************************************ ****** **************

SELg.seq2       GAGAATAACACTCAGGGGCTTGACGCTTTCTGGCTGAAGCAGGGGGACTTCGGTAATGCTA 1140
SELg.seq1       GAGAATAACACTCAGGGGCTTGACGCTTTCTGGCTGAAGCAGGGGACTTCGGTAATGCTA 1140
SELg.Sc26       GAGAATAACACTCAGGGGCTTGACGCTTTCTGGCTGAAGCAGGGAACTTCGGTAATGCTA 1140
SELg.Sc24       GAGAATAACACTCAGGGGCTCGACGCTTTCTGGCTGAAGCAGGGGACTTCGGTAATGCTA 1140
                ******************* ********************** ***************

SELg.seq2       AGGCTGATCAAGAGCATGCAAGAAGACGTGCAGGAGAGAGCAGCCACAGCTTTAGCGACT 1200
SELg.seq1       AGGCTGATCAAGAGCATGCAAGAAGACGTGCAGGAGAGAGCAGCCACAGCTTTAGCGACT 1200
SELg.Sc26       AGGCTGATCAAGAGCATGCAAGAAGACGTGCAGGAGAGAGCGGCCACAGCTTTAGCGACT 1200
SELg.Sc24       AGGCTGATCAAGAGCATGCAAGAAGACGTGCAGGAGAGAGCGGCCACAGCTTTAGCGACT 1200
                ************************************** ******************

SELg.seq2       TTCGTAGTGGTTGACGATGAGAATGCTACTGTTGATTCTTCGAGGGCGGAGGCTGTTATG 1260
SELg.seq1       TTCGTAGTGGTTGACGATGAGAATGCTACTGTTGATTCTTCGAGGGCGGAGGCTGTTATG 1260
SELg.Sc26       TTCGTAGTGGTTGACGATGAGAATGCTACTGTTGATTCTTCGAGGGCGGAGGCTGTTATG 1260
SELg.Sc24       TTCGTAGTGGTTGACGATGAGAATGCTACTGTTGATTCTTCGAGGGCGGAGGCTGTTATG 1260
                ************************************************************

SELg.seq2       CACGGTGGAGGAATTCGCTCTCTTCTGGACCTTGCCAGGTCTTCTCGGGAAGGGGTGCAG 1320
SELg.seq1       CACGGTGGAGGAATTCGCTCTCTTCTGGACCTTGCCAGGTCTTCTCGGGAAGGGGTGCAG 1320
SELg.Sc26       CACGGTGGAGGAATTCGCTCTCTTCTGGACCTTGCCAGGTCTTCTCGGGAAGGGGTGCAG 1320
SELg.Sc24       CACGGTGGAGGAATTCGCTCTCTTCTGGACCTTGCCAGGTCTTCTCGGGAAGGGGTGCAG 1320
                ************************************************************

SELg.seq2       TCCGAAGCAGCGAAGGTGAGTTTCTCTTGAGTTGGATTCTTTTGAGCTTCCGTTTACTCT 1380
SELg.seq1       TCCGAAGCAGCGAAGGTGAGTTTCTCTTGAGTTGGATTCTTTTGAGCTTCCGTTTACTCT 1380
SELg.Sc26       TCCGAAGCAGCGAAGGTGAGTTTCTCTTGAGTTGGATTCTTTTGAGCTTCCGTTTACTCT 1380
SELg.Sc24       TCCGAAGCAGCGAAGGTGAGTTTCTCTTGAGTTGGATTCTTTTGAGCTTCCGTTTACTCT 1380
                ************************************************************

SELg.seq2       TCTCGCTTATAGGCAATTGCAAACTTGTCCGTAAATGCCGAAGTTGCCAAGGCCGTGGCA 1440
SELg.seq1       TCTCGCTTATAGGCAATTGCAAACTTGTCCGTAAATGCCGAAGTTGCCAAGGCCGTGGCA 1440
SELg.Sc26       TCTCGCTTACAGGCAATTGCAAACTTGTCCGTAAATGCCGAAGTTGCCAAGGCCGTGGCA 1440
SELg.Sc24       TCTCGCTTATAGGCAATTGCAAACTTGTCCGTAAATGCCGAAGTTGCCAAGGCCGTGGCA 1440
                ********* ************************************************
```

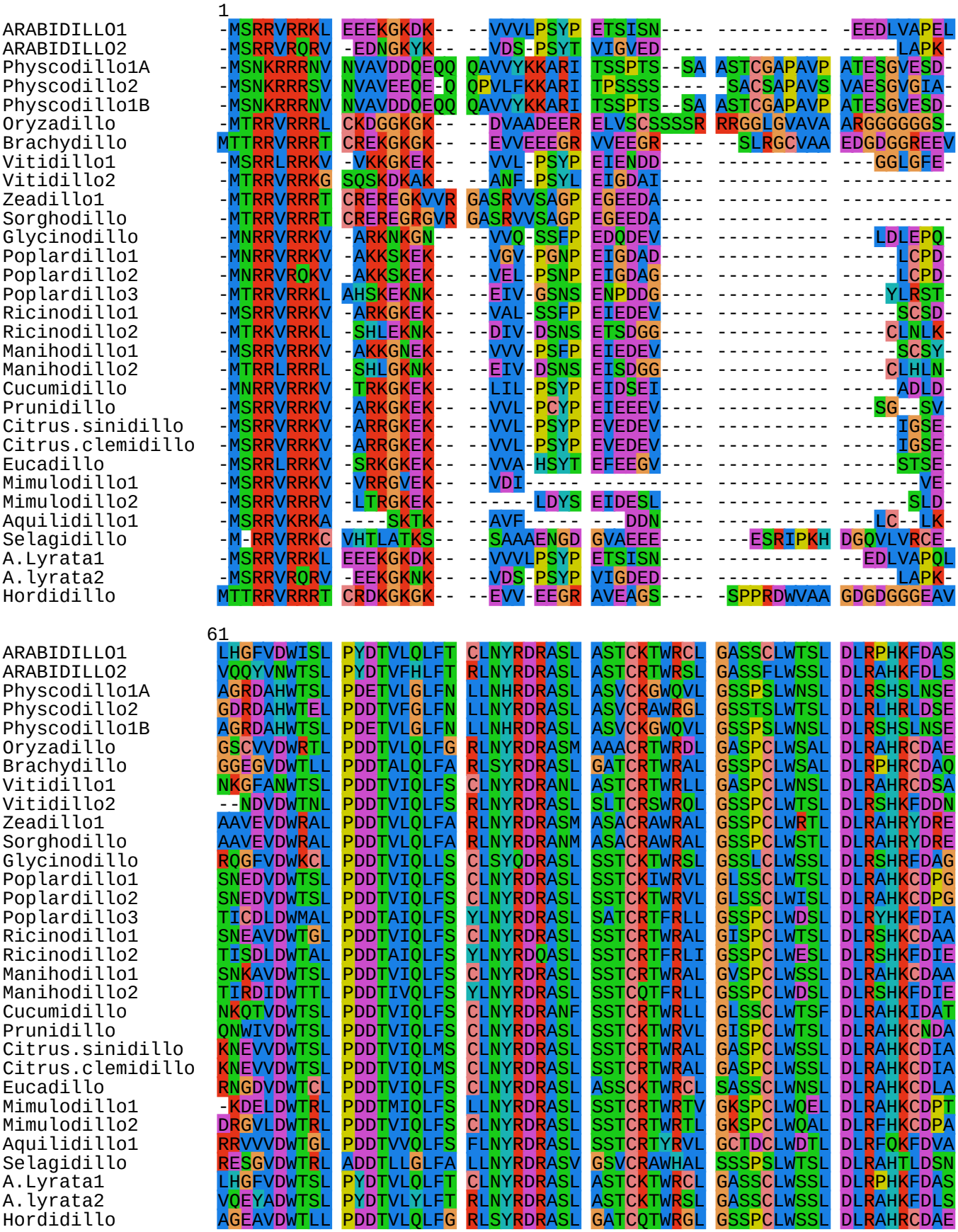
```
SELg.seq2       ACCGAGGGTGGAATCAACATCCTTGCAGGCTTGGCCAGATCTCCCAATCGTTGGGTTGCA 1500
SELg.seq1       ACCGAGGGTGGAATCAACATCCTTGCAGGCTTGGCCAGATCTCCCAATCGTTGGGTTGCA 1500
SELg.Sc26       ACCGAGGGTGGAATCAACATCCTTGCAGGCCTAGCCAGATCTCCCAATCGTTGGGTTGCA 1500
SELg.Sc24       ACCGAGGGTGGAATCAACATCCTTGCAGGCTTGGCCAGATCTCCCAATCGTTGGGTTGCA 1500
                ****************************** * **************************

SELg.seq2       GAAGAGGCAGCTGGAGGACTATGGAACCTTTCAGTTGGCGAAGAGCACAAGGTTTGTTTT 1560
SELg.seq1       GAAGAGGCAGCTGGAGGACTATGGAACCTTTCAGTTGGCGAAGAGCACAAGGTTTGTTTT 1560
SELg.Sc26       GAAGAGGCAGCTGGAGGACTCTGGAACCTTTCAGTTGGCGAAGAGCACAAGGTTTGTTTT 1560
SELg.Sc24       GAAGAGGCAGCTGGAGGACTCTGGAACCTTTCAGTTGGCGAAGAGCACAAGGTTTGTTTT 1560
                ******************** ****************************************

SELg.seq2       TTTCTGATACTTCCAGTATGACTATAGTCGAGCTTTTCAGGGCGCTATCGCTGACGCTGG 1620
SELg.seq1       TTTCTGATACTTCCAGTATGACTATAGTCGAGCTTTTCAGGGCGCTATCGCTGACGCTGG 1620
SELg.Sc26       TTTCTGATACTTCCAGTATGACTATAGTCGAGCTTTTCAGGGCGCTATCGCTGACGCTGG 1620
SELg.Sc24       TTTCTGATACTTCCAGTATGACTATGGTCGAGCTTTTCAGGGTGCTATCGCTGACGCTGG 1620
                ************************* *************** *****************

SELg.seq2       TGCGATTGAGGCTCTCGTCGACTTGGCCTTGAAGTGGCCCGCCGGTGGCGAAGGAGTTTT 1680
SELg.seq1       TGCGATTGAGGCTCTCGTCGACTTGGCCTTGAAGTGGCCCGCCGGTGGCGAAGGAGTTTT 1680
SELg.Sc26       TGCGATTGAGGCTCTCGTCGACTTGGCCTTGAAGTGGCCCGCCGGTGGCGAAGGAGTTTT 1680
SELg.Sc24       TGCGATTGAGGCTCTCGTCGACTTGGCCTTGAAGTGGCCCGCCGGTGGCGAAGGAGTTTT 1680
                ************************************************************

SELg.seq2       GGTACGGAGACGTGCTTCTCGTTCCAATCATATGTTGAGCTCGACATCTTTATGTTCAGG 1740
SELg.seq1       GGTACGGAGACGTGCTTCTCGTTCCAATCATATGTTGAGCTCGACATCTTTATGTTCAGG 1740
SELg.Sc26       GGTACGGAGACGTGCTTCTCGTTCCGATCTTATGTTGAGCTCGACATCTTTATGTTCAGG 1740
SELg.Sc24       GGTACGGAGACGTGCTTCTCGTTCCGATCTTATCTTGAGCTCGACATCTTAATGTTCAGG 1740
                ************************* *** *** *************** *********

SELg.seq2       AGCGAGCAGCCGGCGCTCTTGCCAATCTAGCTGCCGACGATAAATGCAGCATGAAGGTTG 1800
SELg.seq1       AGCGAGCAGCCGGCGCTCTTGCCAATCTAGCTGCCGACGATAAATGCAGCATGAAGGTTG 1800
SELg.Sc26       AGCGAGCAGCCGGCGCTCTTGCCAATCTAGCTGCCGACGATAAATGCAGCATGAAGGTTG 1800
SELg.Sc24       AGCGAGCAGCCGGCGCTCTTGCCAATCTAGCTGCCGACGATAAATGCAGCATGAAGGTTG 1800
                ************************************************************

SELg.seq2       CAAATGCTGGGGGTGTGAACGCTCTCGTCAATCTGGCTCGCTTTTGCAAGCATGAAGGTG 1860
SELg.seq1       CAAATGCTGGGGGTGTGAACGCTCTCGTCAATCTGGCTCGCTTTTGCAAGCATGAAGGTG 1860
SELg.Sc26       CAAATGCTGGGGGTGTGAACGCTCTCGTCAATTTGGCTCGCTTTTGCAAGCATGAAGGTG 1860
SELg.Sc24       CAAATGCTGGGGGTGTGAACGCTCTCGTCAATTTGGCTCGCTTTTGCAAGCATGAAGGTG 1860
                ******************************** ****************************

SELg.seq2       TCCAAGAGCAGGTAGAATTTTCTTACATTTGTATGCTGATCATAACTCCCTAATTCTTTG 1920
SELg.seq1       TCCAAGAGCAGGTAGAATTTTCTTACATTTGTATGCTGATCATAACTCCCTAATTCTTTG 1920
SELg.Sc26       TCCAAGAGCAGGTAGAATTTTCTTACATTTGTCTGCTGATCATAACTCCCTAACTCTTTG 1920
SELg.Sc24       TCCAAGAGCAGGTAGAATTTTCTTACATTTGTATGCTGATCATAACTCCCTAATTCTTTG 1920
                ******************************** ******************** ******

SELg.seq2       CAGGCTGCCCGAGCGCTTGCAAATCTAGCAGCTCACGGTGATAGCAATGGGAATAATGCT 1980
SELg.seq1       CAGGCTGCCCGAGCGCTTGCAAATCTAGCAGCTCACGGTGATAGCAATGGGAATAATGCT 1980
SELg.Sc26       CAGGCTGCCCGAGCGCTTGCAAATCTAGCAGCTCACGGTGATAGCAATGGGAATAATGCT 1980
SELg.Sc24       CAGGCTGCCCGAGCGCTTGCAAATCTAGCAGCTCACGGTGATAGCAATGGGAATAATGCT 1980
                ************************************************************

SELg.seq2       GCAGTAGGGCGGGAGGCGGGTGCGTTAGAGGCGCTGGTAAAGTTAACATGTTCCAACCAC 2040
SELg.seq1       GCAGTAGGGCGGGAGGCGGGTGCGTTAGAGGCGCTGGTAAAGTTAACATGTTCCAACCAC 2040
SELg.Sc26       GCAGTAGGGCGGGAGGCGGGTGCGTTAGAGGCGCTGGTAAAGTTAACATGTTCCAACCAC 2040
SELg.Sc24       GCAGTAGGGCGGGAGGCGGGTGCGTTAGAGGCGCTGGTAAAGTTAACATGTTCCAACCAC 2040
                ************************************************************

SELg.seq2       GAAGGGGTCCGGTGAGGTCGTTTCGTTTCAAGTCTATCATGGATTTAGTTTGCGTTTCTT 2100
SELg.seq1       GAAGGGGTCCGGTGAGGTCGTTTCGTTTCAAGTCTATCATGGATTTAGTTTGCGTTTCTT 2100
SELg.Sc26       GAAGGGGTCCGGTGAGGTCGTTTCATTTCAAGTCTATCATGGATTTAGTTTGCGTCTCTT 2100
SELg.Sc24       GAAGGGGTCCGGTGAGGTCGTTTCGTTTCAAGTCTATCATGGATTTAGTTTGTGTTTCTT 2100
                ************************ *************************** ** ****

SELg.seq2       CTAGGCAAGAAGCTGCTGGTGCCCTTTGGAACTTGTCATTTGACGACAGAAATCGGGAGG 2160
SELg.seq1       CTAGGCAAGAAGCTGCTGGTGCCCTTTGGAACTTGTCATTTGACGACAGAAATCGGGAGG 2160
SELg.Sc26       CTAGGCAAGAAGCTGCTGGCGCCCTTTGGAACTTGTCATTTGACGACAGAAATCGGGAGG 2160
SELg.Sc24       CTAGGCAAGAAGCTGCTGGCGCCCTGTGGAACTTGTCATTTGACGACAGAAATCGGGAGG 2160
                ******************* ***** **********************************
```

```
SELg.seq2       CAATTGCTGCAGCTGGTGGTGTTGAGGCATTGGTAAGTTTCAAGTAACTTCGTTCTCTTT 2220
SELg.seq1       CAATTGCTGCAGCTGGTGGTGTTGAGGCATTGGTAAGTTTCAAGTAACTTCGTTCTCTTT 2220
SELg.Sc26       CAATTGCTGCAGCTGGTGGTGTTGAGGCATTGGTAAGTTTCAACTAACTTCGTTCTCTTG 2220
SELg.Sc24       CAATTGCTGCAGCTGGTGGTGTTGAGGCATTGGTAAGTTTCACCTTACTTCGTTCTCTTG 2220
                ****************************************  * *************

SELg.seq2       TCTATGGTCCTCACTGTGAACTGCTTGTGAAGGTTGCTTTGGCACAAGGATGCAGCAATG 2280
SELg.seq1       TCTATGGTCCTCACTGTGAACTGCTTGTGAAGGTTGCTTTGGCACAAGGATGCAGCAATG 2280
SELg.Sc26       TCTTTGGTCCTCACTGTGAACTGCTTGTGAAGGTTGCTTTGGCACAAGGATGCAGCAATG 2280
SELg.Sc24       TCTATGGTCCTCACTGTGAACTGCTTGTGAAGGTTGCTTTGGCACAAGGATGCAGCAATG 2280
                *** ************************************************************

SELg.seq2       GATCTCAAGGATTGCAAGAAAGAGCAGCGGGTGCTCTCTGGGGGGTTGTCTGTATCGGAGG 2340
SELg.seq1       GATCTCAAGGATTGCAAGAAAGAGCAGCGGGTGCTCTCTGGGGGGTTGTCTGTATCGGAGG 2340
SELg.Sc26       GATCTCAAGGATTGCAAGAAAGAGCAGCGGGTGCTCTCTGGGGGGTTGTCTGTATCGGAGG 2340
SELg.Sc24       GATCTCAAGGATTGCAAGAAAGAGCAGCGGGTGCTCTCTGGGGGGTTGTCTGTATCGGAGG 2340
                ************************************************************

SELg.seq2       AAAACAGGTATTGTGTGTCACCCTTGGTAAAATAGGACTGATTCTTGTGACGTATGTACA 2400
SELg.seq1       AAAACAGGTATTGTGTGTCACCCTTGGTAAAATAGGACTGATTCTTGTGACGTATGTACA 2400
SELg.Sc26       AAAACAGGTATTGTGTGTCACCCTTGGTAAACTAGGACTGATTCTTGTGACGTGTGTACA 2400
SELg.Sc24       AAAACAGGTATTGTGTGTCACCCTTGGTAAACTAGGACTGATTCTTGTGACGTATGTACA 2400
                ****************************** ********************* ******

SELg.seq2       GCATTGCTATTGGCCGTGAAGGAGGCGTTGCTCCTCTTGTTGCATTAGCGAGATCGGATG 2460
SELg.seq1       GCATTGCTATTGGCCGTGAAGGAGGCGTTGCTCCTCTTGTTGCATTAGCGAGATCGGATG 2460
SELg.Sc26       GCATTGCTATTGGCCGTGAAGGAGGCGTTGCTCCTCTTGTTGCATTAGCGAGATCGGATG 2460
SELg.Sc24       GCATTGCTATTGGCCGTGAAGGAGGCGTTGCTCCTCTTGTTGCATTAGCGAGATCGGATG 2460
                ************************************************************

SELg.seq2       CCGAGGTATGCACAACGCCTCTTCTCGTGTGTGACATGACATTGTGGGTGTTTTTAGGAC 2520
SELg.seq1       CCGAGGTATGCACAACGCCTCTTCTCGTGTGTGACATGACATTGTGGGTGTTTTTAGGAC 2520
SELg.Sc26       CGGAGGTATGCACAACGCCTCTTCTTGTGTGTGACATGACATTGTGGGTGTTTTTAGGAC 2520
SELg.Sc24       CCGAGGTATGCACAACGCCTCTTCTTGTGTGTGACATGACATTGTGGGTGTTTTTAGGAC 2520
                * ********************** ***********************************

SELg.seq2       GTTCATGAAACTGCTGCTGGAGCTTTGTGGAATCTTGCTTTCAATCCAGGAAACGCGTTG 2580
SELg.seq1       GTTCATGAAACTGCTGCTGGAGCTTTGTGGAATCTTGCTTTCAATCCAGGAAACGCGTTG 2580
SELg.Sc26       GTTCATGAAACTGCTGCTGGAGCTTTGTGGAATCTTGCTTTCAATCCAGGAAACGCGTTG 2580
SELg.Sc24       GTTCATGAAACTGCTGCTGGAGCTTTGTGGAATCTTGCTTTCAATCCAGGAAACGCGTTG 2580
                ************************************************************

SELg.seq2       AGGATAGTGGAAGAAGACGGTGTTTCTGCACTGGTGCGGCTGTGCTCGTCCTCGCGCTCT 2640
SELg.seq1       AGGATAGTGGAAGAAGACGGTGTTTCTGCACTGGTGCGGCTGTGCTCGTCCTCGCGCTCT 2640
SELg.Sc26       AGGATAGTGGAAGAAGACGGTGTTTCTGCGCTGGTGCGGCTGTGCTCGTCCTCGCGCTCT 2640
SELg.Sc24       AGGATAGTGGAAGAAGACGGTGTTTCTGCACTGGTGCGGCTGTGCTCGTCCTCGCGCTCT 2640
                ***************************** ******************************

SELg.seq2       AAGATGGCACGATTCATGGCCGCTCTGGCTCTGGCATACATGTTCGATGGCAGGTACAAA 2700
SELg.seq1       AAGATGGCACGATTCATGGCCGCTCTGGCTCTGGCATACATGTTCGATGGCAGGTACAAA 2700
SELg.Sc26       AAGATGGCACGATTCATGGCCGCTCTGGCTCTGGCATACATGTTCGATGGCAGGTACAAA 2700
SELg.Sc24       AAGATGGCACGATTCATGGCCGCTCTGGCTCTGGCATACATGTTCGATGGCAGGTACAAA 2700
                ************************************************************

SELg.seq2       ACTTTTCTTAAGTGTTCTCGAGCAGCTTTCGTGTATTCCTGGGCCATTTGTCTGAGTGCT 2760
SELg.seq1       ACTTTTCTTAAGTGTTCTCGAGCAGCTTTCGTGTATTCCTGGGCCATTTGTCTGAGTGCT 2760
SELg.Sc26       ACTTTTCTTA-GTGTTCTCGATCAGCTTTCGTGTATTCCTGGACCATTTGTCTGAGTGCT 2759
SELg.Sc24       ACTTTTCTTA-GTGTTCTCGATCAGCTTTCGTGTATTCCTGGACCATTTGTCTGAGTGCT 2759
                ********** ********** ******************** ******************

SELg.seq2       CGTTTGCCTGCTTGTCAGGATGGATGAGGTTACTACAAACGAAGTTGTTTACTGTGATAG 2820
SELg.seq1       CGTTTGCCTGCTTGTCAGGATGGATGAGGTTACTACAAACGAAGTTGTTTACTGTGATAG 2820
SELg.Sc26       CGTTTGCCTGCTTGTCAGGATGGATGAGGTTACTACAAACGAAGTTGTTTACTGTGATAG 2819
SELg.Sc24       CGTTTGCCTGCTTGTCAGGATGGATGAGGTTACTACAAACGAAGTTGTTTACTGTGATAG 2819
                ************************************************************

SELg.seq2       CATTACCAAAAACGGCGTGGCAAGGCAGTCGGCCATGAAGAACATTGAGGCGTTTGTACA 2880
SELg.seq1       CATTACCAAAAACGGCGTGGCAAGGCAGTCGGCCATGAAGAACATTGAGGCGTTTGTACA 2880
SELg.Sc26       CATTACCAAAAACGGCGTGGCAAGGCAGTCGGCCATGAAGAACATTGAGGCGTTTGTACA 2879
SELg.Sc24       CATTACCAAAAACGGCGTGGCAAGGCAGTCGGCCATGAAGAACATTGAGGCGTTTGTACA 2879
                ************************************************************
```

```
SELg.seq2       AGCGTTCTCGGATCAACCGTCTCTTGCGGCAGTCCCGGCTTCGCAATGGGGGCCTTCGGC 2940
SELg.seq1       AGCGTTCTCGGATCAACCGTCTCTTGCGGCAGTCCCGGCTTCGCAATGGGGGCCTTCGGC 2940
SELg.Sc26       AGCGTTCTCGGATCAACCGTCTCTTGCGGCAGTCCCGGCTTCGCAATGGGGGCCTTCGGC 2939
SELg.Sc24       AGCGTTCTCGGATCAACCGTCTCTTGCGGCAGTCCCGGCTTCGCAATGGGGGCCTTCGGC 2939
                ************************************************************

SELg.seq2       GCTGCAACAAGTCTCCGACTCCGCCCGCATACAAGAAGCTGGGCATCTGCGCTGCAGGTT 3000
SELg.seq1       GCTGCAACAAGTCTCCGACTCCGCCCGCATACAAGAAGCTGGGCATCTGCGCTGCAGGTT 3000
SELg.Sc26       GCTGCAACAAGTCTCCGACTCCGCCCGCATACAAGAAGCTGGGCATCTGCGCTGCAGGTT 2999
SELg.Sc24       GCTGCAACAAGTCTCCGACTCCGCCCGCATACAAGAAGCTGGGCATCTGCGCTGCAGGTT 2999
                ************************************************************

SELg.seq2       TGAGTCTTAGTTTTCTTTCCTCTCTCTTCTCTCTTGCTAAAGTTGAGGTGTGTTTGACAG 3060
SELg.seq1       TGAGTCTTAGTTTTCTTTCCTCTCTCTTCTCTCTTGCTAAAGTTGAGGTGTGTTTGACAG 3060
SELg.Sc26       TGAGTTTTAGTTTTCTTTCCTCTCTCTTCTCTCTTGCTAAAGTTGAGGTGTGATTGACAG 3059
SELg.Sc24       -GAGCTTTAGTTTTCTTTCCTCTCTCTTCTCTCTTGCTAAAGTTGAGGTGTGATTGACAG 3058
                 ***  ********************************************** *******

SELg.seq2       TGGCGCCGAGATAGGAAGGTTCGTGGCCATGCTTCGAAACGGTTCCTCGGTTCTAAGGTC 3120
SELg.seq1       TGGCGCCGAGATAGGAAGGTTCGTGGCCATGCTTCGAAACGGTTCCTCGGTTCTAAGGTC 3120
SELg.Sc26       TGGCGCCGAGATAGGAAGGTTCGTGGCCATGCTTCGAAACGGTTCCTCGGTTCTAAGGTC 3119
SELg.Sc24       TGGCGCCGAGATAGGAAGGTTCGTGGCCATGCTTCGAAACGGTTCCTCGGTTCTAAGGTC 3118
                ************************************************************

SELg.seq2       GTGTGCCGCGTTTGCTCTCCTCCAGGTACGTGAAAGATTTAATAGAAGAGAGAAAGGTTT 3180
SELg.seq1       GTGTGCCGCGTTTGCTCTCCTCCAGGTACGTGAAAGATTTAATAGAAGAGAGAAAGGTTT 3180
SELg.Sc26       CTGTGCCGCGTTTGCTCTCCTCCAGGTACGTGAAAGATTTAATAGAAGAGAGAAAGGTTT 3179
SELg.Sc24       CTGTGCCGCGTTTGCTCTCCTCCAGGTACGTGAAAGATTTAATAGAAGAGATAAAGGTTT 3178
                 ************************************************* ********

SELg.seq2       CGAGA-AAACCTGGTGTTTCCTTGCAGTTTACAATGCCTGGTGGAAGGCATGCTTTGCAC 3239
SELg.seq1       CGAGA-AAACCTGGTGTTTCCTTGCAGTTTACAATGCCTGGTGGAAGGCATGCTTTGCAC 3239
SELg.Sc26       CGAGA-AAACCTGGTGTTTCCTTGCAGTTTACAATGCCTGGTGGAAGGCATGCTTTGCAC 3238
SELg.Sc24       CGAAAGAAACGTGGTGTTTCGTTGCAGTTTACAATGCCTGGTGGAAGGCATGCTTTGCAC 3238
                *** * **** ********* ****************************************

SELg.seq2       CACGCCAATCTCCTCCAAAGGTCCGGCGCGGCGCGAGTTCTCCGTGGAGCTGCTGCGTCC 3299
SELg.seq1       CACGCCAATCTCCTCCAAAGGTCCGGCGCGGCGCGAGTTCTCCGTGGAGCTGCTGCGTCC 3299
SELg.Sc26       CACGCCAATCTCCTCCAAAGGTCCGGCGCGGCGCGAGTTCTCCGTGGAGCTGCTGCGTCC 3298
SELg.Sc24       CACGCCAATCTCCTCCAAAGGTCCGGCGCGGCGCGAGTTCTCCGTGGAGCTGCTGCGTCC 3298
                ************************************************************

SELg.seq2       ACAACAGCTCCACTCCAGGCGAGGGTGTTTGCGCGACTAGTTTTGAGGAATCTGGAGCTG 3359
SELg.seq1       ACAACAGCTCCACTCCAGGCGAGGGTGTTTGCGCGACTAGTTTTGAGGAATCTGGAGCTG 3359
SELg.Sc26       ACAACAGCTCCACTCCAGGCGAGGGTGTTTGCGCGACTAGTTTTGAGGAATCTGGAGCTG 3358
SELg.Sc24       ACAACAGCTCCACTCCAGGCGAGGGTGTTTGCGCGACTAGTGTTGAGGAATCTGGAGCTG 3358
                ************************************* ******************

SELg.seq2       TGCCAGAGCGAGAAAAGCTGA 3380
SELg.seq1       TGCCAGAGCGAGAAAAGCTGA 3380
SELg.Sc26       TGCCAGAGCGAGAAAAGCTGA 3379
SELg.Sc24       TGCCAGAGCGAGAAAAGCTGA 3379
                *********************
```

8.6 ARABIDILLO homologue protein alignment

Alignment: ARABIDILLO homologues
Seaview [blocks=10 fontsize=10 A4] on Thu Jul 21 15:34:25 2011



```
                    121
ARABIDILLO1         MAASLASRCV NLHYLRFRGV ESADSLIH-L KARNLIEVSG DYCKKITDAT LSMIVARHEA
ARABIDILLO2         MAASLATRCV DLQKIRFRGV DSADAIIH-L KARSLLEISG DYCRKITDAT LSMIAARHEA
Physcodillo1A       MVSALAGRCS NLEALKFRRG ASASSIVG-L QAKGLRELSG DCCSQLSDAT LSMVVARHAN
Physcodillo2        VVSALAGRCS NLQCLKFRGG AFANSIVG-L HARELRELSG DWCSLLSDAT LSMVVARHGN
Physcodillo1B       MVSALAGRCS NLEALKFRRG ASASSIVG-L QAKGLRELSG DCCSQLSDAT LSMVVARHAN
Oryzadillo          VASSLSSRCG SLRRLRLRGH EAAAAASG-L RARGLREVVA DGCRGLTDAT LAVLAARHEA
Brachydillo         VASSLASRCG GLRRLRLRGH EAAAAVASCL RARDLREVVA DGCRGLTDAT LAVLAARHEA
Vitidillo1          AAASLASRGM NLQKLRFRGQ ETADAIIH-L QARGLREISG DYCRKINDAT LSVIAARHEQ
Vitidillo2          AADYLSSQCA NITKLRFRGA ESANAIIR-L QARGLREISG EFCRDINDAT LSVIAARHEA
Zeadillo1           VASSLASRCG SLRRIRLRGH EAAEAVLG-L RARGLREVVA DGCRGLTDAT LAVLAARHED
Sorghodillo         VASSLASRCG SLRRLRLRGH EAAEAVPG-L RARGLREVVA DGCRGLTDAT LAVLAARHED
Glycinodillo        MASSLAPRCV HLQKLRFRGA ESADAIIH-L RARNLRELSG DYCRKITDAT LSVIVARHEL
Poplardillo1        MAVSLASRCV NLQKLRFRGA ECADAIIH-L QARNLREISG DYCRKITDAT LSMIVARHEA
Poplardillo2        MAVSLASRCV NLQKIRFRGA ESADAIIH-L QARNLREISG DYCRKITDAT LSMIVARHEA
Poplardillo3        AAQSLSSRSK NLRKLRFLGA ESADAIIS-L EARDLREISG DFCRDITDAT LSMIAARHEM
Ricinodillo1        TATSLAPRCI QLQKLRFRGA ESADAIIH-L QAKNLREISG DYCRKITDAS LSVIVARHEL
Ricinodillo2        AAASLSSRSK NLRKLRFRGA ESADAVIQ-L QARGLREISG DFCRDITDAT VSVIAARHEM
Manihodillo1        MAASLASRCV NLQKLRFRGA ESADAIIH-L QDRNLREISG DYCRKITDAT LSVIAARHEL
Manihodillo2        AAESLSSRSR NLQKLRFRGA DSACAIIN-L QARGLREISG DFCPDITDAI VSVIAARHEM
Cucumidillo         MAGSLALRCE NLQKLRFRGA ESADAIIL-L LAKNLREISG DYCRKITDAT LSAIAARHQA
Prunidillo          MAASLAARCV NLQKLRFRGA ESADAILH-L QARNLREISG DYCRKITDAT LSVIVARHEA
Citrus.sinidillo    MAASLASRCM NLQKLRFRGA ESADSIIH-L QARNLRELSG DYCRKITDAT LSVIVARHEA
Citrus.clemidillo   MAASLASRCM NLQKLRFRGA ESADSIIH-L QARNLRELSG DYCRKITDAT LSVIVARHEA
Eucadillo           MASSLASRCR NLQKLRFRGA ESADAIIF-L QARNLREISG DYCRKITDAT LSVIAARHEL
Mimulodillo1        ATSSLASRCK NLQKLYFRGP DSADAVIS-L KAKNLKEISG DSCRKMTDST LCVLAARHEA
Mimulodillo2        SASSLVSRCE NLHKLRFRGP DSVDAIIN-L QAKNLREISG DSCRKMTDAT LSVLAARHES
Aquilidillo1        TAIFLAPRCA NLKKIRFRGA ESASAIML-L QARGLREISG EFCREITDAI LSVMAARHEA
Selagidillo         MASALASRCA KLSKLKFRGA SGASLIID-L QARQLKGLIG DGCKDLTDAT LSMLVARHEN
A.Lyrata1           MAASLASRCV NLHNLRFRGV ESADSLIH-L KARNLLEVSG DYCRKITDAT LSMIVARHEA
A.lyrata2           MAASLATRCA DLQKIRFRGV DSADAIIH-L KARTLLEISG DYCRKITDAT LSMIAARHEA
Hordidillo          VASSLASRCG GLQRLRLRGH EAAAAVASAL RARDLREVVA EGCRGLTDAT LAVLAARHEA

                    181
ARABIDILLO1         LESLQLGPDF CERITSDAIK AVAFCCPKLK KLRLSGIRDV TSEAIEALAK HCPQLNDLGF
ARABIDILLO2         LESLQLGPDF CERITSDAIR VIAFCCPKLK KLRVSGMRDV SSEAIESLAK HCPQLSDLGF
Physcodillo1A       LESLLLGSD- CERVTSEALK VIAVCCPKLR RLCVSGVLKV ERDAIQALFQ HCKGLTELGF
Physcodillo2        LESLQLGSD- CERVTSEALK VVAVCCPKLR RLCISGIRDV DRDAIQAMFQ HCQGLTELGF
Physcodillo1B       LESLLLGSD- CERVTSEALK VIAVCCPKLR RLCVSGVLKV ERDAIQALFQ HCKGLTELGF
Oryzadillo          LESLQIGPDP LERISSDALR QVAFCCSRLR RLRLSGLRDA DADAIGALAR YCPLLEDVAF
Brachydillo         LESLQIGPDP LEHVSSDALH HVALCCSRLR RLRLSGLREA TADAIGALAR HCPHLEDVAF
Vitidillo1          LESLQLGPDF CEKITTDAIK AIAVCCPKLN KLRLSGVKDV HGDAIDALAK HCRNLTDLGF
Vitidillo2          LESLQLGPDA CDKITSDAIK AVAFCCPKLK RLRISGVQVV TGDAINALGK HCGQLVELGF
Zeadillo1           LQSLQIGPDP LERISSDALR HVALCCSQLC RLRLSGLREV DAEAVGALAR CCPLLEDVAF
Sorghodillo         LQSLQIGPDP LEHISSDALR HVALCCSQLR RLRLSGLREA DADAVGALAR CCPLLEDVAF
Glycinodillo        LESLQLGPDF CERISSDAIK AIAHCCPKLN KLRLSGIRDV NADAINALAK HCPKLTDIGF
Poplardillo1        LETLQLGPDF CERISSDAIK ATAFCCPKLK KLRLSGLRDV SAEVINALAK HCPNLIDIGL
Poplardillo2        LETLQLGPDF CEKVSSDAIK AIAFCCPKLK KLRLSGLRDV SADVINALAK HCPNLIDIGF
Poplardillo3        LECLQLGPDV CERITSFAIR VIALCCPKLK RLQISGVKEV TGEAINALAK HCRQLVEVAF
Ricinodillo1        LESLQLGPDF CERISSDAIK AIAFCCPKLK KLRVSGIRDV SADAINALAK HCPNLIDIGF
Ricinodillo2        LESLQLGPNA CERISSDAIK AVALCCPNLK RLQLSGVREV RIDAINALAR HCGQLLEVAF
Manihodillo1        LESLQLGPDF CERISSDAIK AIAFCCPQLR KLLLSGIRDV SADAINALAK HCPNLFDIGF
Manihodillo2        LESLQLGPNA CERISSDAIK AIALCCPKLR RLRLSGVREV SKDAINALAM HCGNLLEVAL
Cucumidillo         LESLQLGPDF CERISSDAIK AIAICCHKLK KLRLSGIKDV SAEALNALSK HCPNLLDIGF
Prunidillo          LESLQLGPDF CERISSDAIK AIAICCPKLK KLRLSGIRDV HADAIIALTK HCQNLTDIGF
Citrus.sinidillo    LESLQLGPDF CERITSDAVK AIALCCPKLK KLRLSGIRDI CGDAINALAK LCPNLTDIGF
Citrus.clemidillo   LESLQLGPDF CERITSDAVK AIALCCPKLK KLRLSGIRDI CGDAINALAK LCPNLTDIGF
Eucadillo           LECLQLGPDF CERITSDAIK QIAFCCPKLT KLRLSGIRDV QGDAINALGK HCLHLIDVGF
Mimulodillo1        LECLQIGPDF CERISSDAVR AIAICCPKLK KLRISGIQEV DASAINALAR NCPNLNDIGL
Mimulodillo2        LECLMIGPDF CERISSDALK AVAICCPKLR KLRLSGMHEV DADAINALAK HCPKLTDIGF
Aquilidillo1        LECLQLGPDF CEKISSDAIK AVALCCPNLR KLRLSGIRDI DADAINTLAK HCTKLEEVAF
Selagidillo         LESLQLGPE- LEKITNEAIK VVAVCCRRLK CLRLAGIRDV DSEAIGDLVK HCPSLTELAL
A.Lyrata1           LESLQLGPDF CEKITSDAIK AVAFCCPKLT KLRLSGIRDV TSEAIEALAK HCPQLSDLGF
A.lyrata2           LESLQLGPDF CEKITSDAIR VIAFCCPKLK KLRVSGMRDV SSEAIESLAK HCPQLSDLGF
Hordidillo          LESLQIGPDP LERISSDALR HVALCCSRLR RLHLSGLREA DSDAIGALAR YCPLLEDVAF
```

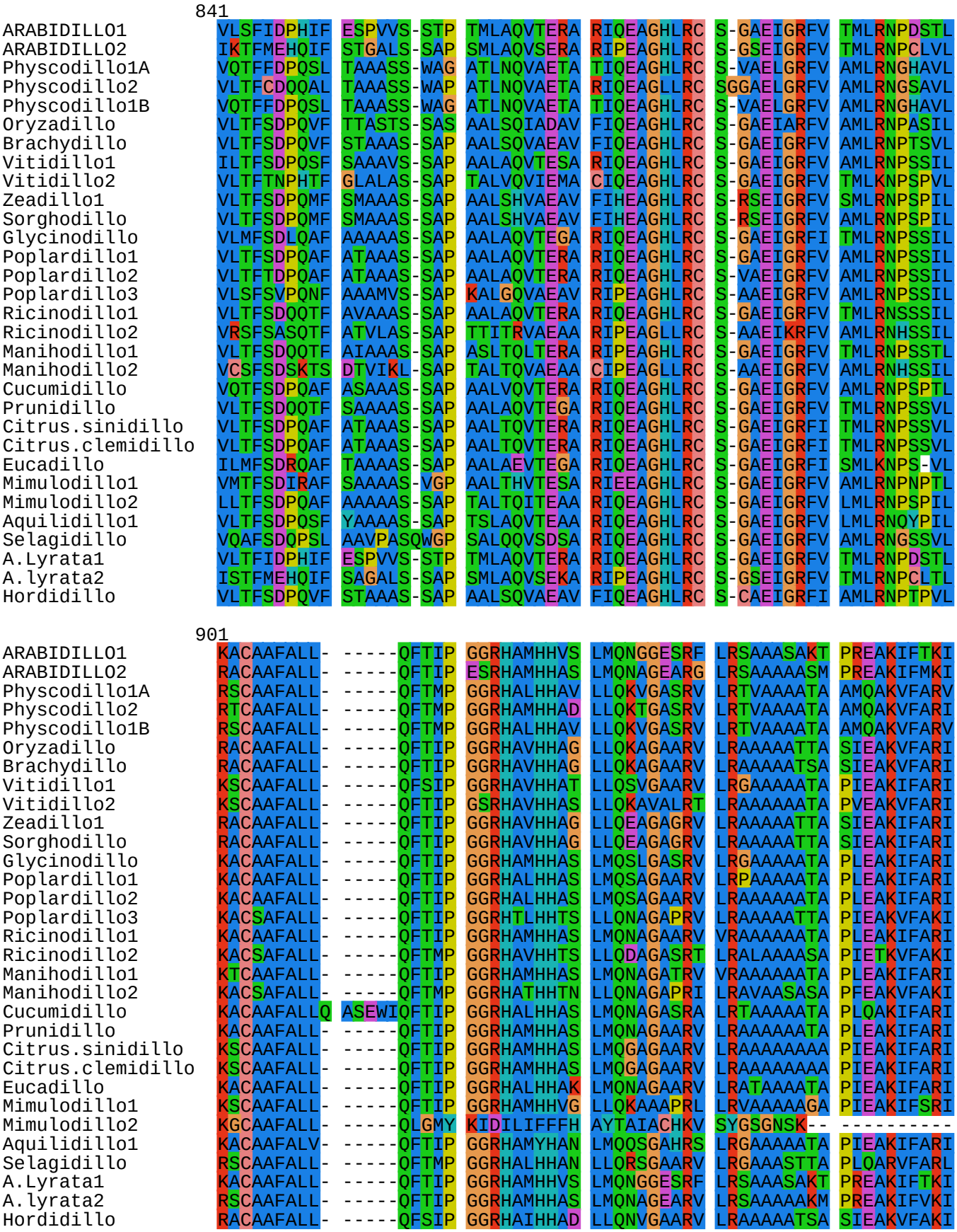
```
                    241
ARABIDILLO1         LDCLNIDEEA LGKVVSVRYL SVAGTSNIKW SIASNNWDKL PKLTGLDVSR TDIGPTAVSR
ARABIDILLO2         LDCLNINEEA LGKVVSLRYL SVAGTSNIKW KVALENWEKL PKLIGLDVSR TTIDHIAVSR
Physcodillo1A       LDSHTIDEGA FGGASSLRFL SVAGCRCIVW STAAHWWSKL PNLAGLDVSR TDITPTALMQ
Physcodillo2        LDSHTIDEGA FGAARNLRFL SVAGCRCIAW STAAQCWSKL PNLSGLDVSR TDITPSTLAQ
Physcodillo1B       LDSHTIDEGA FGGASSLRFL SVAGCRCIVW STAAHWWSKL PNLAGLDVSR TDITPTALMQ
Oryzadillo          LDCGSVDEAA IAGILSLRFL SVAGCHNLKW ATASTSWAQL PSLVAVDVSR TDVSPSAISR
Brachydillo         LDCVVVDESA LGDIHSLRFL SVAGCLNMKW ATASASWAQL PSLVAVDVSR TDVSPNAISR
Vitidillo1          MDCLKVEELA LGNILSLRFL SVAGTTNLKW GLISHLWGKL PNLTGLDVSR TDITPNAASR
Vitidillo2          IDGDNVDGAA LGNLKSVRFL SVAGTRNMKW GSAVQPLCRL NSLIGIDVSR TDISLSSVTR
Zeadillo1           LDCGTVDEAA LAGIHSVRFL SVAGCRNLKW ATASTCWTQL PSLIALDVSR TDVPPSAVSR
Sorghodillo         LDCGTVDEAA LAGIHSLRFL SVAGCRSLKW ATASTSWTQL PSLIALDVSR TDVPPSAVSR
Glycinodillo        IDCLNVDEVA LGNVLSVRFL SVAGTSSMKW GVVSHLWHKL PNLIGLDVSR TDIGPSALLR
Poplardillo1        LDCLKVDEVA LGNVVSVLFL SVAGTSNMKW GVVSHLWHKL PKLIGLDVSR TDIGPSAVSR
Poplardillo2        LDCLKVDEAA LGNVVSVHFL SVAGTSNMKW GVVSHLWHKL PKLIGLDVSR TDIDPSAVSR
Poplardillo3        MESNSVDELA LGNLTSVQFL SLAGTKNLKW NSASCVWSKL PKLVGLDVSR TDITFSSVMR
Ricinodillo1        LDCLNVDEVA LGNVVSVRFL SVAGTSNMKW GVISHLWHKL PKLIGLDVSR TDIGPTAVSR
Ricinodillo2        MESDFVDEVA LGNLASVQFL SIAGTRNVKW GLASQVWSNL PKLVGLDVSR TDISLSSVRK
Manihodillo1        LDCLNVDEVA LGNVVSVCFL SVAGTSNIKW EMI-HLWHKL PNLIGLDVSR TNIPPTAVSG
Manihodillo2        MESENVDEMA LGNLTTVQFL SLAGTRNLKW GIASQVLSKL PKLEGLDVSR TDISLSFIRR
Cucumidillo         IDCFNIDEMA LGNVSSVRFL SVAGTSNMKW GAVSHQWHKL PNLIGLDVSR TDIGPVAVSR
Prunidillo          IDCLNIDEMA LGNVLSVRFL SVAGTSNMKW GVVSHLWHKL PNLTGLDVSR TDIGSAAVSR
Citrus.sinidillo    LDCLNVDEVA LGNVLSVRFL SVAGTSNMKW GVVSQVWHKL PKLVGLDVSR TDVGPITISR
Citrus.clemidillo   LDCLNVDEVA LGNVLSVRFL SVAGTSNMKW GVVSQVWHKL PKLVGLDVSR TDVGPITISR
Eucadillo           LDCLKVDELA LGNVASLRFL SVAGTSNLKW GVVSHVWHKL PNLVGLDVSR TDVGPAAISR
Mimulodillo1        IDCRKVDEAA LGNIASLRFL SLAGSTNMKW SLFV-----L PHLIGLDVSR TDISPSNVSR
Mimulodillo2        IDCRKVDETA LGNVASVRFL SVAGTTHMKW HLISQHWSKL RDLVALDVSR TDITPTIISR
Aquilidillo1        MDCLNVDVEA LGNMVSVRFL SVAGTRSINW TLASQLWSKL PNLIGLDVSR TDINASAVST
Selagidillo         LDCAVVDEAA LGEAKSLRYL SVAGSRNIMW TQAMQAWSKL ENLVALDVSR TEVTPAAVMS
A.Lyrata1           LDCLNIDEEA MGKVVSVRYL SVAGTSNIKW STASNSWDKL PKLTGLDVSR TDIGPTAVSR
A.lyrata2           LDCLNINEEA LGKVVSLRYL SVAGTSNIKW KVALDNWEKL PKLTGLDVSR TTIDHIAVSR
Hordidillo          LDCGTVDEAA LGDIHSLRFL SVAGCYSVKW ATASASWAQL PLLVAVDVSR TDASPNAVAR

                    301
ARABIDILLO1         FLTSSQSLKV LCALNCHVLE EDESLISYN- -----RFKGK VLLALFTNVF DGLASIFADN
ARABIDILLO2         LLKSSQSLKV LCALNCPYLE EDKSYSSN-- -----RFKGK VLLAVFTDTF DELASIFADN
Physcodillo1A       VLA-GPELRV VCALNCPVLE EGSNPVTLP- -----SSKKT VVLARFTDVM EGLDALLSPS
Physcodillo2        VLA-CPELKV VCGLNCPLLE EGSNPIPSP- -----SSKTA VLLARFTDLM EGLEVLLNPS
Physcodillo1B       VLA-GPELRV VCALNCPVLE EGSNPVTLP- -----SSKKT VVLARFTDVM EGLDALLSPS
Oryzadillo          LISHSKTLKL ICTLNCKSVE EEQAHNPGAF S----NSKGK LVLTITSHIF KSVVSLFPDK
Brachydillo         LISHSKTLEL ICAVNCKSVE EEQAHDPAVF R----NSKGK LVLTITSDIF KSIASLFPGK
Vitidillo1          LFASSQSLKV LCALNCSALE QDVTFFATYN NNNNINNKGK LLLAQFSDIF KGIASLFADT
Vitidillo2          LLSFSQNLKV FFALNCPKFE ADVNNSTSY- -----NYKGK LLVALFSDIF KGVASLFADK
Zeadillo1           LISHAKTLKL ICTLNCISVE EEQLHNPAVF S----NSKGK VVLTINSDIF KSFETMFPVV
Sorghodillo         LISHAKTLKL ICTLNCISVE EEQLHNPAVF S----NSKGK VVLTINSD-F KTFETMFPVV
Glycinodillo        MLSLSQNLRV LIALNCPILE EDTSFSAS-- -----KYKNK LLISLRTDIF KGLASLLFDN
Poplardillo1        LLSLSPSLKV LCAMNCPVLE EDNSFSVN-- -----KYKGK LLLALFTDIF KGLASLFADT
Poplardillo2        LLSLSPSLKV LCAMNCPVLE EDNAFSVN-- -----KYKGK LLLALFNDIF KGLASLFADI
Poplardillo3        LFLSSQYLKV LVALNCPVFE AEVDNHMTY- -----NHKGK ILLTVFNDIF KAVGSLFVDI
Ricinodillo1        LLSSSHSLKV LCALNCSVLE EDATFSAN-- -----RYKGK LLIALFTDIF KGLSSLFADT
Ricinodillo2        FLSLSKNLKV LFAFNCPVFE AEVDNDMVY- -----VCKGK ILLTVFNDIF KAVASLFVDT
Manihodillo1        LLSSCHRLKV LCALNCSVLE ADTTFNAN-- -----MCKGK LLISLFTDIF KGLASLFAVT
Manihodillo2        LISWSENLKV LFVLSCPVFE AEVDNDMVY- -----IHKGK ILLTVCNDIF KEVSSLFADA
Cucumidillo         LMSSSQSLKV LCAFNCSVLE DDAGFTVS-- -----KYKGK LLLALFTDVV KEIASLFVDT
Prunidillo          LLSSSQSLKV LCALNCPVLE EDTNFASR-- -----KYKNK LLLACFTEIM EEIAFLLVDI
Citrus.sinidillo    LLTSSKSLKV LCALNCPVLE EENNISAV-- -----KSKGK LLLALFTDIF KALASLFAET
Citrus.clemidillo   LLTSSKSLKV LCALNCPVLE EENNISAV-- -----KSKGK LLLALFTDIF KALASLFAET
Eucadillo           LLNLSLSLKV LCALNCPVLE GDVNFIVP-- -----KIKGK SLLAHFTDIF EGIASLFVDI
Mimulodillo1        FFSSSLSLKV LCALNCPLLE ADPTFVYNNY -----NHKGK VLVSIFSDIL K---------
Mimulodillo2        FFSSSISLKV LCALNCPLLD TDPTYLSNK- -----NHKGK VLLAFTTDIL KGVSALFADT
Aquilidillo1        LLSSTQSLKV LCALNCPLLE TEGNHATY-- -----CTTGR LLLALFKDTF KGVASLFADN
Selagidillo         FLS-APRLRV LCALSCSALE DGSNSVSY-- -----VSKDR VLLARFTELM NGLACISSLE
A.Lyrata1           FLTSSQSLKV LCALNCHVLE EDTSFFSSN- -----RFKGK VLLALFTNVF DGLASIFADK
A.lyrata2           LLKSSQSLKV LCALNCPYLE EDKSYSSN-- -----RFKGK VLLAIFTDTF DELASIFADN
Hordidillo          LISHSKTLEL ICALNCKFVE EEQAHSPTAF S----NSKGK LVLTITCPIF KSLASLFPGK
```

```
                          361
ARABIDILLO1               T--KKPKDIF --AYWRELMK TTKDKTINDF IHWIEWIISH TLLRTAE-CN PEGLDDFWLN
ARABIDILLO2               S--KKPKNIF --SYWRDLIR --KDKSIDEI MLWIEWIISH TLLRIAESSN SQGLNDFWLN
Physcodillo1A             N--WKEEAGS --CARSRCGG --RARADSEV AKWTEWMLSH AVLKIAE-CN APSL----LK
Physcodillo2              N--SKEEACS --SGRSRYGR --RMRVDPEV AKWTERMLSH ALLKIAE-SN APSLDSFWLK
Physcodillo1B             N--WKEEAGS --CARSRCGG --RARADSEV AKWTEWMLSH AVLKIAE-CN APSL----LK
Oryzadillo                V--VKENEVF NECNWKG--- --KDNALGDM MSWLEWILSQ TLLRIAE-SN PQGMDDFWLQ
Brachydillo               A--VKEHGVF NECNWRD--- --KNKVLGHM MSWLEWILSQ SLLRIAE-SN PYGMDEFWLQ
Vitidillo1                S--KNKRDVF --FEWRNGKN --KDKNLDMI MNWLEWALSH TLLRIAE-SN PQGLDTFWLK
Vitidillo2                I--ENQREVF --SHWRKLKN --RDNNLDEI VTWIEWILSH SLLRISE-NN PEEFNEFWLR
Zeadillo1                 D--VKEHEFF NQCNWSH--- --KDKIPGDT MTWLEWILSQ SLLRIAE-SN PQGMDGFWLQ
Sorghodillo               D--VKAHEVF NQCSWSH--- --KDKIAGDT MTWLEWILSQ SLLRIAE-SN PQGMDSFWLQ
Glycinodillo              T--RRGNNVF --LDWRTSKN --NDKDLNEI IPWLEWMLSH TLLRSAE-SP QQGLDSFWVE
Poplardillo1              T--KTGKNVL --LDWRNLKT --KDKNLDEI MTWLEWILSH TLLRTAE-SN PQGLDAFWLK
Poplardillo2              T--KMGKNVL --LEWRNLKT --KDKNVDEI MSWLEWILSH TLLRTAE-SN PQGLDVFWLK
Poplardillo3              T--ENESNNL --SYWRKVKS --IDRSMDVI ATWIEWIFSH FLLRIAE-NN PKELDAFWIK
Ricinodillo1              TNTKKGKNVF --LDWRSSKT --QDKNLDDI MTWLEWILSH TLLPTAE-SN PQGLDDFWLK
Ricinodillo2              T--KNGSNAV --ACWRKLKI --RGRSLDEK VVWIEWVLSH SLLRIAE-NN PKELDVFWLK
Manihodillo1              TNSRKGKNVF --LDWRNSKN --KDKNFDDI MTWLEWILSH TLLRTAE-SN PQGLDDFWLK
Manihodillo2              S----ESNML --PYWRSLKI --REQSLDKI VLWIEWVLSH SLLRVAE-NN QKELDIFWVK
Cucumidillo               T--TKGENML --LDWRNLKI --KNKSLDEI MMWLEWILSH NLLRIAE-SN QHGLDNFWLN
Prunidillo                T--KKGKNVF --LDWRNSKN --KDKNLDDI MTWIEWILSH TLLRIAE-SN QQGLDDFWPK
Citrus.sinidillo          T--KNEKNVF --LDWRNSKN --KDKNLNEI MTWLEWILSH ILLRTAE-SN PQGLDDFWLK
Citrus.clemidillo         T--KNEKNVF --LDWRNSKN --KDKNLNEI MTWLEWILSH ILLRTAE-SN PQGLDDFWLK
Eucadillo                 T--SKGKTVF --LDWRNSKK --THKCMDEM MSWFEWILSH SLLRSAE-NN PQGLDDFWLK
Mimulodillo1              ---ENETNIF --LHWRNSE- --KDKRLDEV LNWLEWIVSN SLLRISE-SN PPGLDNFWLK
Mimulodillo2              P-MTSKTNCF --LDWRNTKV --KDGRTDEV LNWLEWDHFE SLLRVPR--- ----------
Aquilidillo1              T---TEKSVF --LDWRNGKI --GDKHLNVI MIWLEWILAH SLLRIAE-SN PQGLDEFWLS
Selagidillo               Q--QDESRV- ---------- ---------L VCWTEWVLSH ALLRIAE-NN TQGLDAFWLK
A.Lyrata1                 T--KKPKDIF --AYWRELMK TTKDKTVDDF MHWIEWIISH TLLRTAE-CN PQGLDDFWLN
A.lyrata2                 S--KKPKDMF --SYWRDLIR --KDKSTDEI MLWIEWIISH TLLRIAESSN SQGLNDFWLN
Hordidillo                A--VEEHGVF NECNWRN--- --KRKILGVM MNWLEWILSH SLLRISE-CN PYGMDDFWLQ

                          421
ARABIDILLO1               EGAALLLNLM QSSQEDVQER SATGLATFVV VDDENASIDC GRAEAVMKDG GIRLLLELAK
ARABIDILLO2               QGATLLLSLM QSAQEDVQER AATGLATFIV VDDENASIDC GRAEAVMRDG GIRLLLELAK
Physcodillo1A             QGIAMMLRLV QSAQEDVQER AASALATFVV VDDENATVDS ARAEAVMNGG GIALLLGLAK
Physcodillo2              QGTAMMLRLV QSAQEDVQER AAAALAVFVL VDDENATVDS ARAEAVMNGG GIALLLGLAK
Physcodillo1B             QGIAMMLRLV QSAQEDVQER AASALATFVV VDDENATVDS ARAEAVMNGG GIALLLGLAK
Oryzadillo                QGADMLLSLV KSSQEDVQER AATTLATFVV IDDESANVDA ARSEAVMRVG GIPMLLDLAR
Brachydillo               QGTSMLLSLV KSSQEDVQER AATTIATYVV IDDETANVDA ARSEAVMRDG GIPLLLDLAR
Vitidillo1                QGAALLLSLM QSSQEDVQEK AATALATFVV IDDENASIDC GRAEAVMRDG GIRLLLNLAR
Vitidillo2                QGAALLLSLM QSSQEDVQER AATAVATFVV IDDDNATVDC RRAEAVMQDG GVELLLDLAS
Zeadillo1                 KGTTLLLRLL KSLQEDVQER AATSLATFVV MDDESANVDP ARSEAVMQNG GIRMLLDLAR
Sorghodillo               KGTALLLRLL KSLQEDVQER AATALATFVV MDDESANVDP ARSEAVMQNG GIRMLLDLAR
Glycinodillo              QGGALLLSLM QSSQEDVQER AATGLATFVV IDDENASIDC GRAEAVMRDG GIRLLLGLAK
Poplardillo1              QGATILLSLM QSSQEEVQER AATGLATFVV IDDENASIDC GRAEAVMRDG GIRLLLNLAK
Poplardillo2              LGAPILLSLM QSSQEEVQER AATGLATFVV IDDENASIDC GRAEAVMRDG GIRLLLNLAK
Poplardillo3              QGAALLLDLL QSSQEDVQER AANSIATFVV IDDENATVDS QRAEVVMQNG GIQLLLDLAR
Ricinodillo1              QGAAILLSLM QSSQEDVQER AATGLATFVV IDDENASIDC GRAEAVMRDG GIRLLLDLAK
Ricinodillo2              QGAKLLLHLL QSSEEDVQER AATAIATFVV IDDENATIDS RRAEAIVQNG GIQLLLDHAR
Manihodillo1              QGAPLLLILM QSSQEDVQER AATGLATFVV IDDENASIDC GRAEAVMRDG GIRLLLDLAR
Manihodillo2              QGTNLLLHLL QSSEEDVQER AATALATFVV IDDENATIDC QRAEAVMQNG GIQLLLDLAK
Cucumidillo               QGAALLLSLM QSSQEDVQER AATGLATFVV IDDENASIDS GRAEEVMRRG GIRLLLNLAK
Prunidillo                QGASLLLNLM QSSQEDVQER AATGLATFVV IDDENASIDC RRAEAVMRDG GIRLLLNLAK
Citrus.sinidillo          QGAGLLLSLM QSTQEDVQER AATGLATFVV INDENASIDC GRAEAVMKDG GIRLLLDLAK
Citrus.clemidillo         QGAGLLLSLM QSTQEDVQER AATGLATFVV INDENASIDC GRAEAVMKDG GIRLLLDLAK
Eucadillo                 QGAPLLLILM QSSQEDVQER AATGLATFVV IDDEHASIDR GRAEAVMRDG GIRLLLDLAK
Mimulodillo1              QGAPLLLSLV QSSNEEVQER AATAIATFVV TDDESASIDP LRAEAVMQNN GLRLLLDLAR
Mimulodillo2              ATARLLLTFM QSPQEEVQER AATALATFVV IDDENACIDT GRAEAVMRDG GIRLLLDIAR
Aquilidillo1              QGAALLLSLV QSFQEDVQER AATGLATFVV IDDENATVDC KRAEAVMCDG GIRLLLDLAR
Selagidillo               QGTSVMLRLI KSMQEDVQER AATALATFVV VDDENATVDS SRAEAVMHGG GIRSLLDLAR
A.Lyrata1                 QGAALLLNLM QSSQEDVQER SATGLATFVV IDDENANIDC GRAEAVMKDG GIRLLLELAK
A.lyrata2                 QGATLLLSLM QSAQEDVQER AATGLATFIV VDDENASIDC GRAEAVMRDG GIRLLLELAK
Hordidillo                QGTSMLLSLV KSSQESVQER AATTIAIFVV IDDETANVDA ARSEAVMRDG GIPLLLDLAR
```

```
                          481
ARABIDILLO1               SWREGLQSEA AKAIANLSVN ANIAKSVAEE GGIKILAGLA KSMNRLVAEE AAGGLWNLSV
ARABIDILLO2               SWREGLQSEA AKAIANLSVN AKVAKAVAEE GGISVLADLA KSMNRLVAEE AAGGLWNLSV
Physcodillo1A             SCREGVQSEA AKAIANLSVN TEVAKRVALE GGISILAGLA RSRNRWVAEE AAGGLWNLSV
Physcodillo2              SCGEGVQSEA AKAIANLSVN TEVAKRVALE GGISILAALA RSPNRWVAEE AAGGLWNLSV
Physcodillo1B             SCREGVQSEA AKAIANLSVN TEVAKRVALE GGISILAGLA RSRNRWVAEE AAGGLWNLSV
Oryzadillo                CSRESAQSEA AKAIANLSVN AKVAKAVADE GGITILTNLA RSMNRLVAEE AAGGLWNLSV
Brachydillo               CSRVSAQSEA AKAIANLSVN AKVAKVVADE GGITIFTNLA KSTNRLVAEE AAGGLWNLSV
Vitidillo1                SWREGLQSEA AKAIANLSVN ANVAKAVADE GGINILSSLA RSMNRSVAEE AAGGLWNLSV
Vitidillo2                SCQEGLQSEA AKAIANLSVN SKVAKAVAEN GGIDILSNLA RSMNRLVAEE AAGGLWNLSV
Zeadillo1                 CSRESAQSEA AKAIANLSVN TKVAKAVADE GGITILINLA KSMNRLVAEE AAGGLWNLSV
Sorghodillo               CSRESAQSEA AKAIANLSVN TKVAKAVAEE GGITILTNLA KSMNRLVAEE AAGGLWNLSV
Glycinodillo              SWREGLQSEA AKAIANLSVN ANVAKAVAEE GGIQILAGLA RSMNKLVAEE AAGGLWNLSV
Poplardillo1              SWREGLQSEA AKAIANLSVN ANVAKAVAEE GGIQILAGLA SSMNRLVAEE AAGGLWNLSV
Poplardillo2              SWREGLQSEA AKAIANLSVN ANVAKAVAEE GGIEILAGLA RSMNRLVAEE AAGGLWNLSV
Poplardillo3              SCREGLQSEA AKAIANLSVD SKVAKAVADI GGINILVGLA RSVNRLVAEE AAGGLWNLSV
Ricinodillo1              SWREGLQSEA AKAIANLSVN ANVAKAVAEE GGINILAGLA RSMNRLVAEE AAGGLWNLSV
Ricinodillo2              SCQEGLQSEA AKAIANLSVD SKVAKAVAEI GGIKILANLA RSMNRLVAEE AAGGLWNLSV
Manihodillo1              SWREGLQSEA AKAIANLSVN ANVAKAVAEE GGINVLAGLA RSMNRLVAEE AAGGLWNLSV
Manihodillo2              SCHEGLQSEA AKAIANLSVD SKVAKTVAEI GGINILASLA RSMNRLVAEE AAGGLWNLSV
Cucumidillo               SWREGLQSEA AKAIANLSVN ANVAKAVAEE GGIDILAGLA RSMNRLVAEE AAGGLWNLSV
Prunidillo                SWREGLQSEA AKAIANLSVN ANVAKAVAEE GGINILAGLA RSMNRLVAEE AAGGLWNLSV
Citrus.sinidillo          SWREGLQSEA AKAIANLSVN AKVAKAVAEE GGINILAVLA RSMNRLVAEE AAGGLWNLSV
Citrus.clemidillo         SWREGLQSEA AKAIANLSVN AKVAKAVAEE GGINILAVLA RSMNRLVAEE AAGGLWNLSV
Eucadillo                 SWREGLQSEA AKAIANLSVN VNVAKAVAEE GGIDILASLA RSMNRLVAEE AAGGLWNLSV
Mimulodillo1              SWHEGLQSEA AKAIANLSVN AKVAKFVAEE GGIGIIVNLA RSVNRLVAEE AAGGLWNLSV
Mimulodillo2              SWRDGLQSEA AKAIANLSVN ANVAKAVADE GGINVLVNLV SSVNRMVAEE AAGGLWNLSV
Aquilidillo1              SSREGLQSEA AKAIANLSVN AEVAKSVAEE GGIKILSNLA RSMNRLVAEE AAGGLWNLSV
Selagidillo               SSREGVQSEA AKAIANLSVN AEVAKAVATE GGINILAGLA RSPNRWVAEE AAGGLWNLSV
A.Lyrata1                 SWREGLQSEA AKAIANLSVN ANVAKSVAEE GGIKILAGLA KSMNRLVAEE AAGGLWNLSV
A.lyrata2                 SWREGLQSEA AKAIANLSVN AKVAKAVAEE GGISVLADLA KSMNRLVAEE AAGGLWNLSV
Hordidillo                CSRVSAQSEA AKAIANLSVN AKVAKVVVDE GGIAIFTNLA KSTNRLVAEE AAGGLWNLSV

                          541
ARABIDILLO1               GEEHKNAIAQ AGGVKALVDL IFRW--PNGC DGVLERAAGA LANLAADDKC SMEVAKAGGV
ARABIDILLO2               GEEHKNAIAQ AGGVNALVDL IFRW--PHGC DGVLERAAGA LANLAADDKC SMEVARAGGV
Physcodillo1A             GEEHKGAIAG AGAIEALVGL AFKW--PAGG EGVLERAAGA LANLAADDKC SMEVAVAGGV
Physcodillo2              GEEHKGAIAE AGAIEALVDL AFKW--PAGG EGVLERAAGA LANLAADDKC SMEVAVAGGV
Physcodillo1B             GEEHKGAIAG AGAIEALVGL AFKW--PAGG EGVLERAAGA LANLAADDKC SMEVAVAGGV
Oryzadillo                GEEHKAAIAA AGGIKALVDL ILRW--PAGT DGVLERAAGA LANLAADDKC SMEVAKAGGV
Brachydillo               GEEHKASIAA AGGIKALVDL IFRW--PAGT DGVLERAAGA LANLAADDKC SLEVAKAGGV
Vitidillo1                GEEHKGAIAE AGGVKSLVDL IFKW--SAGG DGVLERAAGA LANLAADDKC SMEVALAGGV
Vitidillo2                GEEHKGAIAE TGGIRALVDL IFKW--QSAG DGVLERAAGA LANLAADDKC SMEVAMVGGV
Zeadillo1                 GEDHKAAIAV SGGIKALVDL IFRW--PAGT DGVLERAAGA LANLAADDKC SLEVAKAGGV
Sorghodillo               GEDHKAAIAV SGGIKALVDL IFRW--PAGT DGVLERAAGA LANLAADDKC SLEVAKAGGV
Glycinodillo              GEEHKGAIAE AGGIQALVDL IFKW--SSSG DGVLERAAGA LANLAADDKC STEVALAGGV
Poplardillo1              GEEHKGAIAE AGGVKALVDL IFKW--FSGG DGVLERAAGA LANLAADDKC SMEVALAGGV
Poplardillo2              GEEHKGAIAE AGGVKALVDL IFKW--SSGS DGVLERAAGA LANLAADDKC SMEVALAGGV
Poplardillo3              GEEHKGAIAE AGGIKVLIDL IYKW--HAGN DGVLERAAGA LANLAADDSC SMEVAVAGGV
Ricinodillo1              GEEHKGAIAE AGGIKALVDL IFKW--SSGG DGVLERAAGA LANLAADDKC SMEVALAGGV
Ricinodillo2              GEEHKGAIAE AGGIKALVDL IFKW--PSSS DGVVERAAGA LANLAADDKC SMEVAMAGGV
Manihodillo1              GEEHKGSIAE AGGVKALVDL IFKW--SSGG DGVLERAAGA LANLAADDKC SMEVALAGGV
Manihodillo2              GEEHKGAIAE AGGIKALVDL IFKW--PFCN DGVLERAAGA LANLAADDKC SMEVAMAGAV
Cucumidillo               GEEHKGAIAE AGGVRALVDL IFKW--SSGG DGVLERAAGA LANLAADDRC STEVALAGGV
Prunidillo                GEEHKGAIAE AGGVKALVDL IFKW--SSGG DGVLERAAGA LANLAADDKC STEVAVAGGV
Citrus.sinidillo          GEEHKGAIAD AGGVKALVDL IFKW--SSGG DGVLERAAGA LANLAADDKC SMEVALAGGV
Citrus.clemidillo         GEEHKGAIAD AGGVKALVDL IFKW--SSGG DGVLERAAGA LANLAADDKC SMEVALAGGV
Eucadillo                 GEEHKGAIAE AGGVKALVDL IFKW--SSGG DGVLERAAGA LANLAADDKC SMEVALAGGV
Mimulodillo1              GEEHKGAIAE AGGVKALVDL IYKWSQSSCG EGVLERAAGA LANLAADEKC STAVASMGGV
Mimulodillo2              GDDHKGTIAE AGAVKALVDI IYKWSRSGGG DGVLERAAGA LANLAADDGC SREVASAGGV
Aquilidillo1              GEEHKGAIAV AGGIKALVDL IFRW--PSGI DGVLERAAGA LANLAADDKC STEVALAGGV
Selagidillo               GEEHKGAIAD AGAIEALVDL ALKW--PAGG EGVLERAAGA LANLAADDKC SMKVANAGGV
A.Lyrata1                 GEEHKNAIAL AGGVKALVDL IFRW--PNGC DGVLERAAGA LANLAADDKC SMEVATAGGV
A.lyrata2                 GEEHKNAIAQ AGGVNALVDL IFRW--PNGC DGVLERAAGA LANLAADDKC STEVARAGGV
Hordidillo                GEEHKAAIAA AGGIKALVDI IFRW--PAGT DGVLERAAGA LANLAADDKC SLEVAKAGGV
```

```
                          601
ARABIDILLO1               HALVMLARNC KYEGVQEQAA RALANLAAHG DSNNNNAAVG QEAGALEALV QLTKSPHEGV
ARABIDILLO2               HALVMLARNC KYEGAQEQAA RALANLAAHG DSNGNNAAVG QEAGALEALV QLTQSPHEGV
Physcodillo1A             RALVRLARFC NHEGVQEQAA RALANLAAHG DSNGNNAAVG REEGALEALV QLTCSNHEGV
Physcodillo2              RALVRLAQFC NHEGVQEQAA RALANLATHG DSNGNNAAVG REAGALEALV RLTGSNHEGV
Physcodillo1B             RALVRLARFC NHEGVQEQAA RALANLAAHG DSNGNNAAVG REEGALEALV QLTCSNHEGV
Oryzadillo                HALVMLARSC KLEGVLEQAA RALANLAAHG DNNNNNAAVG QEAGALEALV QLTSSQNEGV
Brachydillo               HALVTLARSC KLEGVLEQAA RALANLAAHG DNNNNNAAVG QEAGALEALV QLTCSQNEGV
Vitidillo1                HALVMLARNC KFEGVQEQAA RALANLAAHG DSNSNNAAVG QEAGALEALV LLTKSPHEGV
Vitidillo2                HALVMLARSC KFEGVQEQAA RALANLAAHG DSNNNNSAVG QEAGALEALV QLTCSQHEGV
Zeadillo1                 HALVTLARSC KLDGVLEQAA RGLANLAAHG DNNDNNAAVG QEAGALEALV QLTSSQNEGV
Sorghodillo               HALVTLARSC KLDGVLEQAA RGLANLAAHG DNNDNNAAVG QEAGALEALV QLTGSQNEGV
Glycinodillo              HALVMLARNC KFEGVQEQAA RALANLAAHG DSNSNNAAVG QEAGALEALV QLTCSPHEGV
Poplardillo1              HALVMLARNC KFEGVQEQAA RALANLAAHG DSNTNNAAVG QEAGALEALV QLTRSLHEGV
Poplardillo2              HALVMLARNC KFEGVQEQAA RALANLAAHG DSNSNNAAVG QEAGALEALV QLTRSLHEGV
Poplardillo3              HALVMLARSC KFEGVQEQAA RALANLAAHG DNNNDNAAVR REAGALEALV QLTSSQHEGV
Ricinodillo1              HALVMLARNC KFEGVQEQAA RALANLAAHG DSNTNNAAVG QEAGALEALV QLTRSPHEGV
Ricinodillo2              HALVMLARSC KFEGVQEQAA RALANLAAHG DSNSNNAAVG QEAGALEALV QLTCSQHEGV
Manihodillo1              HALVMLARNC KFEGVQEQAA RALANLAAHG DSNTNNAAVG REAGALEALV QLTRSPHEGV
Manihodillo2              HALVMLVRHC NFEGVQEQAA RALANLAAHG DSNNNNAAIG QEAGAIEALV RLTYSQHEGV
Cucumidillo               HALVMLARNC KFEGVQEQAA RALANLAAHG DSNTNNSAVG QEAGALEALV QLTHSPHEGV
Prunidillo                QALVMLARNC KFEGVQEQAA RALANLAAHG DSNSNNAAVG QEAGALEALV QLTQSPHEGV
Citrus.sinidillo          HALVMLARSC KFEGVQEQAA RALANLAAHG DSNSNNSAVG QEAGALEALV QLTRSPHEGV
Citrus.clemidillo         HALVMLARSC KFEGVQEQAA RALANLAAHG DSNSNNSAVG QEAGALEALV QLTRSPHEGV
Eucadillo                 HALVMLARNC KFEGVQEQAA RALANLAAHG DSNGNNAAVG QEVGALEALV QLTRSPHEGV
Mimulodillo1              HALVTLARNC KIEGVQEQAA RALANLAAHG DSNSNNTAVG KEAGAIDALV QLTRAPHDGV
Mimulodillo2              QALVMLARTY KVEGVQEQAA RALANLAAHG DSNTNNAAVG QEAGALEALL QLTRSTHDGV
Aquilidillo1              HALVTIARSC KFEGVQEQAA RALANLAAHG ESNGNNAAIG QEAGALEALV QLTCSNHEGV
Selagidillo               NALVNLARFC KHEGVQEQAA RALANLAAHG DSNGNNAAVG REAGALEALV KLTCSNHEGV
A.Lyrata1                 HALVMLARNC KYEGVQEQAA RALANLAAHG DSNNNNAAVG QEAGALEALV QLTQSLHEGV
A.lyrata2                 HALVMLARNC KYEGAQEQAA RALANLAAHG DSNDNNAAVG QEAGALEGLV QLTQSPHEGV
Hordidillo                HALVTLARSC KLEGVLEQAA RALANLAAHG DNNNNNAAVG QEAGALEALV QLTCSQNEGV

                          661
ARABIDILLO1               RQEAAGALWN LSFDDKNRES ISVAGGVEAL VALAQSCSNA STGLQERAAG ALWGLSVSEA
ARABIDILLO2               KQEAAGALWN LAFDDKNRES IAAFGGVEAL VALAKSSSNA STGLQERVAG ALWGLSVSEA
Physcodillo1A             RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQGCSSG SQGLQERAAG ALWGLSVSEA
Physcodillo2              RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQDCSSG SQGLQERAAG ALWGLSVSEA
Physcodillo1B             RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQGCSSG SQGLQERAAG ALWGLSVSEA
Oryzadillo                RQEAAGALWN LSFDDRNREG IAAAGGVEAL VSLAQECLNA SEGLQERAAG ALWGLSVSEA
Brachydillo               RQEAAGALWN LSFDDRNREA IAAAGGVQAL VSLAQECLNA SEGLQERAAG ALWGLSVSES
Vitidillo1                RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQSCSNA SPGLQERAAG ALWGLSVSEA
Vitidillo2                RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQTCSNA SQGLQERAAG ALWGLSVSEA
Zeadillo1                 RQEAAGALWN LSFDDRNREA IAAVGGVEAL VALVQQCLNA SEGLQERAAG ALWGLSVSEA
Sorghodillo               RQEAAGALWN LSFDDRNREA IAAVGGVEAL VALVQQCLNA SEGLQERAAG ALWGLSVSEA
Glycinodillo              RQEAAGALWN LSFDDRNREA IAAAGGVQAL VALAQACANA SPGLQERAAG ALWGLSVSET
Poplardillo1              RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQSCGNA SPGLQERAAG ALWGLSVSEA
Poplardillo2              RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQSCANA SPGLQERAAG ALWGLSVSEA
Poplardillo3              RQEAAGALWN LSFDDKNREA IAAAGGITAL VSLAQSCSNS SQSLQERAAG ALWGLSVSEA
Ricinodillo1              RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQSCSNA SPGLQERAAG ALWGLSVSEA
Ricinodillo2              RHEAAGALWN LSFDDKNREA IAAAGGIVAL VSLAQSCSNS SQGLQERAAG ALWGLSVSEA
Manihodillo1              RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQACSNA SPGLQERAAG ALWGLSVSEA
Manihodillo2              RHEAAGALWN LSFDDKNREA IAAAGGIVAL VSLAQSCSNS SQGLQERAAG ALWGLSVSEA
Cucumidillo               RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQSCSNA SPGLQERAAG ALWGLSVSEA
Prunidillo                RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQGCSNA SPGLQERAAG ALWGLSVSEA
Citrus.sinidillo          RQEAAGALWN LSFDDRNREA IAAAGGVEAL VVLAQSCSNA SPGLQERAAG ALWGLSVSEA
Citrus.clemidillo         RQEAAGALWN LSFDDRNREA IAAAGGVEAL VVLAQSCSNA SPGLQERAAG ALWGLSVSEA
Eucadillo                 RQEAAGALWN LSFDDRNREA IASAGGVEAL VALAQTCSNA SPGLQERAAG ALWGLSVSEA
Mimulodillo1              RQEAAGALWN LSFHDRNREA IATAGGVEAL VALAHSCSNA SHGLQERAAG ALWGLSVSEA
Mimulodillo2              RQEAAGALWN LSFDDRNREA IAAAGGVEAL VSLAQSCSGS SHGLQERAAG ALWGLSVSET
Aquilidillo1              RQEAAGALWN LSFDDRNREA IAAVGGVEAL VSLAQNCSNA SQGLQERAAG ALWGLSVSEA
Selagidillo               RQEAAGALWN LSFDDRNREA IAAAGGVEAL VALAQGCSNG SQGLQERAAG ALWGLSVSEE
A.Lyrata1                 RQEAAGALWN LSFDDKNRES IAVAGGVEAL VVLAQSCSNA STGLQERAAG ALWGLSVSEA
A.lyrata2                 KQEAAGALWN LAFDDKNRES IAASGGVEAL VALAKSCSNA STGLQERAAG ALWGLSVSEA
Hordidillo                RQEAAGALWN LSFDDRNREA IAAAGGVEAL VSLAQQCLNA SEGLQERAAG ALWGLSVSES
```

```
                    721
ARABIDILLO1         NSVAIGREGG VPPLIALARS EAEDVHETAA GALWNLAFNP GNALRIVEEG GVPALVHLCS
ARABIDILLO2         NSIAIGHEGG IPPLIALVRS EAEDVHETAA GALWNLSFNP GNALRIVEEG GVVALVQLCS
Physcodillo1A       NSIAIGREGG VAPLITLAHS DFEDVHETAV GALWNLVFNP GNALRMVEEE GVPALVHLCS
Physcodillo2        NSIAIGREGG VAPLITLAHS NSEDVHETAV GALWNLAFNP GNALRMAEE- GVPALVHLCS
Physcodillo1B       NSIAIGREGG VAPLITLAHS DFEDVHETAV GALWNLVFNP GNALRMVEEE GVPALVHLCS
Oryzadillo          NSMAIGQEGG VAPLLTLAQS DVEDVHETAA GALWNLAFYS GNALCIVEEG GVPILVRLCS
Brachydillo         NSIAIGQEGG VAPLLTMAQS DAEDVHETAA GALWNLAFYS SNALRIVEEG GVPILVHLCS
Vitidillo1          NSIAIGREGG VAPLIALARS DAEDVHETAA GALWNLAFNP GNALRIVEEG GVPALVHLCA
Vitidillo2          NSIAIGRQGG VAPLIALARS NVEDVHETAA GALWNLAFNP HNALRIVEDG GVQALVNLCS
Zeadillo1           NSIAIGQGGG VAPLLTLARS EVEDVHETAA GALWNLAFYS GNALRIVEEG GVPVLVKICS
Sorghodillo         NSIAIGQGGG VAPLLTLARS EVEDVHETAA GALWNLAFYY GNALRIVEEG GVPVLVKICS
Glycinodillo        NSVAIGREGG VAPLIALARS EAEDVHETAA GALWNLAFNA SNALRIVEEG GVSALVDLCS
Poplardillo1        NSIAIGREGG VVPLIALARS ETEDVHETAA GALWNLAFNP GNALRIVEEG GVPALVDLCS
Poplardillo2        NSIAIGQEGG VAPLIALARS EAEDVHETAA GALWNLAFNR GNALRIVEEG GVPALVDLCS
Poplardillo3        NSIAIGQEGG VAPLIVLACS DIADVHETAA GALWNLAFYP TNALRIVEGG GVPALIHLCS
Ricinodillo1        NSIAIGREGG VAPLIALARS EAEDVHETAA GALWNLAFNP GNALRIVEEG GVPALVHLCS
Ricinodillo2        NSVAIGQVGG VAPLIVLARS DVTDVHETAA GALWNLAFNP GNALRIVEDG GVPALVCLCT
Manihodillo1        NSIAIGREGG VAPLIALARS EAEDVHETAA GALWNLAFNP GNALRIVEEG GVPALVHLCS
Manihodillo2        NSVAIGQEGG VASLIALARS SDADVHETAA GALWNLAFNP GNALRIVEDG GVPALVHICT
Cucumidillo         NSIAIGQQGG VAPLIALARS DAEDVHETAA GALWNLAFNP GNALRIVEEG GVPALVHLCY
Prunidillo          NSIAIGREGG VVPLIALARS EAADVHETAA GALWNLAFNP GNALRIVEEG GVPALVNLCS
Citrus.sinidillo    NCIAIGREGG VAPLIALARS EAEDVHETAA GALWNLAFNP GNALRIVEEG GVPALVHLCS
Citrus.clemidillo   NCIAIGREGG VAPLIALARS EAEDVHETAA GALWNLAFNP GNALRIVEEG GVPALVHLCS
Eucadillo           NSVAIGREGG VAPLIALARS DAEDVHETAA GALWNLAFNP GNALRIVEEG GVPALVHLCS
Mimulodillo1        NSIAIGREGG VAPLIALARS NAVDVHETAA GALWNLAFNS GNALRIVEEG GVPALVHLCS
Mimulodillo2        NSIAIGQEGG VAPLIALARS DAEDVHETAA GALWNLAFNP GNALRIVEEG GVPDLVHLCS
Aquilidillo1        NSIAIGQGGG VAPLIALASS DAEDVHETAA GALWNLAFNA GNAYRIVEEG GVPALVHLCS
Selagidillo         NSIAIGREGG VAPLVALARS DAEDVHETAA GALWNLAFNP GNALRIVEED GVSALVRLCS
A.Lyrata1           NSVAIGREGG VPPLIALARS EAEDVHETAA GALWNLAFNP GNALRIVEEG GVPALVHLCS
A.lyrata2           NSIAIGHGGG IPPLITLALS EAEDVHETAA GALWNLAFNP GNALRIVEEG GVVALVHLCS
Hordidillo          NSIAIGQEGG VAPLLTMAQS EVEDVHETAA GALWNLAFYS SNAQRIVEEG GVPILVHLCS

                    781
ARABIDILLO1         SSVSKMARFM AALALAYMFD GRMDEYALMI --GTSSSEST SKNISLDGAR NMALKHIEAF
ARABIDILLO2         SSVSKMARFM AALALAYMFD GRMDEYA-MI --GT-SLEST SKSVTLNGAR TMALDQIKAF
Physcodillo1A       SSRSKMARFM AALALAYMFD GRMDEVA--V --GLSSVENN GRTVSLEAAR KLALRNIDAF
Physcodillo2        SSRSKMARFM AALALAYMFD GRMDEVA--V --RVSSGENH GRTVNLEAIR KLALRSIDAF
Physcodillo1B       SSRSKMARFM AALALAYMFD GRMDEVA--V --GLSSVENN GRTVSLEAAR KLALRNIDAF
Oryzadillo          SSGSKMARFM SALALAYMFD GRMDEVA-LV --GT-SSEGS SKSVNVEGAR RMALKHIQTF
Brachydillo         SSGSKMARFM SALALAYMFD RRMDEVA-IV --GT-SSEGS SKGATVEGAR RMALKHIQLF
Vitidillo1          SSVSKMARFM AALALAYMFD GRMDEFA-LI --GT-SSEST SKSVSLDGAR RMALKHIETF
Vitidillo2          YSLSKMARFM AALALAYMFD GRMDEVA-LV --GP-SSEGA SKSRNINGAR KMALKNIEAF
Zeadillo1           SSRSKMARFM SALALAYMFD GRMDEVA-LV --GA-SSDSS SKSVNVEGAR RIAFKHIETF
Sorghodillo         SSGSKMARFM SALALAYMFD GRMDEVA-LV --GA-SSDSS SKSVNVEGAR RIAFKHIETF
Glycinodillo        SSVSKMARFM SALALAYMFD GRMDEYA-LV --VT-SSESI SKSVSLDGAR RMALKHIEAF
Poplardillo1        LSASKMARFM AALALAYMFD RRMDEVA-PI --GT-LTEST SKSANLDGAR RMALKHIEAF
Poplardillo2        SSVSKMARFM AALALAYMFD GRMDEFA-LI --GT-STESI SKSVNLDGAR RMALKHIEAF
Poplardillo3        SSVSKMARLM AALALAYMFD QRTEEFA-PG --GP-LSEGT SKTMNFDEIK RVALKKIEAF
Ricinodillo1        SSVSKMARFM AALALAYMFD GRMDEFA-LI --GT-STEST SKSVSLDGAR RMALKHIEAF
Ricinodillo2        SSLSKMARFM AALALAYVFD GRMDEIA-PV GPGP-SLEGV LKSVNLATVK RIALKHVEAF
Manihodillo1        SSVSKMARFM AALALAYMFD GRMDEFG-LM --GT-STEST SKSVSLDGAR RMALKHIEAF
Manihodillo2        KSLSKMARFM AALALAYIFD GRMDEIA-AV --GP-SSDGD IKSVTLNVVK RISLKHIEAF
Cucumidillo         ASVSKMARFM AALALAYMFD GRMDECA-LP --GS-SSEGI SKSVSLDGAR RMALKNIEAF
Prunidillo          SSVSKMARFM AALALAYMFD GRMDEFA-LI --GT-SSESI SKSVSLDGSR RMALKHIEAF
Citrus.sinidillo    SSGSKMARFM AALALAYMFD GRMDEFA-LI --GT-STEST SKCVSLDGAR RMALKHIEAF
Citrus.clemidillo   SSGSKMARFM AALALAYMFD GRMDEFA-LI --GT-STEST SKCVSLDGAR RMALKHIEAF
Eucadillo           SSVSKMARFM AALALAYMFD GRMDEYA-LV --GT-SSEST SKIVSLDGSR RMALKHIEAF
Mimulodillo1        SSLSKMARFM SALALSYMFD GRMDDIA-PT --GTTSTEGT SKSVNLDGAR RMALKNIEAF
Mimulodillo2        SSVSKMARFM SALALAYMFD GRLDEIA-VV --GT-STESG SKSVNLDGFR RMAMKHIEAF
Aquilidillo1        SSVSKMARFM AALALAYMFD GRMDYIG-VP --GP-YSDSA VKSVNIDDAR KMALRQIEAF
Selagidillo         SSRSKMARFM AALALAYMFD GSITKN---- --GV------ --------AR QSAMKNIEAF
A.Lyrata1           SSVSKMARFM AALALAYMFD GRMDEYALMI --GTSSSEST SKSISLDGAR NMALKHIEAF
A.lyrata2           SSVSKMARFM AALALAYMFD GRMDEYA-MI --GT-SSEST SKSVALNSAR TLALEHIKAF
Hordidillo          SSGSKMARFM SALALAYMFD GRMDEAA-IV --GT--SEGS SKGVNVEGAR RMALKHIETF
```


841
ARABIDILLO1 VLSFIDPHIF ESPVVS-STP TMLAQVTERA RIQEAGHLRC S-GAEIGRFV TMLRNPDSTL
ARABIDILLO2 IKTFMEHQIF STGALS-SAP SMLAQVSERA RIPEAGHLRC S-GSEIGRFV TMLRNPCLVL
Physcodillo1A VQTFFDPQSL TAAASS-WAG ATLNQVAETA TIQEAGHLRC S-VAELGRFV AMLRNGHAVL
Physcodillo2 VLTFCDQQAL TAAASS-WAP ATLNQVAETA RIQEAGLLRC SGGAELGRFV AMLRNGSAVL
Physcodillo1B VQTFFDPQSL TAAASS-WAG ATLNQVAETA TIQEAGHLRC S-VAELGRFV AMLRNGHAVL
Oryzadillo VLTFSDPQVF TTASTS-SAS AALSQIADAV FIQEAGHLRC S-GAEIARFV AMLRNPASIL
Brachydillo VLTFSDPQVF STAAAS-SAP AALSQVAEAV FIQEAGHLRC S-GAEIGRFV AMLRNPTSVL
Vitidillo1 ILTFSDPQSF SAAAVS-SAP AALAQVTESA RIQEAGHLRC S-GAEIGRFV AMLRNPSSIL
Vitidillo2 VLTFTNPHTF GLALAS-SAP TALVQVIEMA CIQEAGHLRC S-GAEIGRFV TMLKNPSPVL
Zeadillo1 VLTFSDPQMF SMAAAS-SAP AALSHVAEAV FIHEAGHLRC S-RSEIGRFV SMLRNPSPIL
Sorghodillo VLTFSDPQMF SMAAAS-SAP AALSHVAEAV FIHEAGHLRC S-RSEIGRFV AMLRNPSPIL
Glycinodillo VLMFSDLQAF AAAAAS-SAP AALAQVTEGA RIQEAGHLRC S-GAEIGRFI TMLRNPSSIL
Poplardillo1 VLTFSDPQAF ATAAAS-SAP AALAQVTERA RIQEAGHLRC S-GAEIGRFV AMLRNPSSIL
Poplardillo2 VLTFTDPQAF ATAAAS-SAP AALAQVTERA RIQEAGHLRC S-VAEIGRFV AMLRNPSSIL
Poplardillo3 VLSFSVPQNF AAAMVS-SAP KALGQVAEAV RIPEAGHLRC S-AAEIGRFV AMLRNPSSIL
Ricinodillo1 VLTFSDQQTF AVAAAS-SAP AALAQVTERA RIQEAGHLRC S-GAEIGRFV TMLRNSSSIL
Ricinodillo2 VRSFSASQTF ATVLAS-SAP TTITRVAEAA RIPEAGLLRC S-AAEIKRFV AMLRNHSSIL
Manihodillo1 VLTFSDQQTF AIAAAS-SAP ASLTQLTERA RIPEAGHLRC S-GAEIGRFV TMLRNPSSTL
Manihodillo2 VCSFSDSKTS DTVIKL-SAP TALTQVAEAA CIPEAGLLRC S-AAEIGRFV AMLRNHSSIL
Cucumidillo VQTFSDPQAF ASAAAS-SAP AALVQVTERA RIQEAGHLRC S-GAEIGRFV AMLRNPSPTL
Prunidillo VLTFSDQQTF SAAAAS-SAP AALAQVTEGA RIQEAGHLRC S-GAEIGRFV TMLRNPSSVL
Citrus.sinidillo VLTFSDPQAF ATAAAS-SAP AALTQVTERA RIQEAGHLRC S-GAEIGRFI TMLRNPSSVL
Citrus.clemidillo VLTFSDPQAF ATAAAS-SAP AALTQVTERA RIQEAGHLRC S-GAEIGRFI TMLRNPSSVL
Eucadillo ILMFSDRQAF TAAAAS-SAP AALAEVTEGA RIQEAGHLRC S-GAEIGRFI SMLKNPS-VL
Mimulodillo1 VMTFSDIRAF SAAAAS-VGP AALTHVTESA RIEEAGHLRC S-GAEIGRFV AMLRNPNPTL
Mimulodillo2 LLTFSDPQAF AAAAAS-SAP TALTQITEAA RIQEAGHLRC S-GAEIGRFV LMLRNPSPIL
Aquilidillo1 VLTFSDPQSF YAAAAS-SAP TSLAQVTEAA RIQEAGHLRC S-GAEIGRFV LMLRNQYPIL
Selagidillo VQAFSDQPSL AAVPASQWGP SALQQVSDSA RIQEAGHLRC S-GAEIGRFV AMLRNGSSVL
A.Lyrata1 VLTFIDPHIF ESPVVS-STP TMLAQVTERA RIQEAGHLRC S-GAEIGRFV TMLRNPDSTL
A.lyrata2 ISTFMEHQIF SAGALS-SAP SMLAQVSEKA RIPEAGHLRC S-GSEIGRFV TMLRNPCLTL
Hordidillo VLTFSDPQVF STAAAS-SAP AALSQVAEAV FIQEAGHLRC S-CAEIGRFI AMLRNPTPVL
901
ARABIDILLO1 KACAAFALL- -----QFTIP GGRHAMHHVS LMQNGGESRF LRSAAASAKT PREAKIFTKI
ARABIDILLO2 RACAAFALL- -----QFTIP ESRHAMHHAS LMQNAGEARG LRSAAAAASM PREAKIFMKI
Physcodillo1A RSCAAFALL- -----QFTMP GGRHALHHAV LLQKVGASRV LRTVAAAATA AMQAKVFARV
Physcodillo2 RTCAAFALL- -----QFTMP GGRHAMHHAD LLQKTGASRV LRTVAAAATA AMQAKVFARI
Physcodillo1B RSCAAFALL- -----QFTMP GGRHALHHAV LLQKVGASRV LRTVAAAATA AMQAKVFARV
Oryzadillo RACAAFALL- -----QFTIP GGRHAVHHAG LLQKAGAARV LRAAAAATTA SIEAKVFARI
Brachydillo RACAAFALL- -----QFTIP GGRHAVHHAG LLQKAGAARV LRAAAAATSA SIEAKVFARI
Vitidillo1 KSCAAFALL- -----QFSIP GGRHAVHHAT LLQSVGAARV LRGAAAAATA PIEAKIFARI
Vitidillo2 KSCAAFALL- -----QFTIP GSRHAVHHAS LLQKAVALRT LRAAAAAATA PVEAKVFARI
Zeadillo1 RACAAFALL- -----QFTIP GGRHAVHHAG LLQEAGAGRV LRAAAAATTA SIEAKIFARI
Sorghodillo RACAAFALL- -----QFTIP GGRHAVHHAG LLQEAGAGRV LRAAAAATTA SIEAKIFARI
Glycinodillo KACAAFALL- -----QFTIP GGRHAMHHAS LMQSLGASRV LRGAAAAATA PLEAKIFARI
Poplardillo1 KACAAFALL- -----QFTIP GGRHALHHAS LMQSAGAARV LRPAAAAATA PLEAKIFARI
Poplardillo2 KACAAFALL- -----QFTIP GGRHALHHAS LMQSAGAARV LRAAAAAATA PLEAKIFARI
Poplardillo3 KACSAFALL- -----QFTIP GGRHTLHHTS LLQNAGAPRV LRAAAAATTA PIEAKVFAKI
Ricinodillo1 KACAAFALL- -----QFTIP GGRHAMHHAS LMQNAGAARV VRAAAAAATA PLEAKIFARI
Ricinodillo2 KACSAFALL- -----QFTMP GGRHAVHHTS LLQDAGASRT LRALAAAASA PIETKVFAKI
Manihodillo1 KTCAAFALL- -----QFTIP GGRHAMHHAS LMQNAGATRV VRAAAAAATA PLEAKIFARI
Manihodillo2 KACSAFALL- -----QFTMP GGRHATHHTN LLQNAGAPRI LRAVAASASA PFEAKVFAKI
Cucumidillo KACAAFALLQ ASEWIQFTIP GGRHALHHAS LMQNAGASRA LRTAAAAATA PLQAKIFARI
Prunidillo KACAAFALL- -----QFTIP GGRHAMHHAS LMQNAGAARV LRAAAAAATA PLEAKIFARI
Citrus.sinidillo KSCAAFALL- -----QFTIP GGRHAMHHAS LMQGAGAARV LRAAAAAAAA PIEAKIFARI
Citrus.clemidillo KSCAAFALL- -----QFTIP GGRHAMHHAS LMQGAGAARV LRAAAAAAAA PIEAKIFARI
Eucadillo KACAAFALL- -----QFTIP GGRHALHHAK LMQNAGAARV LRATAAAATA PIEAKIFARI
Mimulodillo1 KSCAAFALL- -----QFTIP GGRHAMHHVG LLQKAAAPRL LRVAAAAAGA PIEAKIFSRI
Mimulodillo2 KGCAAFALL- -----QLGMY KIDILIFFFH AYTAIACHKV SYGSGNSK-- ----------
Aquilidillo1 KACAAFALV- -----QFTIP GGRHAMYHAN LMQQSGAHRS LRGAAAAATA PIEAKIFARI
Selagidillo RSCAAFALL- -----QFTMP GGRHALHHAN LLQRSGAARV LRGAAASTTA PLQARVFARL
A.Lyrata1 KACAAFALL- -----QFTIP GGRHAMHHVS LMQNGGESRF LRSAAASAKT PREAKIFTKI
A.lyrata2 RSCAAFALL- -----QFTIP GGRHAMHHAS LMQNAGEARV LRSAAAAAKM PREAKIFVKI
Hordidillo RACAAFALL- -----QFSIP GGRHAIHHAD LLQNVGAARV LRAAAAATSA SIEAKVFAKI

```
                    961
ARABIDILLO1         LLRNLE-HHQ AESSI----- -----
ARABIDILLO2         VLRNLE-HQQ AESPEGMKVS YNRI-
Physcodillo1A       VLRNLE-HHQ AESNTVVALS *----
Physcodillo2        VLRNLE-HHQ AESNAVVALL -----
Physcodillo1B       VLRNLE-HHQ AESNTVVALS *----
Oryzadillo          VLRNLE-HHQ TGTST----- -----
Brachydillo         VLRNLE-HHQ AGTSI*---- -----
Vitidillo1          VLRNLE-HHQ MEQSI----- -----
Vitidillo2          VLRNLE-HYQ VEASI----- -----
Zeadillo1           VLRNLE-HHQ LGMST*---- -----
Sorghodillo         VLRNLE-HHQ SGTST*---- -----
Glycinodillo        VLRNLE-YHQ IEQALA*--- -----
Poplardillo1        VLRNLE-YHH IESSI----- -----
Poplardillo2        VLRNLE-FHH IESSI----- -----
Poplardillo3        ILRNLE-HHH LEALN----- -----
Ricinodillo1        VLRNLE-HHQ IEPSI----- -----
Ricinodillo2        VLRNLE-HKH L--------- -----
Manihodillo1        VLRNLE-HHQ MEPSIGRLSN LA*--
Manihodillo2        VIWNLE-HYH VDANSRNNG* -----
Cucumidillo         VLRNLE-HHS VESSL----- -----
Prunidillo          VLRNLE-HHH IEPSL*---- -----
Citrus.sinidillo    VLRNLEHHHH VELSI*---- -----
Citrus.clemidillo   VLRNLEHHHH VELSI*---- -----
Eucadillo           VLRNLE-HHQ AESSKGL*-- -----
Mimulodillo1        VLRNLE-QHQ S--------- -----
Mimulodillo2        -------YID LSPNTA---- -----
Aquilidillo1        VLRNLE-HHF GEVAI*---- -----
Selagidillo         VLRNLE-LCQ SEKS------ -----
A.Lyrata1           ILRNLE-HHQ AESSI*---- -----
A.lyrata2           VLRNLE-HQQ AESSKGKKVS YNRI*
Hordidillo          VLRNLE-HHQ AGTTST---- -----
```