CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD IRAPUATO

Caracterización funcional del gen *AINTEGUMENTA-PLETHORA-LIKE* en *Marchantia polymorpha*

Tesis que presenta

Melissa Dipp Álvarez

Para Obtener el Grado de

Maestro en Ciencias

En la Especialidad de

Biotecnología de Plantas

Director de Tesis: Dr. Luis Alfredo Cruz Ramírez

Irapuato, Guanajuato   Julio, 2017

Acknowledgements

First, I want to express my gratitude to my advisor Dr. Luis Alfredo Cruz Ramírez, for his guidance, encouragement and the effort he makes to always have the time to answer my questions or discuss experiments.

I want to thank my thesis committee: Dr. Nayelli Marsch Martínez and Dr. Stefan de Folter, for their help and interest in this project.

To Consejo Nacional de Ciencia y Tecnología (CONACyT), for the scholarship granted.

To all the Molecular and Developmental Complexity lab members. For their support and everyday laughs. Specially to Alejandro Aragón Raygoza, the other "plant kid", who is always up for a good plant evolution discussion. Thanks for your friendship, lab mate.

Special thanks to Erik Díaz, who challenges me intellectually and inspires me to always give my best.

To my family. My mom, dad and brothers. Your daily good-morning messages and pictures from everyday life make me feel like I'm not 2 500 km away from home. Thank you for all your love and support.

Index

1. Resumen

La genómica comparativa ha revelado que los miembros de linajes basales de plantas comparten un set de factores de transcripción altamente conservados con las plantas modernas. Mientras que el número de copias de genes se ha expandido a lo largo del tiempo, se ha predicho que la diversificación y recableado de redes regulatorias implicadas en el desarrollo está directamente relacionado con las innovaciones morfológicas presentes en plantas terrestres (Bennett et al., 2014; Pires and Dolan, 2012).

La familia AINTEGUMENTA-LIKE (AIL) de factores de transcripción en *Arabidopsis thaliana* incluye ocho genes considerados reguladores maestros de procesos de desarrollo. Las proteínas AIL se expresan principalmente en tejidos en división, donde regulan el mantenimiento de nichos de células madre y el correcto desarrollo del embrión y órganos de la raíz y la parte aérea de la planta. Por ejemplo, en la doble mutante *plt1;plt2* de *A. thaliana* el nicho de células madre de raíz se arresta a causa de la total diferenciación de las células iniciales y el centro quiescente,, lo cual arresta el crecimiento de la raíz principal y el correcto desarrollo de las raíces laterales.

En este estudio, sentamos las bases para contestar la siguiente pregunta: ¿Qué función podría tener un gen AIL en *Marchantia polymorpha*, una planta basal con una de las arquitecturas más simples de las plantas terrestres?

Nuestros resultados muestran que *AINTEGUMENTA-PLETHORA-LIKE* de Marchantia *(MpAPL)* es un ortólogo de la familia génica *AIL* en *A. thaliana*, que la estructura fundamental de la proteína está conservada entre linajes y que la duplicación de genes *AIL* en dicotiledóneas, seguido por la divergencia en sus regiones promotoras, pudo haber resultado en un patrón de expresión y funciones espacio-temporalmente diverso. En nuestro análisis de conservación funcional, el fenotipo de la doble mutante *plt1;plt2* de A. thaliana fue parcialmente complementado por la sobreexpresión de MpAPL. Esto sugiere que la función de los genes AIL en el mantenimiento de los nichos de células madre y su interacción con auxinas está conservada desde la divergencia de las Hepáticas.

2. Abstract

Comparative genomics has revealed that members of basal lineages of land plants share a set of highly conserved transcription factors with the derived flowering plants. While gene copy numbers have expanded through time, it has been predicted that diversification, co-option and reassembly of gene regulatory networks implicated in development is directly related to the innovations in land plant body (Bennett et al., 2014; Pires and Dolan, 2012).

The AINTEGUMENTA-LIKE (AIL) transcription factor family in *Arabidopsis thaliana* includes eight genes considered master regulators of developmental processes. AIL proteins are mainly expressed in dividing tissues where they regulate the maintenance of stem cell niches and the correct development of the embryo, root and shoot organs. For instance, *A. thaliana plt1;plt2* double mutants fail to maintain Quiescent Center cells with correct identity, the size of the root meristem of the primary and lateral roots decreases and all cells eventually differentiate.

With this study, we aim to unravel the following biological question: What function would an *AIL* gene have on *Marchantia polymorpha*, a basal organism with one of the simplest body architectures of land plants?

We show that *AINTEGUMENTA-PLETHORA-LIKE* from Marchantia *(MpAPL)* is an ortholog of the *A. thaliana AIL* gene family, that the protein core structure is well conserved between lineages and that AIL gene duplication in eudicots was followed by divergence in their promoter regions, which may have resulted in a diverse spatiotemporal pattern of expression and functions. In our functional conservation analysis, *A. thaliana plt1;plt2* double mutant phenotype is partially complemented by the overexpression of MpAPL, which is evidence, though preliminarily, that the function of AIL genes in stem cell regulation and their interaction with auxin is conserved since the divergence of Liverworts.

3. Introduction

The plant terrestrialization process, also known as the colonization of land by plants, took place 425-490 million years ago (Sanderson, 2003) within a terrestrial microbial moist biofilm composed of fungi, bacteria, cyanobacteria and green algae (Wellman and Strother, 2015). Prior to land colonization, freshwater green algae related to extant Charophyte algae had already evolved different types of forms; from unicellular organisms to multicellular filaments and thallose forms (Harrison, 2017). It is hypothesized that first land plants originated from one of these types of ancestral Charophyte algae living in the terrestrial biofilm mentioned above and adapting to its aerial conditions (Wellman and Strother, 2015; Ishizaki, 2016). Terrestrialization was critical for the formation of earth's biosphere, a critical change that made possible the colonization of terrestrial environments by animals.

What allowed early plants to diversify and thrive on changing terrestrial environments was the evolution and change from a gametophyte-dominant to sporophyte-dominant life form, the three-dimensional growth form originated by division plane orientations in stem cells and the development of a sporophytic apical meristem and organs such as roots, vascular tissue, leaves, flowers and seeds (Pires and Dolan, 2012; Ishizaki, 2016). Phylogeny, evo-devo and functional studies using plants that diverged early in the evolution of land plants, can help elucidate the genetic change that accompanied such innovations in form.

Among extant land plants, the divisions Anthocerotophyta (hornworts), Bryophyta (mosses) and Marchantiophyta (liverworts) are considered descendants of the first events of embryophyte diversification (Wickett et al., 2014). This groups share common traits considered basal, which include a dominant haploid gametophyte and a sporophyte that depends on the gametophyte form, the dependence on water for sexual reproduction and the lack of roots and vascularization. The liverwort *Marchantia polymorpha* has been an experimental organism for studying the evolution of morphophysiological characteristics among land plants. The recent availability of its genome sequence (https://phytozome.jgi.doe.gov/pz/portal.html) as well as molecular tools has facilitated comparative genomics and functional gene research on this model plant.

In land plants, comparative genomics has revealed that members of basal lineages share a set of highly conserved transcription factors with the derived flowering plants.

While gene copy numbers have expanded through time, Pires *et* al. (2013) predicted that the gene regulatory networks implicated in development of derived plants may have their origin in the ancestors of early groups of plants and that their diversification, co-option and reassembly is directly related to the innovations in land plant body (Bennett et al., 2014; Pires and Dolan, 2012).

The purpose of this study is to characterize the ancestral function of an *Arabidopsis thaliana* AINTEGUMENTA-LIKE (AIL) gene ortholog coded in the *Marchantia polymorpha* genome. Since AIL genes in *A. thaliana* have been shown to play a critical role in stem cell niche maintenance and the development of root and shoot organs, the main biological question of the study is, what function would an *AIL* gene have on *Marchantia polymorpha*, a basal organism with one of the simplest body architectures of land plants?

3.1   AINTEGUMENTA-LIKE gene family of transcription factors

The AINTEGUMENTA-LIKE (AIL) transcription factor family in *Arabidopsis thaliana* has been well studied for over 20 years and includes eight genes considered master regulators of developmental processes: *AINTEGUMENTA (ANT)*, *AINTEGUMENTA-LIKE 1 (AIL1)*, *BABY BOOM (BBM)* and the *PLETHORA (PLT1, PLT2, PLT3, PLT5, PLT7)* genes.

3.1.1   AIL protein structure

AIL genes belong to the APETALA 2/ETHYLENE RESPONSE FACTOR (AP2/ERF) family of transcription factors characterized for containing the AP2/ERF DNA binding domain, discovered first in the *A. thaliana* AP2 protein, which is critical for the establishment of the floral meristem (Kim *et* al., 2006).

AP2/ERF genes are classified depending on the number of AP2 domains they have. The AIL proteins contain two AP2 domains separated by a linker region (25 amino acids) so they fit under the AP2-like subfamily of proteins (Riechmann and Meyerowitz, 1998). The AP2-like subfamily is further divided into two different lineages: the euAP2 lineage, which mRNA has a miR172 binding motif, and the post-domain region and the AINTEGUMENTA (ANT) lineage, characterized by a 10- amino acid insertion in the first AP2 domain (R1) and a 1- amino acid insertion in the second AP2 domain (R2). The ANT lineage divides into two sub lineages, basalANT and euANT (Horstman *et* al., 2014). AIL proteins belong to the latter. What differentiates them is that euANT proteins are defined by a long pre-domain region and four conserved motifs. The first motif, euANT1, consists

of conserved amino acids (NSC[K/R][K/R]EGQ[T/S]R) in the 10-amino acid insertion located in the first AP2 domain. The other three motifs, euANT2 (WLGFSLS), euANT3 (PKLEDFLG) and euANT4 (TFGQR), are located in the pre-domain region of the euANT proteins (Fig. 1) (Kim *et* al., 2006).
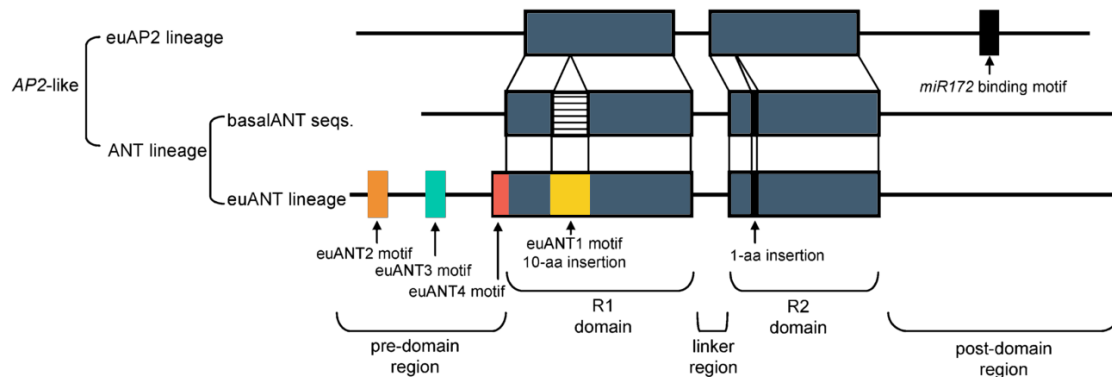
Fig. 1 Structure of *AP2-like* genes. Blue boxes indicate AP2 domains. Orange, blue, red and yellow boxes indicate euANT-specific motifs. Hatched region in basalANT sequences shows portion of the 10-amino acid insertion that is not conserved. Black boxes in the ANT lineage indicate a 1-amino acid insertion. Back box in euAP2 lineage shows site of the post domain region where a miR172 binding motif is located (Kim *et* al., 2006).

3.1.2 Function of AIL proteins among developmental stages of *Arabidopsis thaliana*

AIL proteins are mainly expressed in dividing tissues where they regulate the maintenance of stem cell niches, the correct development of the embryo, the root and shoot organs. Since they have overlapping roles, the best way to summarize their function is to describe its role in the different plant organs and their development.

3.1.2.1 Embryo development

PLT1, PLT2, PLT5 and *BBM* have critical roles during embryogenesis. Studies of *plt1;plt2* double mutants have revealed that *PLT1* and *PLT2* expression in the basal part of the octant stage embryo and in the quiescent center later in embryogenesis is necessary for the correct specification and development of the root meristem. *plt1;plt2* heart stage embryos show enlarged and mis-specified QC progenitor cells and grow into seedlings with defective root development (Aida *et* al., 2004). An interesting effect observed in *plt2;bbm* double mutant indicates their importance in plant development since they

cannot develop after the early embryo stage. On the other hand, overexpression of *BBM* or *PLT5* induces ectopic embryos on the aerial part of the seedling while *PLT1* and *PLT2* induce ectopic root stem cell niches when overexpressed during embryogenesis (Horstman *et* al., 2014).

3.1.2.2 Root development

PLT1, PLT2, PLT3 and BBM are expressed in *A. thaliana* roots. Aida *et* al. (2004) demonstrated that *PLT1* and *PLT2* act in parallel with *SHORT-ROOT (SHR)* and *SCARECROW (SCR),* transcription factors that belong to the GRAS family, to specify QC identity and the maintenance of the root stem cell niche. *A. thaliana* PLT proteins expressed in the root display a gradient distribution with maxima in the stem cell niche, this is also true for *Oryza sativa* PLT genes (Li and Xue, 2011), where they act as a morphogen: high levels of PLT promote stem cell identity, lower levels allow stem cell daughters to have mitotic activity and an even lower concentration of PLTs is necessary for cell differentiation (Galinha *et* al., 2007). Seedlings of the homozygous triple mutant *plt1;plt2;plt3* developed by Galinha *et* al. (2007) show a fully differentiated embryonic root pole at 3 days post germination (d.p.g.) and they never develop roots, such effects are evidence that PLTS/BBM are master regulators of root development.

Hofhuis et al. (2013) showed that *PLT3, PLT5* and *PLT7* are expressed in lateral root primordia, regulating lateral root development and preventing that lateral root primordia are formed close to one another (rhizotaxis) and revealed that the genetic mechanism that regulates rhizotaxis is the same that controls phylotaxis in the shoot is regulated by the same group of transcription factors (Prasad *et* al., 2011).

3.1.2.3 Shoot and floral meristems

PLT3, PLT5, PLT7 and *ANT* are expressed within the shoot and the floral meristems. *ANT, PLT3* and *PLT7* are required for shoot apical meristem maintenance, since *ant;plt3;plt7* triple mutants have shoots that stop growing after producing a few leaves as a consequence of differentiation of the meristematic cells (Horstman *et* al., 2014). A study carried out by Prasad *et* al. (2011) revealed that he PLT3, PLT5 and PLT7 proteins control shoot organ positioning in the shoot apical meristem (SAM). *plt3;plt5;plt7* mutants show changes in phyllotaxis compared with wild-type seedlings, where leaves arise in a spiral pattern maintaining an angle of 137.5°. This pattern is supposed to be preserved through floral transition but *plt3;plt5;plt7* triple mutants develop inflorescences that

depart from the spiral pattern and show flower positioned in an opposite manner, diverging by 180° or 90 °.

3.1.2.4   Flower development

AIL genes activity is not only involved in organ positioning pattern. Research done with *ANT* demonstrates its importance during *A. thaliana* flower development since *ant* mutants suffer a reduction in floral organ number and the ovule does not develop integuments that would later develop into the seed coat (Reviewed in Krizek, B., 1999). Krizek, B. (1999) ectopically expressed *ANT* under the cauliflower mosaic virus 35S constitutive promoter (35S::ANT) in wild-type *A. thaliana* plants and showed that they develop larger floral organs resulting from an increase in cell number in sepals while larger petals, stamens and carpels are a result of an increase in cell size.

ANT and PLT3 also act in parallel with AUXIN RESPONSE FACTOR5/MONOPTEROS (ARF5/MP) to induce the transcriptional expression of LEAFY (*LFY*, a master regulator of the pre-floral to floral inflorescence phase) expression in incipient primordia and regulate the proper timing of flower formation (Yamaguchi *et* al., 2016).

3.2   AIL – Auxin interactions and feedback loop

The studies on AIL transcription factors have unraveled their tight relationship with the phytohormone Auxin (indole-3-acetic acid, IAA), which controls key aspects of plant development.

It has been shown that in angiosperms, a significant amount of IAA is produced from Tryptophan (Trp) via indole-3-pyruvic acid (IPyA). The enzymes that catalyze the conversion of Trp to IPyA belong to the TRYPTOPHAN AMINOTRANSFERASE OF ARABIDOPSIS (TAA) and enzymes that catalyze the IPyA to IAA conversion, to the YUC families (Reviewed in Eklund *et* al., 2015).

Plant cells interpret IAA signals through a pathway where an auxin receptor called TIR1, a subunit of the $SCF^{TIR1/AFB}$ ubiquitin protein ligase E3 complex, binds to IAA. Such binding promotes the ubiquitination and degradation of the Aux/IAAs repressors of transcription, which are attached to Auxin Response Factors (ARFs) located upstream of IAA target genes (binding to Auxin Response Elements, AuxREs) when IAA concentration is low. Therefore, high levels of auxin promote ubiquitination of Aux/IAAs

through SCF$^{TIR1/AFB}$, which guides them to degradation on the proteasome (Fig. 2) This way, ARFs can activate transcription of target genes (Reviewed in Salehin *et al.*, 2015).
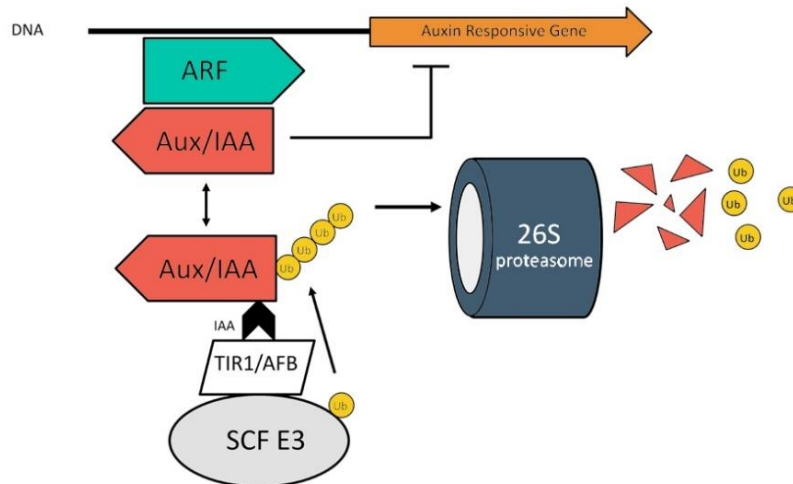
Fig. 2 Auxin signaling and action mechanism. Black arrow, Auxin (IAA). ARF, Auxin Response Factor. SCF E3, SCF$^{TIR1/AFB}$ ubiquitin protein ligase. Yellow circles show ubiquitinations. (Extracted from da Costa *et al.*, 2013)

In the root meristem*, PLT1* and *PLT2* are auxin response genes dependent on the action of ARF5/MONOPTEROS (MP) and ARF7 (Aida *et al.*, 2004; Horstman *et al.*, 2014) and they are in turn required for the expression of *PIN-FORMED 4* (*PIN4*) which codes for an auxin-efflux carrier protein. PIN proteins help to establish the root auxin gradient, where the maximum concentration locates at the QC. This suggests a feedback loop, where PIN proteins direct auxin accumulation towards the root meristem, which restricts PLT1 and PLT2 expression to maintain the location and size of the stem cell niche (Horstman *et al.*, 2014) (Fig 3).

Research on the auxin-AIL pathway has recognized that ARFs act as upstream mediators of AIL gene, such pathway s is also involved in embryogenesis, flower initiation and lateral root formation processes, which are discussed on 3.1.2 (Reviewed by Horstman *et al.*, 2014).
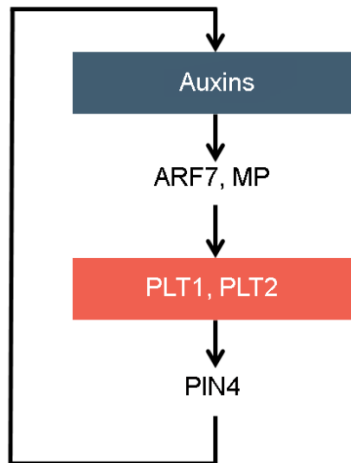
Fig. 3 Root AIL−Auxin feedback loop. ARF7, AUXIN RESPONSE FACTOR 7. MP, MONOPTEROS. PLT1, PLETHORA1. PLT2, PLETHORA2. PIN4, PIN-FORMED 4.

To add another layer to the auxin -AIL pathway, Pinon *et* al. (2013) detected that PLT5 induction stimulated *YUC* expression after only 2 hours, indicating that auxin biosynthesis is activated by PLT expression.

3.3   Function of AIL proteins in earlier land plants

Physcomitrella patens is a leafy moss that became a basal plant model to study the evolution of plant development. There are four AIL genes encoded in the *P. patens* genome, APB1-4 (Aoyama *et* al., 2012). Previous studies demonstrated that they are vital for the transition from a protonemal 2-D filamentous body to a gametophore 3-D leafy form (Reviewed by Harrison, 2017). Aoyama *et* al. (2012) defined that *APB* genes act as a molecular switch for the development of different types of stem cells and that their expression was induced by auxin, similarly to what has been shown for *A. thaliana* AIL genes.

In the C-fern *Ceratopteris richardii,* Bui *et* al. (2017) identified an ortholog of the *A. thaliana* *ANT* gene, *CrANT,* which has an expression pattern similar to *A. thaliana BBM* (in embryo development). However, their phylogenetic analysis positioned *CrANT* in a clade sister to *AthBBM*. The ectopic expression of *CrANT* in *C. richardii* gametophytes originated the spontaneous production of apogamous sporophytes that develop sporophyte only vascular tissue, tracheids and stomata. This suggests that *CrANT* has an important role in sexual reproduction and that *AthBBM* might have evolved from an ancestral AIL gene such as *CrANT*.

According to the above-mentioned studies, it can be proposed that the AthAIL gene interaction with the phytohormone Auxin their expression in stem cells and their action on cell-fate determination is conserved since this early divergent plant lineages.

In order to go even further in the determination of the ancestral gene function of AIL genes, involved in the development of organs that allowed land plants to diversify and colonize terrestrial environments, it is necessary to evaluate their function in plant lineages that show a simpler body form and that diverged earlier than mosses.

3.4   *Marchantia polymorpha*

Among extant land plants, the divisions Anthocerotophyta (hornworts), Bryophyta (mosses) and Marchantiophyta (liverworts) are considered descendants of the first events of embryophyte diversification (Fig. 4). The three groups are characterized by having a dominant gametophyte haploid phase, a multicellular sporophyte dependent on the gametophyte body, lack of vascularization and motile sperm for fertilization. Such features make them perfect for studying the evolution of genetic systems that underlie major changes in land plant morphology and physiology (Ishizaki, 2016).
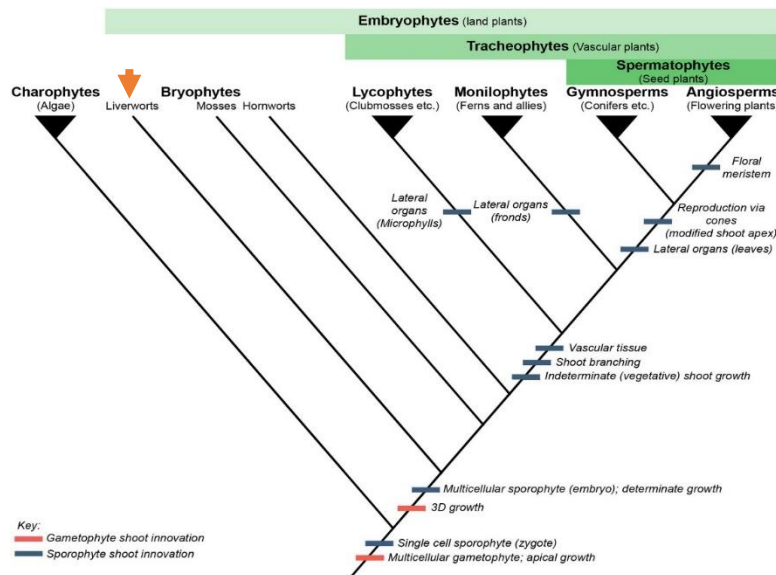

Fig. 4 Phylogenetic tree of extant land plants based on Gametophyte and Sporophyte innovations across land plant evolution. Innovations in gametophyte development are marked by a red rectangle. Blue rectangles mark innovations in sporophyte development. Orange arrow marks *Marchantia polymorpha* position in the plant phylogenetic tree (Extracted from Plackett et al., 2015)

Based on morphology and fossil evidence, it is generally accepted that liverworts diverged earlier than the rest of the embryophytes. However, it is still debated which division originated first and which is the sister group to vascular plants based on molecular phylogenetic analyses (Wickett et al., 2014; Plackett *et* al., 2015).

Marchantia polymorpha is a dioecious liverwort species belonging to the Marchantiopsida group with a long history as an experimental organism. *M. polymorpha* has a complex lobed thalloid body with air chambers, ventral scales, rhizoids, gemma cups and stalked archegoniophore and antheridiophore (Shimamura, 2016) (Fig. 5).

The haploid phase of *M. polymorpha* starts from a single-celled spore generated by meiosis of a mother cell in the sporophyte. Once the spore germinates, it undergoes mitosis and develops into a sporeling, which later becomes a thallus. The thallus develops gemma-cups periodically, which develop clonal propagules called gemmae from single epidermal cells by mitosis (Fig. 5) (Shimamura, 2016). This confers an advantage for molecular research since isogenic lines can be maintained through gemmae subculture. Another trait that makes *M. polymorpha* a good model plant for genetic studies is that the haploid phase of its life cycle is dominant, which makes homozygous mutant lines easy to obtain.
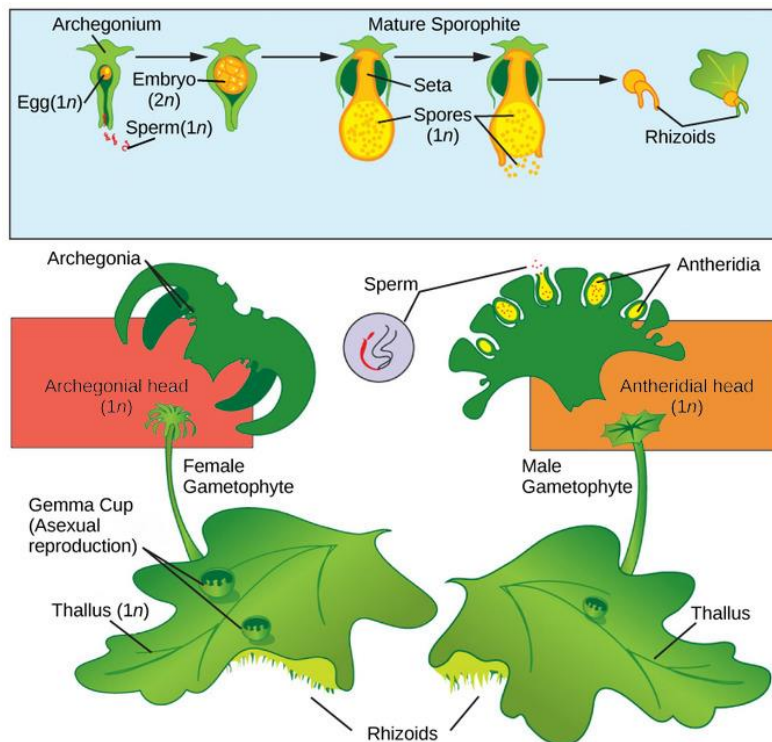
Fig. 5 Structure and Life Cycle of *Marchantia polymorpha*. Modified from Boundless, 2016.

M. polymorpha has a growing point with meristematic cells at the apices of the thallus in a zone called the apical notch. The thalloid form of *Marchantia* has indeterminate growth and develops from the activity of a lens-shaped apical cell surrounded by merophytes, or meristematic cells, that result from periclinal divisions if the apical cell (Fig. 6). Dorsal merophytes give rise to dorsal epidermis, gemmae and gemma cups. Ventral merophytes develop ventral scales and lateral merophytes give rise to all tissue lineages that compose the thallus (Shimamura, 2016).
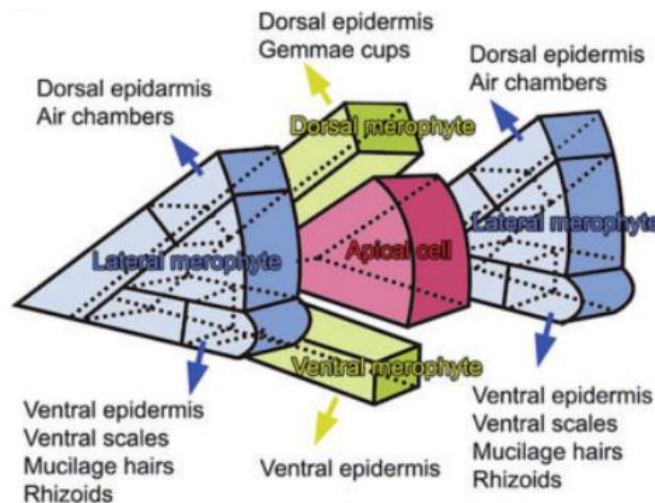
Fig. 6 Structure of the apical meristem in *Marchantia polymorpha*. Extracted from Shimamura, 2016.

Recently, wild-type male and female accessions, Takaragaike-1 (Tak-1) and Takaragaike-2 (Tak-2), respectively, have been used in molecular genetic studies. The whole-genome project at DOE-JGI has been publically available since 2016 and its analysis has revealed that *M. polymorpha* has low genetic redundancy. This trait positions *Marchantia* as a rising model for evo-devo and forward genetic studies. Molecular techniques for *Agrobacterium*-mediated transformation and vectors for heterologous gene expression in *M. polymorpha have* already been developed (Ishizaki *et* al., 2008; Kubota *et* al., 2013; Tsuboyama-Tanaka *et* al., 2015).

Auxin research carried out in *Marchantia polymorpha* suggests that the activity of auxin is crucial for the correct development of the gametophyte and sporophyte. On the other hand, studies with the GH3 promoter, which contains AuxREs, reveal that auxin-mediated transcriptional activation is conserved since the divergence of liverworts (Ishizaki *et* al., 2012). Eklund *et* al. (2015) demonstrated that auxin imposes dormancy

on gemmae within the cup and that *M. polymorpha* harbors the smallest IPyA-dependent auxin synthesis toolkit identified in land plants, with one *TAA* and two *YUC* genes. The sites of auxin biosynthesis in the thallus were also identified, being the apical notch and the Mid-rib the places with a highest concentration of auxin, followed by de gemmae cups and the thallus margins.

4. Hypothesis

The AIL-PLT Like gene (MpAPL) is involved in the maintenance of the stem cell niche of *Marchantia polymorpha.*

5. Objectives

5.1 General objective

Characterize the single AINTEGUMENTA-LIKE gene in *Marchantia polymorpha*.

5.2 Specific objectives

a. Identify and analyze *in silico* the single AIL gene ortholog in *Marchantia polymorpha, MpAPL.*
b. Determine the functional conservation by genetic complementation of *A. thaliana plt1;plt2* double mutant with MpAPL.
c. Generate overexpression and transcriptional fusions for *Marchantia polymorpha* genetic transformation.
d. Establish transformant M. polymorpha lines for functional characterization.

6. Materials and Methods

6.1 Plant Material

Marchantia polymorpha Takaragaike-1 (Tak-1; Japanese male accession) thalli derived from gemmae were cultured on half-strength Gamborg's B5 (Phyto Technology Laboratories) containing 0.5 g/L MES, 1% sucrose and 1.3% phytoagar (Phyto Technology Laboratories) (pH 5.7), unless stated otherwise. Plants were grown under a 16 h light/ 8 h dark photoperiod at 22° C.

Arabidopsis thaliana (Col-0) wild type plants and a *plt1;plt2* double mutant line (*plt1-4 plt2-2* in a ws background, Aida *et* al., 2004) were used.

Seeds were sterilized (70% ethanol for 5 minutes, 20% chlorine for 5 minutes followed by five washes with sterile Milli-Q water under a laminar flow hood) and kept in Eppendorf tubes with sterile Milli-Q water for 3 days at 4° C. After cold treatment, seeds were sown on Murashige & Skoog media (Phyto Technology Laboratories) containing 1.3% phytoagar (Phyto Technology Laboratories), MES (pH 5.8) in a growth chamber at 22° C under photoperiod (16 h light, 8 h dark). Seedlings were later transferred to soil and finished their life cycle in a growth room at 23° C with photoperiod (16 h light / 8 h dark). Plants were fertilized once a week with MiracleGro following the product's instructions for indoor plants.

6.2 *In silico* analysis

6.2.1 Identification of an AIL gene ortholog in *Marchantia polymorpha*

To find orthologs for the AIL gene family in *Marchantia polymorpha,* Arabidopsis AIL coding sequences were BLASTed one by one against the published genome of *Marchantia* polymorpha (V3.1) in Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). Just one protein sequence which contains two AP2 domains was retrieved and named AINTEGUMENTA-PLETHORA LIKE (refered as *MpAPL* in this study) based on the phylogenetic analysis results. I characterized its AA composition in the pre-domain and domain regions to make sure it has the characteristic structure of an AIL protein: two AP2 domains, three motifs in the pre-domain region (euANT2, euANT3, euANT4), a 10-amino acid insertion (euANT1 motif) on the first AP2 domain, and a conserved linker region between the AP2 domains (Kim, 2006).

The JBrowse (Skinner *et* al., 2009) genome browser in Phytozome was used to determine the scaffold of the genome in which APL is located, the length and structure of the genomic region (5' UTR, coding region, 3' UTR), the exon-intron arrangement and the putative promoter region.

6.2.2 *Cis*-Region elements analysis

To identify Auxin Response Elements (AuxREs) in the 5' region of both Mp*APL* and Arabidopsis AIL genes, the fragment spanning promoter and 5' UTR region (usually contains regulatory elements; Barret et al., 2012) of the eight *AIL* Arabidopsis genes and the 4.9 kb intergenic region upstream of *MpAPL* start codon were analyzed with MatInspector (Genomatix Software Suite v 3.8). Prior to the analysis, the AREF and ARF3 matrix families were selected from the MatInspector library. A pairwise alignment between all sequences was carried out using DiAlign (Genomatix software suite v3.8) to evaluate similarity and percentage of identical nucleic acids of the *cis*-region. The results of the DiAlign analysis were used to build a pairwise similarity matrix and a heatmap with the pheatmap R package (Kolde, 2015). A clustering multivariate analysis was carried out based on Euclidean UPGMA distances using Past software V 3.15 (Hammer *et* al., 2001) to reflect the structure of the pairwise similarity matrix.

6.2.3 Phylogenetic analyses

For the phylogenetic analysis of the eight *A. thaliana* AIL gene family members and the *M. polymorpha* ortholog MpAPL, a FASTA format database was built with the amino acid (AA) sequences retrieved from TAIR (https://www.arabidopsis.org) and Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html).

For the phylogeny reconstruction of AIL proteins among the Streptophyta infrakingdom (ITIS, http://www.itis.gov), which includes all organisms under the Charophyta and Embryophyta divisions, a FASTA format dataset of AA sequences from different land plant species was constructed by querying the coding sequence of every *Arabidopsis thaliana* AIL gene family member against Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html), The Conifer Genome Integrative Explorer (ConGenIE.org) and The 1000 plants (oneKP, https://sites.google.com/a/ualberta.ca/onekp/) (organism, sequence name, length and database shown on Supplementary Table 1). The amino acid sequences that resulted from the search were analyzed in Pfam (Finn *et* al., 2016) to determine present AP2

domains. Then, sequences were manually analyzed to determine the characteristic motif structure of AIL proteins. To expand the phylogenetic analysis, a search against the *Klebsormidium nitens* genome (*Klebsormidium nitens* NIES-2285 genome project) was made and a single sequence was found that could be an ortholog. Also, two bryophyte WRINKLED sequences were added to the database. The final database contains 31 AA sequences, see Supp. Table 1 for their characteristics.

For both phylogenetic analyses, the sequences were aligned with MAFFT (Katoh and Standley, 2013) (http://mafft.cbrc.jp/alignment/server/) selecting the L-INS-i alignment strategy, which can align sequences around alignable domains. Alignments were visualized using BioEdit alignment editor v7.2.6.1 and checked for the correct alignment of the two AP2 domain regions. To find the best-fitting model of protein evolution, the resulting alignment matrix was analyzed with ProtTest (Abascal *et* al., 2005) selecting the default parameters using the MAZORKA Linux cluster at the LANGEBIO bioinformatics facility. Based on the AIC and BIC selection criteria, the JTT+G+F substitution model was selected for both datasets. I then ran a Maximum Likelihood-based phylogenetic analysis with RAxML V 8.0.0 (Stamatakis, 2014) via the CIPRES Science Gateway V. 3.3 (Miller *et* al., 2010) for both datasets, previously converted to phylip format (.phy), indicating the JTTF model of AA substitution and with 1000 bootstrap replicates to compute support values of the final tree branches. I also ran a Maximum Parsimony phylogenetic analysis for the Streptophyta dataset with the bootstrap method using 500 replicates in MEGA version 7 (Kumar *et* al., 2015). Resulting trees were edited for presentation using FigTree v 1.4.2 (Rambaut, 2012).

6.3 Genetic constructs for transgenesis

MpAPL coding sequence without stop codon was amplified from a pCRII plasmid in which it was cloned (generated by Eduardo Flores-Sandoval at John Bowman's lab) with primers that contain attB sequences for Gateway Cloning System.

The *MpAPL* promoter fragment was cloned from Tak-1 DNA using primers designed with attB tails for Box 1 and Box 2 of the Gateway Cloning System (Supplementary table 2), the forward primer aligns with a sequence 2.5 kb upstream of APL start codon while the reverse primer hybridizes just before the APL start codon.

Both sequences were PCR-amplified with Accuprime High Fidelity polymerase and the PCR products were checked by electrophoresis on agarose gel (Supplementary Figure

1) and then purified with QIAquick PCR Purification Kit (Qiagen). PCR products were quantified on NanoDrop.

To set up the Gateway® Recombination System for the construction of expression plasmids, BP reactions between the PCR products and donor plasmids (pDONR™221 and pDONR™ P4-P1r) were made to form entry vectors. Five microliters previously dialyzed of the BP reaction were electroporated into *Escherichia coli* competent cells (DH5α). After culturing *E. coli* in liquid LB medium with Kanamycin (50 mg/L), the plasmids were purified using the QIAprep Spin Miniprep Kit (Qiagen). Entry vectors were verified by PCR with M13 and specific primers and then by Sanger sequencing.

$_{pro}$MpAPL:GUS expression vector was constructed fusing a 2.5 kb APL promoter fragment in front of GUS in a Gateway® LR equimolar reaction between pDONR™221 harboring $_{pro}$APL and the pMpGWB204 binary destination vector (Ishizaki *et* al., 2015) (Figs.13A and 13B).

The $_{pro}$35S:MpAPL-Citrine expression vector for constitutive expression was constructed fusing the coding sequence of *MpAPL* without a stop codon under the CamV 35S promoter and Citrine as a reporter fluorescent protein at the N-terminal. To do this, an equimolar LR Gateway® reaction was made between pDONR™221 harboring the MpAPL coding sequence and the pMpGWB206 binary destination vector (Ishizaki *et* al., 2015), making the OE APL vector (Figs. 13C and 13D).

Each reaction was transformed into electro-competent *E. coli* cells (DH5α), which were then cultured in liquid LB medium with 100 mg/L spectinomycin. Expression vectors were purified with GeneJET Plasmid Miniprep Kit (Thermo Scientific™) and verified by PCR with M13 universal primers (Figs. 13B and 13D) and then by Sanger sequencing carried out in Laboratorio de Servicios Genómicos (Labsergen) at LANGEBIO.

To create the MpAPL knock-down *M. polymorpha* lines, a pKART plasmid carrying an artificial micro RNA (amiRMpPLT1^{MpMIR160}, made by Eduardo Flores-Sandoval) was also used to transform *E. coli* cells (DH5 α). amiRMpPLT1^{MpMIR160} is driven by a constitutive promoter from an endogenous gene called elongation factor 1-α ($_{pro}$EF1) and targets a 24 bp fragment of the *MpAPL* transcript (Fig. 13E) located 75 bp downstream of ATG. Cell cultures and plasmid purification was carried out as mentioned above.

6.4 Agrobacterium transformation

Agrobacterium tumefaciens strain GV2260 was used in both *M. polymorpha* and *A. thaliana* transformation experiments. Each of the expression vectors was electroporated into *A. tumefaciens* competent cells and then cultured at 28 °C for 48 h in LB plates containing 100 mg/L spectynomicin, 100 mg/L ampicillin and 100 mg/L rifampicin. Plates with *A. tumefaciens* carrying the expression vectors were stored at 4 °C until cells were needed to start liquid cultures for plant transformation.

6.5 Agrobacterium-mediated transformation of *Marchantia polymorpha*

Two methods of *Agrobacterium* mediated genetic transformation of *M. polymorpha* were tested with $_{pro}$APL:GUS, $_{pro}$35S:APL-Citrine and amiRMpPLT1^{MpMIR160}.

First, the T-AgarTrap method developed by Tsuboyama-Tanaka *et* al. in 2015 was used. In this method transformation and selection solutions were poured over mature (one month old) thalli pieces of *M. polymorpha*. In addition to the Tak-1 thalli pieces that were transformed with $_{pro}$APL:GUS, $_{pro}$35S:APL-Citrine and amiRMpPLT1^{MpMIR160}, some pieces were not soaked in transformation solution and were used as negative controls for Gentamicin and Kanamycin.

The other is a modification of the protocol developed by Kubota *et* al. in 2013. Instead of making the 14-day-old *M. polymorpha* pieces co-culture with *A. tumefaciens* in liquid M51C medium, it was made on plastic plates with the same medium formula containing 1.3% phytoagar (Phyto Technology Laboratories) and under continuous white light for three days. Pre-culture, washing and selection (100 mg/L Gentamicin or 150 mg/L Kanamycin, 100 mg/L Cefotaxime) steps were carried out as in the original protocol.

6.6 *Arabidopsis thaliana* genetic transformation

A. thaliana Col-0 and the *plt1;plt2* homozygous line were genetically transformed with *A. tumefaciens* carrying $_{pro}$35S:APL-Citrine.

The transformation method used was published by Martinez-Trujillo *et al*. in 2004, an improvement of the floral dip method. Here, the *Agrobacterium* solution is applied as a drop and directly onto the flower buds and just-opened flowers. All steps were performed as in the protocol of the mentioned article.

Plants were grown and seeds were collected when the siliques matured. Then, seeds were surface sterilized as mentioned and after cold treatment, seeds were sown and grown on Murashige & Skoog media (Phyto Technology Laboratories) containing 1.3% phytoagar (Phyto Technology Laboratories), MES (pH 5.8) and Gentamicin 100 mg/L for transformant selection. Transformant lines are being maintained to take them to a homozygous state for the transgene.

6.7   Microscopy

Plants were observed under a stereoscopic microscope with a Leica camera.

To analyze Citrine expression in the $_{pro}$35S:APL-Citrine transgenic lines, *A. thaliana* seedlings were mounted on slide cover glasses with water or a Propidium Iodide (PI) solution to stain DNA and visualize the nucleus of cells. An inverted Zeiss LSM710 confocal laser-scanning microscope (Zeiss, Oberkochen, Germany) was used. The plates with $_{pro}$35S:APL-Citrine *M. polymorpha* transgenic plants on Gamborg ½ medium were placed upside-down on a LSM-800 Zeiss confocal laser-scanning microscope (Zeiss, Oberkochen, Germany). Fluorescence of Citrine was excited at 514 nm, emitted light was collected between 520 nm and 560 nm. PI was excited at 514 nm.

7. Results

7.1 *In silico* analyses

7.1.1 *Marchantia polymorpha* has a single euANT gene family ortholog

The search for orthologs of the AIL gene family, in the *M. polymorpha* genome (V3.1) resulted in the same hit for each sequence with which the query was done. The hit sequence is located on the scaffold_8:895126..905363 and the primary transcript is named Mapoly0008s0071.1. The primary transcript is annotated as an AP2-LIKE ethylene-responsive transcription factor. The genomic region of Ma Mapoly0008s0071.1 is 10 239 base pairs (bps) long and is composed by a 5' UTR region of 2 667 bps, 9 exons, 8 introns and a 3' UTR 886 bps long (Fig. 7A).

The protein sequence, which is 824 amino acids (aa) long, was analyzed in Pfam (Finn et al., 2016). It was determined that the protein contains two AP2 motifs, a linker region between the motifs and large pre- and post-domain regions (Fig. 7B)

A

10 238 bp's

B

824 aa's
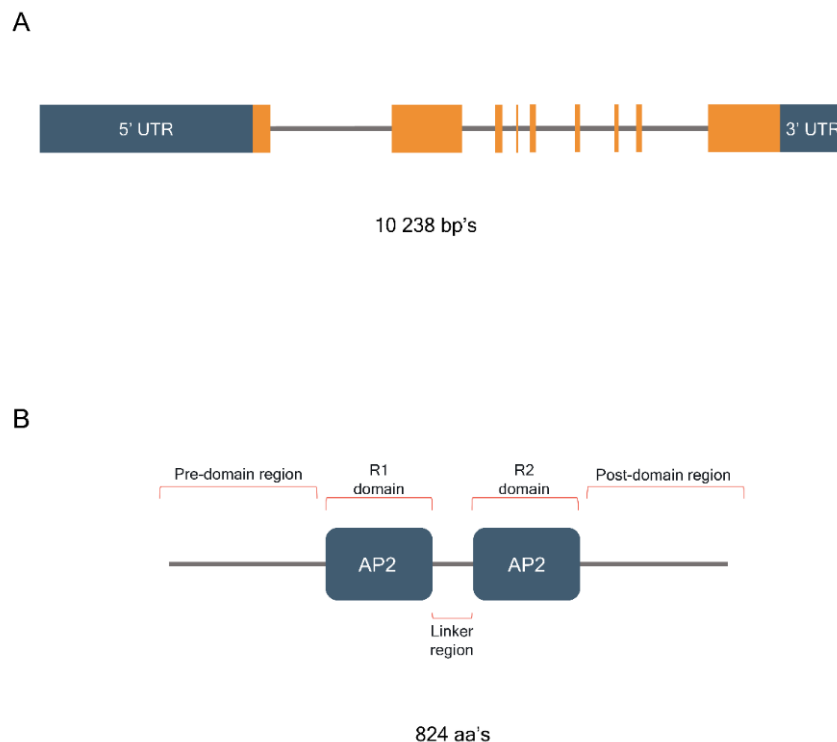
Fig 7 (A) The genomic region of *M. polymorpha AINTEGUMENTA-PLETHORA-LIKE* (Mp*APL*) spans 10 238 base pairs. 5' UTR and 3' UTR regions in blue. Exons are represented as an orange box

while introns are shown as a gray line. (B) The MpAPL protein has the typical structure of an AP2-like gene. Blue boxes show AP2 domains.

To confirm that Mapoly0008s0071.1 is a homolog of the *A. thaliana* genes in the AIL family, the four characteristic motifs of the family were located by hand in the sequence. In Figure 8, euANT1, euANT2, euANT3 and euANT4 motifs present in the AIL proteins from *A. thaliana* and MpAPL are shown as colored boxes as well as the logo or consensus sequence of each motif generated using WebLogo 3 (http://weblogo.threeplusone.com/).

Even though MpAPL is a much larger protein, the AP2 domain sequence and length is conserved, this is also true for linker region and the motifs in the pre-domain region of the protein (Supp. Fig. 2).
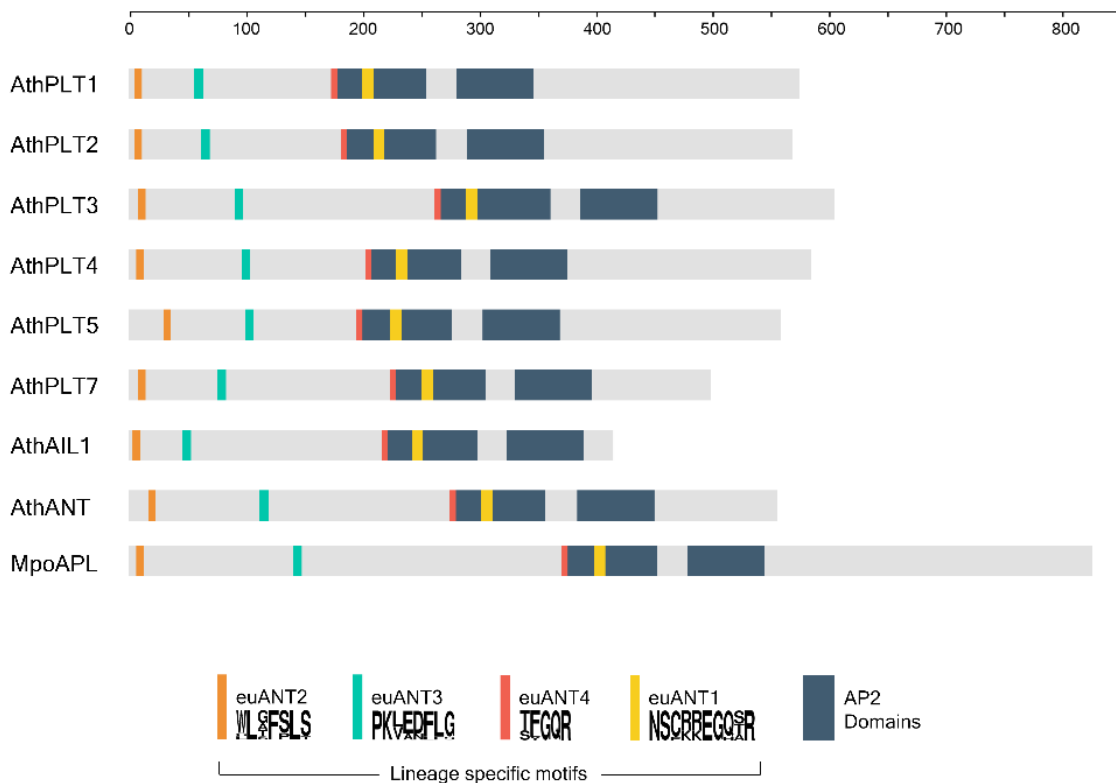

Fig 8 Structure of AIL genes from *Arabidopsis thaliana* and the single AIL type gene from *Marchantia polymorpha.* AP2 domains are indicated with blue boxes. Orange, cyan, red and yellow boxes mark euANT exclusive motifs for which the consensus sequence is shown as a logo. euANT1 motif (yellow) consists of a 10 aa insertion in the first AP2 domain. euANT2 (orange),

euANT3 (cyan) and euANT4 (red) motifs are sequences 7, 8 and 5 amino acids long located in the pre-domain region of euANT proteins. Ruler shows length in number of amino acids.

In terms of the exon-intron arrangement, MpAPL possesses the same number of exons as AthAIL1, AthPLT1, AthPLT4 and AthPLT5. However, APL intronic sequences are larger (Fig. 9)

7.1.2 Promoter analysis for AuxREs Cis-elements in AIL genes.

To detect the number of auxin response elements (AuxREs) to which auxin response factors (ARFs) will potentially bind to both *A. thaliana* AIL genes and *M. polymorpha* APL, a comparative cis-regulatory region analysis was made. The analysis revealed the position, orientation and sequence of all the AuxREs identified in the sequences (Supp. Table 3).
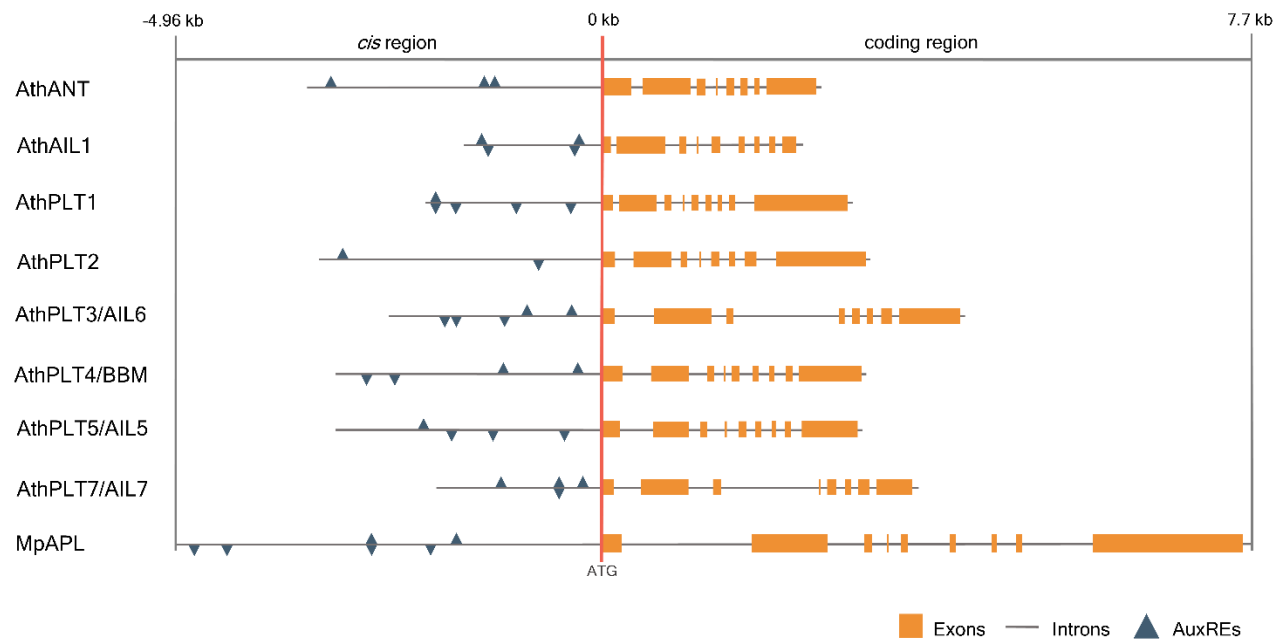
Fig 9 (A) AuxREs Cis-elements in the promoter regions of MpAPL and *Arabidopsis thaliana* AIL genes. Number and sequence of motifs are shown in Supplementary Table 3. (B) Exon-intron arrangement after ATG in MpAPL and *A. thaliana* AIL genes.

MpAPL has six AuxREs with a core similarity of 1. Four of them are distributed in the region between -2700 to -1706 bp upstream of ATG site. AthPLT2 has two AuxREs, preceded by AthANT, with three AuxREs in its cis-regulatory region (Fig. 9).

The DiAlign analysis results in a pairwise similarity matrix that includes two kinds of data: similarity, relative to maximum similarity found (value goes from 0 to 1) and the number of identic nucleic acids in percentage of shorter sequence (values from 0 to 100). A similarity of 1.000 does not mean that the compared sequences are identical, but that the pair of sequences are the most similar. In this case, the *cis*-region of *AthPLT1* and *AthPLT2* were the most similar (1.000 of similarity, 26% of identic nucleic acids), followed by the *cis*-region of *AthPLT3* and *AthPLT7* (0.446 of similarity, 21% of identic nucleic acids). This analysis was carried out to find which cis-region of the *A. thaliana* AIL genes is most similar to the *MpAPL* cis-region, this could hint a function in accordance with the elements that bind to the upstream region of the gene and promote its transcription. Just considering the results for the *MpAPL* cis-region, the alignment with *AthPLT1* cis-region had the highest similarity value (0.133) while the highest percentage of identic nucleic acids was obtained from the alignment with *AthAIL1* (8%) (Fig. 10B) (Supp. table 4).
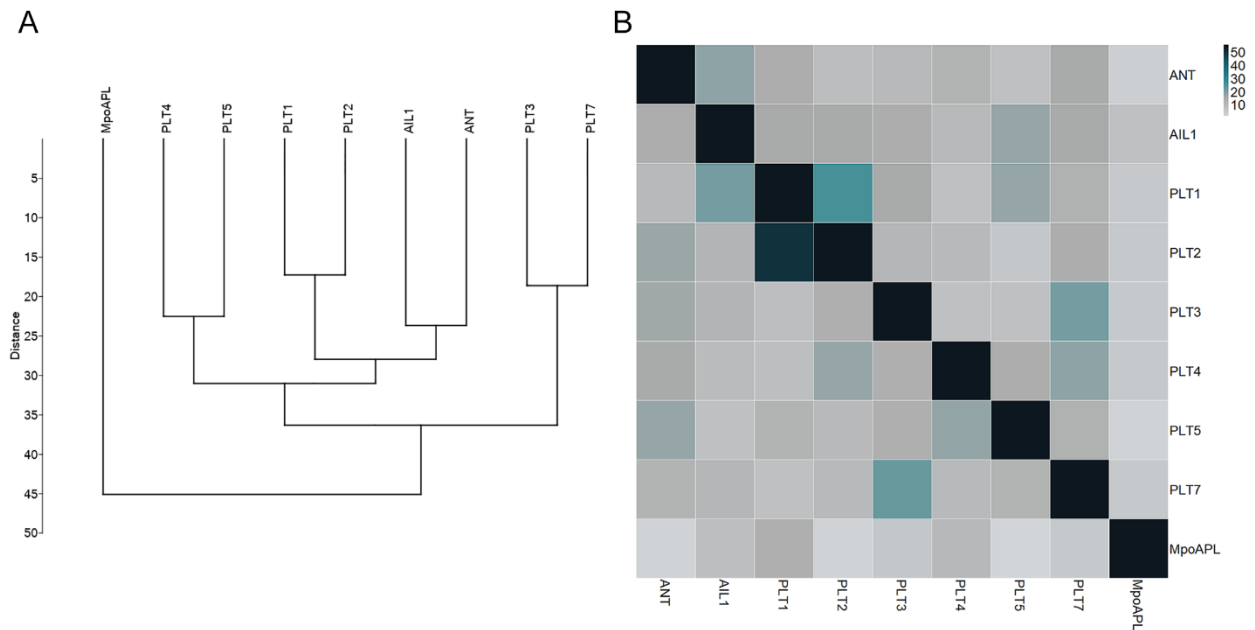

Fig 10 (A) Dendogram of the cis-region similarity of *Arabidopsis thaliana* AIL genes and MpAPL which reflects structure of the pairwise similarity matrix obtained with DiAlign (Supplementary table 4). Clustering analysis based on Euclidean UPGMA distances. Cis-regions that cluster together had a lower distance between them (B) Heatmap shows DiAlign analysis matrix data

for each pairwise alignment. Below the diagonal, the similarity (relative to maximum similarity) is given. The number of identical nucleic acids (in percentage of shorter sequence) is shown on the cells above the diagonal. Color gradient goes from dark blue, which indicates maximum similarity to light gray, that shows lower similarity between the pairs.

The Euclidean distances (UPGMA) dendrogram revealed the structure of the cis-region similarity matrix. All the cis-regions of AIL genes from *A. thaliana* formed a cluster separate from *MpAPL* in which the distance was shorter between them, this means they are more similar between them than with *MpAPL* (Fig. 10A).

7.1.3   Phylogenetic analyses of MpAPL

To explain the relationship between *Arabidopsis thaliana* AIL protein sequences and the single *Marchantia polymorpha* sequence MpAPL, a Maximum Likelihood phylogenetic analysis was carried out with the alignment matrix. The alignment matrix included AthSMZ and AthTOE3, added as an outgroup as they are AP2-like genes but belong to the euAP2 lineage (Kim *et* al., 2006). In the resulting phylogenetic tree (Fig. 11), all the *A. thaliana* AIL protein sequences and MpAPL form a distinct clade (supported with a bootstrap value of 100) separate from the one that includes the euAP2 sequences. MpAPL groups in a clade with all the AthPLT sequences (bootstrap value of 76), which is positioned as a sister group to the AthANT sequence.
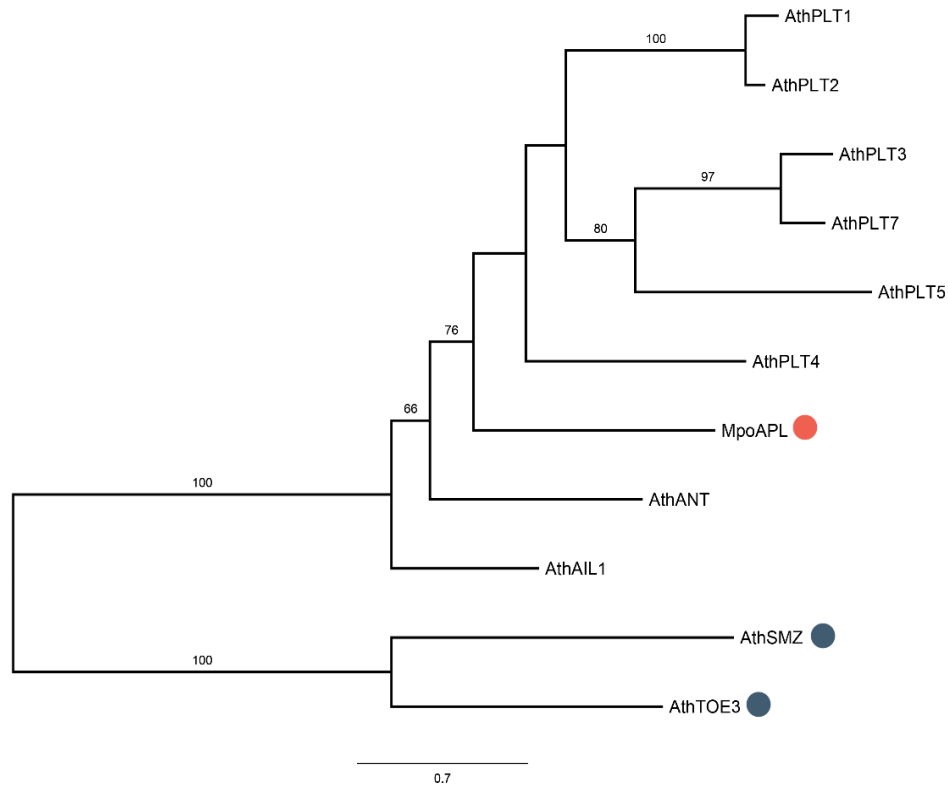
Fig 11 Maximum likelihood phylogenetic tree of the *Arabidopsis thaliana* AIL genes and the single *Marchantia polymorpha* AIL sequence. Bootstrap support values over 60 are shown at the nodes. Bootstrap support was calculated using 1000 replicates. An orange circle marks the position of MpAPL (MpoAPL) within the PLT clade. Dark blue circles indicate euAP2 sequences added as outgroup.

In the phylogenetic reconstruction of the AIL gene family among Streptophyta, two large clades are formed when using the *Klebsormidium flaccidum* (*Klebsormidium nitens*) aa sequence as outgroup; one, with the protein sequences from *Physcomitrella patens*, *Megaceros tosanus* and *Marchantia polymorpha* groups of early divergent lineages in the plant phylogeny and the second, which includes the sequences from taxa belonging to the Embryophyta division (Fig. 12). This grouping scored a bootstrap value of 100.

In the clade that has the bryophyte sequences, which scored a Bootstrap value of 92, can be observed that MpAPL is a sister sequence to the group with the aa sequences from *P. patens* y *M. tosanus.* It can also be seen that two later duplications possibly occurred in *P. patens* because of the way the four PLT-like copies grouped.

On the other hand, the clade that groups Embryophyte sequences is divided, although with low support (less that 60 of Bootstrap), in two sister groups that descend from a common ancestor sequence. One of the clades includes two PLT-like aa sequences

from *P. abies* as well as the AIL1/ANT from *A. thaliana*, *O. sativa* and *A. tricopoda*. The other clade, includes all the PLT sequences from the Embryophyta taxa (bootstrap value of 96) divided in two, a group with BBM sequences from *O. sativa*, *A. thaliana* and *A. trichopoda* and a group with the PLT1, PLT2, PLT3, PLT5 and PLT7 sequences from *A. thaliana* and their homologs in *O. sativa* and *A. trichopoda.*
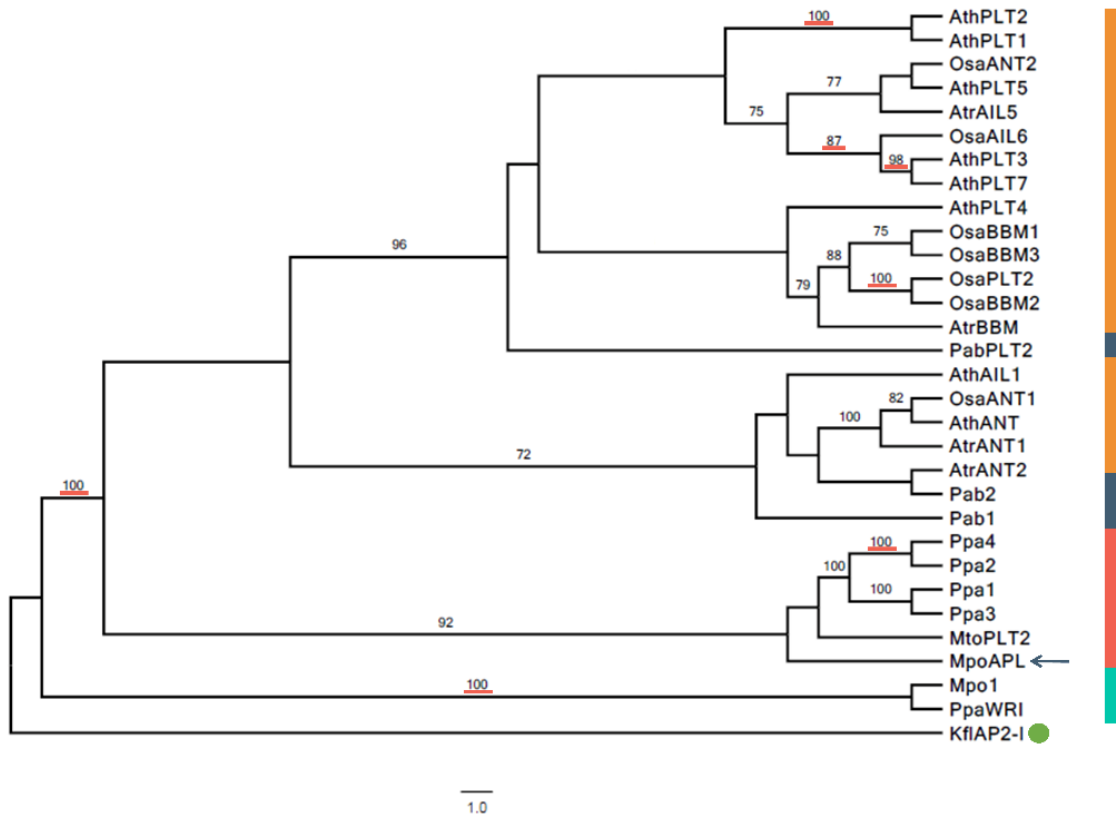
Fig 12 Phylogeny of the AIL gene family across Streptophyta. Maximum Likelihood tree constructed on RAxML with bootstrap support calculated using 1000 replicates. Only bootstrap values over 60 are shown on branches. Values underlined in red mean clades supported on the Maximum Parsimony analysis. Abbreviations: *Arabidopsis thaliana* (Ath), *Oriza sativa* (Osa), *Amborella trichopoda* (Atr), *Picea abies* (Pab), *Physcomitrella patens* (Ppa), *Megaceros tosanus* (Mto), *Marchantia polymorpha* (Mpo) and *Klebsormidium flaccidum* (Kfl). Colored vertical line left to branch tip labels shows in cyan, euAP2 homologs, in red, euANT Bryophyte sequences, in dark blue, euANT sequences from *P. abies*, and in orange, euANT Mono and Dicot sequences. Blue arrow indicates MpoAPL position. Green circle marks outgroup position of the *K. flaccidum* sequence.

7.2 Over expressor and transcriptional fusions for *Agrobacterium*-mediated transformation of *Marchantia polymorpha* and *Arabidopsis thaliana*

To unravel the spatio-temporal expression pattern of *MpAPL* in *M. polymorpha* a transcriptional fusion called $_{pro}$APL:GUS was constructed. This was achieved by cloning a 2.5 kb fragment located upstream of the *MpAPL* start codon into the pMpGWB204 destination vector (Ishizaki *et* al., 2015), which contains a promoterless β-glucuronidase (GUS) coding sequence. In $_{pro}$APL:GUS, GUS expression will be directed by the transcription of the promoter region of MpAPL. The T-DNA region of $_{pro}$APL:GUS and the electrophoresis gel made to verify the construction are shown in Fig. 13A and 13B.

To perform functional analyses of *MpAPL*, a gain of function expression plasmid was created by recombining the pMpGWB206 destination vector (Ishizaki *et* al., 2015) with pDONR™221 containing the coding sequence of *MpAPL.* In the resulting $_{pro}$35S:APL-Citrine plasmid, MpAPL-Citrine expression will be under the control of the Cauliflower mosaic virus 35S promoter (35S). The T-DNA region of OE APL and the electrophoresis gel made to verify the construction are shown in Fig. 13C and 13D.

Both plasmids were verified by Sanger sequencing using the primers with which the promoter region and the coding sequence of *MpAPL* was amplified (data not shown).
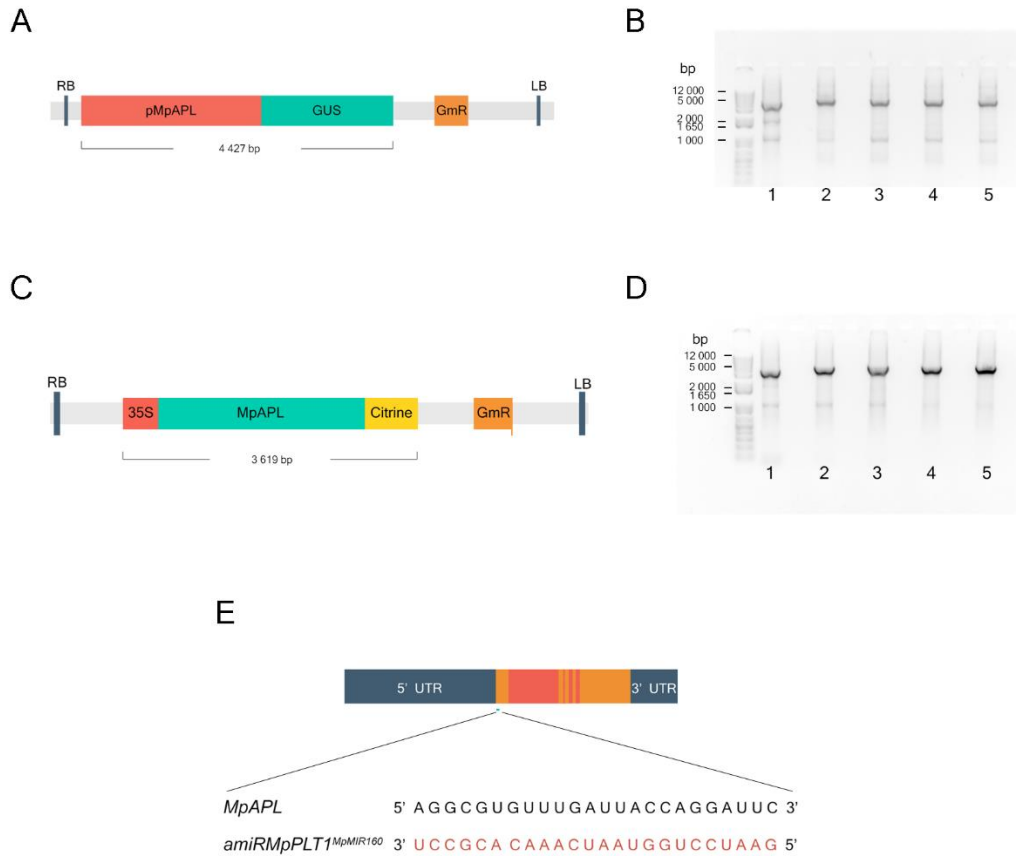
Figure 13 (A) proAPL:GUS T-DNA region (4 427 base pairs) (B) Electrophoresis gel of PCR products amplified using M13 primers. 1: PCR product of amplification with M13 primers using the empty MpGWB204 destination vector as template (3 955 bp). 2, 3, 4 and 5: Products of amplification using plasmid extracted from *E. coli* transformed with LR reaction between proAPL and MpGWB204. The expected PCR product size is 4 823 bp. (C) OE APL T-DNA region (4 427 base pairs) (D) Electrophoresis gel of PCR products amplified using M13 primers. 1: PCR product of amplification with M13 primers using the empty MpGWB206 destination vector as template (3 635 bp). 2, 3, 4, 5 and 6: Products of amplification using plasmid extracted from *E. coli* transformed with LR reaction between MpAPL and MpGWB204. The expected PCR product size is 4 133 bp. (E) Target sequence of amiRMpPLT1[MpMIR160] on the MpAPL mRNA. The binding site of amiRMpPLT1[MpMIR160] is located 75 bp downstream of ATG, in a fragment coded by the first exon of MpAPL.

7.3 Genetic transformation of Marchantia polymorpha trying two different methods

To obtain *M. polymorpha* transformant lines for the functional characterization of *MpAPL*, construcs generated (proAPL:GUS, pro35S:APL and amiRMpPLT1[MpMIR160])were

transformed into *Agrobacterium tumefaciens* and were used then for transformation of *M. polymorpha* thalli following two different methods.

When the genetic transformation of *M. polymorpha* was carried out by the modified protocol developed by Kubota *et* al. (2013), thalli pieces showed signs of regeneration in apical zones. However, once they were under selection of Gentamicin or Kanamycin, thalli pieces showed signs of death and at 19 days-post-selection, it was evident that thalli were no longer alive. This happened for the transformation of the three constructs. In Figure 14 a sequence of pictures with thalli pieces at 1, 11 and 19 days-post-selection in gentamicin is shown. This shows that for this specific protocol, the co-culture with *A. tumefaciens* in liquid media and with constant shaking is essential to achieve the transformation of *M. polymorpha*.
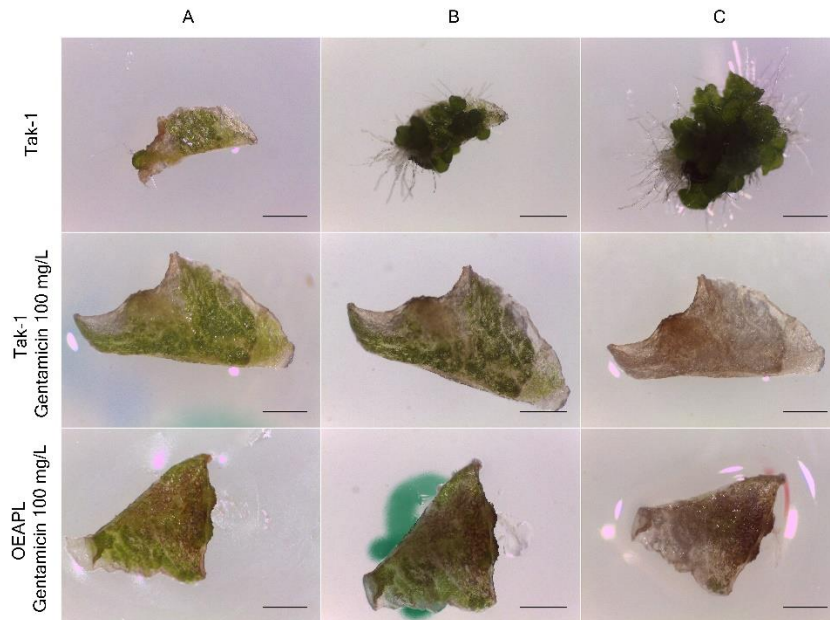
Fig. 14 Agrobacterium-mediated transformation of 14-day-old thalli pieces of *Marchantia polymorpha.* The protocol was modified based on the one developed by Kubota *et* al. (2013). Rows represent different treatments. First row, Tak-1 thalli pieces which were nor transformed neither grown under selection solution. Second row, Tak-1 thalli pieces on which selection solution containing Gentamicin 100 mg/L was poured. Third row, thalli pieces on which both transformation (35S:MpAPL) and selection (Gentamicin 100 mg/L) were poured. Column A shows plants one day after selection solution was poured. Column B, 11 days after selection solution was poured. Column C, 19 days after selection solution was poured. Scale bars, 1 mm.

When the T-AgarTrap method (Tsuboyama-Tanaka *et* al., 2015) was employed, successful transformation signs were observed in *M. polymorpha* thalli fragments. Negative controls for gentamicin had already died completely, while kanamycin controls showed signs of death at 30 days after selection (Fig. 15B and 15C). On the other hand, $_{pro}$APL:GUS, OE APL y amiRMpPLT1^{MpMIR160} potentially-transformed thalli showed green regions where regeneration seemed to take place (Fig. 15D-F). Ideally, these zones developed from one or more transformed cells.
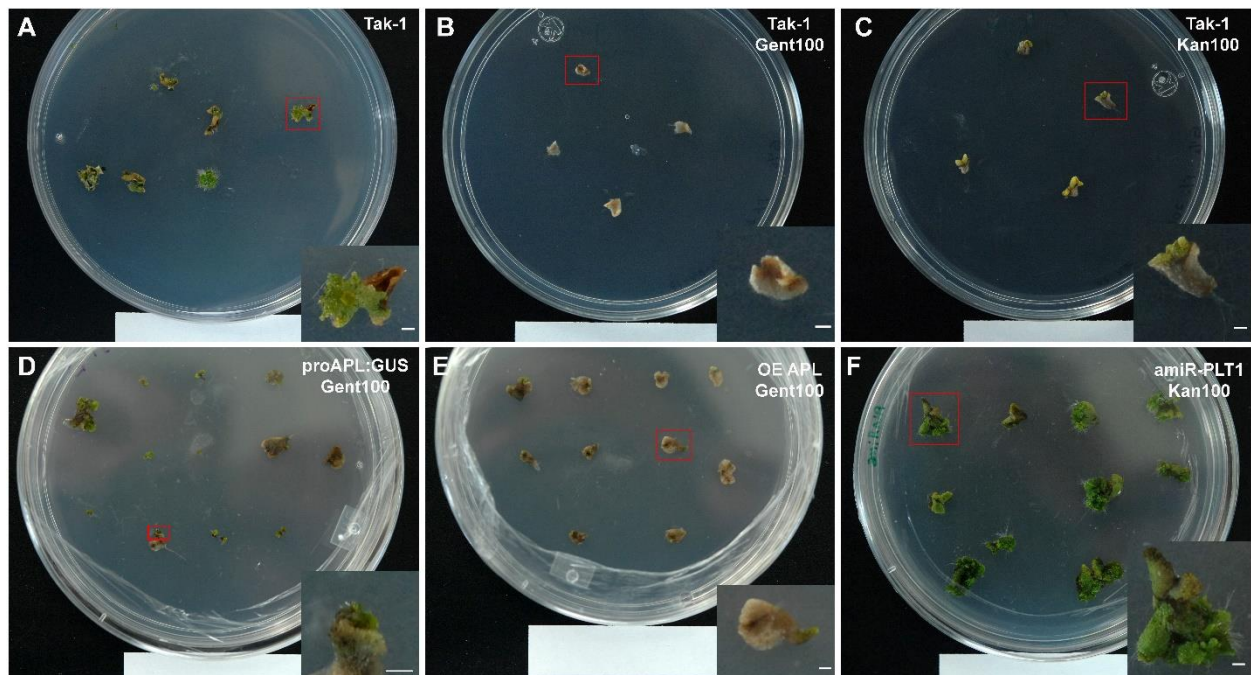
Fig. 15 T-AgarTrap *Agrobacterium*-mediated transformation of *Marchantia polymorpha.* Method consists in pouring transformation and selection solutions over mature thalli pieces (Tsuboyama-Tanaka *et* al., 2015). One-month-after-selection (A) Tak-1 plants. (B) and (C) Tak-1 soaked with Gentamicin and Kanamycin selection solutions, respectively. (D) Thalli pieces transformed with the vector carrying the $_{pro}$APL:GUS construction. (E) Thalli pieces transformed with the OEAPL expression vector. (F) Thalli pieces transformed with the amiRMpPLT1^{MpMIR160} vector. Insets show magnification of one of the thalli pieces from each treatment. Scale bars, 5 mm.

Three gentamicin-resistant primary transformants were obtained from the $_{pro}$APL:GUS genetic transformation and they have already developed gemmae. Once transformant

lines are established, gemmae will be assayed for GUS expression to reveal where in the gametophyte *MpAPL* is being transcribed.

35S:APL-transformed thalli were checked for citrine expression. Regions where citrine expression co-localized with auto fluorescence were observed (Supp. Fig. 4). Still, cells that appeared to be transformed died. It could be that prolonged gentamicin exposure possibly prevents regeneration of this transformants or that the MpAPL overexpression has effects that will not let cell proliferation take place. New events of transformation with this construct are in progress.

From the transformation with amiRMpPLT1^{MpMIR160}, two primary transformants were obtained that, at two months after selection in kanamycin, have developed gemma cups and gemmae (Fig. 16). When compared with Tak-1 thalli, amiRMpPLT1^{MpMIR160} plants have a less expanded thallus that does not branch following the same pattern as the wild-type, amiRMpPLT1^{MpMIR160}transformants are smaller and branch more, which gives them a spherical instead of the proper Wt body form. The most striking phenotype in amiRMpPLT1^{MpMIR160} transformants is the number of gemma cups, which are a lot more compared to the wild type (Fig. 16).
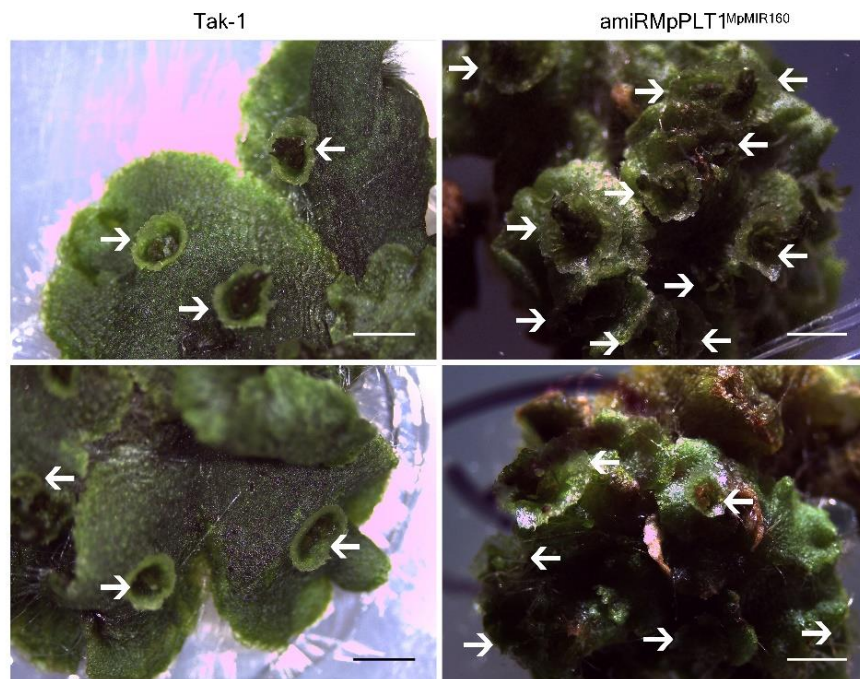
Fig. 16 Tak-1 control and amiRMpPLT1^{MpMIR160} primary transformants 60 days after selection solution was poured over the thalli pieces (no selection on Tak-1 control, Kanamycin 100 mg/L

on amiRMpPLT1^{MpMIR160} plants). Scale bars, 2 mm. White arrows mark the position of gemmae cups.

RT-PCR analyses, to quantify *MpAPL* mRNA levels, will be performed once stable transformant lines are achieved to confirm the action of amiRMpPLT1$_{MpMIR160}$ as well as a detailed phenotypic characterization of the gametophyte development.

7.4   Functional conservation between *MpAPL* and *A. thaliana AIL* genes

Functional conservation analyses provide helpful information to gain insight of conserved developmental mechanisms between plant lineages or to discern if a mechanism was recruited uniquely by a lineage.

With this idea in mind, a complementation approach was carried out. The *A. thaliana plt1;plt2* double mutant (Aida *et* al., 2004) was genetically transformed with 35S:APL-CITRINE construct to explore if the ectopic expression of *MpAPL* recovers the root phenotype characteristic of the double mutant.

The *plt1;plt2* double mutants display morphological defects such as disturbed structure of the columella, reduced root growth compared to wild-type plants and reduced number of meristematic cells which decreases over time and eventually differentiates (Figure 17 A, D).
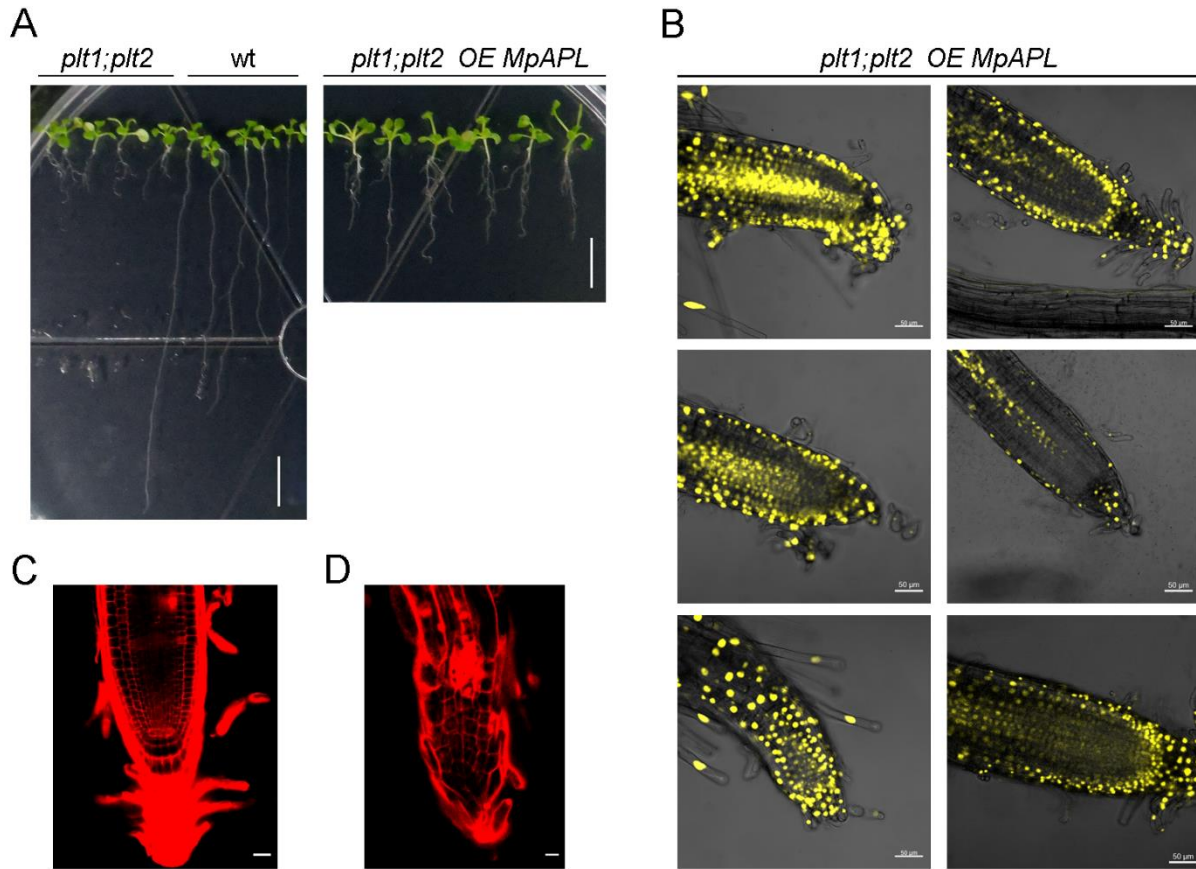
Fig. 17 (A*)* 21 d.p.g. seedlings: *plt1;plt2*, wild-type (wt) and *35S:MpAPL-CITRINE* in the *plt1;plt2* background. Scale bars, 1 cm. (B) Citrine expression on T_1 independent lines of *plt1;plt2 OE MpAPL* at 8 d.p.g. Scale bars, 50 µm. (C, D) Propidium iodide-stained root tips of 21 dpg seedlings from the wt (C) and *plt1; plt2* (D)*.* Scale bars, 20 µm.

T_1 line seedlings show a slight difference in phenotype compared to *plt1;plt2* seedlings at 21 days post germination (d.p.g.), roots appear longer in the *plt1;plt2;* 35S:APL-CITRINE lines but the wild-type phenotype was not fully recovered (Fig 17A). Different patterns of citrine expression were detected in independent T_1 lines when observed using confocal microscopy (Fig. 17B). It can also be observed that some *plt1;plt2;* OE 35S:APL-CITRINE lines have a root phenotype that resembles the wild-type, with a more organized root stem cell niche and longer cell files in the root meristem zone when compared to that of *plt1;plt2* seedlinds (Fig. 17B and C). Other *plt1;plt2* OE APL lines maintained the double mutant phenotype. These results show that while a macroscopic recovering of the wild-type phenotype was not achieved when expressing MpAPL in the

A. thaliana plt1;plt2 double mutant, a slight complementation was obtained in some lines of transgenic plants, which suggests that the function of MpAPL in maintaining the organization and function of stem cell niches might be conserved since the divergence of the liverwort lineage.

Col-0 OE MpAPL
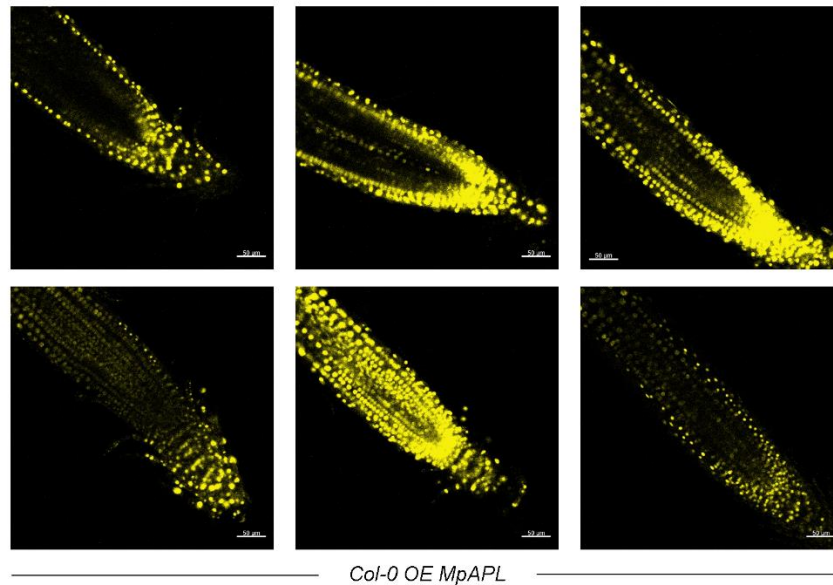
Fig. 18 Confocal images showing citrine expression in the root tips of independent T_2 lines of *Col-0 35S:MpAPL-CITRINE* at 7 and 11 d.p.g. Scale bars, 50 µm.

Overexpression of different AIL genes in *A. thaliana* produces aberrant phenotypes such as development of ectopic organs and formation of larger floral structures (Aida *et* al., 2004; Krizek, 1999). As part as the conservation analysis, OE APL was transformed into wild-type *A. thaliana* plants to evaluate if overexpressing *MpAPL* has a similar effect as another *A. thaliana* overexpressed AIL gene. Resulting transgenic lines show no apparent overexpressing phenotype in the root tips (Fig. 18). Once homozygous lines are obtained, more detailed analyses of the shoot and root will be carried out to confirm or discard such observations.

8. Discussion

8.1 Phylogenetic analysis of *MpAPL*

To hypothesize evolution of gene function it is essential to understand phylogenetic relationships among plant lineages. In the search for an *AthAIL* gene ortholog in *M. polymorpha* only one gene with conserved euANT lineage characteristics was found, *MpAPL,* which has the two AP2 domains and the four pre-domain motifs highly conserved (Fig. 7). It has been stated that the *Marchantia polymorpha* genome shows low levels of gene duplication and a minimal gene family size has been found to be the trend for genes with important roles in *Marchantia* development (Eklund *et* al., 2015). The euANT gene lineage seems to be no exception. In the phylogenetic analysis where the relationship beetween *AthAIL* genes and *MpAPL* is exlpored, MpAPL was positioned as sister to the PLT group, which means that it evolved from an ancestor sequence that also gave rise to the PLT sequences in *A. thaliana*. According to the phylogenetic reconstruction among Streptophyes, AIL genes evolved in a common ancestor of liverworts and vascular plants and then it had lineage-specific duplications in mosses, conifers, dicots and monocots. The search in *Klebsormidium nitens* retrieved just one sequence with basal ANT features. The absence of an euANT-type sequence in *K. nitens* suggests that the appearance of the euANT lineage accompanied terrestrialization, although more sampling in Streptophyte algae is required to strength such hypothesis.

8.2 AIL genes promoter analysis

Gene duplication and expansion of gene families is associated with an increment in the complexity of regulatory networks in which genes are involved. These innovations might co-evolved with an increase in the complexity of plant morphology. Since the MpAPL protein core is conserved in liverworts (Fig. 8), it could be that after gene duplication the promoter regions of *AthAIL* genes diverged, causing the diversity in expression patterns and developmental processes they are regulating. Our promoter similarity analysis revealed that promoter regions of *AthAIL* genes are highly variable between all the members of the family. However, pairs of promoter sequences that are more similar between them, correspond to genes that show similar expression patterns or redundant action (Fig. 10), like *PLT1* and *PLT2* (Aida *et* al., 2004).

AIL genes regulate stem cell niche maintenance and organ growth both in the shoot and root of *A. thaliana* by their interaction with the phytohormone auxin (Horstman et al.,

2014). Comparative promoter analysis for AuxREs Cis-elements in *AIL* genes shows that the *MpAPL* promoter has six putative AuxREs. Since the auxin perception pathway is conserved in *M. polymorpha* (Eklund et al., 2016; Ishizaki *et* al., 2012) and *MpAPL* could potentially bind ARFs, the Auxin-AIL regulatory network could be conserved in liverworts. Experiments with $_{pro}$APL:GUS *M. polymorpha* lines grown in the presence of different concentrations of auxin or auxin inhibitors could help unravel if *MpAPL* responds to auxins in the same way *AIL* genes do to regulate developmental processes in *A. thaliana*.

8.3   Functional conservation of MpAPL

Studies of *AIL* genes in *Physcomitrella patens* and *Ceratoperis richardii* have found that they act as a molecular switch for the development of stem cells and cell-fate determination (Aoyama *et* al., 2012; Bui *et* al., 2017). In *P. patens*, AIL gene expression is also induced by auxin. $_{pro}$APL:GUS and amiRMpPLT1^{MpMIR160} *M. polymorpha* lines obtained in this study (Fig. 15 and 16) will help us evaluate this in a most basal lineage. However, the fact that the *A. thaliana plt1;plt2* double mutant phenotype is partially complemented by the overexpression of MpAPL (Fig. 17) suggests that that the role of *AIL* genes in stem cell regulation might be conserved since the liverworts divergence in plant evolution. A more specific complementation experiment, in which *MpAPL* CDS is driven by the Arabidopsis *PLT1* or *PLT2* promoters, is needed in order to sustain such statement.

9. Conclusion

This study shows that *MpAPL* is an ortholog of the *Arabidopsis thaliana AIL* gene family, that the protein core structure is well conserved between lineages and that AIL gene duplication in eudicots was followed by divergence in their promoter regions, which resulted in a diverse pattern of expression and functions. It also gives evidence, though preliminarily, that the function of AIL genes in stem cell regulation and their interaction with auxin is conserved since the divergence of Liverworts. Further studies with *M. polymorpha* lines obtained will unravel the ancestral function of AIL genes. Undergoing experiments, such as the detailed analyses of the phenotype of Marchantia lines carrying loss and gain of function constructs for *MpAPL*, are needed to unravel the function of MpAPL in Marchantia.

10. References

Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., … Scheres, B. (2004). The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. *Cell*, *119*(1), 109–20.

Aoyama, T., Hiwatashi, Y., Shigyo, M., Kofuji, R., Kubo, M., Ito, M., & Hasebe, M. (2012). AP2-type transcription factors determine stem cell identity in the moss Physcomitrella patens. *Development (Cambridge, England)*, *139*(17), 3120–9.

Barrett, L. W., Fletcher, S., & Wilton, S. D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci.,* 69(21):3613-34.

Bateman, R. M., Crane, P. R., DiMichele, W. A., Kenrick, P. R., Rowe, N. P., Speck, T., & Stein, W. E. (1998). EARLY EVOLUTION OF LAND PLANTS: Phylogeny, Physiology, and Ecology of the Primary Terrestrial Radiation. *Annual Review of Ecology and Systematics*, *29*(1), 263–292.

Bennett, T. A., Liu, M. M., Aoyama, T., Bierfreund, N. M., Braun, M., Coudert, Y., … Harrison, C. J. (2014). Plasma membrane-targeted PIN proteins drive shoot development in a moss. *Current Biology*, *24*(23), 2776–2785.

Boundless. "Liverworts and Hornworts". *Boundless Biology*. Retrieved from: https://www.boundless.com/biology/textbooks/boundless-biology-textbook/seedless-plants-25/bryophytes-156/liverworts-and-hornworts-611-11830/

Bui, L. T., Pandzic, D., Youngstrom, C. E., Wallace, S., Irish, E. E., Szövényi, P., & Cheng, C.-L. (2017). A fern AINTEGUMENTA gene mirrors BABY BOOM in promoting apogamy in Ceratopteris richardii. *The Plant Journal*, *90*(1), 122–132.

da Costa, C. T., de Almeida, M. R., Ruedell, C. M., Schwambach, J., Maraschin, F. S., & Fett-Neto, A. G. (2013). When stress and development go hand in hand: main hormonal controls of adventitious rooting in cuttings. *Frontiers in Plant Science*, *4*, 133.

Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. (2004). WebLogo: A sequence logo generator. *Genome Research*, 14:1188-1190.

Darriba, D., Taboada, G.L., Doallo, R. & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, 27:1164-1165.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J. & Bateman, A. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, 44, D279-D285.

Floyd, S. K., & Bowman, J. L. (2007). The Ancestral Developmental Tool Kit of Land Plants. International Journal of Plant Sciences, 168(1), 1–35.

Galinha, C., Hofhuis, H., Luijten, M., Willemsen, V., Blilou, I., Heidstra, R., & Scheres, B. (2007). PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. *Nature*, *449*(7165), 1053–7.

Graham, L. E., Cook, M. E., & Busse, J. S. (2000). The origin of plants: Body plan changes contributing to a major evolutionary radiation. *Proceedings of the National Academy of Sciences*, *97*(9), 4535–4540.

Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.,* 41: 95-98.

Hammer, Ř., Harper, D.A.T. & Ryan, P.D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica* 4(1): 9.

Horstman, A., Willemsen, V., Boutilier, K., & Heidstra, R. (2014). AINTEGUMENTA-LIKE proteins: Hubs in a plethora of networks. *Trends in Plant Science*, *19*(3), 146–157.

Ishizaki, K., Chiyoda, S., Yamato, K. T., & Kohchi, T. (2008). Agrobacterium-Mediated Transformation of the Haploid Liverwort Marchantia polymorpha L., an Emerging Model for Plant Biology. *Plant and Cell Physiology*, *49*(7), 1084–1091.

Ishizaki, K., Nonomura, M., Kato, H., Yamato, K. T., & Kohchi, T. (2012). Visualization of auxin-mediated transcriptional activation using a common auxin-responsive reporter system in the liverwort Marchantia polymorpha. *Journal of Plant Research*, *125*(5), 643–651.

Ishizaki, K. (2016). Evolution of land plants: insights from molecular studies on basal lineages. *Bioscience, Biotechnology, and Biochemistry*, *8451*(September), 1–8.

Jill Harrison, C. (2017). Development and genetics in the evolution of land plant body plans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1713), 20150490.

Katoh, K. & Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol*, 30 (4): 772-780.

Kim, S., Soltis, P. S., Wall, K., & Soltis, D. E. (2006). Phylogeny and domain evolution in the APETALA2-like gene family. *Molecular Biology and Evolution*, *23*(1), 107−20.

Kolde, R. (2015). pheatmap: Pretty Heatmaps. R package version 1.0.8. http://CRAN.R-project.org/package=pheatmap

Krizek, B. A. (1999). Ectopic expression of AINTEGUMENTA in Arabidopsis plants results in increased growth of floral organs. *Developmental Genetics*, *25*(3), 224−36.

Kubota, A., Ishizaki, K., Hosaka, M., & Kohchi, T. (2013). Efficient Agrobacterium-mediated transformation of the liverwort Marchantia polymorpha using regenerating thalli. *Bioscience, Biotechnology, and Biochemistry*, *77*(1), 167−72.

Kumar, S., Stecher, G. & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol Biol Evol*, 33(7):1870-4.

Li, P., & Xue, H. (2011). Structural characterization and expression pattern analysis of the rice PLT gene family. *Acta Biochimica et Biophysica Hungarica*, *43*(July), 688−697.

Ligrone, R., Duckett, J. G., & Renzaglia, K. S. (2012). Major transitions in the evolution of early land plants: a bryological perspective. *Annals of Botany*, *109*(5), 851−871.

Martinez-Trujillo, M., Limones-Briones, V., Cabrera-Ponce, J.L. & Herrera-Estrella, L. (2004). Improving transformation efficiency of Arabidopsis thaliana by modifying the floral dip method. *Plant Mol Biol Rep*, 22: 63.

Miller, M.A., Pfeiffer, W. & Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop*, 14 Nov. 2010, New Orleans, LA pp 1-8.

Pinon, V., Prasad, K., Grigg, S. P., Sanchez-Perez, G. F., & Scheres, B. (2013). Local auxin biosynthesis regulation by PLETHORA transcription factors controls phyllotaxis in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(3), 1107−12.

Pires, N. D., & Dolan, L. (2012). Morphological evolution in land plants: new designs with old genes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1588), 508−518.

Pires, N. D., Yi, K., Breuninger, H., Catarino, B., Menand, B., & Dolan, L. (2013). Recruitment and remodeling of an ancient gene regulatory network during land plant evolution. *Proceedings of the National Academy of Sciences*, *110*(23), 9571−9576.

Plackett, A. R. G., Di Stilio, V. S., & Langdale, J. A. (2015). Ferns: the missing link in shoot evolution and development. *Frontiers in Plant Science*, *6*, 972.

Prasad, K., Grigg, S. P., Barkoulas, M., Yadav, R. K., Sanchez-Perez, G. F., Pinon, V., … Scheres, B. (2011). Arabidopsis PLETHORA transcription factors control phyllotaxis. *Current Biology: CB*, *21*(13), 1123−8.

Prince, V. E., & Pickett, F. B. (2002). Splitting pairs: the diverging fates of duplicated genes. *Nature Reviews Genetics*, *3*(11), 827−837.

R Core Team (2013). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL http://www.R-project.org/.

Rambaut, A. (2012). FigTree v1.4.2: Tree figure drawing tool. Available: http://tree.bio.ed.ac.uk/software/figtree/.

Riechmann, J. L., & Meyerowitz, E. M. (1998). The AP2/EREBP family of plant transcription factors. *Biological Chemistry*, *379*(6), 633−46.

Salehin, M., Bagchi, R., & Estelle, M. (2015). SCF TIR1/AFB -Based Auxin Perception: Mechanism and Role in Plant Growth and Development. *The Plant Cell Online*, *27*(1), 9−19.

Sarkar, A. K., Luijten, M., Miyashima, S., Lenhard, M., Hashimoto, T., Nakajima, K., … Laux, T. (2007). Conserved factors regulate signalling in Arabidopsis thaliana shoot and root stem cell organizers. *Nature*, *446*(7137), 811−814

Schneider, T.D. & Stephens, R.M. (1990). Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res*. 18:6097-6100

Shimamura, M. (2016). Marchantia polymorpha: Taxonomy, Phylogeny and Morphology of a Model System. *Plant & Cell Physiology*, *57*(2), 230−56.

Skinne, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J. & Holmes, I.H. (2009). JBrowse: A next-generation genome browser. *Genome Res.,* 19:1630-1638.

Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*.

Tsuboyama-Tanaka, S., Nonaka, S., & Kodama, Y. (2015). A highly efficient AgarTrap method for genetic transformation of mature thalli of the liverwort *Marchantia polymorpha* L. *Plant Biotechnology*, *32*(4), 333−336.

Wellman, C. H., & Strother, P. K. (2015). The terrestrial biota prior to the origin of land plants (embryophytes): a review of the evidence. *Palaeontology*, *58*(4), 601−627.

Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., … Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences*, *111*(45), E4859−E4868.

Yamaguchi, N., Jeong, C. W., Nole-Wilson, S., Krizek, B. A., & Wagner, D. (2016). AINTEGUMENTA and AINTEGUMENTA-LIKE6/PLETHORA3 Induce LEAFY Expression in Response to Auxin to Promote the Onset of Flower Formation in Arabidopsis. *Plant Physiology*, *170*(1), 283−93.

11. Apendix

11.1 Supplementary Tables

Organism	ID	Sequence name	Length (aa's)	Database
Klebsormidium nitens	Kf100657_0020_v1.1	KflAP2-l	1026	NIES-2285 genome project
Marchantia polymorpha	Mapoly0008s0071.1	MpoAPL	850	Phytozome
Marchantia polymorpha	Mapoly0009s0013.1	Mpo1	413	Phytozome
Physcomitrella patens	Pp3c9_25570v3.1	Ppa1	869	Phytozome
Physcomitrella patens	Pp3c15_24790v3.1	Ppa2	911	Phytozome
Physcomitrella patens	Pp3c9_2500v3.1	Ppa3	767	Phytozome
Physcomitrella patens	Pp3c15_24980v3.1	Ppa4	910	Phytozome
Physcomitrella patens	Pp3c6_21840v3.1	PpaWRI	469	Phytozome
Megaceros tosanus	UCRN_2011439	MtoPLT2	962	oneKP
Picea abies	MA_98095g0010	Pab1	844	ConGenIE.org
Picea abies	MA_121578g0010	Pab2	655	ConGenIE.org
Picea abies	MA_86195g0010	PabPLT2	887	ConGenIE.org
Amborella trichopoda	scaffold00066.6	AtrBBM	813	Phytozome
Amborella trichopoda	scaffold00019.292	AtrAIL5	505	Phytozome
Amborella trichopoda	scaffold00024.229	AtrANT2	541	Phytozome
Arabidopsis thaliana	AT3G20840	AthPLT1	592	Phytozome
Arabidopsis thaliana	AT1G51190	AthPLT2	586	Phytozome
Arabidopsis thaliana	AT5G10510	AthPLT3	624	Phytozome
Arabidopsis thaliana	AT5G17430	AthPLT4	602	Phytozome
Arabidopsis thaliana	AT5G57390	AthPLT5	576	Phytozome
Arabidopsis thaliana	AT5G65510	AthPLT7	514	Phytozome
Arabidopsis thaliana	AT4G37750	AthANT	573	Phytozome
Arabidopsis thaliana	AT1G72570	AthAIL1	427	Phytozome
Oryza sativa	Os03g56050.1	OsaANT1	672	Phytozome
Oryza sativa	Os01g67410.1	OsaBBM1	717	Phytozome
Oryza sativa	Os04g42570.1	OsaBBM2	678	Phytozome
Oryza sativa	Os02g40070.1	OsaPLT2	722	Phytozome
Oryza sativa	Os11g19060.1	OsaBBM3	577	Phytozome
Oryza sativa	Os04g55970.2	OsaANT2	508	Phytozome
Oryza sativa	Os03g19900.1	OsaAIL6	506	Phytozome

Supp. Table 1 Organism, Sequence ID, Sequence name, Length and Database of sequences used for the phylogenetic reconstruction of *AIL* genes among Streptophyte.

| Primer name	Sequence
MpAPL CDS FW	GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGAGGTCGGTCAACAACTGGTTGGCCTTC
MpAPL CDS RV	GGGGACCACTTTGTACAAGAAAGCTGGGTTATCGTTCCAAACTGCAAAAATGGGCATATG
MpAPL PROM 1-4 FW	GGGGACAACTTTGTATAGAAAAGTTGttAAGCTCAGTGAACTGAAGTGAAGC
MpAPL PROM 1-4 RV	GGGGACTGCTTTTTTGTACAAACTTGtAGCTTGACTGCTTGCGAGAAACAGC
MpAPL PROM 1-2 FW	GGGGACAAGTTTGTACAAAAAAGCAGGCTttAAGCTCAGTGAACTGAAGTGAAGC
MpAPL PROM 1-2 RV	GGGGACCACTTTGTACAAGAAAGCTGGGTtAGCTTGACTGCTTGCGAGAAACAGC
M13 FW	GTAAAACGACGGCCAG
M13 RV	CAGGAAACAGCTATGAC

Supp. Table 2 Name and sequence of primers used in this study.

Seq. name	Family	Matrix	Start	End	Strand	Matrix sim.	Sequence
ANT	P$ARF3	P$ETT.01	288	298	+	0.857	ttGTCGaaaag
ANT	P$ARF3	P$ETT.01	2054	2064	+	0.927	atGTCGggtca
ANT	P$ARF3	P$ETT.01	2162	2172	+	0.976	ttGTCGgatta
AIL1	P$AREF	P$ARE.01	235	247	+	0.942	aatTGTCacatta
AIL1	P$AREF	P$ARE.01	283	295	-	0.942	agtTGTCtccatg
AIL1	P$ARF3	P$ETT.01	1291	1301	-	0.851	ttGTCGtaagc
AIL1	P$AREF	P$ARE.01	1319	1331	+	0.942	tttTGTCtcaaat
PLT1	P$ARF3	P$ETT.02	96	106	-	0.836	atgTCGAtata
PLT1	P$ARF3	P$ETT.02	97	107	+	0.836	ataTCGAcata
PLT1	P$AREF	P$ARE.01	339	351	-	0.971	tttTGTCtccccg
PLT1	P$AREF	P$ARE.01	1082	1094	-	0.952	cggTGTCaccttc
PLT1	P$ARF3	P$ETT.01	1702	1712	-	0.976	ttGTCGgattt
PLT2	P$ARF3	P$ETT.01	270	280	+	0.833	ttGTCGagaat
PLT2	P$ARF3	P$ETT.01	2563	2573	-	0.84	gtGTCGtaaat
PLT3	P$AREF	P$SEBF.01	654	666	-	0.966	attTGTCacgatc
PLT3	P$ARF3	P$ETT.01	792	802	-	0.834	ttGTCGtacat
PLT3	P$AREF	P$ARE.01	1363	1375	-	0.932	ataTGTCacaact
PLT3	P$ARF3	P$ETT.01	1636	1646	+	0.836	ttGTCGaatat
PLT3	P$ARF3	P$ETT.01	2155	2165	+	0.958	gtGTCGgacaa
BBM	P$AREF	P$ARE.01	349	361	-	0.932	catTGTCacattg
BBM	P$ARF3	P$ETT.01	734	744	-	0.869	taGTCGgaagt
BBM	P$ARF3	P$ETT.01	1900	1910	+	0.855	taGTCGgataa
BBM	P$AREF	P$ARE.01	2823	2835	+	0.932	ataTGTCgcctcg
PLT5	P$ARF3	P$ETT.01	1041	1051	+	0.942	ttGTCGgtcat
PLT5	P$ARF3	P$ETT.01	1386	1396	-	0.849	tgGTCGgtata
PLT5	P$AREF	P$ARE.01	1812	1824	-	0.932	gcgTGTCccctct
PLT5	P$ARF3	P$ETT.01	2645	2655	-	0.946	atGTCGgatgt
PLT7	P$AREF	P$ARE.01	772	784	+	0.961	attTGTCacctcg
PLT7	P$ARF3	P$ETT.02	1417	1427	-	0.955	gtgTCGAccgg
PLT7	P$ARF3	P$ETT.02	1418	1428	+	0.913	cggTCGAcaca
PLT7	P$AREF	P$ARE.01	1711	1723	+	0.981	ttgTGTCtccatt
MpAPL	P$ARF3	P$ETT.01	192	202	-	0.914	atGTCGgtcta
MpAPL	P$AREF	P$ARE.01	601	613	-	0.961	tcaTGTCtcctcc
MpAPL	P$ARF3	P$ETT.02	2253	2263	-	0.858	cagTCGAcgtc
MpAPL	P$ARF3	P$ETT.02	2254	2264	+	0.894	acgTCGActgt
MpAPL	P$ARF3	P$ETT.01	2963	2973	-	0.913	ctGTCGgtctt
MpAPL	P$ARF3	P$ETT.01	3244	3254	+	0.948	ctGTCGgactg

Supp. Table 3 AuxREs Cis-elements in the promoter regions of *AthAIL* genes and *MpAPL.*

	ANT	AIL1	PLT1	PLT2	PLT3	PLT4	PLT5	PLT7	MpoAPL
ANT	150	18	14	9	10	12	8	15	3
AIL1	14.1	150	15	15	14	10	17	15	8
PLT1	10.1	20.8	150	26	15	8	17	13	5
PLT2	16.5	11.7	100	150	11	10	6	14	5
PLT3	16	11.9	8.2	13.5	150	8	8	21	5
PLT4	14.8	9.7	8.7	17	13.1	150	14	18	5
PLT5	17.3	7.8	12.1	10.2	13.5	17.8	150	13	2
PLT7	12.1	11.2	8.1	10	44.6	10.3	12.3	150	5
MpoAPL	1.8	8.7	13.3	1.8	5.8	10.7	1.2	5.2	150

Supp. Table 4 Similarity matrix obtained from the DiAlign promoter region analysis of AIL genes.

11.2 Supplementary Figures
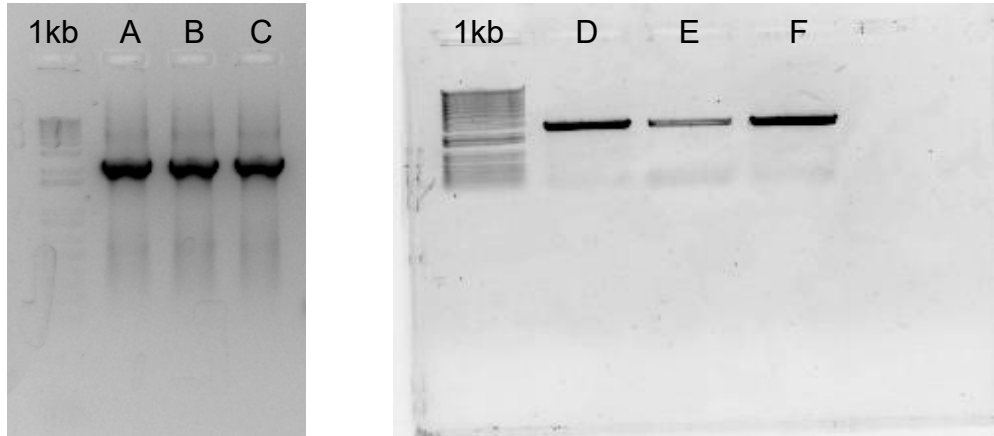
Supp. Fig. 1 Electrophoresis of PCR products from amplification of *MpAPL* coding sequence (A, B and C, expected product size: 2475 bp) and of the promoter sequence (D, E and F, expected product size: 2.5 kb).
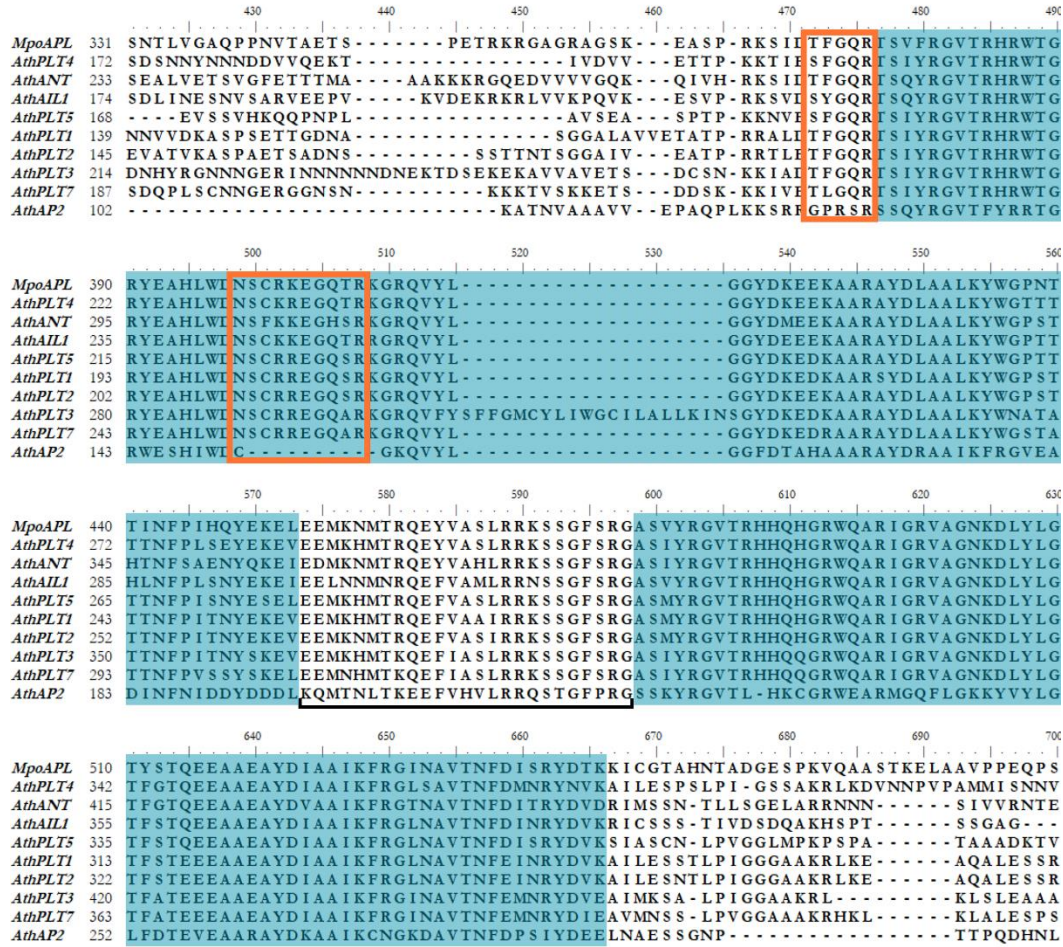
Supp. Fig. 2 *Arabidopsis thaliana* AIL and MpAPL protein sequences alignment. Marked in blue are conserved AP2 domains. Black line under alignment marks linker region. Orange rectangles show motif euANT4, in the pre-domain zone and motif euANT1, inside first AP2 domain.
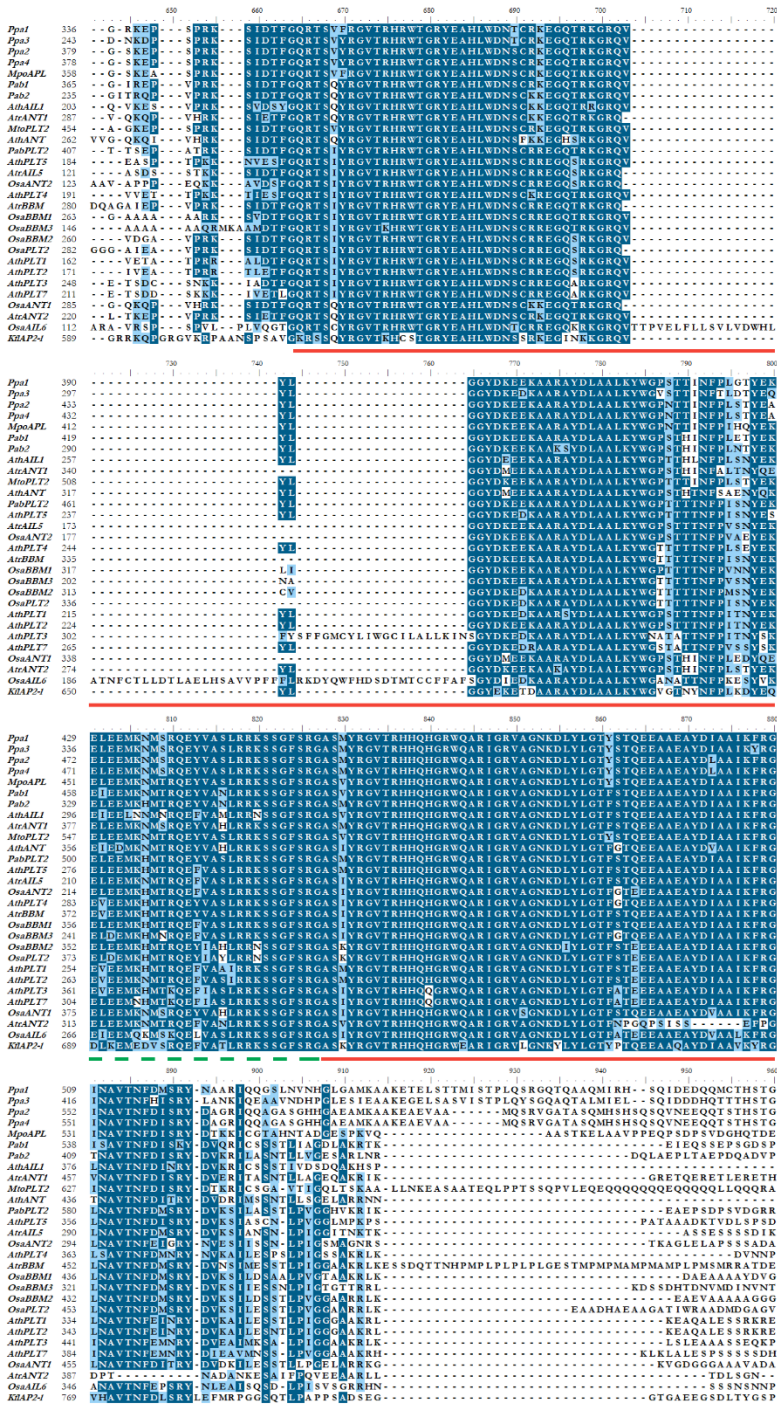
Supp. Fig. 3 Alignment of AIL proteins among Streptophyte taxa. Orange lines correspond to AP2 domains. Dark blue color shows high conservation.
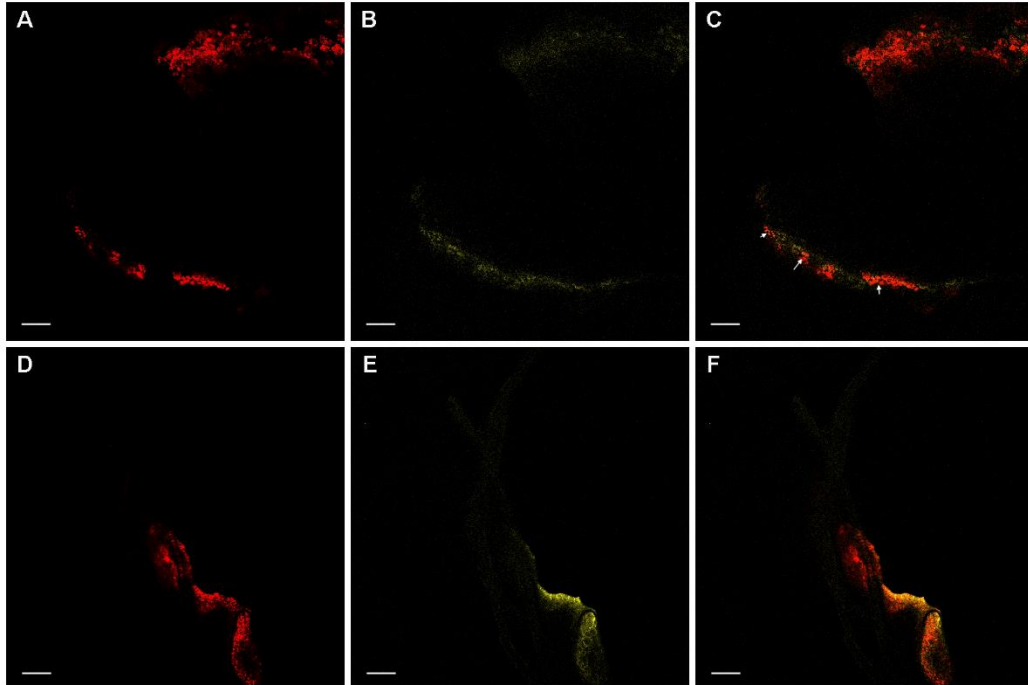
Supp. Fig. 4 Expression of OEAPL ($_{pro}$35S:APL-Citrine) on Tak-1 thalli pieces transformed by the T-AgarTrap method (Tsuboyama-Tanaka *et* al., 2015). (A) and (D), auto florescence of thalli (red). (B) and (E), Citrine expression (yellow). (C) and (F), merge of auto florescence and Citrine expression. White arrows mark places of co-localization which are orange colored because of merging. Scale bars, 100 μm.