ARABIDILLO gene homologues in basal land plants

Moody, L.A.; Saidi, Y.; Smiles, E.J.; Bradshaw, S.J.; Meddings, M.; Coates, J.C.; Winn, P.J.; Coates, Juliet

DOI:

[10.1007/s00425-012-1742-7](https://doi.org/10.1007/s00425-012-1742-7)

License:

Creative Commons: Attribution-ShareAlike (CC BY-SA)

Document Version

Early version, also known as pre-print

Citation for published version (Harvard):

Moody, LA, Saidi, Y, Smiles, EJ, Bradshaw, SJ, Meddings, M, Coates, JC, Winn, PJ & Coates, J 2012, 'ARABIDILLO gene homologues in basal land plants: Species-specific gene duplication and likely functional redundancy', *Planta*, vol. 236, no. 6, pp. 1927-1941. <https://doi.org/10.1007/s00425-012-1742-7>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

A definitive version was subsequently published in PUBLICATION, *Planta* December 2012, Volume 236, Issue 6, pp 1927-1941 DOI: <http://dx.doi.org/10.1007/s00425-012-1742-7>. The final publication is available at link.springer.com.

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

ARABIDILLO gene homologues in basal land plants: species-specific gene duplication and likely functional redundancy.

Journal:	<i>Planta</i>
Manuscript ID:	Planta-2012-06-0407.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Moody, Laura; School of Biosciences, University of Birmingham Saidi, Younousse; School of Biosciences, University of Birmingham Smiles, Emma; School of Biosciences, University of Birmingham Bradshaw, Susan; School of Biosciences, University of Birmingham Meddings, Matthew; School of Biosciences, University of Birmingham Winn, Peter; School of Biosciences, University of Birmingham Coates, Juliet; University of Birmingham, School of Biosciences
Keywords:	Development, Evolution, Physcomitrella, Gene locus, sequencing, promoter activity

***ARABIDILLO* gene homologues in basal land plants: species-specific gene duplication and likely functional redundancy.**

Laura A. Moody[^], Younouss Saidi, Emma J. Smiles, Susan J. Bradshaw, Matthew Meddings, Peter J. Winn[%], Juliet C. Coates*.

School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

[^] Current address: Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK.

[%] Centre for Systems Biology, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

*Author for correspondence: Email: j.c.coates@bham.ac.uk; Tel: +44 121 414 5478; Fax +44 121 414 5925.

Abstract

ARABIDILLO proteins regulate multicellular root development in *Arabidopsis thaliana*. Conserved ARABIDILLO homologues are present throughout land plants, even in early-evolving plants that do not possess complex root architecture, suggesting that ARABIDILLO genes have additional functions. Here, we have cloned and characterised ARABIDILLO gene homologues from two early-evolving land plants, the bryophyte *Physcomitrella patens* and the lycophyte *Selaginella moellendorffii*.

We show that two of the PHYSCODILLO genes (PHYSCODILLO1A and -1B) exist as a tail-to-tail tandem array of two almost identical 12kb sequences, while a third related gene (PHYSCODILLO2) is located elsewhere in the *Physcomitrella* genome. *Physcomitrella* possesses a very low percentage of tandemly arrayed genes compared to the later-evolving plants whose genomes have been sequenced to date. Thus, PHYSCODILLO1A and -1B genes represent a relatively unusual gene arrangement. PHYSCODILLO promoters are active largely in the haploid gametophyte, with additional activity at the foot of the sporophyte. The pattern of promoter activity is uniform in filamentous and leafy tissues, suggesting pleiotropic gene functions and likely functional redundancy: the latter possibility is confirmed by the lack of discernible phenotype in a *physcodillo2* deletion mutant. Interestingly, the pattern of PHYSCODILLO promoter activity in female reproductive organs is strikingly similar to that of an *Arabidopsis* homologue, suggesting co-option of some PHYSCODILLO functions or regulation into both the sporophyte and gametophyte.

In conclusion, our work identifies and characterises some of the earliest-evolving land plant ARABIDILLO homologues. We confirm that all land plant ARABIDILLO genes arose from a single common ancestor and suggest that PHYSCODILLO proteins have novel and pleiotropic functions, some of which may be conserved in later-evolving plants.

Keywords:

Development, evolution, *Physcomitrella*, gene locus, sequencing, promoter activity

Introduction

One of the most important events in the history of life on earth was the colonisation of the terrestrial environment by green, multicellular land plants during the mid-Ordovician approximately 470 million years ago (MYA) (Coates et al. 2011; Kenrick et al. 2012; Pires and Dolan 2012; Wellman et al. 2003). The moss *Physcomitrella patens* is a member of the earliest group of plants to appear on land, the bryophytes. *Physcomitrella* retains many of the characteristics of these ancient land plants, making it a key organism for gaining understanding of plant evolution. *Physcomitrella* has the ability to undergo targeted gene replacement by homologous recombination at high frequency when transforming DNA is introduced, thus making it a uniquely placed land plant for reverse genetic studies (Prigge and Bezanilla 2010; Schaefer 2002). Importantly, *Physcomitrella* recently became the first non-flowering plant to gain an annotated genome sequence (Rensing et al. 2008), with genes so far assembled onto a series of scaffolds, not all of which have been assigned to chromosomes (Kamisugi et al. 2008). The availability of the *Physcomitrella* genome sequence is considered a major step forward for “evo-devo” studies.

The *Physcomitrella* draft genome sequence assembly still requires some refinement, as many minor gaps remain throughout the scaffold sequences. Furthermore, although EST evidence is in place for 98% of predicted protein-coding genes, only 84% of the predicted proteins appear to be complete (Rensing et al. 2008). One curious feature of the *Physcomitrella* genome is the relatively low frequency of tandemly arrayed genes (TAGs). Despite having a considerably larger genome than the model dicot *Arabidopsis* and a similar sized genome to the monocot *Oryza sativa* (rice), only around 1% of *Physcomitrella* protein coding genes are in tandem arrays. This compares to 16% in *Arabidopsis* and 14% in rice (Rensing et al. 2008). This implies that the larger genome of *Physcomitrella* compared to *Arabidopsis* is not due to small-scale gene duplications resulting in tandem arrays. Indeed, a large-scale (possibly whole-) genome duplication occurred in *Physcomitrella* between 30 and 60 million years ago (Rensing et al. 2007; Rensing et al. 2008). The only sequenced flowering plant with a similarly small number of TAGs is cucumber, with around 2% TAGs (Huang et al. 2009).

Within the lycopphyte lineage, the first vascular plants evolved around 410 MYA. The genome of an extant lycopphyte, *Selaginella moellendorffii*, has now been sequenced (Banks et al. 2011). *Selaginella* has the smallest genome of any land plant sequenced to date, and the genome is free from whole-genome duplications or polyploidy (Banks et al. 2011).

The *ARABIDILLO* gene family is found throughout land plants (Coates et al. 2006; Nibau et al. 2011). *ARABIDILLO* proteins possess a land plant-specific domain structure composed of an F-box followed by leucine-rich repeats and an ARMADILLO repeat domain (Nibau et al. 2011). ARMADILLO repeat-containing proteins are important for multicellular development in several eukaryotic kingdoms (Coates 2003; Tewari et al. 2010). The strong conservation of *ARABIDILLO* homologues in all multicellular land plants sequenced to date suggests they have key and conserved functions.

In *Arabidopsis*, there are two *ARABIDILLO* homologues that function redundantly to promote lateral root development and are 77% identical to each other at the protein level (Coates et al. 2006). Phylogenetic analysis suggests that the two *Arabidopsis* *ARABIDILLO* homologues likely arose as a result of an *Arabidopsis*-specific duplication (Nibau et al. 2011), particularly as *ARABIDILLO1* and *ARABIDILLO2* are on areas of chromosome 2 and 3, respectively, that underwent large-scale duplication around 112 MYA (The *Arabidopsis* Genome Initiative 2000). Our initial analyses detected 3 putative *Physcomitrella* *ARABIDILLO* homologues - two of which were incompletely or incorrectly annotated - and a single *Selaginella* homologue (Nibau et al. 2011).

In this paper, we fully characterise the genomic loci of the three *Physcomitrella* *ARABIDILLO* homologues, named *PHYSCODILLO1A*, *-1B* and *-2*. We show that *PHYSCODILLO1A* and *PHYSCODILLO1B* are 100% identical within their coding regions and part of their upstream regions, and exist as tail-to-tail tandemly arrayed genes. *PHYSCODILLO2* is on a separate genomic scaffold, and is probably the result of a whole-genome duplication event in *Physcomitrella* (Rensing et al. 2008). We have cloned the *Selaginella*

homologue, *SELAGIDILLO*, and show that it is more similar in gene structure and protein sequence to other land plant *ARABIDILLO* homologues than indicated by the draft genome sequence (Banks et al. 2011). We demonstrate that *PHYSCODILLO1A* and *PHYSCODILLO2* are widely expressed in *Physcomitrella* and that they may share some conserved regulation or function with their *Arabidopsis* homologues.

Methods

Plant growth and culture

Physcomitrella patens (Gransden wild type strain) protonemata and gametophores were cultured at 22°C in long days (16h light) on solid BCD medium (250mg/l MgSO₄·7H₂O, 250mg/l KH₂PO₄ (pH6.5), 1010mg/l KNO₃, 12.5mg/l FeSO₄·7H₂O, 0.001% Trace Element Solution [0.614mg/l H₃BO₃, 0.055mg/l AlK(SO₄)₂·12H₂O, 0.055mg/l CuSO₄·5H₂O, 0.028mg/l KBr, 0.028mg/l LiCl, 0.389mg/l MnCl₂·4H₂O, 0.055mg/l CoCl₂·6H₂O, 0.055mg/l ZnSO₄·7H₂O, 0.028mg/l KI and 0.028mg/l SnCl₂·2H₂O) with 0.8% agar (Sigma-Aldrich) or in liquid BCD cultures. Gametophore medium was additionally supplemented with 1mM CaCl₂ and protonemal medium was supplemented with 1mM CaCl₂ and 5mM ammonium tartrate.

In order to induce sporulation, gametophyte tissue from homogenate was 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 dissecting microscope (Nikon).

In order to prepare fresh protonemal tissue 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.

Selaginella moellendorffii (sourced from Long Acre Plants, Charlton Musgrove, UK) was grown in Levington M3 compost/vermiculite mix and grown in the greenhouse at 20-22°C under long days (16h light, 8h dark).

Arabidopsis thaliana seeds (*pARABIDILLO1::GUS* (Coates et al. 2006)) were sown in Levington M3 compost/vermiculite mix and grown in the greenhouse at 20-22°C under long days (16h light, 8h dark) until flowers could be harvested.

Physcomitrella transformation

1-2 plates of 7-day old protonemal tissue was digested in 8% mannitol containing 1% Driselase® (Sigma-Aldrich) for 30-40 minutes at room temperature. Tissues were then filtered through a 100µm mesh and left to stand for 15 minutes. Protoplasts were centrifuged for 3 minutes at 120xg and then washed twice with 8% mannitol. Protoplast number was determined using a haemocytometer and adjusted to a final density of 1.5x10⁶ cells/ml in MMM solution (91g/l mannitol, 0.15M MgCl₂, 0.1% MES pH5.6). 15µg linearised transforming DNA was gently mixed with 300µl protoplasts. 300µl PEG solution (8% mannitol; 0.1M Ca(NO₃)₂; 10mM Tris pH8; 40% PEG 6000, Sigma-Aldrich) was then added. Protoplast suspensions were immediately heat-shocked at 45°C for 5 minutes and then recovered at room temperature for 5 minutes. 300µl 8% mannitol was added 5 times at 4-6 minute intervals. 1ml 8% mannitol was then added 5 times at 4-6 minute intervals. Cells were centrifuged for 4 minutes at 140xg and protoplasts resuspended in 5ml liquid BCD containing 10mM CaCl₂, 8% mannitol and 0.5% glucose. Protoplasts were incubated in the dark overnight at room temperature, centrifuged for 4 minutes at 120xg and resuspended in 0.5ml 8% mannitol. 2.5ml warm PRMT medium (BCD supplemented with 5mM ammonium tartrate, 8% mannitol, 10mM CaCl₂, 0.5% glucose and 0.5% agar) was added. 1ml protoplast mixture was plated onto cellophane overlaid PRMB medium (BCD supplemented with 5mM ammonium tartrate, 6% mannitol, 10mM CaCl₂, 0.5% glucose and 0.8% agar). Stable transformants were selected using 50µg/ml G418.

Nucleic acid isolation

Physcomitrella and *Selaginella* genomic DNA was prepared as follows. Tissue was harvested, frozen rapidly in liquid nitrogen and homogenised by hand using a polypropylene pestle (Sigma-Aldrich). 0.6ml (per 500mg tissue) CTAB extraction buffer (100mM Tris-HCl pH8, 1.42M NaCl, 2% CTAB, 20mM EDTA, 2% PVP-40, 0.07% β -mercaptoethanol and 0.1% ascorbic acid), preheated to 65°C, was then added. 100 μ g RNase A was added to the suspension and incubated at 65°C for 5 minutes. An equal volume of chloroform:isoamyl alcohol (24:1) was then added, samples vortexed and centrifuged for 10 minutes at 9,300xg. The upper phase was removed, added to an equal volume of isopropanol and centrifuged for an additional 5 minutes at 9,300g. The resulting pellet was washed in 70% ethanol, allowed to air dry and used directly in PCR reactions and/or Southern blotting.

Physcomitrella and *Selaginella* RNA was prepared using a Qiagen RNeasy plant mini-prep kit according to manufacturer's instructions. RNA was DNase treated according to the manufacturer's specifications (Fermentas) and then converted to cDNA using SuperscriptTM II reverse transcriptase according to the manufacturer's instructions (Invitrogen).

PHYSCODILLO and *SELAGIDILLO* cloning

Genomic DNA and cDNA products were amplified using Long PCR enzyme mix according to the manufacturer's instructions (Fermentas). Full-length *PHYSCODILLO2* was amplified using P1/2-5'Start (5'-ATGTCCAACAAGCGGCGGCG) and P2-3'-1 (5'-CGCGCTGCCACTGGCTTCACC). Full-length *PHYSCODILLO1A/1B* and sequences within the *PHYSCODILLO1A/1B* locus were amplified and sequenced using primers shown in Supplemental Table 1. Full-length *SELAGIDILLO* gDNA and cDNA was amplified using Sel_5'Start (5'-ATGCGTCGCGTTCGCCGGAATGCG) and Sel_3'-1 (5'-ATTGGCCGCTAAAGATCG).

RT-PCR analyses of *PHYSCODILLO* genes

PHYSCODILLO1A/1B tissue-specific expression patterns were determined using P1A/B-GSP.F (5'-GGCGCAATCGAAGCACTGGTGG) and P1A/B-GSP.R (5'-TGTACGTCCTCAAAATCAGAGTGC). *PHYSCODILLO2* tissue-specific expression patterns were determined using P2-GSP.F (5'-CGCAATTGAAGCACTGGTGGATCT) and P2-GSP.R (5'-ACGTCCTCAGAGTTCGAGTGTGC). A positive control tubulin cDNA was amplified using PptubF (5'- TGTGCTGTTGGACAATGAG) and PptubR (5'- ACATCAGATCGAACTTGTG).

Southern blotting

15 μ g *Physcomitrella* genomic DNA was digested to completion using *Eco*RI overnight at 37°C, according to the manufacturer's instructions (NEB). DNA was precipitated using phenol-chloroform purification and run at 50V on a 0.8% agarose gel. The gel was depurinated for 20 minutes in 0.25M HCl, denatured for 1h in denaturing buffer (1.5M NaCl, 0.5M NaOH) and neutralised for 1h in neutralising buffer (1M Tris-HCl pH8.0, 1.5M NaCl). The gel was transferred to N⁺-Hybond membrane (Amersham) overnight by capillarity using transfer buffer (1.5M NaCl, 0.25M NaOH). The membrane was washed using 6xSSC for 5 minutes and dried. DNA was subsequently UV-crosslinked to the membrane.

The *PHYSCODILLO1A/1B* coding region probe was amplified using P1A/B_3'F (5'-GTGGAGCCATGGTTGGGAGATGG) and P1A_3'R (5'-CCTCATCCCAACCTCTTACGACAG). The *PHYSCODILLO1A/1B* promoter probe was amplified using P1A/1B_promF (5'-ATCCCACAGAGAACATGTTAAACA) and P1A_prR (5'-GTCATCAGCTGCTCCTCCGAT). The *PHYSCODILLO2* coding region probe was amplified using P2_5'F (5'- CGTCGTCGCTTGTAGTGGCC) and P2_5'R (5'- CACGCCTACACTCGAGGGAGG).

Blotted nitrocellulose membranes were incubated for 2h at 65°C in hybridisation buffer (1% BSA, 1mM EDTA pH8, 0.5M Na₂HPO₃, 7% SDS). Hybridisation probes were labelled with dCTP[³²P]-3000Ci/mmol (Perkin Elmer) using the Prime-It[®] II Random Primer Labelling Kit, according to the manufacturer's instructions (Stratagene) and then added to the hybridisation buffer. Blots were incubated with radiolabelled probes overnight at 65°C. Blots were subsequently 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°C.

Blots were wrapped in cling film and exposed to Hyperfilm™ ECL (Amersham) for 3-7 days in an autoradiograph cassette at -80°C (Southern, 1975).

Vector construction

The full-length *PHYSCODILLO1A* gene promoter was cloned using P1A_prF (5'-ACGGTGACAAGTGCCGGACGAA) and P1A_prR (as above). The full-length *PHYSCODILLO1B* gene promoter was cloned using P1BprF (5'-TACGGCGAAAAAGTTCCTGGC) and P1A_prR. The identical promoter region shared by both *PHYSCODILLO1A* and *PHYSCODILLO1B* was cloned using P1A/B_3'F (as above) and P1A_prR. The full-length *PHYSCODILLO2* gene promoter was cloned using P2_prF (5'-TTCACAAGCTGAGATGTTTCGGAGC) and P2_prom_R (5'-ATCATCATCACCTCCTCCCACAATC).

Sequence alignment and phylogeny

Putative full-length land plant ARABIDILLO protein sequences were identified from fully sequenced land plant genomes, by using BLASTP with ARABIDILLO1 protein sequence to search both GenBank and Phytozome. Sequences were aligned automatically using Clustal X (Larkin et al 2007) and the alignment was refined manually in SeaView (Gouy et al. 2010). Gblocks (Castresana 2000; Talavera and Castresana 2007) was used (on default settings) to filter the alignment to remove poorly aligned or divergent regions of the sequence alignment that may lead to misleading phylogenies. The phylogenetic tree was calculated using the neighbour joining algorithm in Clustal X applying a correction for multiple substitutions. The tree was displayed using TreeView X (Page 2002).

Promoter::*GUS* fusion transgene construction

To assess the activity of *PHYSCODILLO* promoters, we first generated a transformation vector to allow an easy cloning of promoter::*GUS* fusions. The vector pBS108CH-35SNPTb ((Saidi et al. 2009), NCBI: GQ463722), containing the "35S::G418-35Sterm" resistance cassette and the 108 targeting locus was digested with *NotI* and *SmaI* restriction enzymes. The *pHSP::GUS* cassette (Saidi et al. 2005), where the *GUS* reporter gene is driven by the *Gmhsp17.3* promoter, was then digested with *NotI* and *Asp718*, and half-blunt cloned in the pBS108CH-35SNPTb vector. The resulting vector is called the *pHSP::GUS*-108-35SNPT transformation vector and it has several unique restriction sites to replace the *Gmhsp17.3* promoter with promoters of interest. This vector was then used to clone *PHYSCODILLO* promoters. *PHYSCODILLO1A* and -2 full-length promoters (respectively 2179bp and 2157bp) were amplified to replace the *HSP* promoter using *NotI*-*BamHI* restriction sites. Stable transgenic lines were generated.

Staining of *PHYSCODILLO* promoter::*GUS* transgenic lines

Tissue was harvested and submerged in an X-gluc solution (0.5mg/ml X-gluc, 0.2% Triton X-100, 0.5mM K₄Fe(CN)₆·3H₂O, 0.5mM K₃Fe(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 an SMZ 1000 stereomicroscope (Nikon). More than three independent transgenic lines were examined for each promoter.

Construction of a *physcodillo2* deletion mutant

The *PHYSCODILLO2* deletion construct was generated by cloning 5' and 3' *PHYSCODILLO2* gene sequences from *Physcomitrella* genomic DNA (Supplemental Figure 6) and inserting them into the pAHG1 vector (a kind gift from Drs. Yasuko Kamisugi and Andrew Cuming) either side of a hygromycin resistance cassette. The 5' flanking sequence was amplified using P2.5F.KpnI (5'-AAAGGTACCCGTCGTCGCTTGTAGTGCCC) and P2.5R.XhoI (5'-AAACTCGAGCACGCCTACACTCGAGGGAGG) and ligated into KpnI/XhoI-cut pAHG1 to create pP2KO-5'. The 3' flanking sequence was amplified using P2.3F.XbaI (5'-

AAATCTAGAGGCTGGTGCTTCGTTGGGATGC) and P2.3R.NotI (5'-AAAGCGGCCGCCGCGCTGCCACTGGCTTCACC) and ligated into XbaI/NotI-cut pP2KO-5'. The resulting construct was transformed into *Physcomitrella* protoplasts. Two rounds of hygromycin selection were carried out to obtain putative transformants. Genomic DNA isolated from 68 putative *physcodillo2* deletion mutant lines was screened by PCR using a pair of internal *PHYSCODILLO2* gene-specific primers, (P1/2seq4 forward 5'-CTGGTGCGCTGGAGGCACTG and P2-3'-1 reverse 5'-CGCGCTGCCACTGGCTTCACC); to confirm absence of a wild type *PHYSCODILLO2* gene. Genomic DNA from 27 lines where *PHYSCODILLO2* appeared to be absent was then amplified by PCR using primers flanking the *PHYSCODILLO2* locus (P1/25'Start forward 5'-ATGTCCAACAAGCGGCGGCG and P2-3'-1 reverse; Supplemental Figure 6). PCR products of the expected size for a single-copy insertion were obtained from 5 lines: these were then sequenced using a primer to confirm insertion of a single copy of the hygromycin cassette into the *PHYSCODILLO2* gene, and hence replacement of the locus, in 2 independent lines (Supplemental Figure 6).

physcodillo2 mutant phenotyping

All stages of *Physcomitrella patens* life cycle were examined to assess the effect of *PHYSCODILLO2* deletion compared to wild type plants. Fragmented protonemal tissues were inoculated onto sterile peat plugs and cultured for four weeks at 25°C under continuous light. The peat plugs were then transferred to 15°C under 8h/16h light/dark conditions to induce reproductive organs. Antheridia and archegonia were isolated in 25% glycerol and mounted on a glass slide. Spore capsules were isolated once dark brown, sterilised using 25% bleach and 0.02% Tween solution, washed once with 70% ethanol and then three times with sterile distilled water. After bursting the capsules in 50-100µl water, spores in suspension were plated on BCD medium, supplemented with 5mM ammonium tartrate and 5mM CaCl₂, and placed in growth chambers at 22°C. Pictures of germinating spores were acquired three days after exposure to light. Growth of the mutant was monitored and colony size estimated. Tissues were visualised and pictures were acquired using a Nikon SMZ 1000 stereomicroscope.

Results

The *Physcomitrella* genome contains three putative *ARABIDILLO* homologues and a related pseudogene sequence.

In order to identify *ARABIDILLO* homologues in *Physcomitrella*, BLASTN and BLASTP 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 the *Arabidopsis* *ARABIDILLO* genomic DNA and protein sequences respectively (Coates et al. 2006). The search identified three putative *Physcomitrella* *ARABIDILLO* homologues, which were subsequently named *PHYSCODILLO1A*, *PHYSCODILLO1B* and *PHYSCODILLO2* (Fig. 1). When each of the three putative *PHYSCODILLO* sequences was used in a BLASTP search against the *Arabidopsis* proteome, *ARABIDILLO1* was returned as a reciprocal best hit.

In the published *Physcomitrella* genome sequence (Rensing et al. 2008) a full-length copy of the *PHYSCODILLO1A* gene and a partial *PHYSCODILLO1B* gene were located on scaffold 91 and flanked by two genes, *Phypa_230042*, which encodes a protein of unknown function, and *Phypa_131941*, which encodes a putative metallopeptidase (Fig. 1a). The published *PHYSCODILLO1B* gene sequence was truncated in the contig sequence but was otherwise 100% identical to the equivalent portion of the *PHYSCODILLO1A* gene (Fig. 1a). This identity extends to 1566bp of both *PHYSCODILLO1A* and *PHYSCODILLO1B* upstream (promoter) regions (Fig. 1a). The gap in the contig sequence is followed by a ~4kb intergenic region, which shares no homology with any of the *PHYSCODILLO* genes (Fig. 1a). The annotated *PHYSCODILLO1A* gene is 6251 base pairs long and contains 12 exons and 11 introns (Fig. 3). The corresponding transcript is 2823 base pairs long and encodes a 941 amino acid protein, 64% identical to *ARABIDILLO1* and 63% identical to *ARABIDILLO2*.

The annotated full-length *PHYSCODILLO2* gene resides alone on scaffold 13 and lies between unrelated genes, *Phypa_159475* (encoding a G-protein coupled receptor) and *Phypa_115455* (callose synthase) (Fig. 1b).

Interestingly, a BLASTN search of the *Physcomitrella* genome revealed a partial sequence of a truncated pseudogene. This is located upstream of the putative *PHYSCODILLO1B* gene and *Phypa_230042* on scaffold 91: bases 474945-475467. The pseudogene is 522 base pairs long and is 95% identical to the first 534 base pairs of both *PHYSCODILLO1A* and *PHYSCODILLO1B* genes. The start codon of the pseudogene has been mutated from methionine (ATG) to isoleucine (ATA) (Supplemental Fig. 1). Beyond the 522 base pair sequence, there is no further homology. The upstream region also shares no similarity with any of the *PHYSCODILLO* promoters (Supplemental Fig. 1).

There are three *PHYSCODILLO*-encoding genes in *Physcomitrella*.

In order to determine whether *PHYSCODILLO1A* and *PHYSCODILLO1B* genes were both present within the same locus as an inverted repeat, as the draft genome sequence suggested, and to determine the sequence of the genome “gap”, a methodical sequencing approach was undertaken. Long PCR products were generated from *Physcomitrella* genomic DNA using primers (listed in Supplemental Table 1) specific to the known full-length *PHYSCODILLO1A* (or *-1B*) gene and within the divergent gene sequences either side of the *PHYSCODILLO1A/1B* locus (Fig. 1a and Supplemental Table 1). This was to ensure that *PHYSCODILLO1A* and *PHYSCODILLO1B* could be distinguished from one another from sequencing data produced (Fig. 2a). Both full-length *PHYSCODILLO1A* and *PHYSCODILLO1B* gene and promoter sequences were amplified: sequencing revealed that they were 100% identical to each other across the entire coding region and 1566bp of upstream promoter region. Full-length cDNA corresponding to either *PHYSCODILLO1A* or *1B* was amplified with relative ease, demonstrating that at least one of the genes is expressed in *Physcomitrella*. Moreover, the full-length sequenced *PHYSCODILLO1A* gene and cDNA sequences were identical to that annotated for *PHYSCODILLO1A* in the published genome (Rensing et al. 2008), with 12 exons and 11 introns (Fig. 3).

Our sequencing data shows that *PHYSCODILLO1A* and *PHYSCODILLO1B* are in a tail-to-tail orientation with a 7729bp intergenic sequence separating their stop codons. The intergenic region is made up of two duplicated sequences (3809bp) downstream of their respective genes that also sit in a tail-to-tail orientation (Fig. 2a). The *PHYSCODILLO1A* and *PHYSCODILLO1B* downstream regions are nearly 100% identical, differing only by seven individual nucleotide differences towards the ends furthest from the stop codons of *PHYSCODILLO1A* and *PHYSCODILLO1B* genes. An additional 111 nucleotides is present between the *PHYSCODILLO1A* and *-1B* downstream regions (Fig. 2a,b). This 111-nucleotide sequence was completely unique and aided the genomic assembly of the *PHYSCODILLO1A/1B* locus by overlapping the sequences of cloned products (Fig. 2b). The complete sequence of the *PHYSCODILLO1A/1B* locus is given in Supplemental Fig. 2 and is available at GenBank (accession JQ290353). The *PHYSCODILLO1A/1B* protein sequence was previously used for phylogenetic analysis (Nibau et al. 2011), where it was referred to as *PHYSCODILLO1*.

As mentioned by Nibau et al. (2011), the cloned *PHYSCODILLO2* protein sequence differs from that published in the draft genome (Rensing et al. 2008). The original annotated *PHYSCODILLO2* gene was 5576 base pairs long and contained 11 exons and 10 introns (Rensing et al. 2008). The corresponding predicted transcript was 2802 base pairs long and encoded a 934 amino acid protein, 63% identical to *ARABIDILLO1*, 62% identical to *ARABIDILLO2* and 83% identical to *PHYSCODILLO1A*. However, our sequencing has shown that there was a mis-annotation of the genome sequence leading to the “skipping” of exon 11 and an incorrect start site for exon 12 (Supplemental Fig. 3). Consequently, the correct *PHYSCODILLO2* gene sequence is 5860 base pairs long and contains 12 exons and 11 introns (Fig. 3). The corresponding transcript is 2823 base pairs long and encodes a 941 amino acid protein, more similar than that in the draft genome to the predicted full-length sequence of *PHYSCODILLO1A* (85% identity), and both of the *Arabidopsis* *ARABIDILLO* sequences (64% identity).

Southern blot analyses were carried out to confirm the sequencing data obtained (Fig. 4). Restriction sites within the *PHYSCODILLO1A/1B* locus had already been determined by sequencing. These were used to

confirm the presence of either a full-length *PHYSCODILLO1A* or *PHYSCODILLO1B* gene. *Physcomitrella* genomic DNA was digested using *EcoRI*, a restriction enzyme that cuts both within the genes themselves and within the two 4kb intergenic regions (Fig. 4a). Using a radiolabelled probe specific to the 3' end of either *PHYSCODILLO1A* or *PHYSCODILLO1B* genes (Fig. 4a) a distinguishable band was revealed the expected size of 7kb (Fig. 4c). This result thus confirmed the presence of full-length *PHYSCODILLO1A* and/or *PHYSCODILLO1B* gene.

To confirm the presence of both *PHYSCODILLO1A* and *-1B* genes in a tail-to-tail orientation within the same locus, genomic DNA again digested with *EcoRI* was used. *EcoRI* cuts both within the two genes and in divergent sequences outside the locus, where the restriction sites are located at different distances from the start codons of each gene (Fig. 4a). Radiolabelled probes homologous to the 1.6kb promoter region shared by both *PHYSCODILLO1A* and *PHYSCODILLO1B* genes were used (Fig. 4a). An *EcoRI* digest produced two bands of 2.4kb and 2.7kb: these confirm the sequencing results and clearly indicate the presence of *PHYSCODILLO1A* and *PHYSCODILLO1B* respectively (Fig. 4d).

A radiolabelled probe that specifically hybridises to the 5' end of the *PHYSCODILLO2* gene (Fig. 4b) was then used to confirm the structure of this gene, following genomic DNA digestion with *BglII*. Distinguishable bands with expected size were observed (3.5kb and 4.1kb; Fig. 4e). The presence of these bands confirmed that *PHYSCODILLO2* is present in the genome as a single copy.

An ancient gene structure for *PHYSCODILLO*/*ARABIDILLO* homologues

Our results show that *PHYSCODILLO1A*, *-1B* and *-2* all contain 12 exons and 11 introns (Fig. 3). We also cloned and sequenced cDNA of the putative *Selaginella* *ARABIDILLO* homologue, named *SELAGIDILLO*. In contrast to the annotated theoretical genome sequence, which contains only 11 exons (Banks et al 2011), we found that *SELAGIDILLO* also contains 12 exons and 11 introns (Fig. 3 and Supplemental Fig. 4; GenBank accession JQ406949). Our analysis of other full-length *PHYSCODILLO* homologues in published plant genome sequences shows that the majority of genes possess 12 exons (homologues and their gene identifiers are listed in Supplemental Table 2). This strongly suggests that the ancestral land plant *ARABIDILLO*/*PHYSCODILLO* homologue had 12 exons. *Arabidopsis thaliana* *ARABIDILLO1* and *-2* have 11 and 9 exons, respectively (Fig. 3), as do homologues in their close relatives, namely *Arabidopsis lyrata*, *Capsella rubella* and *Thelungiella halophila* (Supplemental Table 2). A *Brassica rapa* *ARABIDILLO* homologue has even fewer exons, suggesting that multiple *ARABIDILLO* intron losses may have occurred specifically during divergence of the *Brassicaceae* (Supplemental Table 2). A phylogeny of all the full-length *ARABIDILLO* homologues identified to date is shown in Fig. 5. *ARABIDILLO* gene duplication appears to have occurred independently in a bryophyte (*Physcomitrella*), one monocot (maize), more than once in dicots (rosids), but not in a lycophyte (*Selaginella*) or the majority of monocots. This data builds on and is in accordance with our previous phylogenetic analysis with plant *ARABIDILLO* homologue protein sequences, which suggested that there was a single common *ARABIDILLO* ancestor at the dawn of land plant evolution (Nibau et al. 2011).

PHYSCODILLO gene expression and promoter activity

ARABIDILLO genes are expressed in both roots and shoots of *Arabidopsis* seedlings, with specific promoter activity patterns within certain root cell types (Coates et al. 2006). In order to determine the *in planta* activity of *PHYSCODILLO* promoters, transgenic lines expressing *GUS* under the control of *PHYSCODILLO1A* and *PHYSCODILLO2* promoters were generated. To minimise any position effect and to properly compare the activities of both promoters, the expression cassettes were all targeted to the 108 locus, a site within the genome where homologous recombination yields no detrimental phenotypes (Schaefer and Zryd 1997). Multiple independent lines of *pPHYSCODILLO1A::GUS* and *pPHYSCODILLO2::GUS* showed identical expression patterns throughout both the haploid and diploid phases of the *Physcomitrella* life cycle (Fig. 6a-j,m-p). *GUS* expression was clearly visible both in protonemal tissue (chloronema and caulonema), and in leafy tissue and rhizoids (Fig.6a-f). *GUS* expression was also detected in reproductive organs: throughout male antheridia (Fig. 6i,j) but only within upper cells of female archegonia (Fig. 6g,h). Interestingly, the archegonial expression pattern is reminiscent of that observed in *Arabidopsis* female reproductive organs expressing *pARABIDILLO1::GUS*.

GUS expression was visible only at the top of the stigma (Fig. 6k). However, *pARABIDILLO1::GUS* expression in the anther is restricted compared to *pPHYSCODILLO::GUS* expression in antheridia (Fig. 6l). *pPHYSCODILLO::GUS* expression was also detected at the foot of the diploid sporophyte (Fig. 6m-p). This ring of expression appears to be in sporophytic tissue, as it is below the stomatal ring (Chater et al 2011) but above the brown area identified as the top of the gametophyte by Chater et al. (2011). However, we cannot rule out the possibility that the sporophytic *GUS* staining represents leakage from gametophytic tissue.

To confirm that the *PHYSCODILLO1A* and *PHYSCODILLO2* promoter activity observed relates to *PHYSCODILLO* gene expression, we performed RT-PCR on cDNA isolated from protonemal tissue and leafy shoots, using primer pairs specific to *PHYSCODILLO1A/1B* (which cannot be distinguished using gene-specific primers) and *PHYSCODILLO2*. Both genes were detected in the tissues tested (Supplemental Fig. 5).

PHYSCODILLO genes are likely to function redundantly during *Physcomitrella* development

As *PHYSCODILLO* proteins show a high degree of similarity, and *PHYSCODILLO1A* and *PHYSCODILLO2* promoters show identical activity, it seems likely that the genes have redundant functions. To begin to test this possibility, we generated a *physcodillo2* deletion mutant by targeted gene replacement (Supplemental Figure 6). Two independent deletion lines were obtained, both of which appeared to grow and develop similarly to wild type. One line, *p2-105*, in which no *PHYSCODILLO2* expression could be detected by RT-PCR (Supplemental Figure 6), was subjected to further phenotypic analysis: we compared phenotypes of wild type and *p2-105* deletion mutant plants in tissues where *PHYSCODILLO2* is expressed. *p2-105* deletion mutants showed no defects in spore germination, protonemal filament growth, gametophore morphology or morphology of the reproductive organs or sporophyte (Figure 7). Thus, it seems highly likely that *PHYSCODILLO2* functions redundantly with *PHYSCODILLO1A/B* in *Physcomitrella*.

Discussion

Conservation of ARABIDILLO proteins and genes throughout the land plants

Completion of the genomes of the early-evolving land plant *Physcomitrella patens* (Rensing et al 2008) and the non-seed vascular plant *Selaginella moellendorffii* (Banks et al. 2011) have opened up new avenues of research, enabling analysis of the conservation and evolution of protein functions across the plant kingdom. Our discovery of *Physcomitrella* and *Selaginella* proteins highly similar to ARABIDILLO proteins in the model angiosperm *Arabidopsis thaliana* raises questions about the functions of ARABIDILLO homologues early in land plant evolution.

ARABIDILLO1 and ARABIDILLO2 promote root branching in *Arabidopsis* (Coates et al. 2006). However, bryophytes such as *Physcomitrella* lack specialised rooting structures and instead produce 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). *Selaginella* possesses multicellular roots likely to have evolved independently from those of seed plants (Banks 2009). Unlike the roots of seed plants, *Selaginella* roots branch dichotomously at the tip from apical initial cells and do not form lateral roots (Banks 2009). ARABIDILLO homologues must therefore function differently in land plants that predate the origins of ‘true roots’.

We have identified three putative ARABIDILLO gene homologues, named *PHYSCODILLO1A*, *-1B* and *-2*, in the early-evolving bryophyte *Physcomitrella patens*. We have also cloned a single homologue, *SELAGIDILLO*, from the lycophyte *Selaginella moellendorffii*, one of the earliest-evolving vascular plants. Our findings extend and correct the previously annotated genome sequences (Rensing et al. 2008, Banks et al. 2011).

Despite their likely divergent functions, land plant *ARABIDILLO* homologues share a remarkable degree of identity at the amino acid level, with *Physcomitrella* and *Arabidopsis* proteins sharing 64% identity. In addition, most land plant *ARABIDILLO* genes share a conserved exon/intron structure, with only a minority of dicot and monocot homologues showing relatively recent intron loss or gain. This suggests that all *ARABIDILLO* homologues arose from a single common ancestor. It also suggests that regulatory elements necessary for the proper expression and function of *ARABIDILLO* homologues might be embedded within the exon/intron borders.

PHYSCODILLO1A/1B locus and gene duplication

The *PHYSCODILLO1A/1B* locus was incompletely sequenced in the original *Physcomitrella* draft genome (Rensing et al. 2008). We have used a combination of deep sequencing and Southern blotting to confirm the locus structure. *PHYSCODILLO1A* and *1B* share identical coding sequences and 1.6kb of upstream identity. They exist as a tail-to-tail inverted repeat tandem array on Scaffold 91, with no intervening genes. The intergenic region separating the two genes also exists 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, suggests that a very recent gene duplication event had taken place.

Despite having a larger genome size, the genome of *Physcomitrella* contains only 1% tandemly arrayed genes (TAGs) in contrast to 16% in *Arabidopsis*, 14% in rice and 11% in *Poplar* genomes (Rensing et al. 2008). 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 (Lang et al. 2008; Rensing 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).

Most *Physcomitrella* tandemly arrayed genes consist of two recently duplicated genes, with no intervening genes, which have an elevated GC content (52% average) compared to the average for all *Physcomitrella* genes (45%; (Rensing et al. 2008)). *PHYSCODILLO1A* and *-1B* fall into this category, with a GC content of 57%. However, *PHYSCODILLO1A* and *-1B* genes are much longer (~6kb gDNA, ~3kb cDNA) than the average *Physcomitrella* TAG (~2.2kb gDNA, ~1.1kb cDNA), and the average *Physcomitrella* gene overall (~3kb gDNA, ~1.3kb cDNA). *PHYSCODILLO1A/1B* also have far more exons (12) than both TAGs (4) and *Physcomitrella* genes overall (6.7). The most likely scenario is that *PHYSCODILLO1A* and *PHYSCODILLO1B* are the result of relatively recent gene duplication, whereas *PHYSCODILLO2* is the result of the earlier whole-genome duplication proposed by Rensing et al. (Rensing et al. 2008).

The most similar example to the *PHYSCODILLO1A/1B* gene duplication is the *Chalcone Synthase (CHS)* gene family with 17 members (Koduri et al. 2010). 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. 2010). 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 (Koduri et al. 2010). The CHS3 and CHS5 genes were very similar to each other, with 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 (Koduri et al. 2010). 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. 2010). In contrast to the intergenic regions separating both *PHYSCODILLO1A* and *PHYSCODILLO1B*, the sequences between the UTRs of the duplicated CHS transcripts are not repetitive in nature, and show no evidence of duplication (Koduri et al. 2010).

Promoter activity and expression of *PHYSCODILLO* genes in relation to their function

PHYSCODILLO genes show widespread expression and high promoter activity in the gametophyte part of the life cycle, being readily detected in protonemal and leafy tissues. Weaker promoter activity was observed in rhizoids. *Arabidopsis ARABIDILLO1* and *-2* gene expression is detected in all sporophyte plant tissues tested (Coates et al. 2006), with more cell type-restricted and gene-specific promoter activity in the roots (Coates et al. 2006). Whether *ARABIDILLO*s are expressed in the *Arabidopsis* gametophyte is currently unknown.

PHYSCODILLO promoter::*GUS* expression was detected throughout male antheridia but only in the upper cells of female archegonia containing unfertilised egg cells. *pPHYSCODILLO*::*GUS* expression was observed at the base of the spore capsule, below the ring of stomata, but not elsewhere in the diploid sporophyte generation. However, we currently have no evidence to support a role for *PHYSCODILLO* genes in sporophyte development. The pattern of activity of *PHYSCODILLO* promoters in the *Physcomitrella* archegonium is similar to *ARABIDILLO1* promoter activity in the *Arabidopsis* stigma. This suggests that *PHYSCODILLO*s and *ARABIDILLO*s may have been co-opted into both the gametophyte and sporophyte lineages to perform a conserved and as yet uncharacterised function in female reproductive organs. Co-option of genes for related functions into both gametophyte and sporophyte is not without precedent. For example, related transcription factors regulate *Physcomitrella* rhizoid development and *Arabidopsis* root hair development (Menand et al. 2007) and strigolactone signalling regulates filament branching in *Physcomitrella* and shoot branching in *Arabidopsis* (Proust et al. 2011).

PHYSCODILLO functions in growth and development?

The widespread expression of the *PHYSCODILLO* genes suggests that they may have pleiotropic functions during *Physcomitrella* development, in addition to the possible roles in reproduction discussed above. The similarities in gene sequence and promoter activity between *PHYSCODILLO1A/1B* and *PHYSCODILLO2* suggest that, like their *Arabidopsis* homologues (Coates et al. 2006), the genes are likely to be functionally redundant, and this is confirmed by the lack of a discernible morphological phenotype in a *physcodillo2* deletion mutant. We speculate that *PHYSCODILLO*s may function to regulate a fundamental aspect of growth and development that is conserved during angiosperm lateral root formation, for example, cell division or cell elongation. A reverse genetic approach is currently underway in our laboratory to address these questions.

Figure captions

Fig. 1 Three putative *ARABIDILLO* homologues were identified in the published *Physcomitrella patens* draft genome sequence (Rensing et al 2008).

a. A full-length copy of the *PHYSCODILLO1A* gene (ID *Pp1s91_74v6/Phypa_186256*; scaffold 91 488571-494860 for coding region) and a partial *PHYSCODILLO1B* gene (ID *Pp1s91_72V6/Phypa_131917*; scaffold 91 477603-480429 for coding region) are located on scaffold 91 of the published genome sequence. *PHYSCODILLO1B* is truncated by a putative gap in the contig sequence (indicated by a grey line, a run of Ns and arrows) and this is followed by ~4kb of intergenic sequence. *PHYSCODILLO1B* is identical to the equivalent portion of the *PHYSCODILLO1A* gene (indicated by black bars under the genes) and this includes their putative 1.6kb promoter regions (light grey boxes). *PHYSCODILLO1A* and *PHYSCODILLO1B* are surrounded by two unrelated genes, *Phypa_230042* and *Phypa_131941*, as indicated on the diagram.

b. Full-length *PHYSCODILLO2* (ID *Pp1s13_130V6/Phypa_175550*) is located on scaffold 13 (934143-939754) (Rensing et al 2008). The putative promoter sequence is indicated by a dark grey box. The *PHYSCODILLO2* gene lies adjacent to genes *Phypa_159475* and *Phypa_113455* as indicated.

Fig. 2 *PHYSCODILLO1A* and *PHYSCODILLO1B* exist as a near-identical inverted repeat in a tail-to-tail orientation, with *PHYSCODILLO2* in a separate genomic location.

a. Different primer pairs were used to amplify various portions of the *PHYSCODILLO1A/1B* locus. The amplicons are indicated by a series of coloured lines below the diagram: Roman numerals indicate PCR products referred to in Fig. 2b. 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. *PHYSCODILLO1A* and *PHYSCODILLO1B* genes exist as an inverted repeat and are 100% identical in their coding sequences (black filled arrows on the gene diagram) and their promoter regions (grey boxes on the gene diagram). The 4kb intergenic regions are also duplicated, although not as a 100% perfect repeat (see also Fig. 2B and Supplemental Fig. 2): two near identical regions of 3809bp (differing only at 7 nucleotides) flank a unique region of 111bp (sizes indicated). Limits of the entire locus inverted repeat are indicated by the two outermost black hashed vertical lines and limits of symmetry in the intergenic region are indicated by the innermost black hashed lines. The unique 111bp sequence between the innermost hashed vertical lines is detailed in Fig. 2b, shaded in grey.

b. *PHYSCODILLO1A* and *PHYSCODILLO1B* intergenic regions were sequenced and aligned to the intergenic sequence annotated in the *Physcomitrella* genome resource, using ClustalW2. 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 italicised and underlined in the alignment (indicated in Fig. 2a). IV represents the sequenced data collected from three independent PCR reactions utilising the same gene-specific primer and the reverse complements of the italicised/underlined primers (indicated in Fig. 2a). V 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 (detected by sequencing I, II and III) are highlighted in black. The additional 111nt unique sequence present between the *PHYSCODILLO1A* and *PHYSCODILLO1B* intergenic regions is highlighted in grey.

Fig. 3 Exon/intron structure of *PHYSCODILLO* genes compared to more recently-evolved homologues. All three *PHYSCODILLO* genes consist of 12 exons and 11 introns, similarly to the *Selaginella* homologue (*SELAGIDILLO*). Exons are indicated by numbered boxes in alternating grey and white. The number of nucleotides in each gene sequence is also shown. *Arabidopsis* *ARABIDILLO* genes have fewer exons, presumably arising from intron loss and exon fusion (putative fused exons are depicted as 2-tone grey/white numbered boxes).

Fig. 4 Southern blot analysis of *PHYSCODILLO1A/1B* and *PHYSCODILLO2* genomic loci.

a. Schematic representation of the *PHYSCODILLO1A/1B* locus showing *Eco*RI restriction sites. Black arrows and grey boxes on scaffold 91 represent coding- and promoter sequences, respectively. Grey and white boxes beneath the locus represent different probes used for Southern blotting. Note the different sized *Eco*RI fragments (2.4kb and 2.7kb) at the 5' end of *PHYSCODILLO1A* and *PHYSCODILLO1B*, respectively.

b. Schematic representation of the *PHYSCODILLO2* locus on scaffold 13 (*PHYSCODILLO2* promoter, grey box; *PHYSCODILLO2* coding region, black arrow; upstream gene, white arrow), showing *Bgl*II restriction sites. Black box beneath the locus represents the probe used for Southern blotting.

c. Southern blot of *Physcomitrella patens* genomic DNA digested with *Eco*RI and probed with the *PHYSCODILLO1A/1B* gene-specific probe (white box) shown in (a). A 7kb fragment is detected, which represents either the *PHYSCODILLO1A* or *PHYSCODILLO1B* gene fragment depicted in (a) (the probe does not distinguish between the two genes).

d. Southern blot of *Physcomitrella patens* genomic DNA digested with *Eco*RI and probed with the *PHYSCODILLO1A/1B* promoter probe (grey box) shown in (a). Both a 2.4kb fragment (representing the *PHYSCODILLO1A* promoter fragment) and a 2.7kb fragment (representing the *PHYSCODILLO1B* promoter fragment) are detected, thus confirming the presence of both *PHYSCODILLO1A* and *PHYSCODILLO1B* in the genome.

e. Southern blot of *Physcomitrella patens* genomic DNA digested with *Bgl*II and probed with the *PHYSCODILLO2* gene probe (black box) shown (b). Both a 3.5kb and a 4.1kb fragment are detected, confirming the existence of a single *PHYSCODILLO2* gene in the genome.

Fig. 5 Phylogeny of land plant *ARABIDILLO* homologues. Putative full-length land plant *ARABIDILLO* homologue protein sequences were aligned and subject to phylogenetic analysis. Bootstrap branch values greater than 700 are shown and the scale bar represents the average number of substitutions per column in

the sequence alignment used for generating the phylogeny. Sequences already identified in Nibau et al (2011) are named as previously. Sequences identified in additional plant species since the publication of Nibau et al (2011) are named as “*Genus species* ARABIDILLO” (where only one ARABIDILLO homologue is present in the species) or as “*Genus species* ARABIDILLO1, -2, -3” where additional homologues (full length or truncated) are present in the genome of the species.

Fig. 6 *PHYSCODILLO* promoters are active throughout the *Physcomitrella* gametophyte life cycle. Histochemical localisation of GUS expression driven by p*PHYSCODILLO1A* promoter (a, c, e, g, i, m, n) or p*PHYSCODILLO2* promoter (b, d, f, h, j, o, p). GUS expression in filamentous tissues ((protonema) is shown in (a, b), whole gametophores and rhizoids (c, d), leaves (e, f), archegonia (g, h), antheridia (i, j), and at the foot of dissected sporophytes (m, n, o, p). n is a close-up view of m, and p is a close-up view of o. Arrows in (g, h) highlight localised GUS expression seen in archegonia. In n and p, the arrow indicates the brown pigmented top of the gametophyte identified by Chater et al. (2011), while the arrowhead indicates diploid sporophyte GUS staining. In (m, n, o, p) the sporophyte (2n) and gametophyte (n) regions are indicated. GUS staining from *Arabidopsis* stigma (k) and anther (l) transformed with p*ARABIDILLO1::GUS* is shown as a comparison to the staining observed in moss reproductive organs. Scale bars: 0.1mm

Fig. 7 *physcodillo2* mutants appear identical to wild type plants. Tissues and organs of wild type (a, c, e, g, i, k, m) and a *physcodillo2* deletion mutant (b, d, f, h, j, l, n) were monitored throughout the life cycle. (a, b) dissected archegonia; (c, d) dissected antheridia; (e, f) dark brown spore capsules; (g, h) three day old germinating spores; (i, j) three week old colonies; (k, l) protonemal filaments with caulonemal apical cell and chloronemal side branch initial: arrow heads indicate cell walls. (m, n) young gametophores. Scale bars: 30µm (a-d), 50µm (g, h, k, l), 100µm (e, f, i, j, m, n).

Acknowledgements

We thank Chris Franklin for useful discussions. This work was funded by a Leverhulme Trust Research Project Grant (F/00094/BA) and a Royal Society Research Grant (RG081075) to JC, a Biotechnology and Biological Sciences Research Council Doctoral Training Grant to LAM and a Nuffield Foundation Undergraduate Research bursary to ES. Sequencing was carried out at the University of Birmingham Functional Genomics and Proteomics Unit.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Banks JA (2009) Selaginella and 400 million years of separation. *Annu Rev Plant Biol* 60: 223-238
- Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA, Ashton NW, Axtell MJ, Barker E, Barker MS, Bennetzen JL, Bonawitz ND, Chapple C, Cheng C, Correa LG, Dacre M, DeBarry J, Dreyer I, Elias M, Engstrom EM, Estelle M, Feng L, Finet C, Floyd SK, Frommer WB, Fujita T, Gramzow L, Gutensohn M, Harholt J, Hattori M, Heyl A, Hirai T, Hiwatashi Y, Ishikawa M, Iwata M, Karol KG, Koehler B, Kolukisaoglu U, Kubo M, Kurata T, Lalonde S, Li K, Li Y, Litt A, Lyons E, Manning G, Maruyama T, Michael TP, Mikami K, Miyazaki S, Morinaga S, Murata T, Mueller-Roeber B, Nelson DR, Obara M, Oguri Y, Olmstead RG, Onodera N, Petersen BL, Pils B, Prigge M, Rensing SA, Riano-Pachon DM, Roberts AW, Sato Y, Scheller HV, Schulz B, Schulz C, Shakirov EV, Shibagaki N, Shinohara N, Shippen DE, Sorensen I, Sotooka R, Sugimoto N, Sugita M, Sumikawa N, Tanurdzic M, Theissen G, Ulvskov P, Wakazuki S, Weng JK, Willats WW, Wipf D, Wolf PG, Yang L, Zimmer AD, Zhu Q, Mitros T, Hellsten U, Loque D, Otiillar R, Salamov A, Schmutz

J, Shapiro H, Lindquist E, Lucas S, Rokhsar D, Grigoriev IV (2011) The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332: 960-963

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17: 540-552

Chater C, Kamisugi Y, Movahedi M, Fleming A, Cuming AC, Gray JE, Beerling DJ (2011) Regulatory mechanism controlling stomatal behaviour conserved across 400 million years of land plant evolution. *Curr Biol* 21: 1025-9

Coates JC (2003) Armadillo repeat proteins: beyond the animal kingdom. *Trends Cell Biol* 13: 463-471

Coates JC, Laplace L, Haseloff J (2006) Armadillo-related proteins promote lateral root development in *Arabidopsis*. *Proc Natl Acad Sci U S A* 103: 1621-1626

Coates JC, Moody LA, Saidi Y (2011) Plants and the Earth system - past events and future challenges. *New Phytol* 189: 370-373

Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27: 221-224

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EA, Wu Y, Guo J, He J, Jia Z, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan, Wu Z, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Fang L, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Li S, Zhang X, Wang J, Sun R, Zhang B, Jiang S, Du Y (2009) The genome of the cucumber, *Cucumis sativus* L. *Nat Genet* 41: 1275-1281

Kamisugi Y, von Stackelberg M, Lang D, Care M, Reski R, Rensing SA, Cuming AC (2008) A sequence-anchored genetic linkage map for the moss, *Physcomitrella patens*. *Plant J* 56: 855-866

Kenrick P, Wellman CH, Schneider H, Edgecombe GD (2012) A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. *Philos Trans R Soc Lond B Biol Sci* 367: 519-536

Koduri PK, Gordon GS, Barker EI, Colpitts CC, Ashton NW, Suh DY (2010) Genome-wide analysis of the chalcone synthase superfamily genes of *Physcomitrella patens*. *Plant Mol Biol* 72: 247-263

Lang D, Zimmer AD, Rensing SA, Reski R (2008) Exploring plant biodiversity: the *Physcomitrella* genome and beyond. *Trends Plant Sci* 13: 542-549

Menand B, Yi K, Jouannic S, Hoffmann L, Ryan E, Linstead P, Schaefer DG, Dolan L (2007) An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* 316: 1477-1480

Nibau C, Gibbs DJ, Bunting KA, Moody LA, Smiles EJ, Tubby JA, Bradshaw SJ, Coates JC (2011) ARABIDILLO proteins have a novel and conserved domain structure important for the regulation of their stability. *Plant Mol Biol* 75: 77-92

Page RD (2002) Visualizing phylogenetic trees using TreeView. *Curr Protoc Bioinformatics* Chapter 6: Unit 6 2

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

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

Proust H, Hoffmann B, Xie X, Yoneyama K, Schaefer DG, Nogue F, Rameau C (2011) Strigolactones regulate protonema branching and act as a quorum sensing-like signal in the moss *Physcomitrella patens*. *Development* 138: 1531-1539

Rensing SA, Ick J, Fawcett JA, Lang D, Zimmer A, Van de Peer Y, Reski R (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. *BMC Evol Biol* 7: 130

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

Saidi Y, Finka A, Chakhporanian M, Zryd JP, Schaefer DG, Goloubinoff P (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: 697-711

Saidi Y, Finka A, Muriset M, Bromberg Z, Weiss YG, Maathuis FJ, Goloubinoff P (2009) The heat shock response in moss plants is regulated by specific calcium-permeable channels in the plasma membrane. *Plant Cell* 21: 2829-2843

Schaefer DG (2002) A new moss genetics: targeted mutagenesis in *Physcomitrella patens*. *Annu Rev Plant Biol* 53: 477-501

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

Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56: 564-577

Tewari R, Bailes E, Bunting KA, Coates JC (2010) Armadillo-repeat protein functions: questions for little creatures. *Trends Cell Biol* 20: 470-481

Wellman CH, Osterloff PL, Mohiuddin U (2003) Fragments of the earliest land plants. *Nature* 425: 282-285

1

2

3 **a**

4

5

6

7

8

9

10

11

12

13 **b**

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

b

I	CTATATGCCTTAGTACCTTCACAACAATCAAGAAATATACACTTTGTACCTTTTGCATCA	3660
II	CTATATGCCTTAGTACCTTCACAACAATCAAGAAATATACACTTTGTACCTTTTGCATCA	3660
III	CTATATGCCTTAGTACCTTCACAACAATCAAGAAATATACACTTTGTACCTTTTGCATCA	3660
IV	CTATATGCCTTAGTACCTTCACAACAATCAAGAAATATACACTTTGTACCTTTTGCATCA	3660
V	CTATATGCCTTAGTACCTTCACAACAATCAAGAAATATACACTTTGTACCTTTTGCATCA	3660

I	AGTTTGGACCTCATACAATCTAGTATTTTTTTTATATGCAATGCACCCAAAGGTCCTGATA	3720
II	AGTTTGGACCTCATACAATCTAGTATTTTTTTTATATGCAATGCACCCAAAGGTCCTGATA	3720
III	AGTTTGGACCTCATACAATCTAGTATTTTTTTTATATGCAATGCACCCAAAGGTCCTGATA	3720
IV	AGTTTGGACCTCATACAATCTAGTATTTTTTTTATATGCAATGCACCCAAAGGTCCTGATA	3720
V	AGTTTGGACCTCATACAATCTAGTATTTTTTTTATATGCAATGCACCCAAAGGTCCTGATA	3720

I	TGAGAAATGCAAAGTTGTCTCTCACTCCAAGCTTCTTGTGGTGTGTCATTGATGCAAG GGCA	3780
II	TGAGAAATGCAAAGTTGTCTCTCACTCCAAGCTTCTTGTGGTGTGTCATTGATGCAAG GGCA	3780
III	TGAGAAATGCAAAGTTGTCTCTCACTCCAAGCTTCTTGTGGTGTGTCATTGATGCAAG GGCA	3780
IV	TGGAAATGCAAAGTTGTCTCTCACTCCAAGCTTCTTGTGGTGTGTCATTGATGCAAGAGCA	3780
V	TGGAAATGCAAAGTTGTCTCTCACTCCAAGCTTCTTGTGGTGTGTCATTGATGCAAGAGCA	3780
	** ***** *	
I	ATTGTTGGATATCTATTTCTTGTGTACAT-----	3809
II	ATTGTTGGATATCTATTTCTTGTGTACAT-----	3809
III	ATTGTTGGATATCTATTTCTTGTGTACAT-----	3809
IV	ATTGTTGGACACGTATTTTCATGTGTACAT TGGATTACAGGACAACTTTCATCCAAGAGTCA	3840
V	ATTGTTGGACACGTATTTTCATGTGTACAT TGGATTACAGGACAACTTTCATCCAAGAGTCA	3840
	***** * *****	
I	-----	
II	-----	
III	-----	
IV	TAGCCAA GTTTTGGACGAAAGCTTAATCCAAGAG TCAAAACCTTAGCTATAAGTT TTGGA	3900
V	TAGCCAA GTTTTGGACGAAAGCTTAATCCAAGAG TCAAAACCTTAGCTATAAGTT TTGGA	3900

I	-----	
II	-----	
III	-----	
IV	CGAAAGCTATGGTGAATGCA	3920
V	CGAAAGCTATGGTGAATGCA	3920

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

Supplemental Figure 1

pseudogene <i>PHYSCODILLO1A/1B</i>	ATATCCAACAAGCGGCGGCGGAAGTCAATGTGGCGGTGGACGATCAGGAGCAGCAGCAG 60 ATGTCCAACAAGCGGCGGCGCAATGTCAATGTGGCGGTGGACGATCAGGAGCAGCAGCAG 60 ** ***** *
pseudogene <i>PHYSCODILLO1A/1B</i>	ACTGTTGTCTATAAGAAAGCCAGGATAACATCGTCGCCGAC-----GTCGCT 108 GCTGTTGTATATAAGAAAGCCAGGATAACATCGTCGCCGACATCGTCGGCGCGTCCACT 120 ***** ***** **** *
pseudogene <i>PHYSCODILLO1A/1B</i>	TGTGGTGCATCGGCGGTGCCCTTATAGAATCAGGTGTTAAATTTGGCGCGGTTCGCGAT 168 TGTGGTGCACCGGCGGTGCCGCTACAGAATCAGGTGTTGAAAGTGACGCGGTTCGCGAT 180 ***** ***** ** *
pseudogene <i>PHYSCODILLO1A/1B</i>	GCCCATTTGGACAAGTCTGCCAGATGAAACCGTCTTGGGTTGTTAATCTGCTGAACCAC 228 GCGCATTTGGACAAGTCTTCCAGATGAAACCGTCTTGGGTTGTTAATCTGCTGAACCAC 240 ** ***** *
pseudogene <i>PHYSCODILLO1A/1B</i>	AGGGACATGGCGAGCCTTGCGTCCGTGTGAAAGGATGGCAGGTGCTGGGGAGCTCGCAG 288 AGGGACAGGGCGAGCCTTGCGTCCGTGTGCAAGGATGGCAGGTGCTGGGGAGCTCGCCA 300 ***** ***** *
pseudogene <i>PHYSCODILLO1A/1B</i>	TCGCTGTGGAATTCGTCGGATCTTCGGTTGCACACGTTGAATTCGGAGATGGTGTCTGCG 348 TCGCTGTGGAATTCGTCGGATCTTCGGTTCGACTCGTTGAATTCGGAGATGGTGTCTGCG 360 ***** ***** *
pseudogene <i>PHYSCODILLO1A/1B</i>	CTGGCAGGACGATGCTCGAATTTGGAGGCGCTGAAGTTCCGCAGAGGTGCCTTTGCGAGC 408 CTGGCAGGACGATGCTCGAATTTGGAGGCGCTGAAGTTCCGCAGAGGTGCCTTCGCGAGC 420 ***** ***** *
pseudogene <i>PHYSCODILLO1A/1B</i>	TCAATTGTGGGGTTGCAGGCGAAGGGGCTTAGGGAAATGAGCGGCGATTGCTGCAGCCAG 468 TCAATTGTGGGGTTGCAGGCGAAGGGGCTTAGGGAGCTGAGCGGCGATTGCTGCAGCCAG 480 ***** ***** *
pseudogene <i>PHYSCODILLO1A/1B</i>	CTGTCGGATGCAACCTGTCCATGGTGGTTCGCGCGCATGCGAACCTGGAGAGCA----- 523 CTGTCGGATGCAACCTGTCCATGGTGGTTCGCGCGCATGCGAACCTGGAGAGCTTGCTG 540 ***** ***** *
pseudogene <i>PHYSCODILLO1A/1B</i>	-----AAGCCTT-----TTG----- 534 CTGGGTTTCGGATTGCGAGAGAGTGACGAGCGAGGCATTTGAAGGTGATTGCAAGTGTGTTGT 600 **** *
pseudogene <i>PHYSCODILLO1A/1B</i>	----AACTAC-----CGATCTCTCT-----CCCATCT-- 558 CCGAAATTGCGACGGCTGTGCGTGTGCGGGCTCTTAAAGTGGAGAGAGCGCCATCCAG 660 ** * * * * *
pseudogene <i>PHYSCODILLO1A/1B</i>	-CGATCTTCC-----TCCCT----- 572 GCGTTGTTCAGCATTGCAAGGGTCTGACGAGCTGGGGTTCTCTGGACAGTACACCATC 720 * * * * *
pseudogene <i>PHYSCODILLO1A/1B</i>	-----CTCG-----CTTTCGT-----GAT----- 586 GACGAAGGGGCGTTCGGGGTTCGAGCAGCCTGCGATTGTTGTCGGTAGCGGGATGCAGG 780 *** ** *
pseudogene <i>PHYSCODILLO1A/1B</i>	---TGATGTTAATCCAG-----AGAATAAAGTCC---ACCTGAC---CCA 623 TGTATAGTGTGGAGCAGCGCGCATTTGGTGGAGCAAATTGCCAACCTGGCAGGGCTG 840 * * * * *
pseudogene <i>PHYSCODILLO1A/1B</i>	A---TTTTGTAAAA-----CAACTGCTTTGA-----AAAGCTCCAG 656 GATGTTTCGCGGACAGATATCACTCCGACTGCGCTGATGCAGGTCCTGGCAGGGCCGAG 900 *** * * * * *
pseudogene <i>PHYSCODILLO1A/1B</i>	C-----ATTCTCACACAGAA-----TGAAGCATTTGTG-----CGTG--- 687 CTGAGGGTGGTTTTCGCTCTGAAGTGCCTGTTCTGGAGGAGGCGAGCAACCCCGTGACA 960 * * * * *

Supplemental Figure 1. Alignment of *PHYSCODILLO* pseudogene with the first 960bp of the *PHYSCODILLO1A/1B* coding sequence. BLASTN searching of the *Physcomitrella* genome with *PHYSCODILLO* gene sequences revealed a related 522bp pseudogene sequence on Scaffold 91, 474945-475467.

Supplemental Figure 2.

AGACTCCATGTCCACAGCCTTTATTTATAGACCAGCGGGGTACCACTCACACCACAACAATAAAACGCA
TGACAAACATTGCAAGGACACCATCACTGATGTTTACATGAACGAGAAAAATAAAATAATAAAAAATAA
AAAAAATCCAAAAACAAAAACAAAAAATACAATAAAAAACAAAAACACACAGGACCCATCCACGCCACT
TTCGCTGAGCCACATCTGCACTCTTCCAGCCGACGAGACACCATCGACAAACACTACCTCCTCACAAA
CCACCAAACGGACCTATCATCGGTATTAAAGCTTACAGAGCTCATAATCCTAGAAATAGCATCAGGAGAA
GCAAAGCAGTGAAATGATTTCTAGCTCTCACCGTTGTGGCTCTGCCCCCGACAAACTGAGGTGCAAAT
CACGCACAAAGAGTGACCAAGTCCAGGTAAGTGAACATCTTACCCAAACCAAAACGGACTGTACACACT
AAGCACACCGGCTTGCCGCAACGCTCTTTGTGTTATTAACTCTAGCTTACTTAGCTCACCTGTTTCCA
CCACCACGCCGCTACTGCAGTGAAAACCTTTTAGTCACTCAAAACTGAACGGACTGTAGGCGATAGCCGC
CGGGACGTTAGTGGGTGCGTAGCGGTGGGTGTCAGAGATTGGATTGCAGTTTCCAGGTTCCGGAGATT
CAATACGGGAAATTTGTACCCCTGATTCTTGAGCTTGAGAGCTGGTGTGATGTGTTCCGGAGAGAGTAG
AGACGAGAGTTTGAACGATCAGCAGCACTGGCTCCGGTCGGAACGAGTGTGACAGTGGAAGGCTTGA
TCCAAATGTCTTGTGGCCTGAGTGATGTTTTCTGTGGTTGAGGGCTTTTCAACTCCGACAGACAGTTGG
TGCTGAAAAGTCTGGGGAAGCAAGTGGGTAGATTTCGAATTTGTGCCGTTAGGTCTCCGTCGACACTGGA
ATTAGCGGTTACACGCCGAGTGAACCTGTTAGTGAGCATTTCGGACAGACCATGCCCTAGCAATCTCGCT
CGGTTAGGCTCGTCATTTGGAAACGCTGAATCCGTCCGCAAAAGCTTTTTCTCTTTCTCCCACTACCCCT
ACTCGTAATTCTCTCCCCACACACCTTCATTTGGGCAGTATGGATGCCAATTTTCTTCCCCTTGACTAC
TTCAAATTTCCGTACTCAAAGTGTCTCTGACCTGTGACGCGTTGCCGAAGATATTGCTGGCTGAATT
CGGCGTCTCTGCAACGTAGGACAAATGCTTCAGAGTGCCCGGTCAAAGTAATACGCAGATGGATATC
CCGTTTCTTCAACACGAAATTTCAAAGTACCCACACGATAGTTCAAAACGGAGAAAGTGAGAGGAGAAAA
CGACACCACAACGGCTCATCTCGAAGTTTGCAGAGATCCGAAACAAAGAGTTGGGGCATAGATATGAG
GGGGCGCAGGATAGTGGATTCAAAGTGATATCTGGATTGGAGCGATCAACTAGAGGCAACAGTCAGATT
GATCATGGGACAAGAAACATGAGGTCAATTATAAAACCAGTACCACTGAAAAATATTCAACGATTTTGCT
TCTCAACTTCTGGAACAAGTTTGGAAAGTATGTAGTATAATCAATTAATGGGAGGGACAGCACATGAAA
TATCCAACAAGCGGCGGCGCAAGTCAATGTGGCGGTGGACGATCAGGAGCAGCAGCAGACTGTTGTCT
ATAAGAAAGCCAGGATAACATCGTCGCCGACGTCCGCTTGTGGTGATCGGCGGTGCCCCCTATAGAAT
CAGGTGTTAAATTTGGCGCGGGTCGCGATGCCATTGGACAAGTCTGCCAGATGAAACCGTGCTTGGGT
TGTTTAATCTGCTGAACACAGGGACATGGCGAGCCTTGCGTCCGTGTGAAAGGGATGGCAGGTGCTGG
GGAGCTCGCAGTCTGCTGGAATTCGTCGGATCTTCGGTTGCACACGTTGAATTCGGAGATGGTGTCTG
CGCTGGCAGGACGATGTGCAATTTGGAGGCGCTGAAGTTCCGCGAGAGGTGCCCTTTCGAGCTCAATTG
TGGGGTTGCAGGCGAAGGGGCTTAGGGAATGAGCGCGGATTGCTGCAGCCAGCTGTGCGATGCAACCC
TGTCATGTTGGTTCGCCCCGCGATGCGAACCTGGAGAGCAAAAGGCTTTTGAAGTACCGATCTCTCTCCC
CATCTCGATCTTCTCCCTCTCGCTTTCTGTGATTGATGTTAATCCCAGAGAATAAAGTCCACCTGACCC
AATTTTGTAAACAACACTGCTTTGAAAAGCTCCAGCATTTCTCACACAGAATGAAGCATTTGTGCGTGCTCG
CGGTCCCAACAAGCACTGGCGAGTTTACGGCGAAAAAGTTCTTGGCGCAGGTACAACGTGAAAAAGTAG
GAGAGTGCAATTTACTACAGAGAATTTCCGAGCGAAGTCCATCAAATTTTGGGTTCTGGCACTGCCTC
ATCGCGACGAACGACGGTCCAACATGGTCAACGAAACCGAGAGAACAACTTTCTTGTTCGTATCAGAG
CTTCATTTCCGTTCTTACCAGACGCCGTTCAAGGGCGAAGTCAACTCCAACAACAACAGCTTCCCT
GAGATAAATGCCTTCTAAAAGAGAGTACTATTACCAGCTTGTGACAGCCGAGAACATTAAAGAATTATAC
TACTTTAATTTCTCGAATTTATCAAAATAAGTCCAACAACATATGCATAAAAAATAAATATAAATCATGC
ATATGCATAAAAAATAAATCCCACAGAGAACATGTTAAACATCCCTCTCGAGTGTGCAACTACTGAGAA
TTACGAACACGACTCCGTGCATCTCAGCCACGCATGTATTCTTATACAGTTTCACTGCAATTTTTCAGTTC
CTCCCTGCCCCGCTCTCGGACTTGCACTCGTCAAAATCCTTTGTTCCCATCAAAAAGATACGAACCCATA
TCGTGGGCACAACGTGCACAAAACAGAGACAATTTACTGAAAAACCACCGATGTAAACCAAAACGCACA
AAGAGCAGAGGTGGAAGGGGGTGAAGCAAACGATTGCAACACACAAACACAGACGAGTTCTGATTGGCT
TGTGAGCATGAGGTGGACAAGGCATGCACACAGATTTGGAAGGGAAGCAAATGGGGTAATCAGCCTTGA
GGGTGGTGGGTGTGTTGAAGTGGGCACAGATTCCATATTCCGCGGTGGTGGCATTTGATTCAGATTCC
AGACGGGGCAAAGAAAGGAAATAAACAGAAGAAAGGGAAAGGAAAGAAAAGAAAAGAAAGAAAGAA
GAAATATAAAGGGCGCCAATTATAGGGTGCTTTTCTGGGCGAGAAGGGCAGGCGGTCTCGGTCAACGC
AGGGGAACACAAAAACAGCGAGCAGCGAGGAAGAGGCGACGAAAGGCGACACAGGGATGACGACGACGAC
GAGAAGGAGGACGAGAGAAGAAGAGCGGCGGGGGGGGGTGGTTCGTGGACTCCCTAAGGAAGGAAGGA
AGAGGAAGCTGCCACATGAAAACCAAGCTTCTTTCTCTCCATGAAGAATGGAACAGACTCTCTCTG
CAGGAAAAACTCAACCCCACTCGCTCGCTCCCTCACTCACTCACTCACTCACTCACTCTGTCTGCTGTT
TCTCATTCGTGTCCCACGCTCTCTTGTCTTGTCTCTGCTCTCGCTCTCACTCGTCTGCTACGCCTG
CCCTTCCCCCTCCCCCTCCCGTGTCTATCCCGCAATCCTTAGCGCTTCTCTCCCTTCTCGCCACACCCCC
TCTTTTTCTTTCTCTTTCTCTCTACTTCTGCTTCCGCTCTTCCGTTTCTTCTCATCTGCTCTCTTC
CCTCTTCGATGCCATGCTGCTTGGCCGCTCTTTTCGGTTGGAAGTGAAGAATGATGCCGAGTTTGGG
GTTTGGAGCTGGTTTTCGGTGGTGGCTGCTTAGTATTCAAGGATGATTGTGGGTGATGCCGCCAGGGGAT
GGTTTCTTTTCGAGGAAGAGGGGAGGTGTAGATCGGTGCAGGGTTATGCAGCAAGAATATATTCGATTTG

GGCCGCGTGATGCAATGCTTGAGGATAGGTAGAACACGAAGCGGGGGATGTGACCTTTGATTGCGTGCA
GGAGCGGGTGACGTTCTGATGGGCGAGGACGGAGGAATTGCAGAGTGC GCGGAGTGGAAGGAATAATCG
GCGCGGATATTGGGAGTTGGCAACCCCTAAGAGAGGTTGAAGCGGCAATTTGTGTGGGTGATGACGAAGC
AGTGACGGCACGAAGATGGTGCATGCGCGGTGCGCATGCAAAATGACACAGATCATCGGAGGAGCAGCTG
ATGACATGTCCAACAAGCGGCGGCGCAATGTCAATGTGGCGGTGGACGATCAGGAGCAGCAGCAGGCTG
TTGTATATAAGAAAGCCAGGATAACATCGTCGCCGACATCGTCGGCGGCGTCCACTTGTGGTGCACCGG
CGGTGCCGGCTACAGAATCAGGTGTTGAAAGTGACGCGGGTCGCGATGCGCATTGGACAAAGTCTTCCAG
ATGAAACCGTGCTTGGGTGTTTAATCTGCTGAACCACAGGGACAGGGCGAGCCTTGCCTCCGTGTGCA
AGGGATGGCAGGTGCTGGGGAGCTCGCCATCGCTGTGGAATTCGCTGGATCTTCGGTCGCACCTCGTTGA
ATTCGGAGATGGTGTCTGCGCTGGCAGGACGATGCTCGAATTTGGAGGCGCTGAAATTCCGCAGAGGTG
CCTCTGCGAGCTCAATTTGTGGGTTG CAGGCGAAGGGGCTTAGGGAGCTGAGCGGCGATTGCTGCAAGCC
AGCTGTGCGATGCAACCCGTGTCATGGTGGTGC CCGCATGCGAACCCTGGAGAGCTTGCTGCTGGGTT
CGGATTGCGAGAGAGTGACGAGCGAGGCATTGAAGGTGATTGCAGTGTGTTGTCCGAAATTGCGACGGC
TGTGCGTGTGCGGCGTCTTAAAGTGAGAGAGACGCCATCCAGGCGTTGTTCCAGCATTGCAAGGGTC
TGACGGAGCTGGGGTTCTTGACAGTCACACCATCGACGAAGGGGCGTTGCGGGGTGCGAGCAGCCTGC
GATTTTTGTGCGTAGCGGGATG CAGGTGTATAGTGTGGAGCACGGCGGCGCATTGGTGGAGCAAATTGC
CGAACCCTGGCAGGGCTGGATGTTTCGCGACAGATATCACTCCGACTGCGCTGATGCAGGTCTTGGCAG
GGCCCCGAGCTGAGGGTGGTTTGCCTCTGA ACTGCCCTGTTCTGGAGGAGGGCAGCAACCCCGTGACAC
TGCCGTGCTCGAAGAAGACGGTGGTGTGCTGCTCGGTTACCGGATGTGATGGAAGGGTTGGATGCTTTGC
TGAGCCCCGAGCAATTGGAAGGAGGAGGCAGGCTCGTGCGGAGGTGAGGTGCGGCGGGCGCGCGCGCG
CGGATTCCGAAGTGGCGAAGTGGACGGAGTGATGTTGTGCGATGCGGTTTTGAAAGATCGCGGAGTGCA
ATGCACCGAGTCTGCTGAAGCAGGGGATCGCGATGATGCTTCGACTGGTTCAAAGCGCGCAGGAGGATG
TACAGGAGAGGGCTGCGTGGGCTTTGGCGACGTTTGTAGTGGTGGACGACGAGAACGCGACTGTAGATT
CTGCGCGGGCGGAAGCTGTGATGAACGGGGGCGGGATCGCGCTGCTGTTGGGGCTGGCGAAGTCTTGCC
GCGAGGGGGTG CAGTCGGAAGCTGCGAAGGTAGGGGGGTTGGTGGGGAATGTAGTGGAGTGTGAGTGTG
CACAAGGCAGGCGTGTGCGCAGGAAGGATTGTAGGGTAGCGGGGTTATTGCGTGAGTTGTGGAGCGTGG
GATTGGAGCAGGCTCGGGGAGAGCAGGTGGACGACGATGGTGGTATTTGCGGAATGGGTTGATGTTTGT
GGGAGTGTGGGTCGATGCGGGCAGCTGCGTGTGGGGTG CAGAAGCTGTGGGGAGTTGGACTGACTGG
TAGCTGTTGGGTTGGTTGGTGCAGGCCATCGCAAATTTGTTCGTGAACACGGAGGTTGCCGAAAAGGGTT
GCTTTGGAAGGGGGCATCAGTATCCTGGCTGGTTTGGCGAGGTCCCGGAACCGGTGGGTTGCTGAGGAG
GCGGCGGGGGGTTTATGGAATCTGTGCGTGGGAGAGGAGCACAAGGTGAGGCGCTGTGGTTGGCAATTG
CGATGAAGTGTTTTGTGCGAATTGGTGGTGTGAGCGCGTGGTGTGGACATTGTAGAGCGAGGGGTGA
GGAAGGCAGGTGGTGTGGGAGGTGCGGGTTGTGGAAGGGACAGGGCGCGGCACGGATGTTGAGGTGGG
AACAGTGTAGAGTTGGAAGAGGAACGTGTTTGTGCGAGTCGGTGAGGAAAGCGGAGGGTGGAGAGTG
TAGTGAGGAGCCAGGGGTTAGGGGGTGGGAGATGGATGGTGGGTTGGTTGTCATGGTAATGGTCATGG
TAGCGCACTGACAAATGAGGGCGTGAATGCAAACAGGGTCCGATAGCTGGCGCAGGCGCAATCGAAGCA
CTGGTGGGATTGGCATTCAAGTGGCCCCCGGAGGGGAGGGTGCTGTTGGTGAGCACGTGCATCGCCAG
TAGCTTTAGTGTTGTGGCATGATCCGCACAGCTAGGTTTGACGTTGGGAGTGGTTGACGGGTGGTTGCT
GTGATGGTTGGCATTGTTGGGTGTTGGGATCTTGTGGAATGAATGGTGTGGTGCGGCGAGAGGAAGAAG
GGATGCGTGAGGCTACTACGTTTGGTGTGTGGCATGGGGTGGTGCAGGTTGTTTGGAAAGGCTGGTGG
TGGGCATTGGTGTGTTGGTTTGGGGTGGGTTGTGATGAGCATGTAGATGAGAGAGTG CAGCGGCGAAAGC
GGAAGCGGGGTTGGGATGCGGTTGGGATTGACGTTGAGCATTTTGTTCGGATAGCAGGAGAGAGCAG
CAGGTGCGTTGGCCAACCTAGCAGCCGATGACAAGTG CAGCATGGAAGTCG CAGTGCGGGGGGTGTGC
GGGCTCTGGTGAGGCTGGCGCGGTTTGTAAACACGAAGGCGTGCAAGAGCAGGTACGAGGCAACGGGA
AGCTGAGAGCAGCGGTTGGTAAAGGGGGATGGGGTTTGGAAAGGCTGAGGTTGGGTCGTGATGAGGGGT
AGCGGAGTGAGAAACACACAGCATGGGGTTTGGAAATGAGTGGCAGGGGCGCAGAGTGGAAGGAGGGTTG
GAGGTTTGCAGAGTTAAGTGGTGCAGGACATGGTTGAGTTGGCAGAGAAGAAGTGATTGAGACAGTGGC
GTTGCGGTGAAGGTGTTGCGTGGGGATGGATTTGGTTAGCGTGATTAGCATTGAGGATTAATGAGGGCG
TGACGATAGTGAGGGTTGCGCGTAAGTGAGAGTG CATTGTGTGGGTCTGGAGGCGTTTGGGGCGTGGA
GTGGAGTGTGTTGCACGATCGGTGGAGCCGGTTTTAGGTGCAGCGTTGGCGAAGAGCATGCGGAATGGGA
GAGGAGGTGACGAGTTGCGAGGTGTGACTGGTGTGTTTGGAACTGTGTTTGTAGAGCAGCACGCGCGTTG
GCGAATCTTGCGGCTCATGGAGACAGCAACGGTAACAACGCTGCTGTAGGGCGGGAGGAAGGCGCGCTG
GAAGCTCTGGTGCAGCTGACTTGCTCAAATCATGAAGGTGTCAGGTATGCGGATTTGTTGGGGTGGTGG
ACAGCAGGGAGGTGTTTTTGTGTG CAGGGGTGATGTGAAGGGTTGAGGGTAGAGATGGGCCGGTGTTC
TAAGAGTGTTTTTGGGCGTCCCCCTTGACGGCGCTTG CAGACAAGAAGCAGCGGGAGCATTGTGGAATCT
GTCGTTG CAGACAGAAACCGAGAAGCTATTGCAGCGGCTGGTGGCGTGGAGGCACTGGTGAGTGAGAG
TGGACACGAACCGTGGGCGTAGGAAGAGATACGAAAGGGGCGTTCTGGGGGCAGAGGTGAATGGGGGTG
ATTCGAGAGCGTGGCCGGGGGGTTGAGCAGAGCTGGAGGTGAAGATGAAGGATGCAGGCTTAGTGCGGA
CATGTAGAGTGATGGGGGGCTGCATGACAGTTGAGCGTGGGGTGGTTGTGGAGGGTAGGATCGGGGTGC
GTTGATGTTGAGATCGGGTTGGGGGGCAGTCGTGTGGCGAGTTGTGCGAGGTTTTGCGAGGCGTGAACA
GGCTGTGTTGAGGCAAGGGCGCACGGTGGTGTGGTGTGGAGCAGAGTGGGGGTGGCATTGGGTGCATGA

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

CGATGGTGTGGTGGTTGTGCAGGTGGCTTTGGCTCAAGGGTGTAGCAGCGGATCTCAAGGTCTGCAAGA
GAGGGCTGCGGGTGCATTATGGGGGCTGTCTGTGTCTGGAGGCGAACAGGTGAGGGCATTGAGCGCGATG
GGAGGTGGTGTGTTGGCGGAAGCGTTGGGCGTGGGTTGAAGGTCTGCAGGGGTAGAGTCGTCGGGGATG
GAGCGGATTGTGTGAAGTATTGCGCGGAGTGAGAGTCGATGGGTTGGATTTGGCGATTATGGGGATTTT
GGTTGCGTAGCAGAGGTTGGAGGTGGGTGACGGGGTGTGGTGCATGACCGTGTGTCTGACGTGTGTGT
GTGTGTGTGTGTGTGTTGGGAACCTTGCCTGGTGGCGCAGCATTGCTATCGGGCGCGAAGGCGGAGTG
GCTCCACTGATCACCTGGCGCACTCTGATTTTGGAGTGAGTTGTGGGATGAGTTGCAGGGTTTTGGGG
GACGAGGTTTTGGGTTTGGTTAGCGCAGCGTGCATGTCATGTGAGTGACACGGTTAGACGCGAGTGGTT
GGGGTGTGATGGTGGAGGCTGGGTGGTGGTGGCATACAAAGTGCAAAGTGACAGGAGGCGAGAGCTGGA
TGCATGGGTGAGTGGGAGGAATGGGGTGTAGTGGAATGGAAAGGGGGCGTGGTGTGGAGCCATGGTTG
GGAGATGGTTGTACGTGGCGGGGCATTGGTGGTGGTATCGGTGTGGTGGGACGATGATGGGCGTTTAGG
AGGTGGGTTTTCTGAAGTTTTCCGGCGATGGGAGTGACAGGACGTACATGAAACAGCGGTGGGAGCGTTGTG
GAACCTGGTTTTTCAATCCTGGCAACGCACTGCGGATGGTGGAAAGAGGAGGGAGTCCCGGCGCTGGTGCA
TTTGTGCTCGTCGTCGAGATCTAAGATGGCGCGGTTTTATGGCAGCTCTGGCTCTGGCGTACATGTTCTGA
TGGCAGGTAAAGCGGTGGCGTGTGGGTGTGTGAGGGTGCAGTTCGTAGGACGTGGGGGTGGTGAGATGAC
GTGAGACGGGGAGGAGGAAGTTTCGTAGGCGTTGGATAGGTGCGGTGAGGGAGAGTCGGGTGAAGTGGAG
CAGTGGGGAGTGGGTGACGTGGTTTTGGAGAACCGGAAGATATGGAAGAAGGGCGCGGTTAGGAATGAG
GAGATTGCTGTTGCGTTGTTGGGACGAGGTTGGGGCGGGGTGAGAAAGTGGAGGGGGTTGAGGTGCGTA
GGAGCGCGCGTTTGTGTGCATGATTTGAAGTGGTTGGGCGCGCTTTGGAGCGGTTGAGTGGGTGTGGCA
AAGCAGTGAGTGGGAGTGATGGAGAGTGAACGTGATCGGAGTGGTTGACGTCATAGTTGGGCTCCTCTG
ACGCAGGATGGATGAGGTTGCAGTTGGATTGTGCGAGTGTGGAGAACAACGGGAGGACAGTGAGCTTGGGA
GGCAGCAAGGAAACTGGCGTTGCGAAACATCGATGCGTTTCGTGCAAACGTTTTTTTGATCCGCAGTCTTT
GACAGCTGCTGCGTCATCGTGGGCCGGTGCGACGTTGAACCAGGTTCGCGGAGACCGCCACAATCCAAGA
AGCCGGGCATCTGAGATGCAGGTAGGCGATGCGTACGTAGGGCGTGAATGTGTGGATGGTACAGAATTA
GATGCAACGGATTTGGATGCATGGGCAGCGGAAGCAACGTGGGCGCGTGGTAAGGAATGGACGCGAGTG
GGTGTGGGCATGACAGGCGGTTCGGGGTGTGGGACAGTGGCGGTGATGAGGTAGCGGATAGTGAAGGGC
GGTGTGGAGTGCAGTGCCTGGAGTGTGTTTGTATGGACGAGGCTAAGAGCGTGCGGGTGGGTGGTTGGTG
CAGTGTAGCCGAGCTCGGGCGGTTTTGTGGCCATGTTGAGGAACGGACATGCAGTTTTGCGATCGTGTGC
GGCATTGCGTCTTCTGCAGGTAAGCAGTGTGAGTGAGCATGAAGAGGACAAAGCGGTGGTGAGGGACGC
GAGGTGGGTGTGATGGTAGGGTTGTTATTGATTGCGCCGTTGGAGTGTGTCGTGAGAGAGAGGAAGTG
TTGGAGCGAGGTTGGAGGTTGGTTTTGGCGTGATGGATTGGGATGTGGTTGCGTGTACAGTTTACGAT
GCCTGGAGGCAGGCATGCGCTGCACCACGCAGTTCTGCTGCAGAAAGTTGGTGCAGTTCGGGTTCTGCG
GACTGTGGCTGCAGCAGCGACGGCAGCAATGCAGGCCAAAGTGTTCGCCCGAGTGGTTCTTCGAAACCT
GGAGCATCATCAGGCGGAGAGCAACACGGTGGTGGCACTGTCGTAAAGAGTTGGGATGAGGTGGAAGAA
GCGGGGTAGCACACGAGGGGGTTCGAGTTGGTGTGAAGTAAATTTGAAAAGTGCATGCCATTGGAGGATT
CTTCTTTTTTAATCTTCTTTTTTTAAAATGTTTTCTTTTTTCTTTTTTTTTGGGTGAGGTAGTCCTTG
GTTTGCAGGATAGGTTTGTACGATTGAAAGCTAAAAGACGAATGTAGGAGAGAAGTGGCGGTCCGCAA
TGCAGGTTGAAAACCGCATTGGTGTGAAACCGCATTTCATGGTGTGATGAAAGAGCACTGAGAGTTTGT
GTGCAGCTTGGTTGTATGCCACAGTCACACGAGTATTTGAGTAATTCACAAGGCCACGTTAATTGTAGT
CGGGCATGGTTGTTGGAGTGGGACTCCGGTGGATGCATTTTCGGTTTTGGTTCTAGTTTTTTTACGCAGCT
TGGTTAGAATACTAATTTCCCAAAGTAGATAGCTAATGAGATTTTCATCCTATTTGTTACCGTGTTTTTT
TTTCTGGATTATTTTGTAGCGAATGTAGACTGTTGCACCTTTCGCAATTTCTGTAGCGATTATATTAT
TCACTCTTTCACTGAATTAGCTGTTTTTTCATTGTTTCTTGTATATCTGCCATTTTAAGGATTTATTAG
TTTCAAGTTCATATCATTTGCACAAGCAAATAAATGAACAACCTAAATGGACAATCTAAATACACCACT
GATGCAATATCCCCAATTGTTGATATAAACACATGTTGCTAGTGAGTCAGCAAACAAAATGATCTTTTT
GACCATCAACACAACCTCATTCTTAAGATATGCTTATCAAATAAACTACTACAAAAGTTACAAAGTT
ACATTATAAAGCCTCCCAACAATACTACAAGCAAGGTCAAAGAGAAAACTAAAAATCCATGTTCT
ATTCTTTTCGACAACAACATGAGAAAAAGCAAAAAATTTATTCCAAGTAGTAGTGTTTTTGTCCCCCTAG
AGGTTACATTGGGTCTATTTGTTTCTTGTGTTGGCTATAGTGAGAGGCAATTGAATTGGTGGTTTTGTT
GTATCACTTTGTAATAACTTGCAATACTTTGAGTGGTCATAAGATCCATATGGTAATAGAAGCCATCTC
AATTGTAGGGGCTTCTCAGCACCCATAAAGATAATTACAAAAGAAATAATTATAAATATTCTACCTA
GAATTCTTTATTGAAGAAAAGAAGTCTAAAAGCTACAAACACTTAGCATCAGCAAGGATTGTATTCTAA
GAGAAGTGTGGAGAAAGTTATCTCTTACTTCCAACACCTTTTACTTATAGATGGAGTTTACATCCCAAG
TACATAGACTTCATCTAGAGTCACCTTCATCCTTTTAACTAGATCTAGATTCATCTTAACCTAATTGCC
ACCTTAAAACCTAAGTGTACCATAGTAACCTCTTTTAACTCCTTGGGTGCCTTTTGACTATTAGCAAC
TTATTATGCTTAATCTTTATGTGGTGGTAACTAAGTAAGAACTTGTGATCTTCAACACTCCCACTTTA
CTAATTTTCAAAAGGTTTTTAATCCTAATCTCTGAATCAAGTGATGATGTCTATTTTTTCACTAATGCCT
TTGTAAGCACATCAACAAGCATGTCATTTGTAGCAAAAATACTTCATACCTATTATTCCATCTTGAATCA
TCTCTCATATGAAATGATGTTGAATATCAATGTGTTTGGTTCCTGAGTGGTGTGTTTTGGATTCTTGGCAA
ATGCAAGATATCTTTGATTGTACACATAATTAGAATTGCTTCAATTGCACTAGACCAATGTCTTCTA
AGAATTGCCTTATCCAACACGCTTCCATTGCACAATGACTTGCTGCCATGTACTCTGCTCCATCGTTG

ATCTTGCAATGGTAGATTGTCTCTTGTAATTCCATGAAATGACTCCCACTCCAATGAAGAAGACATATC
CCATTGTGGATCTTCTATCATTTGCATCTCCAGTCCAATCCGCATCACAATAACCTTGCAAATCAATGT
TCTTACCACCAAGGCAAAATTTGAAGTCTTTTGTACTTTTAAGTACCTCAATATCCTTTTTACTATTGT
CCAATGCATTGGACTAGCTTTAGACATGTGTTGGCTTACAACACTTATCGCATATGCAAGATCCACTCT
TGTGGCCACCATTGCATACATGAGTGAGCCTACTGCTACTTTGTATGGAACACTTTGTATCTCACCTTC
AACTTCTACAAATTCCTCATCCGTGAGCTTCATCAACTTCAAAATTTGCCTCCAAGGGTGTTCATGGG
CTTGCATCCTCCATGTTGAATCGTTTCAACACTTCTTCAATGTACTTGATTGTTGGCTCATTGTAATGGT
CTTGTTACCACGATCTCTTACAAATTCCTCCCAAGCAGTAGTGCAATTCTTCCAAATCACTCATCTC
ATATTCCTTCTCTAATTTGGCCTTCATCCAATCCATCTTTTCCATATTGTTAGCCAAGATGATCAAATC
ATCAACATAAAGGATTAAAGTCAGCAAAAACCTCTAGTTTGCATCACATACAATGAATGATCAGCCTC
ATTCCTCACAAACCTTCATTGATAAAAAAAGGTGTCAATTTGTTGGTACCATACCCT
TGGTGATTGTTTCAATCCATATAAGGATTTCTTCAATTTGCACACAAGATGTTCTTTGCTTTTTTGTAT
AAATTCTTGAGGTTGGTCCATGTAGATGTCTTCTTATAATTCTCCATTCAAAAAGGCCGTCTTTACATC
CATTTGGTGTATTTCCAAATCCATTGCTGCACCAATTGCTACAATGGTCCTAATGGTGGTGAATTTGGC
AACGGGTGCATAAGTCTCATTAAATCAATTCCAACAACCTGTGAATATCCTTTAGCAATTAGTCTTGC
CTTGTGCCTCACCATCACATCCATTGCATCTCTTTTAGTCTAAACATCCACTTACAACCAACACTAGA
TCGATTTGTTGGAAGAGGTACCAATTTCCATGTACCATTGTCCATAAGAAATTGGTATTCTTCTTCCAT
AGCCTTCTCCCACTTGATTGCATCTTCACACTTCATAGCATCACACAAAGACAATGGATCAACAAGAAG
ACAACATTTGCATGTTCTTCACTTACTTGAGGGAAGATGTGATCCTTCTACCATTGCGCCAGTCGCT
TCTTCTCTCAAAAGATATCTCCTTCACTTATAGAAGAATATTCACCTTGTGCATCTATGTTCTTT
TTCAACCTTCTTCTCTCTTAAACCTTGTCGTTTCTCACTCAATGGTTGATACTCTTTATCATTATCATC
ATTCCAATCTCCAACATCTTCATTTCTTTTCTTGTGTTGGTCTTAGCAATTGTAGGAGCCTTCATCTTTTC
AACCTTCTTTGGCTTTTCATCTTCATCATCTTCATTATCTTCATCTTCTTGTTTTTTTGCTGAATTCAA
AGATGTATCCACAAACACACCATTACTCCCACTTGACACATCTCCAAGTGCTCACAAGCACTCTTATG
CTCCAAAACTTAGCATCACAACATTTGGTGATCTTCTTGGTTTCCAAGCACATTAGCCTATATGCCTT
AGTACCTTCACAACAATCAAGAAATATACACTTTGTACCTTTTGCATCAAGTTTGACCTCATACAATC
TAGTATTTTTTATATGCAATGCACCAAGGTCTGATATGAGAAATGCAAAGTTGTCTCTCACTCCA
AGCTCTTGTGTGTCATTGATGCAAGGCAATTTGTTGGATATCTTATTTCTGTGTACATTGCATTAC
CATAGCTTTTCGTCCAAAACCTTATAGCTAAGGTTTGTGACTCTTGGATTAAAGCTTTTCGTCCAAAACCTTGGC
TATGACTCTTGGATGAAAGTTGTGCTGAATCCAATGTACACATGAAATACGTGTCCAACAATTGCTCTT
GCATCAATGACACCACAAGAAGCTTGGAGTGAGAGACAACCTTTGCATTTTCCATATCAGGACCTTTGGG
TGCATTGCATATAAAAAAATACTAGATTGTATGAGGTCCAACTTGATGCAAAAGGTACAAAGTGTATA
TTTCTTGATTGTTGTGAAGGTACTAAGGCATATAGGCTAATGTGCTTGGAAACCAAGAAGATCACC
AGTTGTGATGCTAAGTTTTTGGAGCATAAGAGTGCTTGTGAGCACTTGGAGATGTGTCCAAGTGGGAGT
AATGGTGTGTTTGTGGATACATCTTTGAATTCAGCAAAAAACAAGAAGATGAAGATAATGAAGATGAT
GAAGATGAAAAGCCAAAGAAGGTTGAAAAGATGAAGGCTCTACAATTGCTAAGACCAACAATGAAAGA
AATGAAGATGTTGGAGATTGGAATGATGATAATGATAAAGAGTATCAACCATTGAGTGAGAAACGACAA
GGTTTAAGAGAGAAGAAGGTTGAAAAGAACATAGAGTGCAACAAGGTGAATATTCTTCTATAAGTGAA
AGGAGATATCTTTTGGAGAGAAAGAAGGCGACGTGGCGAATGGTAGAAGGATCACATCTTCCCTCAAGTA
AGTGAAGAACATGCAATGTTGTTCTTCTTGTGATCCATTGTCTTTGTGTGATGCTATGAAGTGTGAA
GATGCAATCAAGTGGGAGAAGGCTATGGAAGAAGAATACCAATTTCTTATGGACAATGGTACATGGA
TTGGTACCTCTTCCAACAAATCGATCTAGTGTGTTGGTGTAAAGTGGATGTTTAGGACTAAAAGAGATGCA
ATGGATGTGATGGTGAGGCACAAGGCAAGACTAATTGCTAAAGGATATTCACAAGTTGTTGGAATTGAT
TTTAATGAGACTTATGCACCCGTTGCCAAATTCACCACCATTAGGACCATTTGTAGCAATTGGTGCAGCA
ATGGATTTGGAAATACACCAATGGATGTAAAGACGGCTTTTGAATGGAGAATTATAAGAAGACATC
TACATGGACCAACCTCAAGAATTTATACAAAAAAGCAAGAACATCTTGTGTGCAAAATGAAAGAAATCC
TTATATGGATTGAAACAATCACCAGGGTATGGTACCAACAAATGACACCTTTTTTTTTTTTTTTTTTT
ATCAATGAAGGGTTTGTGAGGAATGAGGCTGATCATTGATGTATGTGATGCAAACTAGAGAGTTTTTG
CTGACTTTAATCCTTTATGTTGATGATTTGATCATCTTGGCTAACAATATGGAAAAGATGGATTGGATG
AAGGCCAAATTAGAGAAGGAATATGAGATGAGTGATTTGGAAGAATTGCACTACTGCTTGGGAGTGGAA
TTTTGTAAGAGATCGTGGTAACAAGACCATTACAATGAGCCAAATCAAGTACATTGAAGAAGTGTGAAA
CGATTCAACATGGAGGAGTGCAAGCCCATTGGAACACCTTGGAGGCAATTTGAAGTTGATGAAGCTC
ACGGATGAGGAATTTGTAGAAGTTGAAGGTGAGATACAAAGTGTCCATACAAAGTAGCAGTAGGCTCA
CTCATGTATGCAATGGTGGCCACAAGAGTGGATCTTGCATATGCGATAAGTGTGTAAGCCAACACATG
TCTAAAGCTAGTCCAATGCATTGGACAATAGTAAAAAGGATATTGAGGTACTTAAAAGTACAAAAGACT
TCAAAATTTTGCTTGGTGGTAAAGAACATTGATTTGCAAGGTTATTGTGATGCGGATTGGACTGGAGATG
CAAATGATAGAAGATCCACAATGGGATATGTCTTCTTCAATTGGAGTGGGAGTCATTTTATGGAATTACA
AGAGACAATCTACCATTGCAAGATCAACGATGGAAGCAGAGTACATGGCAGCAAGTCATTGTGCAATGG
AAGCGTGTGGAATAAGGCAATTTCTAGAAGACATTGGTCTAGTGCAATTAGAAGCAATTTCTAATTATGT
GTGACAATCAAAGATATCTTGCAATTTGCCAAGAATCCAAAACACCACTCAAGGACCAACACATTGATA
TTCAACATCATTTTATATGAGAGATGATTCAAGATGGAATAATAGGTATGAAGTATTTTGTACAAATG

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ACATGCTTGTGATGTGCTTACAAAGGCATTAGTGAAAAATAGACATCATCACTTGATTGAGAGATTAG
GATTAAAAACCTTTTGAATAATTAGTAAAGTGGGAGTGTGAAGATCAACAAGTTCTTACTTAGTTACCA
CCACATAAAGATTAAGCATAATAAGTTGCTAATAGTCAAAAGGCACCCAAGGATTAAAAGAAGTTACTA
TGGTGACAGTTAGGGTTTTAAGGTGGCAATTAGGTTAAGATGAATCTAGATCTAGTTAAAAGGATGAAG
GTGACTCTAGATGAAGTCTATGTACTTGGGATGTAAACTCCATCTATAAGTAAAAGGTGTTGGAAGTAA
GAGATAACTTTCTCCACACTTCTCTTAGAATACAATCCTTGCTGATGCTAAGTGTGTTGTAGCTTTTAGA
CTTCTTTTCTTCAATAAAGAATTCTAGGTAGAATATTTATAATTATTTCTTTTGTAAATTATCTTTATGG
GTGCGTGAGGAAGCCCCTACAATTGAGATGGCTTCTATTACCATATGGATCTTATGACCACTCAAAGTA
TTGCAAGTTATTACAAAGTGATACAACAAACCACCAATTCAATTGCCTCTCCACTATAGCCAACACAAG
AAACAAATAGACCCAATGTAACCTCTAGGGGGACAAAACACTACTACTTGGAAATAAATTTTTTGGCTTT
TCTCATGTTGTTGTGCGAAAAGAATAGAACATGGATTTTTAGTTTTTCTCTTTGACCTTGCTGTAGATT
TTGTTGTGGGAGGCTTTATAATGTAACCTTTGTAACCTTTGTGTAGTAGTTTATTTTGTATAAGCATATCTT
AAGAATGAGGTTGTGTTGATGGTCAAAAAGATCATTTTGTTTGTGCTGACTCACTAGCAACATGTGTTTAT
ATCAACAATTGGGGATATTGCATCAGTGGTGTATTTAGATTGTCCATTTAGGTTGTTTATTTATTTGCT
TGTGCAAATGATATGAACCTGAAACTAATAAATCCTTAAAAATGGCAGATATACAAGAAACAATGAAAAA
ACAGCTAATTGAGTGAAGAGTGAATAATATAATCGCTACAAGAATTCGCAAGGGTGCAACAGTCTACA
TTCGCTAGCAAAATAATCCAGAAAAAAAACACGGTAACAAATAGGATGAAATCTCATTAGCTATCTAC
TTTGGGAAATTAGTATTCTAACCAAGCTGCGTAAAAAACTAGAACCAAAACCGAAATGCATCCACCGGA
GTCCCCTCCAACAACCATGCCCCGACTACAATTAACGTGGCCTTGTGAATTACTCAAATACTCGTGTGA
CTGTGGCATAACAACCAAGCTGCACAACAACTCTCAGTGCTCTTTCATCACACCATGAATGCGGTTTCA
CACCAATGCGGTTTTCAACCTGCATTGCGGACCGCCACTTCTCTCTACATTCGTCTTTTAGCTTTCAA
TCGTGACAAACCTATCTTGCAAAACCAAGGACTACCTCACCCAAAAAAAAGGAAAAAAGAAAAACATTTT
AAAAAAGAAAGATTAAAAAGAAAGAATCCTCCAATGGCATGCACTTTTCAAATTTACTTCACACCAACT
CGACCCCCCTCGTGTGCTACCCCGCTTCTTCCACCTCATCCCAACCTCTTACGACAGTGCCACCACCGTG
TTGCTCTCCGCCTGATGATGCTCCAGGTTTGAAGAACCCTCGGGCGAACACTTTGGCCTGCATTGCT
GCCGTGCTGCTGCAGCCACAGTCCGCAGAACCCGACTCGCACCAACCTTCTGCAGCAGAACTGCGTGG
TGCAGCGCATCCCTGCCTCCAGGCATCGTGAATGTGACACGCAACCACATCCCAATCCATCAGCCAA
AACCAACCTCCAACCTCGCTCCAACACTTCTCTCTCTCACGCACATCACTCCACCGGCGAATCAATAA
CAACCCTACCATCACACCCACCTCGCGTCCCTCACCAACCGCTTGTCTCTTTCATGCTCACTCACACTG
CTTACCTGCAGAAGAGCGAATGCCGCACACGATCGCAAAAACGATGTCCGTTCTCTCAACATGGCCACA
AACCGCCCGAGCTCGGCTACACTGCAACCAACCCACCCGACGCTCTTAGCCTCGTCCATACAAACAAC
ACTCCACGCACCGCACTCCACACCGCCCTTCACTATCCGCTACCTCATCACCGCCACTGTCCCATCACC
CCGACCGCCTGTCATGCCACACCCACTCGCGTCCATTCTTACCACGCGCCACGTTGCTTCCGCTGC
CCATGCATCCAAATCCGTTGCATCTAATTCTGTACCATCCACACATTCACGCCCTACGTACGCATCGCC
TACCTGCATCTCAGATGCCCGGCTTCTTGATTGTGGCGGTCTCCGCGACCTGGTTCAACGTCGCACCG
GCCACGATGACGCAGCAGCTGTCAAAGACTGCGGATCAAAAACGTTTGCACGAACGCATCGATGTTT
CGCAACGCCAGTTTCTTGTGCTGCCTCCAAGCTCACTGTCTCTCCGTTGTTCTCCACACTCGACAATCCA
ACTGCAACCTCATCCATCCTGCGTCAGAGGAGCCCAACTATGACGTCAACCACTCCGATCACGTTCACT
CTCCATCACTCCCCTCACTGCTTTGCCACACCCACTCAACCGCTCCAAACGCCGCCCAACCACTTCAA
ATCATGCACACAAACGCGCGCTCCTACGCACCTCAACCCCTCCACTTCTCACCCCCGCCCAACCTCG
TCCCAACAACGCAACAGCAATCTCCTCATTCCTAACCGCGCCCTTCTTCCATATCTTCCGGTTCTCCAA
AACCACGTACCCACTCCCCTGCTCCACTTACCCGACTCTCCCTCACCGCACCTATCCAACGCCTA
CGAACTTCTCTCTCCCGTCTCACGTATCTCACCACCCCAACGTCCTACGACCGCACCTCACACACC
CACACGCCACCGCTTTACCTGCCATCGAATGTACGCCAGAGCCAGAGCTGCCATAAACCGCGCCATC
TTAGATCTGACGACGAGACAAATGCACCAGCGCCGGGACTCCCTCCTCTTCCACCATCCGCAGTGGC
TTGCCAGGATTGAAAACAGGTTCCACAACGCTCCCAACCGCTGTTTCATGTACGTCCTGCACTCCCATC
GCCGGAACCTTCAGAAACCCACCTCCTAAACGCCCCATCATCGTCCCACCACACCGATACCACCACCAAT
GCCCCGCCACGTACAACCATCTCCCAACCATGGCTCCACACCACGCCCCCTTTCCATTTCCACTACACC
CCATTCTCTCCACTCACCCATGCATCCAGCTCTCGCTCCCTGCACTTTGCACTTTGTATGCCACCACC
ACCCAGCCTCCACCATCACACCCCAACCACTCGCGTCTAACCGTGTCACTCACATGACATGCACGCTGC
GCTAACCAAAACCCAAAACCTCGTCCCCAAAACCTGCAACTCATCCACAACCTCACCTCAAAATCAGA
GTGCGCCAGGGTGATCAGTGGAGCCACTCCGCTTCGCGCCCGATAGCAATGCTGCGCCACCACGCAAG
TTCCCAACACACACACACACACACACACACGTCAGACACACGGTCATCGCACCACACCCCGTCAACC
ACCTCCAACCTCTGCTACGCAACCAAAATCCCCATAATCGCCAAATCCAACCCATCGACTCTCACTCCG
CGCAATACTTCACACAATCCGCTCCATCCCCGACGACTCTACCCCTGCGACCTTCAACCCACGCCAAC
GCTTCCGCAACAACACCACCTCCCATCGCGCTCAATGCCCTCACCTGTTCGCTCCGACACAGACAGC
CCCCATAATGCACCCGACGCCCTCTCTTGAGACCTTGAGATCCGCTGCTACACCCCTTGAGCCAAAGCC
ACCTGCACAACCACACACCATCGTCATGCACCCAATGCCACCCCACTCTGCTCCACACCACACCACC
GTGCGCCCTTGCTCAACACAGCCTGTTACGCCTCGCAAAACCTCGCACAACCTCGCCACACGACTGCC
CCCCAACCCGATCTCAACATCAACGCACCCCGATCCTACCTCCACAACCACCCACGCTCAACTGTCA
TGCAGCCCCCATCACTCTACATGTCCGCACTAAGCCTGCATCCTTCATCTTCACTCCAGCTCTGCTC

AACCCCCCGGCCACGCTCTCGAATCACCCCCATTACCTCTGCCCCAGAACGCCCCCTTTCGTATCTCT
TCCTACGCCCACGGTTTCGTGTCCACTCTCACTCAACAGTGCCCTCCACGCCACCAGCCGCTGCAATAGCT
TCTCGGTTTTCTGTGTCGAACGACAGATTCCACAATGCTCCCGCTGCTTCTTGTCTGCAAGCGCCGTCC
AAGGGGACGCCCCAAAACACTCTTAGAACACCGGGCCATCTCTACCCCTCAACCCCTTACATCACCCCTGC
ACAACAAAAACACCTCCCTGCTGTCCACCACCCCAACAAATCCGCATACCTGACACCTTCATGATTTGA
GCAAGTCAGCTGCACCAGAGCTTCCAGCGCGCTTCTCCCGCCCTACAGCAGCGTTGTTACCGTTGCT
GTCTCCATGAGCCGCAAGATTGCGCAACGCGGTGCTGCCTATCAAAACAGTTCCAAAACACCAGTCA
CACCTCGCAACTCGTCACCTCCTCTCCCATTCGCGATGCTCTTCGCCAACGCTGCACCTAAAACCGGCT
CCACCGATCGTGCAAAACACTCCACTCACGCCCCAAACGCCTCCAGGACCCACACAATGCACTCTCCACT
TACGCGCAACCCCTCACTATCGTCACGCCCCATTAATCCTCAATGCTAATCAGCTAACCAATCCATC
CCACGCAACACCTTCACCGCAACGCCACTGTCTCAATCACTTCTTCTCTGCCAACTCAACCATGTCCT
GCACCACTTAACCTCGCAACCTCCAAACCCCTCCTTCCACTCTGCGCCCCGCGCACTCATTCCAAACCC
CATGTGTGTGTTTTCTCACTCCGCTACCCCTCATCACGACCCAAACCTCAGCCTTCCAAACCCCATCCCC
CTTTACCAACCGCTGCTCTCAGCTTCCCGTTGCCTCGTACCTGCTCTTGCAAGCCTTCGTGGTTACAAA
ACCGCGCCAGCCTCACAGAGCCCGCACACCCCCCGCCACTGCGACTTCCATGCTGCACCTTGTATCGG
CTGCTAGGTTGGCCAACGCACCTGCTGCTCTCTCTGCTATCCGGAACAAAATGCTCAACGTCAATCC
CAACCGCATCCCAACCCCGCTTCGCTTTCGCGCTGCACTCTCTCATCTACATGCTCATCACAACCCA
CCCTCAAACCAACACCAATGCCACCACCAGCCTTTCACACACCTCGCACCACCCCATGCCACACAC
CAAACGTAGTAGCCTCACGCATCCCTTCTTCTCTCGCCGACCACACCATTCAATTTCAAACAAGATCC
CAACACCCAAATGCCAACCATCACAGCAACCACCCGTCACCACTCCCAACGTCAAACCTAGCTGTGCG
GATCATGCCACAACACTAAAGCTACTGGGCGATGCACGTGCTCACCAGGACACCCCTCCCTCCGCGGG
CCACTTGAATGCCAATCCCACAGTGCTTCGATTGCGCCTGCGCCAGCTATCGCACCCCTGTTTGCATT
ACGCCCTCATTTGTGAGTGCGCTACCATGACCATTACCATGACAACCAACCCGACCATCCATCTCCAC
CCCCTAACCCCTGGCTCCTCACTACACTCTCCACCCTCCGCTTTCCTCACCGACTCGCACAAACACGTT
CCTCTTTTCCAACCTCTACACTGTTCCCAACCTCAACATCCGTGCGCGCCCTGTCCTTCCACAACCC
GACCTCCCACACCACCTGCCTTCTCTACCCCTCGCTCTACAATGTCCACACCGACGCGCTCACACCAC
AATTGCGCAAAAACACTTCATCGCAATTGCCAACACAGCGCCTCACCTTGTGCTCCTCTCCACCGA
CAGATTCCATAAAGCCCCCGCCCTCCTCAGCAACCCACCGGTTCCGGGACCTCGCCAAACACAGCCAG
GATACTGATGACCCCTTCCAAAGCAACCCCTTTTCGCAACCTCCGTGTTACGGAACAAATTTGCGATGGC
CTGCACCAACCAACCCAAACAGCTACCAGTCAGTCCAACTCCCCACAGCTTCTGCACCCACACGCACTG
TGCCCGCATGCGACCCACACTCCCAACAAATCAACCCATTCCGCAAAATACCACCATCGTCGTCCACCT
GCTCTCCCCGACCCCTGCTCCACCCACGCTCCACAACCTCACGCAATACCCCGCTACCTTACAATCCT
TCCTGCGCACACGCTGCTTGTGCACACTCACACTCCACTACATTCCCCACCAACCCCTTACCTTCG
CAGCTTCCGACTGCACCCCTCGCGCAAGACTTCGCCAGCCCCAACAGCAGCGGATCCCGCCCCCGT
TCATCACAGCTTCCGCCCCGCGAGAATCTACAGTCGCGTTCCTGCTGTCCTCCACTACAAACGTCGCCA
AAGCCGACGACGCTCTCCTGTACATCCTCCTGCGCGCTTTGAACAGTCGAAGCATCATCGCGATCC
CCTGCTTCAGCAGACTCGGTGCATTGCACTCCGCGATCTTCAAACCGCATGCGACAACATCCACTCCG
TCCACTTCGCCACTTCCGAATCCGCGCGCGCGCGCCCGCCGCACTCGACCTCGCGCACGAGCCTGCCT
CCTCCTTCCAATTGCTCGGGCTCAGCAAAGCATCCAAACCTTCCATCACATCCGTGAACCGAGCCAGCA
CCACCGTCTTCTTCGACGACGGCAGTGTCACGGGGTTGCTGCCCTCCTCCAGAACAGGGCAGTTAGAG
CGCAAACACCCCTCAGCTCGGGCCCTGCCAGGACCTGCATCAGCGCAGTCGGAGTGATATCTGTCCGCG
AAACATCCAGCCCTGCCAGGTTTCGGCAATTTGCTCCACCAATGCGCCGCGGTGCTCCACACTATACACC
TGCATCCCGCTACCGACAAAATCGCAGGCTGCTCGCACCCCGAACGCCCCCTTCGTGATGGTGTGAC
TGTCCAGGAACCCAGCTCCGTGACACCTTGAATGCTGGAACAACGCCTGGATGGCGTCTCTCTCCA
CTTTTAGGACGCCCCGACGACACAGCCGTGCAATTTGCGACAACACACTGCAATCACCTTCAATGCCT
CGCTCGTCACTCTCTCGCAATCCGAACCCAGCAAGCTCTCCAGGTTGCGATAGCCGGCTGACCCACCA
TGGACAGGTTGTCATCCGACAGCTGGCTGCAGCAATCGCCGCTCAGCTCCCTAACCCCTTCGCCTGCA
ACCCACAAATTGAGCTCGCAGAGGCACCTCTGCGGAACCTTCAGCGCCTCCAAATTGAGCATCGTCCTG
CCAGCGCAGACACCATCTCCGAATTCACAGAGTGCGACCGAAGATCCAGCGAATTCACAGCGATGGCG
AGCTCCCCAGCACCTGCCATCCCTTGCACACGGACGCAAGGCTCGCCCTGTCCCTGTGGTTTACAGAT
TAAACAACCCAAAGCACGGTTTTCATCTGGAAGACTTGTCCAATGCGCATCGCGACCCGCGTCACTTTCAA
CACCTGATTCTGTAGCCGGCACCGCCGGTGCACCACAAGTGAGCGCCGCGACGATGTCGGCGACGATG
TTATCCTGGCTTTCTTATATACAACAGCCTGCTGCTGCTCCTGATCGTCCACCGCCACATTGACATTGC
GCCGCGCTTGTGGACATGTCATCAGCTGCTCCTCCGATGATCTGTGTCAATTTGCATGCGCACCGCGC
ATGCACCATCTTCGTGCCGTCACTGCTTCGTATCACCCACACAAATTGCCGCTTCAACCTCTCTTAGG
GTTGCCAACTCCCAATATCCGCGCCGATTATTCTTCCACTCCGCGCACTCTGCAATTCCTCCGTCTCT
GCCCATCAGAACGTACCCGCTCCTGCACGCAATCAAAGGTACATCCCCCGCTTCGTGTTCTACCTAT
CCTCAAGCATTCATCACGCGGCCCAATCGAATATATTCTTGCTGCATAACCCCTGCACCGATCTACAC
CTCCCTCTTCTCGAAAGAAACCATCCCTGGCGGCATCACCCACAATCATCCTTGAATACTAAGCAG
CCACCACCGAAACAGCCTCAAACCCCAACTGCGGCATCATCTTCACTTCAAACCGAAAAGACGCGG
GCAAGCAGCATGGCATCGAAGAGGGAAGAGAGCAGATGAGAAGAAACGGAAGAGCGGAAGCAGGAAGTA

GAGAGAGAAAGAGAAAGAAAAAGAGGGGGTGTGGCGAGAAGGGAGGAAGCGCTAAGGATTGCGGGATAG
ACACGGGAGGGGGAGGGGGAAGGGCAGGCGTAGCGACGACGAGTGAGAGCGAGAGCGAGAACAAGAGCA
AGAGAGCGTGGGACACGAATGAGAAACAGCAGACAGAGTGAGTGAGTGAGTGAGTGAGTGAGGGAGCGA
GCGAGTGGGGTTGAGTTTTTCTGCAGAGAGAGTCTGTTCCATTCTTCATGGAGGAGGAAAGAAGCTTT
GGTTTTTCATGTGGCAGCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTCTCTCT
TTCTTCTTCTGCGTCT
TGCTCGCTGTTTTGTGTTCCCTGCGTTGACCGATGACCGCTGCCCTTCTCGCCAGAAAGGCACCC
TATAATTGGCGCCCTTATATTTCT
TTATTTCTCTTCTTTGCCCCGTCTGGAATCTGGAATCAATGCCACCACCGCGGAATATGGAATCTGTGTC
CCACTTCAACACACCCACCCCTCAAGGCTGATTACCCCATTTGCTTCCCTTCCAAATCTGTGTGCA
TGCTTGTCTCCACCTCATGCTCACAAGCCAATCAGAAGCTCGTCTGTGTTTGTGTGTTGCAATCGTTTGCT
TCACCCCTTCCACCTCTGCTCTTTGTGCGTTTTTGCTTTTACATCGGTGGTTTTTTCAGTAAATTGTCTCT
GTTTTGTGCAGTTGTGCCCCAGATATGGGTTTCGTATCTTTTGATGGGAACAAAGGATTTGACGAGTGC
AAGTCCGAGAGTACGGGCAGGGAGGAACTGAAAATTGCACTGAACTGTATAAGAATACATGCGTGGCTG
AGATGCACGGAGTCGTGTTTCGTAATTCTCAGTAGTTGCGACACTCGAGAGGGATGTTTAAACATGTTCCC
TGTGGGATGACTCTTGTGCTTTTTCTGCATCTGTCCATGGTAAGTTAAATTGTAGTTTTGTATCAAGTT
CTGTATTTTATGAAAATGTGTATTCTATGGATTAAGTGCTATTTGTTCTGTAGTCAACTCTTTAGTTAA
TCCTGCATTACTTTTGATTTAAAATTCAGTGGATTGTTTCGAGATGGTTTCGGATTTTTCGAGCCAATT
TGGAGTTACTGTTTATTTAATTTTTCTTTTACATCGTGTAGTGCTTTTTTTTTCTTTTTTTTTTATCTTT
TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTCAAGGTGGAATAGGTATGGATTGGAAGTCGATAACGGATT
GTTGTGCATTTGCGTGATGAATTTCCACCAACTTATCAAAGGACTTATGTTCTTGAACGTATTTCTGAA
AATACTGGTACTCTGCTTGGGCGGCTTTTAAATGCATCTGCAACTGATCTTGACCACTCGACTTTTCCG
TCCCTGGAATTCGGGAGAAGCTGTGGGGTTGAGTGGGGGCTAGAATTGGGTCGGTGAGGAACTGAGAAG
TTTACTCCTTGACCATCGAAAGTGTTTCGTCTCTTGTTCGTCCGGCACTTGTACCCGTCGTTTCGCGGC
GACGAGGGAGTGCCAAAACCCAAATTCCGAGGGACACCGGTCGGCGCAGTTTTGCGGTGTGAATATTAC
ACCCCTACTTGTTGAAGTTGTACTCGCGTTTGAAATTTGTTGTTACAATTTGTGCAAGCTTGTAAAGAA
CCGCGAGCTCGTAACATGACATAATTTTGTGAGAGAATGCTGGGTTTTGTATAGAAGTTGGTGGAAGAA
GTTGATTGTAGATGGTCTCCACCGTCTTGACCAACATTTATTACGAAAGTGAGAGGGAGGAAGAGAGA
GAGAGACGAGAGAGAGAGCGGGAGAGAGATCGGCAATTCTTAAACCTTTTCCGATCATTCGTTGT
CTGGAGCAATCAACAGAGAGAGCTCAAGGAGCGTGAGGAATGCGGGTTCGACCTGAAATAAGTTCGTTA
TTTTCGTTTTAGGTCGGCTTAGCAGATGAAGTCCGCTGAATTGCTGCTTGGACTGTCAATTGGTTGCTCC
AGAGGCTCGATGGGAGGAAGGTGTTTGGATGCGAATGAAGTCGTGTAAGGAGGTTGAACGACTTTTGA
GTGGCAAATTCGGTCATATGCTGGCAAGGATGGCTGCTTGGAGGAGTTTGGTTCCAGTGTCTTGGAGAT
TGTGTTGCAGGGAAGTCTCTTCGGTAAGCAAATCATTAGTTTGTGGTTTGATCCAAATCGGGCAACATG
TTGCCCTATCAAACCTCGCTTGTCAAGGCACAGCGAAGATGTTTTTACTCTTTTTAGGGTCACCATTCGT
TGCTGGCTCATCACACAATGCCAAGAGTTGTATTGTAATCGCGACTTAATCCATACACGTCTGCAGA
CACAGAAAATTGGTCAGAAAATGGCCAGTATTTTGTACCAGGCGATGGTATTCTCAACCATGGACGG
CTCCCTACGATGCGGGGCGAGCTACGTCTGAATCGCACCTCAGGTATGCCCTTTGCGTGGGTCTGAA
TGATGGAAGCGGAGAATCTGTTCTGGCATTTTGGAAGTTTCAGGTAACGAATTTGTGCAATGCTGGAAA
GTGATGTTAATCTGCATAAAGGTGGCAGATGCTGTGAGGGGATGAGCTGTTCCCGGCATTAGTCTTCG
AGATTTCAAGGAAAGAAGAACAGCACTTCGGTCTCGTTTACCCACAGAGCATGGTCATATTGTCAGC
TGCTTCTGTCAAACAA

Supplemental Figure 2. Complete sequence of the *PHYSCODILLO1A/1B* genomic locus from 5' to 3' as shown in Figure 2.
Black text: Scaffold 91 sequence 5' to *PHYSCODILLO1B* gene.
Dark blue text: *PHYSCODILLO1B* promoter; underlined promoter sequence is shared with *PHYSCODILLO1A*.
Red text: *PHYSCODILLO1B* gene, ATG to TAA
Purple text: *PHYSCODILLO1B* intergenic region; the 7 nucleotide differences with *PHYSCODILLO1A* intergenic region are marked in black.
Dark green text: *PHYSCODILLO1A* intergenic region
Dark grey text: unique between *PHYSCODILLO1A* and *PHYSCODILLO1B* intergenic regions.
Light blue text: *PHYSCODILLO1A* gene, TAA to ATG
Maroon text: *PHYSCODILLO1A* promoter; underlined promoter sequence is shared with *PHYSCODILLO1B*.
Orange text: Scaffold 91 sequence 3' to *PHYSCODILLO1A*.

Supplemental Figure 3.

a

<i>PHYSCODILLO2.theoretical</i>	GTTGAGGCTTGGACAGTGGGTAACGACAGCAGTGGTTCCTCAATGCAA---	5347
<i>PHYSCODILLO2.experimental</i>	GTTGAGGCTTGGACAGTGGGTAACGACAGCAGTGGTTCCTCAATGCAATTG	5350

<i>PHYSCODILLO2.theoretical</i>	-----	
<i>PHYSCODILLO2.experimental</i>	GACGTGGACAAAATTTGAGAGTTCGGCGTGGTTGGTTGCAGTGGAGCGGA	5400
<i>PHYSCODILLO2.theoretical</i>	-----	
<i>PHYSCODILLO2.experimental</i>	GCTTGGACGGTTTGTGGCCATGTTGAGGAATGGTTCTGCAGTTCTCCGAA	5450
<i>PHYSCODILLO2.theoretical</i>	-----	
<i>PHYSCODILLO2.experimental</i>	CATGTGCTGCGTTTGTCTCTCTGCAGGTAATCACTGGTGCTGATCGACAG	5500
<i>PHYSCODILLO2.theoretical</i>	-----	
<i>PHYSCODILLO2.experimental</i>	GACAGAGTTGACGGAGCGATGGGGTGGTTGTATCATTTTTAGGTTCTGTTT	5550
<i>PHYSCODILLO2.theoretical</i>	-----	
<i>PHYSCODILLO2.experimental</i>	TGTTACGGGGAGTGTGGACTGAGGTTTAAGGTACCTTTGACTTAGGTT	5600
<i>PHYSCODILLO2.theoretical</i>	-----GGTTTACTGACCCGTAGGG	5366
<i>PHYSCODILLO2.experimental</i>	TACTGACCCGTAGGGTGTGGGTGTGTGGAACGGTTTACTGACCCGTAGGG	5650

<i>PHYSCODILLO2.theoretical</i>	TGTGGGTGTGTGGAACAGTTTACTATGCCTGGAGGTAGGCATGCGATGCA	5416
<i>PHYSCODILLO2.experimental</i>	TGTGGGTGTGTGGAACAGTTTACTATGCCTGGAGGTAGGCATGCGATGCA	5700

b

Supplemental Figure 3. *PHYSCODILLO2* was misannotated in the *Physcomitrella* genome.

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.

Supplemental Figure 4.

a

<i>SELAGIDILLO</i> .theoretical_gDNA	TGTTTCTGCACTGGTGC	2650
<i>SELAGIDILLO</i> .theoretical_cDNA	TGTTTCTGCACTGGTGC	2215
<i>SELAGIDILLO</i> .experimental_gDNA	TGTTTCTGCACTGGTGC	2650
<i>SELAGIDILLO</i> .experimental_cDNA	TGTTTCTGCACTGGTGC	2215

<i>SELAGIDILLO</i> .theoretical_gDNA	GATTCATGGCCGCTCTGGCTCTGGC	2700
<i>SELAGIDILLO</i> .theoretical_cDNA	GATTCATGGCCGCTCTGGCTCTGGC	2259
<i>SELAGIDILLO</i> .experimental_gDNA	GATTCATGGCCGCTCTGGCTCTGGC	2700
<i>SELAGIDILLO</i> .experimental_cDNA	GATTCATGGCCGCTCTGGCTCTGGC	2258

<i>SELAGIDILLO</i> .theoretical_gDNA	ACTTTTCTTA-GTGTTC	2749
<i>SELAGIDILLO</i> .theoretical_cDNA	-----	
<i>SELAGIDILLO</i> .experimental_gDNA	ACTTTTCTTAAGTGTTC	2750
<i>SELAGIDILLO</i> .experimental_cDNA	-----	
<i>SELAGIDILLO</i> .theoretical_gDNA	TCTGAGTGCTCGTTTGCCTGCTTGT	2799
<i>SELAGIDILLO</i> .theoretical_cDNA	-----ATGGATGAGGTTACTACAAAC	2280
<i>SELAGIDILLO</i> .experimental_gDNA	TCTGAGTGCTCGTTTGCCTGCTTGT	2800
<i>SELAGIDILLO</i> .experimental_cDNA	-----	
<i>SELAGIDILLO</i> .theoretical_gDNA	GAAGTTGTTTACTGTGATAGCATTACCA	2849
<i>SELAGIDILLO</i> .theoretical_cDNA	GAAGTTGTTTACTGTGATAGCATTACCA	2330
<i>SELAGIDILLO</i> .experimental_gDNA	GAAGTTGTTTACTGTGATAGCATTACCA	2850
<i>SELAGIDILLO</i> .experimental_cDNA	-----CATTACCAAAACGGCGTGGCAAGGCAGTC	2288

<i>SELAGIDILLO</i> .theoretical_gDNA	GGCCATGAAGAACATTGAGGCGTTTGTACAAGCGTTCTCGGATCAACCGT	2899
<i>SELAGIDILLO</i> .theoretical_cDNA	GGCCATGAAGAACATTGAGGCGTTTGTACAAGCGTTCTCGGATCAACCGT	2380
<i>SELAGIDILLO</i> .experimental_gDNA	GGCCATGAAGAACATTGAGGCGTTTGTACAAGCGTTCTCGGATCAACCGT	2900
<i>SELAGIDILLO</i> .experimental_cDNA	GGCCATGAAGAACATTGAGGCGTTTGTACAAGCGTTCTCGGATCAACCGT	2338

b

Supplemental Figure 4. *SELAGIDILLO* was misannotated in the *Selaginella* genome.

a) Experimental sequences were aligned to theoretical sequences obtained from the *Selaginella* genome resource using ClustalW2. Nucleotides missing from the cloned cDNA (compared to the annotated cDNA sequence) are in red. 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 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 the shorter exon 10 present in the cloned *SELAGIDILLO* cDNA sequence.

Supplemental Figure 5.

Supplemental Figure 5. RT-PCR detection of *PHYSCODILLO* gene expression in protonemata and gametophores.

RNA was prepared from 7-day old protonemata and 3-week old gametophores and cDNA generated.

Left hand panel: full length *PHYSCODILLO1A/1B* (indistinguishable) and *PHYSCODILLO2* (both ~2.8kb) cDNAs were amplified from protonemal tissue.

Right hand panel: partial fragments (approximately 500bp) of cDNA unique to either *PpTubulin* (control), *PHYSCODILLO1A/1B* or *PHYSCODILLO2* were successfully amplified by PCR. In order to look at semi-quantitative expression, 10µl aliquots were removed from a 50µl PCR reaction after 22, 25, 28, 31 and 35 cycles. The presence of a band indicated the relative expression of each of the genes amplified.

Supplemental Figure 6

Supplemental Figure 6. Generation of a *physcodillo2* knock-out mutant by targeted gene replacement.

a) A *PHYSCODILLO2* gene deletion construct was generated by amplifying two *PHYSCODILLO2* homologous sequences (5': nucleotides 98-1628 and 3': nucleotides 4103-5900, denoted as 5' and 3', respectively (grey boxes in the diagram)) and inserting them into the vector pAHG1 either side of a hygromycin resistance cassette driven by the rice actin promoter and terminated by the 35S terminator (*Act1p-hph-35St*; white box). *physcodillo2* deletion mutants were generated by using this cassette to replace part of the open reading frame of the *PHYSCODILLO2* gene by homologous recombination (indicated using X).

b) An initial PCR screen of 68 putative *physcodillo2* deletion mutants was carried out using two *PHYSCODILLO2* gene-specific primers (red arrows; expected size 2.7kb). Lines that produced no 2.7kb PCR product, were putative *physcodillo2* deletion mutants, which were subjected to further analysis detailed in (c).

c) Correct-sized genomic PCR products (5901bp) were amplified from 5 of the remaining lines using a *PHYSCODILLO2* upstream primer (green arrow) and the 3' gene-specific primer used previously (right hand red arrow). As shown in (d) below, 2 lines (105 and 110 – asterisks) were confirmed as single-insertion events.

d) The PCR products from (c) were sequenced using a primer designed to detect an integration event (blue arrow in (a)). Sequence alignments between the sequenced mutant locus, *PHYSCODILLO2* genomic DNA sequence, and hygromycin cassette sequence showed that there were two lines named *physcodillo2-105* (P2-105) and

physcodillo2-110 (*P2-110*) in which *PHYSCODILLO2* had successfully been replaced (red indicates identity with hygromycin cassette; underlining indicates identity with *PHYSCODILLO2* locus). Together, the 6kb genomic DNA PCR product in (c) plus the sequencing of the 5' junction show that a single insertion event had taken place in each of these two lines.

e) No *PHYSCODILLO2* cDNA expression is detected in the *physcodillo2-105* mutant (*P2-105*), using the purple primers in (a), which detect a cDNA product in wild type. *PHYSCODILLO1* cDNA is still detected in the mutant and *TUBULIN* is detected in wild type and mutant.

For Peer Review

Supplemental Table 1

Primer pair	Primer A	Primer A sequence	Primer B	Primer B sequence
1	P1Bupstream3F	5'-AGACTCCATGTCCACAGCCT	P1-3'-1	5'-CTCATCCCAACCTCTTACGACAG
2	P1-3'-1	5'-CTCATCCCAACCTCTTACGACAG	P1Adownstream1R	5'-TTGTTTGACAGAAGCAGCTGAC
3a, 3b	P1A/1B.promF.NotI	5'-AAGCGGCCGCATCCACAGAGAACATGTTAAACA	P1A/B.prR.BamHI	5'-AAAGGATCCGTCATCAGCTGCTCCTCCGAT
4	P1/2seq5	5'-GCGTACATGTCGATGGCAGG	P1A/1B_D.RC	5'-TTGGACGAAAGCTATGGTGAATGC
5a, 5b	P1/2-5' Start	5'-ATGTCCAACAAGCGGCGGCG	P1-3'-1	5'-CTCATCCCAACCTCTTACGACAG
6	P1A/1B_D	5'-GCATTACCATAGCTTTCGTCCAA	P1/2seq5	5'-GCGTACATGTCGATGGCAGG
7	P1A_prF.NotI	5'-AAGCGGCCGCACGGTGACAAGTGCCGGACGAA	P1A/B.prR.BamHI	5'-AAAGGATCCGTCATCAGCTGCTCCTCCGAT
8	P1B_prF.NotI	5'-AAGCGGCCGCTACGGCGAAAAAGTTCCTGGC	P1A/B.prR.BamHI	5'-AAAGGATCCGTCATCAGCTGCTCCTCCGAT
9	P1A/1B_G	5'-TTTGGCGTGATGGATTGGGAT	P1A/1B_D.RC	5'-TTGGACGAAAGCTATGGTGAATGC
10	P1Bupstream3F	5'-AGACTCCATGTCCACAGCCT	P1A/B.prR.BamHI	5'-AAAGGATCCGTCATCAGCTGCTCCTCCGAT
11	P1A/1B_D	5'-GCATTACCATAGCTTTCGTCCAA	P1A/1B_G	5'-TTTGGCGTGATGGATTGGGAT
12	P1Bupstream3F	5'-AGACTCCATGTCCACAGCCT	P1/2seq1RC	5'-TCTCGCAATCCGAACCCAGCAGC
13	P1Bupstream3F	5'-AGACTCCATGTCCACAGCCT	P1/2seq4RC	5'-CAGTGCCTCCACGCCACCAG

Supplemental Table 1.

Table of primers used for *PHYSCODILLO1A/1B* locus cloning and sequencing. Each numbered pair of primers in the table corresponds to the numbered PCR products shown in the locus diagram below the table.

Supplemental Table 2.

ARABIDILLO gene homologue	Gene sequence identifier (GenBank/NCBI/Phytozome v8.0) and reference(s) for previously published sequence	Exons	Introns
<i>Physcomitrella patens</i> PHYSCODILLO1A	ACQ55239/EDQ67952 [1,2]	12	11
<i>Physcomitrella patens</i> PHYSCODILLO2	EDQ80735 [1]; ACQ55238 [2]	12	11
<i>Physcomitrella patens</i> PHYSCODILLO1B	EDQ67893 [1]	12	11
<i>Selaginella moellendorffii</i> SELAGIDILLO	EFJ24022.1 [2,3]	12	11
<i>Arabidopsis thaliana</i> ARABIDILLO1	AAC31834/AEC10482 [4,5]	11	10
<i>Arabidopsis thaliana</i> ARABIDILLO2	AE80050 [4,5]	9	8
<i>Arabidopsis lyrata</i> ARABIDILLO1	EFH58247 [2,6]	11	10
<i>Arabidopsis lyrata</i> ARABIDILLO2	EFH54588 [2,6]	9	8
<i>Thelungiella halophila</i> ARABIDILLO1	Thhalv10001300m	11	10
<i>Thelungiella halophila</i> ARABIDILLO2	Thhalv10005776m	9	8
<i>Capsella rubella</i> ARABIDILLO1	Carubv10025681m	11	10
<i>Capsella rubella</i> ARABIDILLO2	Carubv10016656m	9	8
<i>Brassica rapa</i> ARABIDILLO1	Bra040336	7	6
<i>Phaseolus vulgaris</i> ARABIDILLO	Phvulv091021074m	12	11
<i>Linum usitatissimum</i> ARABIDILLO1	Lus10042900	12	11
<i>Manihot esculenta</i> MANIHODILLO1	cassava4.1_001220m [2]	12	11
<i>Manihot esculenta</i> MANIHODILLO2	cassava4.1_001242m [2]	12	11
<i>Ricinus communis</i> RICINODILLO1	EEF50270 [2,7]	12	11
<i>Ricinus communis</i> RICINODILLO2	EEF51610 [2,7]	12	11
<i>Populus trichocarpa</i> POPLARDILLO1	EEE73805 [2,8]	12	11
<i>Populus trichocarpa</i> POPLARDILLO2	EEE80501 [2,8]	12	11
<i>Populus trichocarpa</i> POPLARDILLO3	EEE82729 [2,8]	12	11
<i>Glycine max</i> GLYCINODILLO	NCBI Gene ID: 100805530 [2,9]	12	11
<i>Glycine max</i> ARABIDILLO1-like	NCBI Gene ID: 100819086 [2,9]	12	11
<i>Cucumis sativus</i> CUCUMIDILLO	Cucsa.259690.1 [2,10]	12	11
<i>Prunus persica</i> ARABIDILLO homologue	ppa001073m [11]	12	11
<i>Vitis vinifera</i> VITIDILLO1	NCBI gene ID 100263231 [2,12]	12	11
<i>Vitis vinifera</i> VITIDILLO2	gene ID 100247439 [2,12]	13	12
<i>Mimulus guttatus</i> MIMULODILLO1	mgv1a001095m [2]	13	12
<i>Mimulus guttatus</i> MIMULODILLO2	mgv1a001313m [2]	12	11
<i>Aquilegia coerulea</i> ARABIDILLO homologue 1	Aquca_013_00826	12	11
<i>Aquilegia coerulea</i> ARABIDILLO homologue 2	Aquca_067_00022	12	11
<i>Eucalyptus grandis</i> ARABIDILLO homologue	Eucgr.D01517.1	12	11
<i>Citrus clementina</i> ARABIDILLO homologue	clementine0.9_001792m	12	11
<i>Oryza sativa</i> ORYZADILLO	EEC67450/EEE51407/BAF37336/AAG60190 [2,13 14]	12	11
<i>Sorghum bicolor</i> SORGHODILLO	EER94481 [2,15]	12	11
<i>Brachypodium distachyon</i> BRACHYDILLO	Bradi3g33510 [2]	12	11
<i>Zea mays</i> ZEADILLO1	GRMZM2G079031 [2,16]	12	11
<i>Zea mays</i> ZEADILLO2	GRMZM2G028796 [16]	11	10
<i>Setaria italica</i> ARABIDILLO homologue	Si034193m	12	11

Supplemental Table 2. Full-length land plant ARABIDILLO homologues for which exon/intron number in the coding sequence has been determined. Where available, published references to the gene sequences are given. Database identifiers are provided for each gene/protein: where present the GenBank identifier is shown in blue, or in red for other NCBI gene identifiers. Where the sequence is not obtainable via NCBI, the Phytozome v8.0 (www.phytozome.net; [17]) identifier given is in black. The identifier given for PHYSCODILLOs/SELAGIDILLO is that available prior to this publication.

References

1. Rensing *et al.* (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319 p.64-69

2. Nibau *et al.* (2011) ARABIDILLO proteins have a novel and conserved domain structure important for the regulation of their stability. *Plant Molecular Biology* 75 p.77-92

3. Banks *et al.* (2011) The *Selaginella* genome reveals genetic changes associated with the evolution of vascular plants. *Science* 332 p.960-933
4. The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408 p.795-815
5. Coates *et al.* (2006) Armadillo-related proteins promote lateral root development in *Arabidopsis*. *PNAS* 103 p.1621-1626
6. Hu *et al.* (2011) The *Arabidopsis lyrata* genome sequence and the basis of rapid genetic change. *Nature Genetics* 43 p.476-481
7. Chan *et al.* (2010) Draft genome sequence of the oilseed species *Ricinus communis*. *Nature Biotechnology* 28 p.951-956
8. Tuskan *et al.* (2006) The genome of the black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313 p.1596-1604
9. Schmutz *et al.* (2012) Genome sequence of the palaeopolyploid soybean. *Nature* 463 p.178-183
10. Huang *et al.* (2009) The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics* 41 p.1275-1281
11. Ahmad *et al.* (2011) Whole genome sequencing of peach (*Prunus persica* L.) for SNP identification and selection. *BMC Genomics* 12: 569
12. Jaillon *et al.* (2007) The grapevine genome sequence suggests ancient hexaploidization in major angiosperm phyla. *Nature* 449 p.463-467
13. Goff *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296 p.92-100
14. International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature* 436 p.793-800
15. Paterson *et al.* (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457 p.551-556
16. Schnable *et al.* (2009) The B73 maize genome: complexity, diversity and dynamics. *Science* 326 p.1112-1115
17. Goodstein *et al.* (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* 40 p.D1178-1186