UNIVERSITY OF CALIFORNIA
RIVERSIDE

Functional Analysis of Candidate Genes to Encode Rhamnosyl 3-*O*-Methyltransferase in the Moss *Physcomitrella patens*

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Plant Biology

by

Lei Zhu

December 2013

Dissertation Committee:

Dr. Eugene Nothnagel, Chairperson
Dr. Anthony Huang
Dr. Thomas Girke

The Dissertation of Lei Zhu is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Eugene A. Nothnagel. I really appreciate his encouragement, training and support during my Ph.D. study at University of California, Riverside. I would like to thank my other dissertation committee members Dr. Anthony Huang and Dr. Thomas Girke, for their discussions and suggestions for my dissertation project.

I would like to thanks Dr. Martha Orozco-Cardenas and the staff of the UCR Plant Transformation Research Center who generated the transgenic tobacco plants and provided materials and advice for the preparation of the binary vector system for the transformation. I would like to thank my graduate program academic advisor Dr. Linda Walling for her suggestions on analysis of transgenic plants. I would like to thank Dr. Mitsuyasu Hasebe from National Institute for Basic Biology (Japan) for his generous gift of moss transformation vector pTN80.

I would like to thank undergraduate students in Dr. Eugene A. Nothnagel lab, including Lisa Thai, Wynter Hernandez, Jessica Toth, Jane Kim, Yu-Chien Tsai, Candace Fajardo, Thinn Zaw, Jessica Kim, Rutaraj Kanase, Jessica Toth, Robbin Melo, Megan Riley, David Nguyen, and Ryan Brown, Gabriel Juloya, Lauren Kivlen, Jacob Moreno, Allison Cid, Ken Nguyen, Seung Seo, Benjamin Lew, Rutaraj Kanase, Brooke Weston, and Michelle Wurch for their contributions to glycosyl composition analysis of AGPs of knockout moss and transgenic tobacco. Robbin Melo also performed the acetyl bromide assays of lignin-like contents of the moss cell walls. This project was supported by the

National Research Initiative competitive grant No. 2008-35318-04599 from the USDA National Institute of Food and Agriculture.

DEDICATION

This dissertation is dedicated to my family, especially to

My father: Shouxiang Zhu

and

My mother: Yuandi Zhang

Their encouragement and unconditional love support me during my Ph.D. study.

ABSTRACT OF THE DISSERTATION

Functional Analysis of Candidate Genes to Encode Rhamnosyl 3-*O*-Methyltransferase in the Moss *Physcomitrella patens*

by

Lei Zhu

Doctor of Philosophy, Graduate Program in Plant Biology
University of California, Riverside, December 2013
Dr. Eugene A. Nothnagel, Chairperson

Studies of cell walls have long contributed to understanding plant function. Relevance to the production of biofuels from biomass has recently added urgency to the study of plant cell walls. Arabinogalactan proteins (AGPs) are highly glycosylated glycoproteins at the plant plasma membrane and cell wall. Consisting of a core polypeptide surrounded by glycans, AGPs contain abundant galactosyl and arabinosyl residues and often some glucuronosyl, rhamnosyl, or other residues. Previous study revealed that AGPs in the moss *Physcomitrella patens* contain unusual 3-*O*-methyl-rhamnosyl residues (3-*O*-Me-Rha) in amounts as high as 15 mole percent. The goals of this dissertation were to identify and evaluate *Physcomitrella* genes that are candidates to

encode the rhamnosyl 3-*O*-methyltransferase that forms 3-*O*-Me-Rha. A *Mycobacteria* gene encoding a rhamnosyl 3-*O*-methyltransferase and an *Arabidopsis* gene encoding a glucuronosyl 4-*O*-methyltransferase were used as queries in bioinformatics searches of the *Physcomitrella* genome. Eleven candidate genes were identified and, through homologous recombination in *Physcomitrella,* targeted knockouts of these genes were generated. Based on reductions in 3-*O*-Me-Rha/Rha content ratio in AGPs of the knockouts, two genes, *KO1* and *KO9,* were selected for further study. A third gene, *KO11*, was selected for further study because it shares a domain with the *Arabidopsis* glucuronosyl 4-*O*-methyltransferase. Further study of *KO1* led to revision of its gene model. A phylogenic tree suggested that the revised KO1 protein contains a LpxB domain, which is characteristic of bacterial lipid A synthesis. The *Physcomitrella ko1* knockout exhibited phenotypic effects including abnormal growth in protonema and rhizoids, reduced 3-*O*-Me-Rha/Rha in AGPs, and reduced lignin-like content in cell walls. Heterologous expression of *KO1* did not produce detectable 3-*O*-Me-Rha in tobacco AGPs. Further bioinformatics study of KO9 showed it to be an ortholog of caffeoyl-CoA *O*-methyltransferase of lignin biosynthesis. The *Physcomitrella ko9* knockout did not exhibit altered lignin-like content in its cell walls. Heterologous expression of *KO9* in tobacco did not produce detectable 3-*O*-Me-Rha in tobacco AGPs. The *Physcomitrella ko11* knockout exhibited no effect on 3-O-Me-Rha/Rha in AGPs, and no 3-*O*-Me-Rha was detected in AGPs from transgenic *KO11* tobacco. It remains uncertain whether any of the candidate genes encodes rhamnosyl 3-O-methyltransferase in *Physcomitrella*.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 1 - LITERATURE REVIEW

PLANT CELL WALL

The plant cell wall, located outside of the cell membrane, is one of the most important and unique features of the plant cell. Cell walls provide mechanical and structural support, thus protecting the protoplast and cellular organelles and enabling maintenance of turgor pressure within the cell. Moreover, the cell wall provides an effective barrier to help the plant cell to fight against pathogens (Underwood, 2012). Similarly, animal cells have an extracellular matrix that links cells to tissues and serves some of the same functions as a cell wall. The structure and composition of macromolecules in plant cell walls and animal extracellular matrices are, however, totally different. Different from plant cell wall, animal extracellular matrix is typically composed of a gel-like polysaccharide matrix with fibrillar proteins that are associated with a sheet-like basement membrane (Alberts et al., 2002).

Polysaccharides in plant cell walls are the major source of carbohydrates in some plant-derived foods. Ripe fruits and vegetables usually contain large amount of pectins, such as in tomato (Chun & Huber, 1998). In wheat, the cell wall polysaccharide matrix together with lignin provides biomechanical support enabling the stem to resist strong winds during wheat maturation (Zhu et al., 2004). Gum arabic, a natural product that has various uses ranging from an emulsifier to herbal medicine, consists mainly of arabinogalactan proteins (Qi et al., 1991; Goodrum et al., 2000). Thus, studies of plant cell wall and related polysaccharides are important for plant physiology, crop and fruit agriculture, and human nutrition and health.

CELL WALL POLYSACCHARIDES AND BIOSYNTHESIS

Anatomical and other studies have revealed that plant cell walls can be of two distinct types, those being primary cell wall and secondary cell wall. Primary cell wall is principally formed shortly after plant cell division and cell plate deposition. Some types of plant cells, such as leaf mesophyll and root cortex, have only primary cell walls. Other types of plant cells, such as fibers, tracheids, and vessel elements, have both primary and secondary cell wall. Secondary cell wall, if present, is principally deposited and formed only after cell expansion has ceased. Both primary and secondary cell walls contain cellulosic microfibrils interlinked with hemicellulosic polysaccharides, but most primary cell walls additionally have pectic polysaccharides as a third major component. The plant cell wall is also composed of other macromolecular polymers, including proteins and phenolic compounds, the latter being especially abundant in secondary cell walls (Alberts et al., 2002). All of these macromolecules interact with each other to form a dynamic network that may change at various developmental stages and in different environmental conditions.

Cellulose is the major component in most primary and secondary cell walls. Cell walls build strength from cellulose polysaccharide chains, which are 2,000-20,000 glucosyl residues linked together by β (1,4) glycosidic bonds. Adjacent cellulose polysaccharide chains are associated with each other by hydrogen bonds and van der Waals forces. As the most abundant polysaccharide in most cell walls, cellulose accounts for 15% of the total dry weight of leaf cell walls and up to 33% of the dry weight of stem

cell walls of dicots like *Arabidopsis* (Zhong et al., 2005). The composition of cell walls of monocot stems, where considerable secondary wall is present in fiber and xylem cells, can include up to 20-40% cellulose (Gibeaut et al., 2005). In contrast, cellulose accounts for as little as 9-14% of cell wall mass in monocot tissue which only contains primary cell wall (Burke et al., 1974).

Cellulose is synthesized at the plasma membrane by a large rosette protein complex called cellulose synthase (CESA) and then secreted into the cell wall. Cellulose synthesis genes in plants were first characterized in 1996 (Pear et al., 1996). The *CelA* genes in cotton and in rice are homologs of bacterial genes that encode catalytic subunits of cellulose synthase and include coding for a domain that binds the donor substrate UDP-glucose. Movement of the cellulose synthase complex in the plasma membrane is guided by cortical microtubules, and UDP-glucose is the donor of glucosyl residues to synthesize growing cellulose molecules to form a cellulose fiber of about 2-3 nm diameters. Roughly 36 glucan chains are assembled to form a cellulose microfibril, after six of these 2-3 nm fibers bundle together. Following the initial discovery, CESA homologs were soon characterized in the *Arabidopsis* genome, the poplar genome, and the rice genome (Richmond & Somerville, 2000; Somerville, 2006). There are ten *CESA* genes encoding proteins in *Arabidopsis,* but not all of these ten proteins are functionally equivalent. CESA1, CESA3 and CESA6 function together in CESA for synthesis of cellulose in primary cell wall (Arioli et al., 1998; Fagard et al., 2000). CESA4, 7 and 8 are required for biosynthesis of cellulose in secondary cell wall (Taylor et al., 2000; Taylor et al., 2003). Genetic studies show that when *CESA1* and 3 are disrupted, plant

growth is severely retarded. Null alleles of *CESA1* and *CESA3* are gametophyte lethal, but the null allele of *CESA6* shows only a mild phenotype, which indicates that other CESA proteins such as 2, 5, 9 are similar to CESA6 (Hematy et al., 2007).

Hemicelluloses are composed of β-D-1,4-linked pyranosyl polysaccharides that are structurally homologous to cellulose. Because of this structurally similarity, hemicelluloses can interact with cellulose through hydrogen bonding. Hemicelluloses are generally much more abundant in secondary cell wall than in primary cell wall in monocots, dicots and gymnosperms (Darvill et al., 1980). Hemicellulose types include xylan, mannan, and xyloglucan. Xylan is composed of a backbone of β-D-1,4-xylose, and may have side chain residues of glucuronic acid or arabinose, or both, attached to the xylose backbone. Arabinoxylan (AX) has been characterized as having arabinose attached at the O-2 and O-3 positions of xylose in rye (Vinkx et al., 1995). Glucuronoabinoxylan (GAX) accounts for 25% of the primary cell wall in monocots (Darvill et al., 1980) and about 5% of the primary cell wall in dicots. As a further variation, 4-O-methylglucuronic acid has been found attached to xylan in dicot secondary cell wall, but not in monocot cell wall or dicot primary cell wall (Ebringerova & Heinze, 1999). Mannans, which occur in the two categories of galactomannans (GMs) and galactoglucomannans (GGMs), are among the important components in some primary and secondary cell walls and are structurally similar to cellulose. Xyloglucan (XG) is a type of hemicellulose that is particularly abundant in dicot primary cell walls (Bauer et al., 1973; Thomas et al., 1987).

The core of XG has a structure that is very similar to the structure of cellulose. It

has a backbone of β-D-1,4-glucopyranose, the same as cellulose, and about 75% of these glucosyl residues have α-D-xylose attached at the C6 position. The main function of XG is to crosslink to cellulose microfibrils, thereby forming a XG-cellulose network in primary cell walls (Keegstra et al., 1973; Valent & Albersheim, 1974). Hemicelluloses are bonded to cellulose when cellulose molecules are synthesized and secreted to walls. The binding between XG and cellulose is very strong but non-covalent, involving principally hydrogen bonds. This linkage causes cellulose microfibrils to remain small in size, rather than forming even larger aggregates (Hayashi et al., 1987; Whitney et al., 2006), thereby making the cell wall more expandable (Chanliaud et al., 2002).

Pectin is a family of structurally very complex polysaccharides containing galacturonic acid as the principal sugar residue. Pectin variants include homogalacturonan (HG), xylogalacturonan (XGA), apiogalacturonan (AGA), and rhamnogalacturonan I and II (RG-I, RG-II). These pectic polysaccharides account for approximately 90% of the galacturonsyl residues found in the cell wall (Selvendran & O'Neill, 1987). The most abundant form of pectic polysaccharides is HG (Ridley et al., 2001), which consists of a linear backbone polymer of α-1,4 linked galacturonic acid. Pectin is abundant in walls of fruits and many other plant organs in dicots and non-cereal monocots. Pectin accounts for 35% of the mass of dry cell walls of tomato fruits, 52% of the mass of mango fruit walls, and 23% of the mass of *Arabidopsis* leaf walls (Seymour et al., 1990; Zablackis et al., 1995; Muda et al., 1995). Galacturonsyl residues may have either a simple carboxyl group at the C6 position or may be methylated by an ester linkage at the C6 carboxyl group. Some galacturonsyl residues are also acetylated at C2

or C3 position (Ridley et al., 2001). The content of methylesterification changes both temporally and spatially during growth and development (Willats et al., 2001). For those consecutive galacturonic acid residues remaining unmethylated, calcium ions may bind at the C6 carboxyl group. Purified, unesterified HG mixed with calcium salts forms stable gel structures in test tube experiments, and it has been hypothesized such stable gel structures also form in cell walls. This hypothesis is called the "egg-box model" (Liners et al., 1989). The HG can be substituted at the C2 or C3 position of galacturonosyl residues by apiose to make AGA (Hart & Kindel 1970; Ovodov et al., 1971), or substituted at the C3 position of galacturonosyl residues by xylose to form XGA (Schols et al., 1990; Nakamura et al., 2002). The pectic polysaccharide RG-I has a backbone of repeating disaccharide units of →4)-α-D-GalU-(1→2)-α-L-Rha-(1→, as reported for sycamore and soybean cell walls (Yoo et al., 2003; Nakamura et al., 2002). The abundance of RG-I is only 7% in sycamore walls but as high as 36% in potato tuber walls. The C4 position of the rhamnosyl residues in RG-1 can carry a variety of side chain glycans, including galactan, type I arabinogalactan (AG), and type II aranbinogalactan. Most common among these is the type I AG, which has a single L-Ara, inserted into a β (1,4) galactan backbone with branches of one or more arabinosyl residues or a single terminal arabinosyl residue (Huisman et al., 2001). Type II AG is a glycan similar to the glycans attached to arabinogalactan proteins (AGPs) and consists of a β (1,3), β (1,6) galactan framework carrying arabinosyl, glucuronosyl, and other sugar substituents (Luonteri et al., 2003). The RG-II pectic polysaccharide is highly conserved through evolution, existing in plant cell walls across lower plants to higher plants (Darvill et al.,

1978; Zablackis et al., 1995). It has a complex structure containing 12 rare glycan residues, including 2-*O*-methyl xylose, 2-*O*-methyl fucose, aceric acid, 2-keto-3-deoxy-lyxo heptulosaric acid (Dha), and 2-keto-3-deoxy-manno octulosonic acid (Kdo) (York et al., 1985). In addition, RG-II is uniquely self-associated to form dimers through a boron diester bond to certain apiosyl residues in RG-II side chains (Ishii & Matsunaga, 1996).

LIGNIN

Lignin, next to cellulose, is the second most abundant polymer in some plant cell walls (Battle et al., 2000). Lignin is mainly deposited in the secondary cell wall in all vascular plants, especially in the tracheary woody plants. As a phenolic polymer, lignin is deposited among the cell wall polysaccharides in xylem treachery elements in vascular bundles that function in water transport (Zhong & Ye, 2009). Lignin adds rigidity and thus structural support to the cell walls of tracheids, vessel elements, fibers, and various other differentiated cells. Lignin also functions in plant-pathogen interactions, protecting plant cell wall polysaccharides from degradation by enzymes secreted by invading microorganisms. Lignification is commonly found at the site of pathogen invasion (Lange et al., 1995).

Lignification is an important feature that seems to have originated at the time that plants emerged from aquatic environments and adapted to terrestrial environments (Kenrick & Crane, 1997). Lignin evolution is a controversial topic. Comparative genomics studies have revealed that the complete lignin biosynthesis pathway first appeared in the moss *Physcomitrella* (Xu et al., 2009). Weng and Chapple (2010) have

proposed, however, that mosses and other early terrestrial plants accumulated lignans, soluble dimers formed from lignin monomers, rather than polymeric lignin, and that the function of these lignans was to resist UV irradiation rather than to strengthen the cell wall. Besides lignans, other soluble phenylpropanoids such as flavonoids might have also served to protect against UV irradiation (Basile et al., 1999; Umezawa, 2003).

Due to its importance relative to biofuel production, the biosynthetic pathway of lignin has been widely investigated in woody plants and angiosperms in the past ten to fifteen years (Boerjan et al., 2003). The three most abundant monomers of lignin are the *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) residues, these being derived from *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol, respectively (Weng et al., 2008). Woody angiosperm lignin is composed of mainly G and S residues, with only few H residues. Monocot lignin contains closer to equal amounts of G, S, and H residues, while gymnosperm lignin consists of mostly G residues, some H residues, but generally no S residues. Synthesis of lignin monomers begins with phenylalanine. Eight core enzymes are involved in the monolignol biosynthetic pathway (Weng & Chapple, 2010), these eight being phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-hydroxycinnamoyl-CoA ligase (4CL), hydroxycinnamoyl transferase (HCT), p-coumaroyl shikimate 3'-hydroxylase (C3'H), caffeoyl-CoA O-methyltransferase (CCoAOMT), (hydroxy)cinnamoyl-CoA reductase (CCR) and (hydroxy)cinnamyl alcohol dehydrogenase (CAD). Lignin monomers are transported across the plasma membrane and then polymerized in the cell wall through the action of peroxidases.

GLYCOPROTEINS

Cell wall glycoproteins include hydroxyproline-rich proteins (HRGPs) and proline-rich proteins. Extensins and arabinogalactan proteins are the major subfamilies of HRGPs, and most members of this super family usually contain functional repetitive motifs (Kieliszewski & Lamport, 1994).

ARABINOGALACTAN PROTEINS (AGPs)

Arabinogalactan-proteins (AGPs) are a subfamily of highly glycosylated HRGPs found bound to the cell surface and soluble in the cell wall (Serpe & Nothnagel, 1999; Seifert & Roberts 2007). AGPs usually consist of a hydroxyproline-rich core polypeptide surrounded by arabinose and galactose-rich glycan chains. These glycan chains consist of a β (1,3), β (1,6) galactan framework with peripheral substitutents of especially L-arabinosyl residues but also sometimes including D-glucuronosyl, L-rhamnosyl, D-mannosyl, D-xylosyl, D-glucosyl, L-fucosyl, D-glucosaminosyl, and D-galacturonosyl residues in some species (Nothnagel, 1997). Classical AGPs were found to have a hydrophobic C-terminal tail, which is removed during synthesis in the rough endoplasmic reticulum and replaced by a glycosylphosphatidylinositol (GPI)-anchor (Youl et al., 1998). As much as 40% of *Arabidopsis thaliana* AGPs are proposed to have a GPI anchor (Ellis et al., 2010). Other, nonclassical members of the AGP family, however, contain Hyp-poor core proteins (Mollard et al., 1994). In contrast to classical AGPs, none of the nonclassical AGPs have been found to contain C-terminus hydrophobic tails and thus members of this subfamily are considered to be unlikely to have a GPI anchor.

AGP core polypeptides. Classical AGPs contain roughly 100 amino acyl residues in their core polypeptide backbone. Typically, 10 to 13 amino acyl residues in the polypeptide backbone serve as the sites of attachment of *O*-glycan chains (Schultz et al., 2002). AGPs usually contain a N-terminal cleavage signal peptide that guides the AGP into the endomembrane system for eventual secretion to the cell surface. At the C-terminus, many AGPs contain a hydrophobic domain that encodes posttranslational modification with a GPI lipid anchor. Between the N-terminal domain and the C-terminal domain, some classical AGPs and AG peptides consist of a simple domain rich in Pro, Ala, Ser, and Thr residues (Schultz et al., 2000). Chen et al. (1994) and Du et al. (1994) cloned the first two AGP core proteins of this type through traditional biochemical purification. Not all AGPs, however, contain just a single central domain. As additional whole genome sequences of different species have become available, comparative genomic studies have shown that AGP polypeptide backbones are highly diversified. Some AGPs have a short lysine-rich domain, while other AGPs are chimeric with one or more domains that seem to have other functions. For example, fasciclin-like AGPs (FLA) contain one or two fasciclin domains that are hypothesized to function in protein-protein interaction.

Other AGP-like proteins are also modified by arabinogalactan (AG) glycans. For example, nonspecific lipid transfer proteins (ns-LTP) found to contain AGs are thought to be xylogens, i.e., signaling molecules that function in xylogenesis, in *Zinnia elegans*. These *Zinnia* polypeptides share homology with certain polypeptides in *Arabidopsis* (Motose et al., 2004) and with a glycopeptide of rice that is precipitable by the AGP-

specific β-Yariv phenylglycoside (Mashiguchi et al., 2004). Polypeptide sequences that contain hydroxyproline (Hyp) residues alternating with serine (Ser), alanine (Ala), or other small side chain amino acyl residues are considered to be AG glycomodules, i.e., domains that are likely to be glycosylated by arabinogalactan glycans during synthesis (Shpak et al., 1999; 2001). Hence AGP-like proteins usually contain a secretion signal and AG glycomodules. According to this simple rule, up to 40% of *Arabidopsis* polypeptides were predicted to be GPI-modified, which are also predicted to be AG-modified (Borner et al., 2003). Schultz et al (2002) have identified 47 different AGP genes falling into four groups, including classical AGPs, Lys-rich AGPs, AG peptides, and FLAs.

AGP carbohydrate moiety. The most common amino acyl residue found to be O-glycosylated in AGPs is Hyp. Type II AGs are the most common glycan moiety attached to the core polypeptide backbone. These type II AGs have (1-3)-β-galactan and (1-6)-β-galactan chains that are linked together via (1-3, 1-6) branch points. In some AGPs, additional residues in terminal positions on this galactan framework are only Ara_f, Rha_f, and Gal_p residues, thus giving rise to a neutral glycan chain.

Other types of AG side chains, however, are highly variable from this most simple type II AG. For example, some AGs also contain glucuronic acid, fucose, rhamnose, and xylose, thus giving rise to a negatively charged glycan chain. Divergent types of AG side chains on AGPs are present in many species. In *Artemisia vulgaris* pollen allergen, the polypeptide backbone is *O*-glycosylated at hydroxyproline residues with a glycan consisting of a (1→6)-β-galactan core with arabinan side chains (Leonard

et al., 2005). In the moss *Physcomitrella patens*, the β-galactan framework appears to carry linear (1→5)-α-arabinan side chains in the AGP (Lee et al., 2005a). As judged across a range of species, the size of native AG glycans ranges from 5 to 25 kD, or approximately 30 to 120 sugar residues (Fincher et al., 1983; Gane et al., 1995).

GPI anchor. A GPI anchor has been demonstrated to exist on several AGPs and has been proposed to attach to many other AGPs (Borner et al., 2002; 2003; Oxley et al., 1999; Schultz et al., 2004; Svetek et al., 1999; Youl et al., 1998). A bioinformatics search of the *Arabidopsis* genome reveals that about 248 polypeptides encoded by the genome contain the coding sequence for addition of a GPI anchor (Borner et al., 2002, 2003). These 248 polypeptides include diverse classes of cell surface enzymes and other proteins, with approximately 40% of the 248 containing AG modules. In membrane AGPs from pear (Chen et al., 1994), the C-terminus of the core polypeptide has been shown to be coupled via a phosphoethanolamine linker to a glycan of D-Man-(1→2)-α-D-Man-(1→6)-α-D-Man-(1→4)-α-D-GlcN, which is in turn linked to an inositylphosphoceramide lipid residue. This core GPI linker structure is conserved throughout plants, animals, protozoans, and yeast.

Tools to study AGPs. An early and important tool for binding and detecting AGPs are the β-Yariv phenylglycosides, which are brown-red chemical compounds that have general structure of 1,3,5-*p*-glycosyloxyphenylazo-2, 4, 5- trihydroxybenzene. The (β-glucosyl)$_3$ and (β-galactosyl)$_3$ Yariv phenylglycosides are the two most commonly used tools for the detection and binding of AGPs. The (β-mannosyl)$_3$ and (α-galactosyl)$_3$ Yariv phenylglycosides are structurally very similar but do not bind and precipitate AGPs and

are thus useful as negative controls in experiments. Various authors have studied the mechanism of how β-Yariv phenylglycosides react with AGPs, but a consensus conclusion has not yet drawn. It is thought that AGPs bind to stacked aggregates of Yariv phenylglycosides rather than binding to a single Yariv phenylglycoside (Nothnagel et al., 2000).

Besides their use for detection of AGPs, Yariv phenylglycosides are also useful in purifying AGPs and in studying the function of AGPs in plants. Binding of AGPs by Yariv phenylglycosides induces programmed cell death and wound-like responses in *Arabidopsis* cell cultures (Gao & Showalter, 1999; Guan & Nothnagel, 2004). Binding of AGPs by $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside results in increased lethality rate and stalled development of zygote ovules (Hu et al., 2006), or even complete blockage of embryogenesis when applied at a concentration of 250 μM (Chapman et al., 2000). In some but not all cases, the inhibitory effects are reversible, and removal of Yariv phenylglycoside from the medium results in diminished inhibitory effects and gradual return to normal plant growth.

Monoclonal antibody (mAb) probes of AGPs. Monoclonal antibodies are another important class of tools for the study of AGP expression and function *in vivo*. Frequently used anti-AGP mAbs are JIM8, JIM13, JIM14, LM2, and CCRC-M7. For example, JIM13 has been used as a specific probe to detect localization of AGPs in embryos at different developmental stages (Hu et al., 2006).

Molecular biological probes for the study of AGPs. Green fluorescent protein (GFP) fusion proteins have been used to isolate both heterologously expressed AGP

protein backbones and synthetic peptides by the Kieliszewski/Showalter and Matsuoka laboratories (Shpak et al., 1999; Zhao et al., 2002; Shimizu et al., 2005). This approach has enabled direct tests of the Hyp contiguity hypothesis (Kieliszewski & Lamport, 1994) that predicts glycosylation patterns.

Biological functions of AGPs. Many reports in the literature imply that AGPs function in a variety of developmental processes including seed germination (van Hengel & Roberts 2003), growth (Park et al. 2003; Shi et al. 2003; Sun et al. 2004), reproductive development (Acosta-Garcia & Vielle-Calzada, 2004; Sun et al. 2004), vascular development (Motose et al. 2004), root regeneration (van Hengel & Roberts 2003), and interaction with bacteria (Gaspar et al. 2004). Binding and aggregation of AGPs by Yariv phenoglycoside, which presumably blocks AGP function, results in inhibition of cell division in rose suspension cell cultures (Langan & Nothnagel, 1997; Serpe & Nothnagel, 1994). In other cell cultures, such as *Arabidopsis*, Yariv phenylglycoside treatment results in programmed cell death (PCD), with observance of cell corpses, DNA and chromosome breaks, and nucleosomal DNA cleavage (Gao & Showalter, 1999; Chaves et al., 2002).

In seedless nonvascular plants, AGPs seem to participate in the morphogenesis of *Physcomitrella patens* (Lee et al., 2005b) and *Gymnocoloea inflata*, a leafy liverwort (Basile et al., 2000). AGPs also seem to play a role in cell wall regeneration in the liverwort *Marchantia polymorpha* (Shibaya & Sugawara, 2009) and during protonemal tip growth in *Physcomitrella* (Lee et al., 2005b). AGPs are also deposited to the growing tips of lily pollen tubes and moss protonema (Jauh & Lord, 1996; Lee et al., 2005b),

suggesting a role of AGPs in tip growth. A role of AGPs in signal transduction during plant-microbe interactions is suggested by the observations that reduced expression of *AtAGP17/RAT1* in T-DNA insertional mutants and that treatment of roots with β-Yariv phenylglycoside both caused decreased transformation efficiency with *Agrobacterium tumefaciens* (Gasper et al., 2004; Nam et al., 1999).

THE MOSS *PHYSCOMITRELLA PATENS* AS A MODEL ORGANISM

The moss *Physcomitrella patens* of the Phylum Bryophyta is becoming an increasingly popular model organism for studies of plant evolution, development and physiology. In addition to the mosses of Phylum Bryophyta, the other phyla of nonvascular, seedless plants are Phylum Hepatophyta (liverworts) and Phylum Anthocerotophyta (hornworts). Although all of these primitive plants are potentially useful model systems for discovering mechanisms of life in Kingdom Plantae, the moss *Physcomitrella patens* has some properties that have caused it to emerge as the most favored primitive plant for molecular-genetic studies.

Like ferns and seed plants, mosses exhibit alternation of heteromorphic generations in their life cycle with a haploid gametophyte phase and a diploid sporophyte phase. Unlike ferns and seed plants, however, the mosses exhibit the haploid gametophyte as the dominant phase of their life cycle. Many of the most recognizable features of moss occur in this haploid phase. The life cycle of the moss *Physcomitrella patens* starts with germination of spores to produce protonema, single-cell-width filamentous forms of two types. Both types of protonema extend by tip growth and apical

cell divisions. Chloronemal cells, one type of protonema, are very green because they contain many well-developed chloroplasts. The subapical chloronemal cells usually divide once to generate side branches. Some apical chloronemal cells develop into caulonemal cells, a second type of protonema. Compared to chloronemal cells, caulonemal cells contain fewer and smaller chloroplasts, which make caulonemal cells yellowish. Generally, the apical cells of caulonemal filaments extend at a much faster rate than chloronemal cells. As caulonemal cells age, continuing cell cycles will increase the chance of subapical caulonemal cells becoming polyploidy (Reski, 1998). Antheridia, the male reproductive organs, produce spermatozoids, the male gametes which are motile with flagella, whereas archegonia, the female reproductive organs, produce eggs, the female gametes which remain embedded in the archegonia. Antheridia and archegonia can grow on the same leafy shoot and are usually self-fertilized. Motile spermatozoids swim to the archegonia to fertilize the eggs, which become zygotes that grow into sporophytes that eventually bear mature spore capsules. Completion of the entire life cycle of the moss *Physcomitrella* requires several months, the timing depending upon the strain and environmental conditions. For example, the increasingly popular Gransden strain takes about 3 to 4 months to finish its life cycle (Engel, 1968).

Physcomitrella patens plants are relatively easy to culture and maintain in the laboratory because of their high capacity of regeneration. Small pieces of gametophytes or sporophytes can be regenerated to whole plants in liquid or solid culture medium. Protonemal tissue can be disrupted mechanically with a tissue homogenizer. The resulting fragments can be inoculated onto culture medium for further growth (Grimsley

et al., 1977). The resulting cell divisions and extension patterns of the protonemal fragments are very similar to germination of spores.

Current hypotheses on evolution hold that Bryophytes and flowering plants share a common ancestor, with Bryophytes diverging approximately 450 Mya from the main line of evolution leading to flowering plants. Recent EST sequence data and transcriptome studies reveal that *Physcomitrella* and *Arabidopsis* share a high degree of homology (Reski et al., 1998). More than 66% of *Arabidopsis* proteins have homologs in *Physcomitrella*. A variety of genes that have been cloned from *Physcomitrella* are remarkably homologous to higher plant genes (Reski, 1998).

Homologous recombination (HR), which derives from DNA repair mechanisms, enables insertion of a foreign DNA fragment into a specific gene of interest, i.e., gene targeting. Generally, the foreign DNA fragment consists of a linear, double-strand DNA encoding a selection marker gene with partial sequences of the targeted gene both before and after the selection marker gene. The yeast *S. cerevisiae* is a powerful model organism for eukaryotes, in part because it exhibits high efficiency gene targeting through HR (Cho et al., 1999; Courtice & Cove, 1983). In principle, a foreign DNA can integrate into a genome by either HR or illegitimate recombination (IR), depending upon the dominant mechanism used by the cell in repairing DNA double-strand breaks (DSB). In *S. cerevisiae*, the predominant mechanism used to repair DSB is by HR. Thus, yeast exhibits very highly efficient gene targeting by HR. In most plants and mammals, however, non-homologous end joining (NHEJ) is the dominant mechanism used to repair DSB. In those organisms, IR is the dominant mechanism for insertion of a foreign DNA

fragment inserted into genome. While most plant species exhibit very low frequencies of HR (less than 1%), *Physcomitrella patens* is the only plant known to exhibit a high efficiency of HR (Schaefer & Zryd, 1997; Girke et al., 1998). This high frequency of HR greatly facilitates functional genome studies in *Physcomitrella patens* by the gene targeting technique. The efficiency of gene targeting is up to 90% in *Physcomitrella patens*, and thus this moss was selected as the species for knockout studies of plant genes in this dissertation.

An additional factor encouraging the use of *Physcomitrella patens* in this dissertation is the fact that its genome is already completely sequenced by the Joint Gene Institute (JGI). The availability of a sequenced genome is a particular advantage for molecular genetic studies using *Physcomitrella patens* as a model organism (Rensing et al., 2008). *Physcomitrella patens* has 27 chromosomes (n=27) and, according to current information, a genome size of 511 MB, about twice that of *Arabidopsis*.

High efficiency HR that enables gene targeting is a key feature of *Physcomitrella patens* which, when combined with other advantages such as a completely sequenced genome, short life cycle, easy culture and maintenance, makes this moss a powerful model system for use in this dissertation.

COMPARISON OF MOSS CELL WALLS TO ANGIOSPERM CELL WALLS

As already described, the development of gene targeting by HR and the complete sequencing of the *Physcomitrella patens* genome (Rensing et al., 2008) have stimulated the use of this moss as a model system for studies of plant physiology, growth and

development. When using *Physcomitrella* for studies of the plant cell wall, the starting point is largely defined by our understanding of angiosperm cell walls for which a large, but incomplete, literature on composition, structure, and biosynthesis is available. Compared to the cell walls of angiosperms, the cell walls of *Physcomitrella patens* contain the same major classes of polysaccharides, but with some variations in composition and structural details. Although *Physcomitrella patens* and other Bryophytes diverged approximately 450 Mya from the main line of evolution leading to angiosperms, the genomes of *Physcomitrella* and angiosperms have been shown to share some homologous genes encoding glycosyl transferases and other enzymes involved in cell wall synthesis.

Cellulose. Cellulose is a major component in *Physcomitrella patens* cell walls where it has been detected by a variety of methods including comprehensive microarray polymer profiling (CoMPP), chemical analysis of glycosyl composition and linkage, and staining with various fluorescent histochemical stains for cellulose (Kremer et al., 2004; Lee et al., 2011; Moller et al., 2007; Nothnagel & Nothnagel, 2007; Goss et al., 2012).

As already described in a previous section of this chapter, cellulose is synthesized at the plasma membrane by a large rosette protein complex called cellulose synthase (CESA) and then secreted into the cell wall. Following the first characterization of a plant cellulose synthase gene in 1996 (Pear et al., 1996), gene sequence similarities have been found that indicate the presence of CESA orthologs in many plant species. In *Physcomitrella patens*, seven sequences have been classified as CESA genes and three additional sequences have been classified as CESA pseudogenes (Roberts and Bushoven,

2007; Yin et al., 2009; Wise et al., 2011). In vascular plants, some CESA orthologs have been shown to be involved in primary cell wall synthesis, while others have been shown to be involved in secondary cell wall synthesis. *Physcomitrella* has no secondary cell wall, so CESA functional specialization in the moss has been suggested to involve instead some CESAs for tip growth and other CESAs for diffuse growth (Roberts et al., 2012). Phylogenetic analyses leave some uncertainty, however, with regards to whether CESA proteins in *Physcomitrella* form hetero-oligomeric rosettes, as in more advanced plants, or instead form simpler homo-oligomeric rosettes (Roberts and Bushoven, 2007).

Hemicellulose. Again as described in a previous section of this chapter, hemicelluloses are generally composed of β-D-1,4-linked pyranosyl polysaccharides that are structurally homologous to cellulose and can interact with cellulose through hydrogen bonding, thereby forming the cross-linked matrix of the cell wall. Xyloglucan, the most common hemicellulose in dicots, has been detected in *Physcomitrella* through both immunological (Moller et al., 2007) and chemical (Nothnagel & Nothnagel, 2007; Peña et al., 2008) approaches. Peña et al. (2008) reported, however, that *Physcomitrella* xyloglucan has a XXGGG repeating unit rather than the most common XXXG repeating unit of dicots. *Physcomitrella* xyloglucan is further distinguished by side chains that lack fucosyl residues but instead have branching with galacturonosyl and arabinopyranosyl residues. The side chains of *Physcomitrella* xyloglucans differ somewhat between protonemal and gametophore tissue, and it has been suggested that these differences might relate to the roles of xyloglucan in tip and diffuse growth (Roberts et al., 2012).

Chemical analyses, CoMPP, and other immunological approaches all indicate that *Physcomitrella* and other mosses contain much higher amounts of mannans than do the primary cell walls of angiosperms (Geddes & Wilkie, 1971; Popper & Fry, 2003; Liepman et al., 2007; Moller et al., 2007; Nothnagel & Nothnagel, 2007; Lee et al., 2011). Two cellulose synthase-like genes in *Physcomitrella* have been shown to encode proteins that have activity in mannan synthesis when expressed transgenically in other cells (Liepman et al., 2007). *Physcomitrella* cell walls also contain xylans but apparently lack the mixed-linkage β-glucans that are especially prominent in grasses (Popper & Fry, 2003). Overall, *Physcomitrella* cell walls have hemicelluloses that are both similar and different from those in higher plants, and many details of hemicellulose functions in forming the cross-linked matrix of the moss cell wall remain to be elucidated.

Pectin. As with hemicelluloses, the pectins of *Physcomitrella* cell walls exhibit both similarities and differences when compared to the pectins of higher plants. The homogalacturonan, β-1,4-galactan, α-1,5-arabinan, and RG-I of higher plants appear to be present in *Physcomitrella*, as judged by both chemical and immunological methods (Moller et al., 2007; Kulkarni et al., 2012). Although detectable in some mosses, liverworts, and hornworts, RG-II is present in these seedless, nonvascular plants at only very low levels, about 1% as abundant as in angiosperms (Matsunaga et al., 2004). Sequence analysis suggests that the *Physcomitrella* genome contains homologs of several pectin biosynthetic genes (Harholt et al., 2012; Roberts et al., 2012), but none of the polypeptides encoded by these genes have been confirmed to have the expected enzyme activity.

Glycoproteins. Several distinct approaches, including isolation and chemical analysis, staining and growth inhibition by Yariv phenylglycoside, antibody staining, and core protein gene knockout (Lee et al., 2005a; Fu et al., 2007; Moller et al., 2007) combine together to provide very strong evidence for the presence and function of AGPs, one of the two major subfamilies of HRGPs, in *Physcomitrella*. *Physcomitrella* AGPs are structurally very similar to angiosperm AGPs with the exception of the presence of up to 15 mol% of terminal 3-*O*-methyl-rhamnosyl residues, an unusual sugar that has not been reported in any angiosperm polymers (Fu et al., 2007). This relatively hydrophobic sugar residue considerably reduces the water solubility of the moss AGPs compared to angiosperm AGPs, although the biological significance of this change in polarity has yet to be established (Fu et al., 2007). It is potentially interesting to note, however, that an AGP was among the proteins found to be significantly upregulated during drought in a proteome analysis of dehydration stress in *Physcomitrella* (Cui et al., 2012).

Evidence for the presence of extensins, the other of the two major subfamilies of HRGPs, in *Physcomitrella* is less compelling. In a CoMPP study, Moller et al. (2007) found that just one of three different mAbs directed against angiosperm extensins had a detectable binding to *Physcomitrella* cell wall components, and that binding was relatively weak. Bioinformatics searches of the *Physcomitrella* genome have revealed sequences with homology to some angiosperm genes known to be involved in extensin glycosylation, but no sequences with homology to angiosperm genes encoding extensin polypeptides. As Roberts et al. (2012) noted, a more comprehensive analysis of cell wall proteins is needed to resolve the question of extensin presence in *Physcomitrella*.

Lignin. Moss, as represented by *Physcomitrella*, seems to occupy an interesting point in the evolution of the ability of plants to form lignin. Eight classes of enzymes are needed for the formation of lignin monomers, and, based on gene sequence homologies, it seems that moss is the earliest evolved plant to have all eight of these classes (Weng and Chapple, 2010). Nevertheless, differences of opinion exist as to whether mosses actually make polymeric lignin (Vanholme et al., 2010). Weng and Chapple (2010) argue that bryophytes do not synthesize insoluble polymeric lignin in their cell walls but instead use the lignin monomers as building blocks for lignans, which are soluble dimers in the protoplasm. Weng and Chapple (2010) further argue that mosses make lignans and other soluble phenylpropanoids such as flavonoids, not to strengthen the cell wall, but rather to accumulate them in the vacuoles of the cells to protect against damaging UV-B radiation from the direct sun to which early terrestrial plants were exposed when they moved out of their aquatic environment and onto the land. While *Physcomitrella* seems to have all eight of the enzymes required to make the three lignin monomer types H, G, and S, chemical analysis of lignin monomer composition has revealed that *Physcomitrella* contains only the H building blocks and none of the G or S blocks that carry one or two methyl ether groups (Espiñeira et al., 2011).

Other wall-related structures. Antibody binding in CoMPP analysis and histochemical staining with Aniline Blue both indicate the presence of callose in some tissues of *Physcomitrella* and other mosses (Scherp et al., 2001; Moller et al., 2007; Schuette et al., 2009). It is generally agreed (Mauseth, 2008) that most mosses have cuticle only on the upper surface of gametophore leaves, with the lower surface of the

leaves and other surfaces being without cuticle and thus capable of freely gaining water from, or losing water to, the environment. Wyatt et al. (2008) have reported histochemical staining of cuticle in *Physcomitrella*, and other authors have reported observations of cuticle on other moss gametophores and sporophytes (Cook and Graham, 1998; Budke et al., 2011).

METHYL SUBSTITUENTS ON SUGAR RESIDUES IN PLANT CELL WALL POLYSACCHARIDES

Two types of methyl substituents occur on sugar residues in plant cell wall polysaccharides. One type of methyl substituent is formed via ester linkage between a methyl group and a C6 carboxyl group of a uronic acid sugar. The most studied example of such a methyl ester is that formed at the C6 carboxyl group of galacturonosyl (GalU) residues in pectic polysaccharides. For pectic polysaccharides in lemon peel, 76% of the GalU residues are methyl esterified (Aspinall et al. 1968). The extent of esterification affects cell wall porosity, ion-exchange capacity, cell-cell adhesion, and various other aspects of pectin function (Carpita & McCann, 2000).

The second type of methyl substituent is formed via ether linkage between a methyl group and a secondary alcohol (on sugar carbons C2, C3, C4, or C5) or a primary alcohol (C5 on pentose or C6 on hexose). A variety of such methylated sugars have been found in plant cell walls. Most often, these methylated sugars are of low abundance and have been little studied. The main exceptions are 4-*O*-methyl-D-glucuronosyl (4-*O*-Me-D-GlcU) residues, which occur in xylan-backbone hemicelluloses in some plants,

particularly in secondary cell wall, and in certain plant gums. Enzymic activity of a methyltransferase that synthesizes 4-*O*-Me-D-GlcU was first detected long ago by Kauss and Hassid (1967) and subsequently studied by various other investigators. In an article published during the course of work on this dissertation, Urbanowicz et al. (2012) reported the first identification of a gene encoding a methyltransferase that forms a methyl ether on a plant cell wall polysaccharide. The *GXMT1* gene of *Arabidopsis* encodes a domain of unknown function family579 protein that methylates D-GlcU residues on a 1,4-linked β-D-xylan backbone hemicellulose to form 4-*O*-Me-D-GlcU residues. Approximately one in eight xylosyl residues have a side chain of a single D-GlcU residue, and about 75% of these D-GlcU residues are methylated to 4-*O*-Me-D-GlcU (Urbanowicz et al., 2012).

Some lower plants have been recently found to contain considerable amounts of other methylated sugars. Popper et al. (2001) reported that 3-*O*-methyl-D-galactose (3-*O*-Me-D-Gal) occurs at significant levels (5-10 mg per g of alcohol-insoluble residue) in homosporous and heterosporous lycophytes. Low levels of 3-*O*-Me-Gal residues were previously reported in polysaccharides from the green alga *Chlorella vulgaris* (Ogawa et al. 1994), the red alga *Porphyridium aerugineum* (Percival & Foyle, 1979), the leaves of various dicot trees (Bacon & Cheshire, 1971), the twigs of sassafras (Springer et al., 1965), and in a RG I-like pectic mucilage from the inner bark of the elm tree *Ulmus glabra* (Barsett & Paulsen 1992).

Some bacteria have been found to contain 3-*O*-Me-Rha residues in glycolipids. The cyanobacteria *Anabaena variabilis* (Weckesser et al., 1974) and *Spirulina platensis*

(Shekharam et al., 1987) have 3-*O*-Me-L-Rha in lipopolysaccharides and in polysaccharides, respectively. Residues of 3-*O*-Me-L-Rha have been found attached to *allo*-threonine and to the non-reducing end on alaninol in *Mycobacterium* mycosides (López-Marin et al. 1994). Various lower plants contain 3-*O*-Me-L-Rha in their cell wall polymers. Most significant among these, for the purposes of this dissertation, is the moss *Physcomitrella patens*. The AGPs of this moss have been found to contain 15 mole percent of 3-*O*-Me-L-Rha in non-reducing terminal positions on the glycan chains (Fu et al., 2007), which might represent the highest reported content of an *O*-methyl ether sugar in a plant cell wall polymer. Detailed studies of RG II revealed that 3-*O*-Me-Rha also occurs in this pectic domain in some seedless plants (Matsunaga et al., 2004). Some gymnosperms have been reported to contain 3-*O*-Me-Rha in gums (Vogt & Stephen, 1993), but reports of this sugar in angiosperms are very rare. The only reports in angiosperms are for *Acokanthera* and *Antiaris*, both of these genera producing latexes that contain steroidal glycosides having a single 3-*O*-Me-L-Rha residue as the carbohydrate component (Muhr et al., 1954). Thus, it seems that the ability to synthesize cell wall polymers containing 3-*O*-Me-Rha was lost during the evolution to angiosperms

METHYLATION

Most available evidence suggests that formation of methyl ethers on sugars occurs by the action of methyltransferases, enzymes that remove an activated methyl group from a donor and then add that methyl group to the sugar. Although several different methyl donors have been identified in Nature, the most common methyl donor, by far, is *S*-

adenosyl-methionine (AdoMet). The methyl group from AdoMet is added to sugars, but to many other substrates such as nucleic acids, lipids, proteins, lignin subunits, and various small metabolites as well (Cantoni, 1975; Luka et al., 2009). When the acceptor is an oxygen (as in a hydroxyl group on a sugar) and an *O*-methyl ether is formed, then the enzyme is called an *O*-methyltransferase. Similarly, when the acceptor is a nitrogen (as in phosphatidyl ethanolamine, a lipid), then the enzyme is called a *N*-methyltransferase. When the acceptor is a carbon (as in cytosine, a nucleotide base), then the enzyme is called a *C*-methyltransferase. The methyl donor AdoMet is believed to occur in all living organisms (Luka et al., 2009) and is formed by the joining of methionine and ATP by the enzyme methionine adenosyltransferase. When a methyltransferase removes the activated methyl from AdoMet, the remaining part of AdoMet becomes *S*-adenosyl-*L*-homocysteine (AdoHcy). This reaction product is still important because it is a strong inhibitor of the methyltransferase reaction. Thus, the methyltransferase reaction is regulated by product inhibition. Eventually, AdoHcy is recycled via conversion to homocysteine and then to back to methionine and finally AdoMet (Luka et al., 2009).

Among the substrates other than sugars that are methylated by methyltransferases, nucleic acids are currently the substrate of greatest research interest. This great interest arises because methylation of DNA, and related proteins such as histones, has strong effects on gene expression. Epigenetic gene regulation can occur through methylation of DNA by DNA methyltransferases (DNMTs). This methylation of DNA occurs principally on the fifth carbon of cytosine in the sequence CpG (Denis et al. 2011).

Methylation by DNMTs is important, as evidenced by various disorders that occur in mammals that have defects in DNMTs (Bird, 2002). One such disorder is chromosomal instability, since transposable elements can be silenced by DNA methylation (Goll & Bestor, 2005; Howard et al, 2008). Under- or over-methylation of DNA is also related to human cancer (Jones & Baylin, 2007; Portela & Esteller, 2010), and defects in some DNMTs are embryo-lethal in mice (Li et al., 1992). Regulation of gene expression can also occur through methylation of nucleosomal histones, basic proteins that interact with DNA. In histone H3, for example, di- or trimethylation of lysine 4 causes expression of the associated gene, whereas trimethylation of lysine 27 causes repression of the associated gene (Kouzarides 2007). Methylation of DNA in plants also affects many processes including development, growth, response to stress and genome stability (Zheng et al., 2008; Chinnusamy & Zhu, 2009).

METHYLTRANSFERASES

Because a wide variety of molecules are methylated in living organisms, many different methyltransferase enzymes are required. Thus, the genomes of most organisms contain many genes encoding many methyltransferases in the proteome. The human proteome is currently believed to contain 208 known proteins and 38 putative proteins that are AdoMet-dependent methyltransferases (Petrossian & Clarke, 2011). Even yeast, one of the simplest eukaryotes, has 81 known and putative AdoMet-dependent methyltransferases. These methyltransferases alone account for 1.2% of the yeast genome (Petrossian & Clarke, 2009).

Most authors considering methylation of sugars in the plant cell wall make note of evidence that the Golgi apparatus is the principal organelle that functions in biosynthesis of glycan chains of glycoproteins and essentially all cell wall polysaccharides, except cellulose (Driouich et al., 2012). Thus, most authors expect that methylation of sugars in cell wall polymers occurs in the Golgi apparatus. Ambitious studies to detect all Golgi-localized proteins in *Arabidopsis* have turned up many polypeptides that, by sequence homology, appear to be *O*-methyltransferases (Nikolovski et al., 2012; Parsons et al., 2012).

As noted above in this chapter, formation of methyl esters on galacturonosyl residues in pectic polysaccharides is one form of methylation that occurs in plants. *O*-Methyltransferases that form these pectin methyl esters have been characterized biochemically by enzyme activity and have, in fact, been localized in the Golgi apparatus in several species (Vannier et al., 1992; Goubet & Mohnen, 1998; Baydoun et al., 1999; Ishikawa 2000). Through mutational analyses, TSD2 and QUA2, Golgi-localized proteins with a putative methyltransferase domain, have been identified as very good candidates for a pectin methyltransferase in *Arabidopsis* (Krupkova et al., 2007; Mouille et al., 2007).

Other *O*-methyltransferases are those that act on small, usually phenolic metabolites. Such *O*-methyltransferases are important in both mammals and plants. In mammals, catechol-*O*-methyltransferase is involved in dopamine catabolism, and defects in this enzyme lead to various disorders (Hall et al., 2012; Kambur et al., 2010). In plants, many AdoMet-dependent *O*-methyltransferses of this type have been investigated in the

context of secondary metabolism (Ibrahim et al., 1998; Lu et al., 2010; Zorrilla-Fontanesi et al., 2012), particularly with regard to lignin biosynthesis (Weng & Chapple, 2010; Vanholme et al., 2010). Caffeoyl-CoA-*O*-methyltransferase (CCoAOMT) and caffeic acid *O*-methyltransferase (COMT) are two such lignin biosynthetic enzymes, their actions leading to formation of the G- and S-type lignin subunits. Downregulation of CCoAOMT and/or COMT activities or mutation of these enzymes leads to reduced G- and/or S-lignin content and overall reduced lignin content (Meyermans et al., 2000; Zhong et al., 2000; Guo et al., 2001; Pinçon et al., 2001), which in some cases results in easier wall digestion (Tu et al., 2010), a desirable feature for production of biofuels from lignocellulosic biomass.

At the start of this dissertation, no genes encoding *O*-methyltransferases that work on sugars in plant cell wall polysaccharides had yet been identified. Thus, to gain a starting point it was necessary to examine the literature on such *O*-methyltransferases in other organisms. Functionally close enzymes were found in mycobacteria, where AdoMet-dependent *O*-methyltransferases add methyl groups to a terminal nonreducing L-rhamnosyl residue on a glycopeptidolipid (GPL) (Patterson et al., 2000; Jeevarajah et al., 2002; Jeevarajah et al., 2004). Three rhamnosyl *O*-methyltransferases work sequentially to ultimately produce methylation on the 2-, 3-, and 4-positions of the rhamnosyl residue. The first methyl is added, however, at the 3-position to produce a 3-*O*-Me-L-Rha residue, the exact residue found in *Physcomitrella* AGPs. The *MTF1* gene (also called *RMT3*) of *Mycobacterium smegmatis* which encodes the *O*-methyltransferase transfers a methyl from AdoMet to the terminal nonreducing L-Rha on the

glycopeptidolipid to form this 3-*O*-Me-L-Rha has been cloned (Patterson et al. 2000) and provided a starting point for this dissertation. More generally, a variety *O*-methyltransferases that act on 6-deoxyhexoses in *Mycobacteria* and related organisms have been cloned and share a very high degree of homology in conserved domains (Patterson et al. 2000).

As described previously in this chapter, however, an article published during the course of work on this dissertation reported the first identification of a gene encoding a methyltransferase that forms a methyl ether on a plant cell wall polysaccharide (Urbanowicz et al., 2012). The *gxmt1* gene of *Arabidopsis* encodes a domain of unknown function family 579 protein that methylates D-GlcU residues on a 1,4-linked β-D-xylan backbone hemicellulose to form 4-*O*-Me-D-GlcU residues. Earlier biochemical work on several species resulted in characterization of glucuronosyl-4-O-methyltransferase enzyme activity (Baydoun et al. 1989; Vannier et al. 1992; Baydoun et al. 1999) and localization of this activity to the Golgi apparatus (Vannier et al. 1992; Baydoun et al. 1999). With the cloning of the *gxmt1* gene, transgenic expression of the protein became possible, and with that, further characterization of the enzyme activity. Assays with purified protein revealed that the enzyme was active only with polymeric glucuronoxylan as the substrate, i.e., transfer of the methyl from AdoMet to the 4-position of GlcU did not occur with either UDP-GlcU or GlcU as substrate. Furthermore, the transgenically expressed enzyme was unique in requiring Co^{2+} as a metallic cofactor, rather than Mg^{2+}, Ca^{2+}, or Zn^{2+} that are more common metallic cofactors.

GOALS AND ORGANIZATION OF THIS DISSERTATION

The goal of the overall project is to identify the gene encoding the rhamnosyl 3-*O*-methyltransferase enzyme responsible for the synthesis of 3-*O*-Me-Rha in AGPs of *Physcomitrella*. Within the context of the overall project, this dissertation has the goals of: (1) identifying *Physcomitrella* genes that are candidates to encode the rhamnosyl 3-*O*-methyltransferase, and (2) carrying out functional analyses of the most promising candidate genes to determine if one of those genes actually does encode the rhamnosyl 3-*O*-methyltransferase.

Chapter 2 in this dissertation identifies 11 candidate genes, describes the locations and structures of these genes, and reports on the generation and analyses of *Physcomitrella* knockout mutants of these 11 genes. Two candidate genes, *KO1* and *KO9*, were selected for further study due to statistically very significant reductions in the 3-*O*-Me-Rha/Rha content ratio in AGPs of the knockout mutants. A third candidate gene, *KO11*, was also chosen for further analysis because the KO11 protein is the ortholog of an *Arabidopsis* glucuronosyl 4-*O*-methyltransferase, AtGXMT1. The next three chapters of this dissertation report on further functional analyses of these three selected candidate genes.

Chapter 3 reports on functional analyses of *KO1*. After a search of the moss transcriptome, the *KO1* gene model was revised. Based on the revised gene model, *KO1* was surprisingly found to be an ortholog of the bacterial *lpxB* gene, which encodes lipid A disaccharide synthase. The *Physcomitrella ko1* knockout mutant exhibited abnormal polarized tip growth, including curly protonemal filaments and fewer and shorter lateral

rhizoids. Transgenic *KO1* tobacco was not able to make a detectable amount of 3-*O*-Me-Rha in the tobacco AGPs.

Chapter 4 reports on functional analyses of *KO9* in *Physcomitrella* and in tobacco. Because KO9 was found to be an ortholog of a lignin *O*-methyltransferase of higher plants, the lignin-like contents of *Phycomitrella ko9* and *ko1* knockout mutants and wild type were measured. The lignin-like content of *ko9* was not significantly different from the wild type, but surprisingly the lignin-like content of *ko1* was statistically very significantly less than that of the wild-type. Transgenic *KO9* tobacco was not able to make a detectable amount of 3-*O*-Me-Rha in the tobacco AGPs.

Chapter 5 reports on heterologous expression of *KO11* in tobacco. Although KO11 is the ortholog of AtGXMT1, which is a confirmed glucuronosyl 4-*O*-methyltransferase, no 3-*O*-Me-Rha was detected in the AGPs of transgenic *KO11* tobacco plants.

Chapter 6 presents a summary of the dissertation and an outlook towards future research in this field.

REFERENCES

Acosta-Garcia G, Vielle-Calzada JP (2004) A classical arabinogalactan protein is essential for the initiation of female gametogenesis in *Arabidopsis*. Plant Cell **16**: 2614-2628

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) The plant cell wall. In: B Alberts, A Johnson, J Lewis, M Raff, K Roberts, P Walter, eds, Molecular Biology of the Cell, Ed 4. Garland Science, New York, available from: www.ncbi.nlm.nih.gov/books/NBK26928/

Arioli T, Peng LC, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Höfte H, Plazinski J, Birch R, Cork A, Glover J, Redmond J, Willamson R (1998) Molecular analysis of cellulose biosynthesis in *Arabidopsis*. Science **279**: 717-720

Aspinall GO, Craig JWT, Whyte JL (1968) Lemon peel pectin. Part I. Fractionation and partial hydrolysis of water-soluble pectin. Carbohydr Res **7**: 442-452

Bacon JSD, Cheshire MV (1971) Apiose and mono-*O*-methyl sugars as minor constituents of the leaves of deciduous trees and various other species. Biochem J **124:** 555-562

Barsett H, Paulsen BS (1992) Separation, isolation and characterization of acidic polysaccharides from the inner bark of *Ulmus glabra* Huds. Carbohydr Polymers **17**: 137-144

Basile A, Giordano S, Lopez-Saez JA, Cobianchi RC (1999) Antibacterial activity of pure flavonoids isolated from mosses. Phytochemistry **52**: 1479-1482

Basile DV, Basile MR, Mignone MM (2000) Arabinogalactan-proteins, place-dependent suppression and plant morphogenesis. In: EA Nothnagel, A Bacic, AE Clarke, eds, Cell and Developmental Biology of Arabinogalactan-Proteins. Kluwer Academic/Plenum, New York, pp 169-178

Battle M, Bender ML, Tans PP,White JW, Ellis JT, Conway T, Francey RJ (2000) Global carbon sinks and their variability inferred from atmospheric O_2 and δ13C. Science **287**:2467–2470

Bauer WD, Talmadge, KW, Keegstra K, Albersheim P (1973) The structure of plant cell walls: II. The hemicellulose of the walls of suspension-cultured sycamore cells. Plant Physiol **51**: 174-187

Baydoun EAH, Usta JAR, Waldron KW, Brett CT (1989) A methyltransferase involved in the biosynthesis of 4-*O*-methylglucuronoxylan in etiolated pea epicotyls. J Plant Physiol **135**:81-85

Baydoun EAH, Rizk SE, Brett CT (1999) Localisation of methyltransfeases involved in glucuronoxylan and pectin methylation in the Golgi apparatus in etiolated pea epicotyls. J Plant Physiol **155**:240-244

Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev **16**: 6-21

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol **54**: 519-546

Borner GH, Sherrier DJ, Stevens TJ, Arkin IT, Dupree P (2002) Prediction of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A genomic analysis. Plant Physiol **129**: 486-499

Borner GH, Lilley KS, Stevens TJ, Dupree P (2003) Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A proteomic and genomic analysis. Plant Physiol **132**: 568-77

Budke JM, Goffinet B, Jones CS (2011) A hundred-year-old question: is the moss calyptra covered by a cuticle? A case study of *Funaria hygrometrica*. Ann Bot **107**: 1279-1286

Burke D, Kaufman P, McNeil M, Albersheim P (1974) The structure of plant cell walls. Plant Physiol **54**: 109-115

Cantoni GL (1975) Biological methylation: selected aspects. Annu Rev Biochem **44**: 435-451

Carpita NC, McCann M (2000) The cell wall. In: BB Buchanan, W Gruissem, RL Jones, eds, Biochemistry and Molecular Biology of Plants. Amer Soc Plant Physiologists, Rockville, pp 52-108

Chanliaud E, Burrows KM, Jeronimidis G, Gidley MJ (2002) Mechanical properties of primary plant cell wall analogues. Planta **215**: 989-996

Chapman A, Blervacq AS, Vasseur J, Hilbert JL (2000) Arabinogalactan-proteins in *Cichorium* somatic embryogenesis: effect of beta-glucosyl Yariv reagent and epitope localisation during embryo development. Planta **2211**: 305-314

Chaves I, Regalado AP, Chen M, Ricardo CP, Showalter AM (2002) Programmed cell death induced by (β-D-galactosyl)(3) Yariv reagent in *Nicotiana tabacum* BY-2 suspension-cultured cells. Physiol Plant **116**: 548-553

Chen CG, Pu ZY, Moritz RL, Simpson RJ, Bacic A, Clarke AE, Mau SL (1994) Molecular cloning of a gene encoding an arabinogalactan-protein from pear (*Pyrus communis*) cell suspension culture. Proc Natl Acad Sci USA **91**: 10305-10309

Chinnusamy V, Zhu JK (2009) Epigenetic regulation of stress responses in plants. Curr Opin Plant Biol **12**: 133-139

Cho SH, Chung YS, Cho SK, Rim YW, Shin JS (1999) Particle bombardment mediated transformation and GFP expression in the moss *Physcomitrella patens*. Mol Cells **9**: 14-19

Chun JP, Huber D (1998) Polygalaturonase-mediated solubilization and depolymerization of pectic polymers in tomato fruit cell walls: Regulation by pH and ionic conditions. Plant Physio **117**: 1293-1299

Cook ME, Graham LE (1998). Structural similarities between surface layers of selected charophycean algae and bryophytes and the cuticles of vascular plants. Int J Plant Sci **159**: 780-787

Courtice GRM, Cove DJ (1983) Mutants of the moss *Physcomitrella patens* which produce leaves of altered morphology. J Bryol **12**: 595-609

Cui S, Hu J, Guo S, Wang J, Cheng Y, Dang X, Wu L, He Y (2012) Proteome analysis of *Physcomitrella patens* exposed to progressive dehydration and rehydration. J Exp Bot **63**: 711-726

Darvill A, McNeil M, Albersheim P (1978) Structure of plant cell walls: VIII. A new pectic polysaccharide. Plant Physiol **62**: 418-422

Darvill JE, McNeil M, Darvill AG, Albersheim P (1980) Structure of plant cell walls: XI. Glucuronoarabinoxylan, a second hemicellulose in the primary cell walls of suspension-cultured sycamore cells. Plant Physiol. **66**: 1135-1139

Denis H, Ndlovu MN, Fuks F (2011) Regulation of mammalian DNA methyltransferases: a route to new mechanisms. EMBO Reports **12**: 647-656

Driouich A, Follet-Gueye M, Bernard S, Kousar S, Chevalier L, Vicré-Gibouin M, Lerouxe O (2012) Golgi-mediated synthesis and secretion of matrix polysaccharides of the primary cell wall of higher plants. Front Plant Sci **3**: 79

Du H, Simpson RJ, Moritz RL, Clarke AE, Bacic A (1994) Isolation of the protein backbone of an arabinogalactan-protein from the styles of *Nicotiana alata* and characterization of a corresponding cDNA. Plant Cell **6**:1643-1653

Ebringerova A, Heinze T (1999) Xylan and xylan derivatives – biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties. Macromol Rapid Commun **21**: 542-556

Ellis M, Egelund J, Schultz CJ, Bacic A (2010) Arabinogalactan-proteins: key regulators at the cell surface? Plant Physiol **153**: 403-419

Engel PP (1968) The induction of biochemical and morphological mutants in the moss *Physcomitrella patens*. Am J Bot **55**: 438-446

Espineira JM, Novo Uzal E, Gomez Ros LV, Carrion JS, Merino F, Ros Barcel A, Pomar F (2011). Distribution of lignin monomers and the evolution of lignification among lower plants. Plant Biol **13**: 59-68

Fagard M, Desnos T, Desprez T, Goubet F, Refregier G, Mouille G, McCann M, Rayon C, Vernhettes S, Hofte H (2000) *PROCUSTE1* encodes a cellulose synthase required for normal cell elongation specifically in roots and dark-grown hypocotyls of *Arabidopsis*. Plant Cell **12**: 2409-2424

Fincher GB, Stone BA, Clarke AE (1983) Arabinogalactan-proteins: structure, biosynthesis, and function. Annu Rev Plant Physiol **34**: 47-70

Fu H, Yadav MP, Nothnagel EA (2007) *Physcomitrella patens* arabinogalactan proteins contain abundant terminal 3-*O*-methyl-L-rhamnosyl residues not found in angiosperms. Planta **226**: 1511-1524

Gane AM, Craik D, Munro SLA, Howlett GJ, Clarke AE, Bacic A (1995) Structural analysis of the carbohydrate moiety of arabinogalactan-proteins from stigmas and styles of *Nicotiana alata*. Carbohydr Res **277**: 67-85

Gao M, Showalter AM (1999) Yariv reagent treatment induces programmed cell death in *Arabidopsis* cell cultures and implicates arabinogalactan protein involvement. Plant J **19**: 321-31

Gaspar YM, Nam J, Schultz CJ, Lee L-Y, Gilson PR, Gelvin SB, Bacic A (2004) Characterization of the *Arabidopsis* lysine-rich arabinogalactan-protein AtAGP17 mutant (rat1) that results in a decreased efficiency of *Agrobacterium* transformation. Plant Physiol **135**: 2162-2171

Geddes DS, Wilkie KCB (1971) Hemicelluloses from the stem tissues of the aquatic moss *Fontinalis antipyretica*. Carbohydr Res **18**: 333-335

Gibeaut DM, Pauly M, Bacic A, Fincher GB (2005) Changes in cell wall polysaccharides in developing barley (*Hordeum vulgare)* coleoptiles. Planta **390:** 105-113

Girke T, Schmidt H, Zahringer U, Reski R, Heinz E (1998) Identification of a novel delta 6-acyl-group desaturase by targeted gene disruption in *Physcomitrella patens*. Plant J **15**: 39-48

Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. Annu Rev Biochem **74**: 481-514

Goodrum LJ, Patel A, Leykam JF, Kieliszewski MJ (2000) Gum arabic glycoprotein contains glycomodules of both extension and arabinogalactan-glycoproteins. Phytochemistry **54:** 99-106

Goss CA, Brockmann DJ, Bushoven JT, Roberts AW (2012) A *CELLULOSE SYNTHASE (CESA)* gene essential for gametophore morphogenesis in the moss *Physcomitrella patens*. Planta **235**: 1355-1367

Goubet F, Mohnen D (1998) Solubilization and partial characterization of homogalacturonan-methyltransferase from microsomal membranes of suspension-cultured tobacco cells. Plant Physiol **116**: 337-347

Grimsley NH, Ashton NW, Cove DJ (1977) The production of somatic hybrids by protoplast fusion in the moss, *Physcomitrella patens*. Mol Gen Genet **154**: 97-100

Guan Y, Nothnagel EA (2004) Binding of arabinogalactan proteins by Yariv phenylglycoside triggers wound-like responses in *Arabidopsis* cell cultures. Plant Physiol **135**: 1346-1366

Guo D, Chen F, Inoue K, Blount JW, Dixon RA (2001) Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa impacts on lignin structure and implications for the biosynthesis of G and S lignin. Plant Cell **13**: 73-88

Hall KT, Lembo AJ, Kirsch I, Ziogas DC, Douaiher J, Jensen KB, Conboy LA, Kelley JM, Kokkotou E, Kaptchuk TJ (2012) Catechol-*O*-methyltransferase val158met polymorphism predicts placebo effect in irritable bowel syndrome. PLoS One **7**: e48135

Harholt J, Sorensen I, Fangel J, Roberts A, Willats WGT, Scheller HV, Petersen BL, Banks JA, Ulvskov P (2012) The glycosyltransferase repertoire of the spikemoss *Selaginella moellendorffii* and a comparative study of its cell wall. PLoS ONE **7**: e35846

Hart DA, Kindel PK (1970) Isolation and partial characterization of apiogalacturonans from the cell wall of *Lemna minor*. Biochem J **116**: 569-579

Hayashi T, Marsden MPF, Delmer DP (1987) Pea xyloglucan and cellulose: VI. Xyloglucan-cellulose interactions *in vitro* and *in vivo*. Plant Physiol **83**: 384-389

Hématy K, Sado PE, Van Tuinen A, Rochange S, Desnos T, Balzergue S, Pelletier S, Renou JP, Höfte H (2007) A receptor-like kinase mediates the response of *Arabidopsis* cells to the inhibition of cellulose synthesis. Curr Biol **17**: 922-931

Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A (2008) Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. Oncogene **27**: 404-408

Hu Y, Qin Y, Zhao J (2006) Localization of an arabinogalactan protein epitope and the effects of Yariv phenylglycoside during zygotic embryo development of *Arabidopsis thaliana*. Protoplasma. **229**: 21-31

Huisman MH, Brull LP, Thomas-Oates JE, Haverkamp J, Schols HA, Voragen AJ (2001) The occurrence of internal (1 --> 5)-linked arabinofuranose and arabinopyranose residues in arabinogalactan side chains from soybean pectic substances. Carbohydr Res **330**: 103-114

Ibrahim RK, Bruneau A, Bantignies B (1998) Plant *O*-methyltransferases: molecular analysis, common signature and classification. Plant Mol Biol **36**: 1-10

Ishii T, Matsunaga T (1996) Isolation and characterization of a boron-rhamnogalactan II complex from sugar beet pulp. Carbohydr Res **284**: 1-9

Ishikawa M, Kuroyama H, Takeuchi Y, Tsumuraya Y (2000) Characterization of pectin methyltransferase from soybean hypocotyls. Planta **210**: 782-791

Jauh GY, Lord EM (1996) Localization of pectins and arabinogalactan-proteins in lily (*Lilium longiflorum L*) pollen tube and style, and their possible roles in pollination. Planta **199**: 251-261

Jeevarajah D, Patterson JH, McConville MJ, Billman-Jacobe H (2002) Modification of glycopeptidolipids by an *O*-methyltransferase of *Mycobacterium smegmatis*. Microbiology **148**: 3079-3087

Jeevarajah D, Patterson JH, Taig E, Sargeant T, McConville MJ, Billman-Jacobe H (2004) Methylation of GPLs in *Mycobacterium smegmatis* and *Mycobacterium avium*. J Bacteriol **186**: 6792-6799

Jones PA, Baylin SB (2007) The epigenomics of cancer. Cell **128**: 683-692

Kambur O, Talka R, Ansah OB, Kontinen VK, Pertovaara A, Kalso E, PT Mannisto (2010) Inhibitors of catechol-O-methyltransferase sensitize mice to pain. Br J Pharmacol **161**: 1553-1565

Kauss H, Hassid WZ (1967) Biosynthesis of the 4-O-methyl-D-glucuronic acid unit of hemicellulose B by transmethylation from S-adenosyl-L-methionine. J Biol Chem **242**:1680-1684

Keegstra K, Talmadge KW, Bauer WD, Albersheim P (1973) The structure of plant cell walls: III. A model of the walls of suspension-cultured sycamor*e* cells based on the interconnections of the macromolecular components. Plant Physiol **51**: 188-197

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

Kieliszewski MJ, Lamport DTA (1994) Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny. Plant J **5**: 157-172

Kouzarides T (2007) Chromatin modifications and their function. Cell **128**: 693-705

Kremer C, Pettolino F, Bacic A, Drinnan A (2004) Distribution of cell wall components in *Sphagnum* hyaline cells and in liverwort and hornwort elaters. Planta **219**: 1023-1035

Krupkova E, Immerzeel P, Pauly M, Schmülling T (2007) The *TUMOROUS SHOOT DEVELOPMENT2* gene of *Arabidopsis* encoding a putative methyltransferase is required for cell adhesion and co-ordinated plant development. Plant J **50**:735-750

Kulkarni AR, Pena MJ, Avci U, Mazumder K, Urbanowicz BR, Pattathil S, Yin Y, O'Neill MA, Roberts AW. Hahn MG, Xu Y, Darvill AG, York WS (2012) The ability of land plants to synthesize glucuronoxylans predates the evolution of tracheophytes. Glycobiology **22**: 439-451

Langan KJ, Nothnagel EA (1997) Cell surface arabinogalactan-proteins and their relation to cell proliferation and viability. Protoplasma **196**: 87-98

Lange BM, Lapierre C, Sandermann H (1995) Elicitor induced spruce stress lignin. Structural similarity to early developmental lignins. Plant Physiol **108**: 1277-1287

Lee KJD. Marcus SE, Knox JP (2011) Cell wall biology: perspectives from cell wall imaging. Mol. Plant **4**: 212-219

Lee, KJD, Knight CD, Knox JP (2005a) *Physcomitrella patens*: a moss system for the study of plant cell walls. Plant Biosyst **139**: 16-19

Lee KJD, Sakata Y, Mau SL, Pettolino F, Bacic A, Quatrano RS, Knight CD, Knox JP (2005b) Arabinogalactan proteins are required for apical cell extension in the moss *Physcomitrella patens*. Plant Cell **17**: 3051-3065

Leonard R, Petersen BO, Himly M, Kaar W, Wopfner N, Kolarich D, van Ree R, Ebner C, Duus JØ, Ferreira F, Altmann F (2005) Two novel types of O-glycans on the mugwort pollen allergen Art v 1 and their role in antibody binding. J Biol Chem **280**: 7932-7940

Li E, Bestor TH, Jaenisch R (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell **69**: 915-926

Liepman AH, Nairn CJ, Willats WGT, Sorensen I, Roberts AW, Keegstra K (2007) Functional genomic analysis supports conservation of function among cellulose synthase-like A gene family members and suggests diverse roles of mannans in plants. Plant Physiol **143**: 1881-1893

Liners F, Letesson JJ, Didembourg C, Van Cutsem P (1989) Monoclonal antibodies against pectin: recognition of a conformation induced by calcium. Plant Physiol **91**: 1419-1424

Lopez-Marin LM, Gautier N, Laneelle M-A, Silve G, Daffe M (1994) Structures of the glycopeptidolipid antigens of *Mycobacterium abscessus* and *Mycobacterium chelonae* and possible chemical basis of the serological cross-reactions in the *Mycobacterium fortuitum* complex. Microbiology **140**: 1109-1118

Lu F, Marita JM, Lapierre C, Jouanin L, Morreel K, Boerjan W, Ralph J (2010) Sequencing around 5-hydroxyconiferyl alcohol-derived units in caffeic acid *O*-methyltransferase-deficient poplar lignins. Plant Physiol **153**: 569-579

Luka Z, Mudd SH, Wagner C (2009) Glycine N-methyltransferase and regulation of S-adenosylmethionine levels. J Biol Chem **284**: 22507-22511

Luonteri E, Laine C, Uusitalo S, Teleman A, Siika-aho M, Tenkanen M (2003) Purification and characterization of *Aspergillus* β -d-galactanases acting on β -1,4- and β -1,3/6-linked arabinogalactans. Carbohydr Polym **53**: 155-168

Mashiguchi K, Yamaguchi I, Suzuki Y (2004) Isolation and identification of glycosylphosphatidylinositol-anchored arabinogalactan proteins and novel β-glucosyl Yariv-reactive proteins from seeds of rice (*Oryza sativa*). Plant Cell Physiol **45**: 1817-1829

Matsunaga T, Ishii T, Matsumoto S, Higuchi M, Darvill A, Albersheim P, O'Neill, MA (2004) Occurrence of the primary cell wall polysaccharide rhamnogalacturonan II in pteridophytes, lycophytes, and bryophytes. Implications for the evolution of vascular plants. Plant Physiol **134**: 339-351

Mauseth J.D. (2008) Botany, An Introduction to Plant Biology (4th edition). Jones & Bartlett Learning, Burlington, MA

Meyermans H, Morreel K, Lapierre C, Pollet B, De Bruyn A, Busson R, Herdewijn P, Devreese B, Van Beeumen J, Marita JM, Ralph J, Chen C, Burggraeve B, Van Montagu M, Messens E, Boerjan W (2000) Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A O-methyltransferase, an enzyme involved in lignin biosynthesis. J Biol Chem **275**: 36899-36909

Mollard A, Joseleau JP, Paul J (1994) *Acacia senegal* cells cultured in suspension secrete a hydroxyproline-deficient arabinogalactan-protein. Plant Physiol Biochem **32**: 703-709

Moller I, Sorensen I, Bernal AJ, Blaukopf C, Lee K, Obro J, Pettolino F, Roberts A, Mikkelsen JD, Knox JP, Bacic A, Willats WGT (2007). High-throughput mapping of cell-wall polymers within and between plants using novel microarrays. Plant J **50**: 1118-1128

Motose H, Sugiyama M, Fukuda H (2004) A proteoglycan mediates inductive interaction during plant vascular development. Nature **429**: 873-878

Mouille G, Ralet MC, Cavelier C, Eland C, Effroy D, Hématy K, McCartney L, Truong HN, Gaudon V, Thibault JF, Marchant A, Höfte H (2007) Homogalacturonan synthesis in *Arabidopsis thaliana* requires a Golgi-localized protein with a putative methyltransferase domain. Plant J **50**: 605-614

Muda P, Seymour GB, Errington N, Tucker GA (1995) Compositional changes in cell wall polymers during mango fruit ripening. Carbohydr Polym **26**: 255-260

Muhr H, Hunger A, Reichstein T (1954) Die Glykosider der Samen von *Acokanthera friesiorum* Markgr. Glykoside und Aglykone. Helv Chim Acta **37**: 403-427

Nam J, Mysore KS, Zheng C, Knue MK, Matthysse AG, Gelvin SB (1999) Identification of T-DNA tagged arabidopsis mutants that are resistant to transformation by *Agrobacterium*. Mol Gen Genet **261**: 429-438

Nakamura A, Furuta H, Maeda H, Takao T, Nagamatsu Y (2002) Structural studies by stepwise enzymatic degradation of the main backbone of soybean soluble polysaccharides consisting of galacturonan and rhamnogalacturonan. Biosci Biotechnol Biochem **66**: 1301-1313

Nikolovski N, Rubtsov D, Segura MP, Miles GP, Stevens TJ, Dunkley TPJ, Munro S, Lilley KS, Dupree P (2012) Putative glycosyltransferases and other plant Golgi apparatus proteins are revealed by LOPIT proteomics. Plant Physiol **160**: 1037-1051

Nothnagel EA (1997) Proteoglycans and related components in plant cells. Int Rev Cytol **174**: 195-191

Nothnagel EA, Bacic A, Clarke AE, eds (2000) Cell and Developmental Biology of Arabinoglactan-Proteins. New York: Kluwer Acad./Plenum

Nothnagel AL, Nothnagel EA (2007) Primary cell wall structure in the evolution of land plants. J Integr Plant Biol **49**: 1271-1278

Ogawa K, Yamaura M, Maruyama I (1994) Isolation and identification of 3-*O*-methyl-D-galactose as a constituent of a neutral polysaccharide of *Chlorella vulgaris*. Biosci Biotech Biochem **58**: 942-944

Oxley D, Bacic A (1999) Structure of the glycosylphosphatidylinositol anchor of an arabinogalactan protein from *Pyrus communis* suspension-cultured cells. Proc Natl Acad Sci USA **96**: 14246-14251

Ovodov YS, Ovodova RG, Bondarenko OD, Krasikova IN (1971) The pectic substances of *zosteraceae* : Part IV. Pectinase digestion of zosterine. Carbohydr Res **18**: 311-318

Park MH, Suzuki Y, Chono M, Knox JP, Yamaguchi I (2003) CsAGP1, a gibberellin-responsive gene from cucumber hypocotyls, encodes a classical arabinogalactan protein and is involved in stem elongation. Plant Physiol **131**: 1450-1459

Parsons HT, Christiansen K, Knierim B, Carroll A, Ito J, Batth TS, Smith-Moritz AM, Morrison S, McInerney P, Hadi MZ, Auer M, Mukhopadhyay A, Petzold CJ, Scheller HV, Loqué D, Heazlewood JL (2012) Isolation and proteomic characterization of the *Arabidopsis* Golgi defines functional and novel components involved in plant cell wall biosynthesis. Plant Physiol **159**: 12-26

Patterson J, McConville M, Haites R, Coppel R, Billman-Jacobe H (2000) Identification of a methyltransferase from *Mycobacterium smegmatis* involved in glycopeptidolipid synthesis. J Biol Chem **275**: 24900-24906

Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM (1996) Higher plants contain homologs of the bacterial *CelA* genes encoding the catalytic subunit of cellulose synthase. Proc Natl Acad Sci USA **93**: 12637-12642

Pena MJ, Darvill AG, Eberhard S, York WS, O'Neill MA (2008) Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants. Glycobiology **18**: 891-904

Percival E, Foyle RAJ (1979) The extracellular polysaccharides of *Porphyridium cruentum* and *Porphyridium aerugineum*. Carbohydr Res **72**: 165- 176

Petrossian TC, Clarke SG (2011) Uncovering the human methyltransferasome. Mol Cell Proteomics **10**: M110 000976

Petrossian T, Clarke S (2009) Bioinformatic identification of novel methyltransferases. Epigenomics **1**: 163-175

Pincon G, Maury S, Hoffmann L, Geoffroy P, Lapierre C, Pollet B, Legrand M (2001) Repression of *O*-methyltransferase genes in transgenic tobacco affects lignin synthesis and plant growth. Phytochemistry **57**: 1167-1176

Popper ZA, Sadler IH, Fry SC (2001) 3-*O*-methyl-D-galactose residues in lycophyte primary cell walls. Phytochemistry **57**: 711-719

Popper ZA, Fry SC (2003) Primary cell wall composition of bryophytes and charophytes. Ann Bot **91**: 1-12

Portela A, Esteller M (2010) Epigenetic modifications and human disease. Nat Biotechnol **28**: 1057-1068

Qi W, Fong C, Lamport D (1991) Gum arabic glycoprotein is a twisted hairy rope: a new model based on *O*-galactosylhydroxyproline as the polysaccharide attachment site. Plant Physiol **96**: 848-855

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-I T, 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, Cuming 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

Reski R (1998) Development genetics and molecular biology of mosses. Bot Acta **111**: 1-15

Reski R, Reynolds S, Wehe M, Kleber-Janke T, Kruse S (1998) Moss (*Physcomitrella patens*) expressed sequence tags include several sequences which are novel for plants. Bot Acta **111**: 145-151

Richmond TA, Somerville CR (2000) The cellulose synthase superfamily. Plant Physiol **124**: 495-498

Ridley BL, O'Neill MA, Mohnen D (2001) Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. Phytochemistry **57**: 929-967

Roberts AW, Bushoven JT (2007) The cellulose synthase (*CESA*) gene superfamily of the moss *Physcomitrella patens*. Plant Mol Biol **63**: 207-219

Roberts AW, Roberts EM, Haigler CH (2012) Moss cell walls: structure and biosynthesis. Front Plant Sci **3**: article 166

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

Scherp P, Grotha R, Kutschera U (2001) Occurrence and phylogenetic significance of cytokinesis-related callose in green algae, bryophytes, ferns and seed plants. Plant Cell Rep **20**: 143-149

Schols HA, Posthumus MA, Voragen AGJ (1990) Structural features of hairy regions of pectins isolated from apple juice produced by the liquefaction process. Carbohydr Res **206**: 117–129

Schuette S, Wood AJ, Geisler M, Geisler-Lee J, Ligrone R, Renzaglia KS (2009) Novel localization of callose in the spores of *Physcomitrella patens* and phylogenomics of the callose synthase gene family. Ann Bot **103**: 749-756

Schultz CJ, Ferguson KL, Lahnstein J, Bacic A (2004) Post-translational modifications of arabinogalactan-peptides of *Arabidopsis thaliana*. Endoplasmic reticulum and glycosylphosphatidylinositol-anchor signal cleavage sites and hydroxylation of proline. J Biol Chem **279**: 45503-45511

Schultz CJ, Johnson KL, Currie G, Bacic A (2000) The classical arabinogalactan protein gene family of arabidopsis. Plant Cell **12**: 1751-1768

Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YM, Bacic A (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. Plant Physiol **129**: 1448-1463

Seifert GJ, Roberts K (2007) The biology of arabinogalactan proteins. Annu Rev Plant Biol **58**: 137-161

Selvendran RR, ONeill MA (1987) Isolation and analysis of cell walls from plant material. Methods Biochem Anal **32**: 25-153

Serpe MD, Nothnagel EA (1994) Effects of Yariv phenylglycosides on rosa cell-suspensions-evidence for the involvement of arabinogalactan-proteins in cell-proliferation. Planta **193**: 542-550

Serpe MD, Nothnagel EA (1999) Arabinogalactan-proteins in the multiple domains of the plant cell surface. Adv Bot Res 30: 207-289

Seymour GB, Colquhoun IJ, DuPont MS, Parsley KR, Selvendran RR (1990) Composition and structural features of cell wall polysaccharides from tomato fruits. Phytochemistry **29**: 725-731

Shekharam KM, Venkataraman LV, Salimath PV (1987) Carbohydrate composition and characterization of two unusual sugars from the blue-green alga *Spirulina platensis*. Phytochemistry **26**: 2267-2269

Shi H, Kim YS, Guo Y, Stevenson B, Zhu JK (2003) The *Arabidopsis* SOS5 locus encodes a putative cell surface adhesion protein and is required for normal cell expansion. Plant Cell **15**: 19-32

Shibaya T, Sugawara Y (2009) Induction of multinucleation by beta-glucosyl Yariv reagent in regenerated cells from *Marchantia polymorpha* protoplasts and involvement of arabinogalactan proteins in cell plate formation. Planta **230**: 581-588

Shimizu M, Igasaki T, Yamada M, Yuasa K, Hasegawa J, Kato T, Tsukagoshi H, Nakamura K, Fukuda H, Matsuoka K (2005) Experimental determination of proline hydroxylation and hydroxyproline arabinogalactosylation motifs in secretory proteins. Plant J **42**: 877-889

Shpak E, Barbar E, Leykam JF, Kieliszewski MJ (2001) Contiguous hydroxyproline residues direct hydroxyproline arabinosylation in *Nicotiana tabacum*. J Biol Chem **276**: 11272-11278

Shpak E, Leykam JF, Kieliszewski MJ (1999) Synthetic genes for glycoprotein design and the elucidation of hydroxyproline-O-glycosylation codes. Proc Natl Acad Sci USA **96**: 14736-14741

Somerville C (2006) Cellulose synthesis in higher plants. Annu Rev Cell Dev Biol **22**: 53-78

Springer GF, Takahashi T, Desai PR, Kolecki BJ (1965) Cross reactive human blood-group H(O) specific polysaccharide from *Sassafras albidum* and characterization of its hapten. Biochemistry **4**: 2099-2112

Sun W, Kieliszewski MJ, Showalter AM (2004) Overexpression of tomato LeAGP-1 arabinogalactan-protein promotes lateral branching and hampers reproductive development. Plant J **40**: 870-881

Svetek J, Yadav MP, Nothnagel EA (1999) Presence of a glycosylphosphatidylinositol lipid anchor on rose arabinogalactan proteins. J Biol Chem **274**: 14724-14733

Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR (2003) Interactions among three distinct CesA proteins essential for cellulose synthesis. Proc Natl Acad Sci USA **100**: 1450-1455

Taylor NG, Laurie S, Turner SR (2000) Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in *Arabidopsis*. Plant Cell **12**: 2529-2540

Thomas JR, McNeil M, Darvill AG, Albersheim P (1987) Structure of plant cell walls: XIX. Isolation and characterization of wall polysaccharides from suspension-cultured Douglas fir cells. Plant Physiol **83**: 659-671

Tu Y, Rochfort S, Liu Z, Ran Y, Griffith M, Badenhorst P, G. V. Louie, M. E. Bowman, K. F. Smith, J. P. Noel, A. Mouradov, and G. Spangenberga (2010) Functional analysis of caffeic acid *O*-methyltransferase and cinnamoyl-CoA-reductase genes from perennial ryegrass (*Lolium perenne*). Plant Cell **22**: 3357-3373

Umezawa T (2003) Diversity in lignan biosynthesis. Phytochemistry Reviews **2**: 371-390

Underwood W (2012) The plant cell wall: A dynamic barrier against pathogen invasion. Front Plant Sci **3**: article 85

Urbanowicz BR, Pena MJ, Ratnaparkhe S, Avci U, Backe J, Street HF, Foston M, Li H, O'Neill MA, Ragauskas AJ, Darvill AG, Wyman C, Gilbert HJ, York WS (2012) 4-*O*-methylation of glucuronic acid in *Arabidopsis* glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. Proc Natl Acad Sci USA **109**: 14253-14258

Valent BS, Albersheim P (1974) The structure of plant cell walls: V. On the binding of xyloglucan to cellulose fibers. Plant Physiol **54**: 105-108

Van Hengel AJ, Roberts K (2003) AtAGP30, an arabinogalactan-protein in the cell walls of the primary root, plays a role in root regeneration and seed germination. Plant J **36**: 256-270

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol **153**: 895-905

Vannier MP, Thoiron B, Morvan C, Demarty M (1992) Localization of methyltransferase activities throughout the endomembrane system of flax (*Linum usitatissimum* L) hypocotyls. Biochem J **286**: 863-868

Vinkx CJA, Stevens I, Gruppen H, Grobet PJ, Delcour JA (1995) Physcochemical and functional properties of rye nonstarch polysaccharide. VI. Variability in the structure of water-unextractable arabinoxylans. Cereal Chem **72**: 411-418

Vogt DC, Stephen AM (1993) The gum exudate of *Encephelartos longifolius* Lehm. (female): further hydrolytic studies. Carbohydr Res **238**: 249-260

Weckesser J, Katz A, Drews G, Mayer H, Fromme I (1974) Lipopolysaccharide containing L-acofriose in the filamentous blue-green algae *Anabaena variabilis*. J Bacteriol **120**: 672-678

Weng JK, Li X, Stout J, Chapple C (2008) Independent origins of syringyl lignin in vascular plants. Proc Natl Acad Sci USA **105**: 7887-7892

Weng JK, Chapple C (2010) The origin and evolution of lignin biosynthesis. New Phytol **187**: 273-285

Willats WGT, McCartney L, Knox JP (2001) In-situ analysis of pectic polysaccharides in seed mucilage and at the root surface of *Arabidopsis thaliana.* Planta **213**: 37-44

Whitney SEC, Wilson E, Webster J, Bacic A, Grant Reid JS, Gidley MJ (2006) Effects of structural variation in xyloglucan polymers on interactions with bacterial cellulose. Am J Bot **93**: 1402-1414

Wise HZ, Saxena IM, Brown RM Jr (2011) Isolation and characterization of the cellulose synthase genes *PpCesA6* and *PpCesA7* in *Physcomitrella patens*. Cellulose **18**: 371-384

Wyatt HDM, Ashton NW, Dahms TES (2008) Cell wall architecture of *Physcomitrella patens* is revealed by atomic force microscopy. Botany **86**: 385-397

Xu Z, Zhang D, Hu J, Zhou X, Ye X, Reichel KL, Stewart NR, Syrenne RD, Yang X, Gao P, Shi W, Doeppke C, Sykes RW, Burris JN, Bozell JJ, Cheng ZM, Hayes DG, Labbe N, Davis M, Stewart CN Jr, Yuan JS (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. BMC Bioinformatics **10**: S3

Yin Y, Huang J, and Xu Y (2009). The cellulose synthase superfamily in fully sequenced plants and algae. BMC Plant Biol **9:** 99. doi: 10.1186/1471-2229-9-99

Yoo SH, Fishman ML, Savary BJ, Hotchkiss AT (2003) Monovalent salt-induced gelation of enzymatically deesterified pectin. J Agric Food Chem **51**: 7410-7417

Youl JJ, Bacic A, Oxley D (1998) Arabinogalactan-proteins from *Nicotiana alata* and *Pyrus communis* contain glycosylphosphatidylinositol membrane anchors. Proc Natl Acad Sci USA **95**: 7921-7926

York WS, Darvill AG, McNeil M, Albersheim P (1985) 3-deoxy-D-manno-2-octulosonic acid (KDO) is a component of rhamnogalacturonan II, a pectic polysaccharide in the primary cell walls of plants. Carbohydr Res 138: 109-126

Zablackis E, Huang J, Muller B, Darvill AG, Albersheim P (1995) Characterization of the cell-wall polysaccharides of *Arabidopsis thaliana* leaves. Plant Physiol **107**: 1129-1138

Zhao ZD, Tan L, Showalter AM, Lamport DTA, Kieliszewski MJ (2002) Tomato LeAGP-1 arabinogalactan-protein purified from transgenic tobacco corroborates the Hyp contiguity hypothesis. Plant J **31**: 431-444

Zheng X, Pontes O, Zhu J, Miki D, Zhang F, Li WX, Iida K, Kapoor A, Pikaard CS, Zhu JK (2008) ROS3 is an RNA-binding protein required for DNA demethylation in *Arabidopsis*. Nature **455**: 1259-1262

Zhong R, Morrison WH, Himmelsbach DS, Poole FL, Ye ZH (2000) Essential role of caffeoyl coenzyme A *O*-methyltransferase in lignin biosynthesis in woody poplar plants. Plant Physiol **124**: 563-578

Zhong R, Pena MJ, Zhou GK, Nairn CJ, Wood-Jones A, Richardson EA, Morrison WH III, Darvill AG, York WS, Ye ZH (2005) *Arabidopsis* fragile fiber8, which encodes a putative glucuronyltransferase, is essential for normal secondary wall synthesis. Plant Cell **17**: 3390-3408

Zhong R, Ye ZH (2009) Secondary cell walls. In: Encyclopedia of Life Sciences. John Wiley Sons, Ltd: Chichester DOI: 10.1002/9780470015902.a0021256

Zhu L, Shi GX, Li ZS, Kuang TY, Li B, Wei QK, Bai KZ, Hu YX, Lin JX (2004) Anatomical and chemical features of high-yield wheat cultivars with reference to its parents. Acta Botanica Sinica **46**: 565-572

Zorrilla-Fontanesi Y, Rambla JL, Cabeza A, Medina JJ, Sánchez-Sevilla JF, Valpuesta V, Botella MA, Granell A, Amaya I (2012) Genetic analysis of strawberry fruit aroma and identification of *O*-methyltransferase FaOMT as the locus controlling natural variation in mesifurane content. Plant Physiol **159**: 851-870

CHAPTER 2 - IDENTIFICATION OF CANDIDATE GENES TO ENCODE RHAMNOSYL 3-*O*-METHYLTRANSFERASE IN THE MOSS *PHYSCOMITRELLA PATENS*

ABSTRACT

Arabinogalactan proteins (AGPs) in the moss *Physcomitrella patens* contain up to 15 mol percent of 3-*O*-methyl-L-rhamnosyl (3-*O*-Me-Rha) residues, which is among the highest known abundances of methyl-ether sugars in plant polymers. With the aim of identifying the rhamnosyl methyltransferase and studying the biological roles of the methylated terminal rhamnosyl residues in the moss AGPs, eleven *Physcomitrella* genes were selected as candidates to encode the rhamnosyl methyltransferase. Some of these candidate genes were selected via BlastP and other searches of the *Physcomitrella* protein database using as query *Mycobacterium smegmatis* Mtf1, a proven rhamnosyl-3-*O*-methyltransferase. Other candidates were selected through a conserved domain search focused on TylF, the prototypic member of a superfamily of *O*-methyltransferases including Mtf1. One candidate was selected on the basis of having been identified as the only *Physcomitrella* homolog of *Arabidopsis* AtGXMT1, a recently reported glucuronosyl-4-*O*-methyltransferase gene that is thus far the only gene shown to encode an *O*-methyl ether transferase working on plant cell wall polysaccharides. Stable knockouts of nine of the eleven candidate genes in *Physcomitrella* were obtained. Analysis of glycosyl composition of AGPs purified from knockout mutants showed that some of the knockout lines had reduced 3-*O*-Me-Rha/Rha ratio, a result consistent with the predicted effect of disabling the rhamnosyl methyltransferase. None of the knockout lines, however, showed complete loss of 3-*O*-Me-Rha, leaving open the possibilities that

rhamnosyl 3-*O*-methyltransferase might be encoded by a multi-gene family or that the candidate proteins might influence 3-*O*-Me-Rha content only through indirect mechanisms.

INTRODUCTION

Two types of methyl substituents occur on sugar residues in plant cell wall polysaccharides. One type of methyl substituent is formed via ester linkage between a methyl group and a C6 carboxyl group of an uronic acid sugar. The most studied example of such a methyl ester is that formed at the C6 carboxyl group of galacturonosyl (GalU) residues in pectic polysaccharides. The second type of methyl substituent is formed via ether linkage between a methyl group and a secondary alcohol (on sugar carbons C2, C3, C4, or C5) or between a methyl group and a primary alcohol (C5 on pentose or C6 on hexose). A variety of such methylated sugars have been found in plant cell walls. Most often, these methylated sugars are of low abundance and have been little studied. The main exceptions are 4-*O*-methyl-D-glucuronosyl (4-*O*-Me-GlcU) residues, which occur in xylan-backbone hemicelluloses in some plants, particularly in secondary cell wall, and in certain plant gums. Enzymic activity of a methyltransferase that synthesizes 4-*O*-Me-GlcU was first detected long ago by Kauss and Hassid (1967) and subsequently studied by various other investigators. In an article published during the course of work on this dissertation, Urbanowicz et al. (2012) reported the first identification of a gene encoding a methyltransferase that forms a methyl ether on a plant cell wall polysaccharide. The *GXMT1* gene of *Arabidopsis* encodes a domain of unknown function family 579 protein that methylates D-GlcU residues on a 1,4-linked β-D-xylan backbone hemicellulose to form 4-*O*-Me-GlcU residues. Approximately one in eight xylosyl residues have a side chain of a single D-GlcU residue, and about 75% of these D-GlcU residues are methylated to 4-*O*-Me-GlcU (Urbanowicz et al., 2012).

Some lower plants have been recently found to contain considerable amounts of other methylated sugars. Popper et al. (2001) reported that 3-*O*-methyl-D-galactosyl (3-*O*-Me-Gal) residues occur at significant levels (5-10 mg per g of alcohol-insoluble residue) in cell walls of homosporous and heterosporous lycophytes. Low levels of 3-*O*-Me-Gal residues were previously reported in polysaccharides from the green alga *Chlorella vulgaris* (Ogawa et al. 1994), the red alga *Porphyridium aerugineum* (Percival & Foyle, 1979), the leaves of various dicot trees (Bacon & Cheshire, 1971), the twigs of sassafras (Springer et al., 1965), and in a RG I-like pectic mucilage from the inner bark of the elm tree *Ulmus glabra* (Barsett & Paulsen 1992).

A practical motivation for the present study of methylated sugars is their potential relevance to production of biofuels. The presence of methyl-ether groups on sugar residues might cause cell walls to be more recalcitrant to depolymerization to fermentable sugars for production of ethanol. Such effect was demonstrated by Urbanowicz et al. (2012) who showed that cell walls of *gxmt1* mutant *Arabidopsis* plants had altered lignin structure and enhanced xylan release during mild hydrothermal pretreatment. For other approaches to biofuels production, however, methylated sugars could be a positive factor. In thermochemical conversion of biomass to synthetic diesel fuel, methylated sugars would convert to more fuel than would unmethylated sugars. For each singly methylated hexose sugar, the number of reduced carbon atoms is increased from six to seven, so combustibles are increased approximately 17% compared to unmethylated hexose sugar. Therefore, plants containing higher abundances of *O*-methylated sugar residues in wall polymers would produce biomass yielding more biofuel by thermochemical conversion.

Previous work in this laboratory led to the finding that arabinogalactan proteins in the moss *Physcomitrella patens* contain up to 15 mol% of 3-O-methyl-L-rhamnosyl (3-*O*-Me-Rha) residues in terminal non-reducing positions on the glycan chains (Fu et al., 2007). This 15 mol% is among the highest known abundances of methyl-ether sugars in plant polymers. While 3-*O*-Me-Rha has not been found in AGPs or any other cell wall polymers of angiosperms, it has been reported in wall polymers of relictual plants including the fresh-water red alga *Batrachospermum* sp. (Turvey & Griffiths, 1973); the green colonial microalga *Botryococcus braunii* (Allard & Casadevall, 1990); the green alga *Chlorella vulgaris* (Ogawa et al., 1997); charophytes, liverworts, mosses, and homosporous lycophytes (Popper & Fry, 2003; Matsunaga et al., 2004; Popper et al., 2004); the fern *Osmunda japonica* (Akiyama et al., 1988); the moss *Sphagnum novo-zelandicum* (Kremer et al. 2004); and gymnosperms including *Picea nigra* (Gorrod & Jones, 1954); *Araucaria cumminghamii* (Anderson & Munro, 1969); *Encephalartos longifolius* (Kaplan et al., 1966); and *Welwitschia mirabilis* (Kaplan et al., 1966).

Arabinogalactan-proteins (AGPs) are a family of highly glycosylated hydroxyproline-rich glycoproteins (HRGPs) that are found bound to the plasma membrane, bound to the cell wall, and soluble in the cell wall space of plants (Serpe & Nothnagel, 1999; Seifert & Roberts, 2007). AGPs usually consist of a hydroxyproline-rich core polypeptide surrounded by arabinosyl- and galactosyl-rich glycans that are decorated with other less-abundant sugars which, depending on the species, might include L-rhamnosyl, D-mannosyl, D-xylosyl, D-glucosyl, L-fucosyl, D-glucosaminyl, D-glucuronosyl or D-galacturonosyl residues (Nothnagel, 1997). The high amount of 3-

O-Me-L-Rha residues in terminal positions on the arabinogalactan glycans of *Physcomitrella* AGPs does not occur in any angiosperms so examined and thus motivates the present study to identify the methyltransferase that synthesizes 3-*O*-Me-Rha and to identify the biological roles of this unusual sugar residue.

Because no genes in plants have been identified as encoding a methyltransferase that synthesizes 3-*O*-Me-Rha, it is necessary to look to other organisms for a starting point in this study to identify the *Physcomitrella* gene. *Mycobacterium smegmatis* Mtf1 is an *S*-adenosylmethionine (AdoMet)-dependent methyltransferase that methylates rhamnose to produce 3-*O*-Me-Rha in glycopeptidolipids at the cell surface (Patterson et al., 2000). Sequence comparison with other methyltransferases shows that the *mtf1* gene product Mtf1, also named as Rmt3 (Jeevarajah et al., 2004), shares sequence similarity to several known AdoMet-dependent methyltransferases from other microorganisms, such as *M. avium* MtfB, MtfC, MtfD; *Micromonospora griseorubida* MycF; and *Streptomyces fradiae* TylF (Patterson et al., 2000). Several conserved motifs are revealed in multiple sequence alignment of Mtf1 with these other methyltransferases. The crystal structure of one member of this superfamily was recently reported (Gomez García et al., 2010) and shows that motif I forms part of the secondary structure that binds the methionine moiety of AdoMet while motif II binds to the ribose moiety of AdoMet. Overall sequence similarity among methyltransferases is commonly found to be low, however, sometimes as low as 10% even among members of just one of the five topology classes of methyltransferases (Gomez García et al., 2010).

The moss *Physcomitrella patens* belongs to Phylum (Division) Bryophyta and has been gaining favor as a relictual model organism for studies in plant evolution, development and physiology (Cove et al., 2005; Wood et al., 2000; Vidali et al., 2009). Genomic sequence, expressed sequence tag (EST) and other transcriptome data all show that *Physcomitrella* shares a high degree of homology with angiosperms such as *Arabidopsis* (Reski et al., 1998). More than 66% of *Arabidopsis* proteins have homologs in *Physcomitrella*. *Physcomitrella* has additional value as a model system because it is well-suited for targeted knockout studies of plant genes. While plant species generally exhibit homologous recombination (HR) at only very low frequencies (less than 1%), *Physcomitrella* is a unique plant because it exhibits a high efficiency of HR (Abel & Theologis, 1994; Schaefer & Zryd, 1997; Schaefer, 2001; Girke et al., 1998; Kamisugi et al., 2005). This high efficiency HR enables functional genomic studies in *Physcomitrella* by the targeted gene knockout technique.

The study described in this chapter had the goal of identifying candidate genes to encode the rhamnosyl 3-*O*-methyltransferase that acts in synthesis of AGPs in the moss *Physcomitrella patens*. Searches for sequences with similarity to *Mycobacterium* Mtf1 and other bioinformatics approaches were used to identify candidate *Physcomitrella* genes. Knockout mutants of these candidate genes in *Physcomitrella* were then generated and analyzed.

MATERIALS AND METHODS

Bioinformatics searches. Using the amino acid sequence of *Mycobacterium* Mtf1 rhamnosyl 3-*O*-methyltransferase (Patterson et al., 2000) as the query, BlastP (Altschul et al., 1997) and Needle pairwise alignment (Needleman & Wunsch, 1970) searches were performed against the predicted protein database for the Joint Genome Institute (JGI) v1.1 *Physcomitrella* genome (http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html). The JGI v1.1 *Physcomitrella* genome was also searched for predicted proteins with sequences showing similarity to the conserved domain of the bacterial macrocin-*O*-methyltransferase (TylF) family which includes Mtf1 as a member. This conserved domain search was performed at the website of the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/cdd).

Generation of knockout cassettes. Knockout cassettes were generated based on the moss transformation vector pTN80 (gift from Dr. Mitsuyasu Hasebe, National Institute for Basic Biology, Japan, GI: 124377588) that confers resistance to Geneticin antibiotic (G418). To enable homologous recombination, two partial genomic sequences of targeted genes (called left and right genomic fragments) were cloned into the left and right flanking regions of the *npt* gene in pTN80. The following cloning strategies were used. The left and right genomic fragments of the targeted genes were amplified by polymerase chain reaction (PCR) from the *Physcomitrella* genome with the primers shown in Table 2.2 for each candidate gene. After purification by agarose gel electrophoresis, the PCR fragments were cloned into T-easy vector (Promega, Wisconsin, WI, USA). Positive colonies were selected by colony PCR and enzyme digestion and

confirmed by sequencing (Genomics Core Facility, UC Riverside). All restriction enzymes used for generation of knockout cassettes were from NEB (Ipswich, MA, USA). Through use of digestion sites on the PCR primers and the T-easy vector, left genomic fragments of knockout cassettes of *KO1* to *KO11* were subcloned from T-easy vector to pTN80 with ApaI and EcoRI. Right genomic fragments of *KO1, KO2*, *KO5* and *KO6* were cloned to pTN80 at SpeI and NotI sites, while right genomic fragments of *KO3*, *KO4*, and *KO7* were cloned to pTN80 at SacI and NotI sites. For *KO5*, after the right genomic fragment was cloned into pTN80, the plasmid was cut and spliced to remove polylinker between the two SpeI sites. Left genomic fragment of *KO5* was then incorporated to the plasmid at KpnI and XhoI sites. This plasmid was linearized with KpnI and SacI before being used to transform protoplasts. For *KO8*, a 2001 bp gene fragment, amplified by PCR with primers F-GGTTGGATGTGGAGCAGAGT, R-ATTCCCACCGTTGGTCTACA and exTaq (Takara Bio., Mountain View, CA, USA), was cloned into T-easy vector. After it was in the T-easy vector, this fragment was digested at the XhoI site. The *npt* gene with 35S promoter and NOS terminator from vector pTN80 was cloned into the the XhoI site of the above T-easy vector containing the 2001 bp genomic sequence from KO8.. For *KO9* left genomic fragment, an 893 bp genomic fragment was amplified with primers F- TGGGCCCTAACCAGGCAAAGG and R- TTGCTGATGTCCAGCGCAAT and then cloned into T-easy vector. This fragment was then subcloned into pTN80 at the ApaI and EcoRI sites. Similarly for *KO9* right genomic fragment, an 832 bp genomic fragment was amplified with primers F-AAAAGGCGGGCGTTGC and R- GAGCTCCGGTTACGACCATTTT, cloned into T-

easy vector, and then subcloned from T-easy vector to pTN80 at NotI and SacI sites. For *KO10* left genomic fragment, an 858 bp genomic fragment was amplified with primers F-CAATGCACGGATCTTCAACA and R- TGAATTGGACATGGCGAGTA, cloned to T-easy, and then subcloned to pTN80 at ApaI and SalI sites. Similarly for *KO10* right genomic fragment, an 851 bp genomic fragment was amplified with primers F-GTGTAGCGTCGACAGCAGTG and R- CCACCATATAATTCCGCACA, cloned to T-easy, and then subcloned to pTN80 at NotI and SpeI sites. All of the plasmids except *KO5* were linearized with BssHII before being used to transform protoplasts. Typically 6 to 15 μg of linearized plasmids were used in each transformation.

Named media used in protoplast transformation and moss culture. Uses of the following media are mentioned in subsequent subsections.

MMM solution. 0.5 M D-mannitol, 15 mM $MgC1_2$, 5.12 mM 2-(N-morpholino)ethanesulfonic acid, KOH as needed to give pH 5.6. Solution sterilized by filtration. (Cove et al., 2009).

BCD medium. 1 mM $MgSO_4$, 10 mM KNO_3, 45 μM $FeSO_4$, 1.8 mM, KH_2PO_4 (pH 6.5 adjusted with KOH), 1mM $CaCl_2$, 0.22 μM $CuSO_4$, 0.19 μM $ZnSO_4$, 10 μM H_3BO_3, 0.1 μM, Na_2MoO_4, 2μM $MnCl_2$, 0.23 μM $CoCl_2$, and 0.17 μM KI. Medium sterilized by autoclaving. (Ashton & Cove, 1977; Hiwatashi, 2012).

BCDAT medium. Same as BCD medium but additionally containing 5 mM diammonium tartrate, 0.8% (w/v) agar (Sigma, A1296). Medium sterilized by autoclaving. (Hiwatashi, 2012).

MCT solution. 100 mM $Ca(NO_3)_2$,7.65% (w/v) D-mannitol, 10 mM Tris-HCl, pH 8.0. Solution sterilized by filtration. (Cove et al., 2009).

PEG/T solution. 2 g polyethylene glycol (PEG 6000, Calbiochem 528877) heated to melting in a microwave oven then mixed with 5 ml of MCT solution. Mixed well and allowed to cool at room temperature for 2-3 hr, then immediately used in protoplast transformation. (Cove et al., 2009).

Protoplast liquid medium. 6.6% (w/v) D-mannitol, 27.8 mM glucose, 5 mM $Ca(NO_3)_2$, 1 mM $MgSO_4$, 0.184 mM KH_2PO_4 (pH 6.5 adjusted with KOH), 45 µM $FeSO_4$, 0.27 mM diammonium tartrate, Medium sterilized by autoclaving. (Hiwatashi & Hasebe, 2012).

PRMB medium. 6% (w/v) D-mannitol, 5 mM diammonium tartrate, 0.7% (w/v) agar (Sigma-Aldrich, St. Louis, MO, USA), BCD medium added to volume. Medium sterilized by autoclaving. After sterilizing, just before medium used, $CaCl_2$ added to give 10 mM. (Cove et al., 2009).

Knop medium. 1.84 mM KH_2PO_4, 3.35 mM KCl, 1.0 mM $MgSO_4$, 6.09 mM $Ca(NO_3)_2$, 45 µM $FeSO_4$, KOH or HCl as needed to give pH 5.8. For solid medium, 0.8% (w/v) agar (Sigma-Aldrich, St. Louis, MO, USA) added. Medium sterilized by autoclaving. (Reski & Abel, 1985).

KP medium. A modified Knop's medium prepared exactly as described by Fu et al. (2007).

Protoplast transformation. To generate knockout mutants in *Physcomitrella*, the linearized knockout cassettes were introduced into protoplasts by polyethylene glycol

(PEG 6000, Calbiochem, San Diego, CA, USA) mediated transformation (Schaefer et al., 1991; Cove et al., 2009; Hiwatashi & Hasebe, 2012). Protonema tissue grown on cellophane on BCDAT medium for 7 d after subculture was digested with Driselase enzyme to release protoplasts (Grimsley et al., 1977). To make 100 ml of the digestion solution, 2g of Driselase powder (gift of Prof. Yohichi Hashimoto, Saitama University, Urawa, Japan) was mixed in 100 ml of 8.5% (w/v) mannitol solution, centrifuged, and the supernatant sterilized by filtration. All subsequent steps were performed under aseptic conditions. The protonema tissue (mainly chloronema) was added to the Driselase solution (5 ml of Driselase solution for one 90-mm diameter Petri dish of tissue) and incubated at room temperature for 30 to 60 min with gentle mixing every 5 min. The digestion mix was then filtered through a nylon mesh having 64 µm pores. The protoplast suspension passing through the nylon mesh was centrifuged at 1000 rpm (180 x g) for 5 min, the resulting supernatant was discarded, and the pellet of protoplasts was gently resuspended in 80 ml of 8.5 % (w/v) mannitol. The washing by centrifugation was repeated twice, each time gently taping the centrifuge tube to resuspend the protoplasts in the mannitol solution. An aliquot of the protoplast suspension was placed on a hemacytometer, and the protoplasts were counted to determine number density. The protoplasts were then resuspended at 1.6×10^6 protoplasts ml^{-1} in MMM solution.

For protoplasts transformation, 30 µl of linearized plasmid solution (containing 6 to 15 µg of DNA) was added to a 15 ml conical tube. Next, 300 µl of protoplast suspension was added to the tube, followed by 300 µl of PEG/T solution. After gentle agitation, the transformation mixture was incubated in a 45˚C water bath for 5 min and

then at room temperature for 10 min. The transformation mixture was then diluted by adding 300 µl of 8.5% mannitol solution 5 times at 1-min intervals and then 1 ml of protoplast liquid medium 5 times at 1-minute intervals. The diluted protoplast solution was then poured onto a Petri dish (Hiwatashi & Hasebe, 2012) containing PRMB medium overlaid with cellophane and incubated for 7 d for protoplast regeneration at room temperature.

Knockout mutant selection. After one week of incubation, regenerated protoplasts on cellophane were transferred by moving the entire cellophane sheet to BCDAT medium containing G418 antibiotic (25µg/ml, Life Technologies, Carlsbad, CA, USA) in a Petri dish for selection of stable transformants. The culture was then incubated at room temperature under continuous light (40-60 µmol m^{-2} s^{-1}). After incubation on selection medium for 3 to 4 weeks, the growing colonies were individually lifted from the cellophane and transferred to antibiotic-free, normal growth BCDAT medium for 2 weeks to release selection pressure. After the 2 weeks, tissues from the colonies were then transferred back to BCDAT medium containing G418 antibiotic (25µg/ml) for another round of selection to get stable knockout mutants. This process was repeated again to make a total of three selections (two selections for *ko11*) on the antibiotic-containing BCDAT medium. After three rounds of selection, the surviving colonies were considered to be stable transformants and were propagated thereafter on normal BCDAT or KP medium.

Routine moss culture. Wild-type and stable knockout cultures of *Physcomitrella* in predominantly leafy gametophytic form were routinely cultured on KP medium, either

liquid or agar-solidified, with transfers at approximately monthly intervals. To maintain gametophytic cultures in predominantly protonemal form, tissue was cut to small pieces in a Polytron homogenizer and grown on BCDAT medium. Normal conditions in the culture room were 21°C and 16/8 h light/dark photoperiod with 40-60 µmol m^{-2} s^{-1} intensity during the light period.

Mapping the knockout cassettes in stable knockout mutants. To test whether the knockout cassettes had integrated into the genome at the intended locations, PCR was performed to map the location of insertion. Genomic DNA was prepared from the cultures by the CTAB method (Aono et al., 2003; moss.nibb.ac.jp), and GoTaq Green Master Mix (Promega, Wisconsin, WI, USA) was used as the PCR reagent to amplify the border sequences of the insertions. The PCR temperature program was 94°C for 3 min, followed by 35 cycles of 94°C for 1 min, (Tm-2°C of primers) for 1 min, and 72°C for 1 min, followed by a final extension of 72°C for 5 minutes. PCR products were analyzed by agarose gel electrophoresis.

Analysis of AGP glycosyl composition by gas chromatography (GC). Wild-type and knockout cultures of *Physcomitrella* were processed for purification of total soluble AGPs and analysis of AGP glycosyl composition using the methods of Fu et al. (2007), modified to accommodate micro-scale samples of 0.25-0.5 gfw. Moss gametophytes were homogenized by grinding in a mortar and pestle with liquid nitrogen, and then the frozen powder was added to 4.5 ml of a pH 7.5 buffer of 50 mM tris (hydroxymethyl) aminomethane-HCl, 10 mM KCl, 1 mM EDTA, 0.1 mM $MgCl_2$, 8% (w/v) sucrose, 1 mM phenylmethanesulfonyl fluoride, and 0.1% (v/v) plant protease inhibitor (Sigma-

Aldrich, St. Louis, MO, USA). After stirring for 5 min, the homogenate was centrifuged at 35,600xg for 30 min at 4°C. The resulting supernatant was filtered through Whatman #1 paper and collected in a glass conical centrifuge tube (Kimax 45200 with Teflon-lined screw cap). To each filtrate was added $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside, NaN_3 solution, and dry NaCl to give final concentrations of 60 µM, 0.02% (w/v), and 1%(w/v), respectively. The specimens were then incubated overnight at 4°C to allow formation and precipitation of the AGP-Yariv phenylglycoside complex. The following morning, the specimens were centrifuged at 1,700xg for 10 min at room temperature. The supernatants were discarded, and the pellets were extracted with 0.5-ml aliquots of distilled water. After mixing and centrifugation, the supernatants were poured off and saved. Extraction of the pellets with 0.5-ml aliquots of distilled water, and saving and pooling of the supernatants individually for each specimen, was continued until all of the red color was in the pooled supernatants. To the pooled supernatant for each specimen was added NaN_3 solution and dry NaCl to give final concentrations of 0.02% (w/v) and 1% (w/v), respectively, and the specimens were then incubated overnight at 4°C to allow formation and precipitation of the AGP-Yariv phenylglycoside complex. This process was repeated for a total of three additions of dry NaCl. After the centrifugation after the third addition of NaCl, the supernatants were discarded, and the pellets were dissolved in minimal (0.02 or 0.05 ml) dimethylsulfoxide, with sonication to achieve full solubility. Three volumes (0.06 or 0.15 ml) of acetone and 0.08 volumes (1.6 or 4 µl) of 1% (w/v) NaCl solution were then added to each specimen, the tubes were quickly closed with Teflon-lined caps, vortexed, and incubated overnight at 4°C to allow precipitation of the AGPs while

leaving the Yariv phenylglycoside in the supernatant. The following morning, the specimens were centrifuged at 1,700xg for 10 min at room temperature. The supernatants were discarded, and the pellets were again dissolved in minimal dimethylsulfoxide and then precipitated by addition of acetone and NaCl solution. This procedure for extraction of Yariv phenylglycoside was usually repeated for a total of three additions of dimethylsulfoxide, although for some very small specimens only one or two additions of dimethylsulfoxide were used. After the centrifugation after the final addition of dimethylsulfoxide, acetone, and NaCl solution, the supernatants were discarded, and the pellets were washed twice by addition of 0.3 ml of acetone and centrifugation. The pellets after the second acetone wash were dried and then dissolved in 0.2 ml of distilled water. The 0.2 ml solutions of purified AGPs were then split with 0.1 ml being processed for GC and the other 0.1 ml being stored frozen as back-up. Processing for GC, including methanolysis, trimethylsilylation of methyl glycosides, and GC analysis of the trimethylsilyl methylglycosides were done as described by Fu et al. (2007).

Statistical analyses of the glycosyl composition results were performed using InStat (version 2.0, Graphpad Software, San Diego, CA).

RESULTS

Candidate proteins that might be rhamnosyl 3-O-methyltransferase revealed by bioinformatics searches. Bioinformatics searches against the *Physcomitrella* predicted proteome from JGI (v1.1) were performed to identify candidate proteins that might be the rhamnosyl 3-*O*-methyltransferase that transfers the methyl group to terminal rhamnosyl residues in AGPs in *Physcomitrella*. Using the amino acid sequence of *Mycobacterium* Mtf1 rhamnosyl 3-*O*-methyltransferase as the query in BlastP searches at NCBI resulted in eight hits. These eight hits had expectation (E) values ranging 0.81 to 9.3, however, none of which were even close to significant. These eight matches, as well as matches resulting from other BlastP searches, were examined by the Needle pairwise alignment algorithm. Here again overall similarities were low, falling in the range of 7.6% to 20.8%. The results from the Needle algorithm were somewhat more encouraging when attention was directed to the conserved sequence regions of Mtf1 that appear to interact with the AdoMet substrate (Patterson et al., 2000; Gomez García et al., 2010). Several of the proteins appearing as hits in the BlastP searches showed Needle alignments at key amino acid residues in these conserved domains. Because none of the candidates that arose from the similarity searches based on Mtf1 were outstanding above the others, it was apparent that several of the candidates should be investigated to increase the probability of finding the actual rhamnosyl 3-*O*-methyltransferase. Thus, seven predicted proteins were selected from the BlastP and Needle searches for further study. The seven candidates were PHYPADRAFT_169825, PHYPADRAFT_83743, PHYPADRAFT_134505, PHYPADRAFT_153277, PHYPADRAFT_163311,

PHYPADRAFT_209593, and PHYPADRAFT_139716. For convenience in this dissertation, these seven candidate genes were given the shortened names *KO1* to *KO7*, respectively. Among these seven predicted proteins, five have complete gene sequences available in the JGI (v1.1) genome, but *KO2* and *KO5* are known only by EST sequences.

A different bioformatics approach was used to identify other candidate proteins. The conserved domain search at NCBI was used to search the JGI (v1.1) *Physcomitrella* genome for predicted proteins with a domain similar to the conserved domain of the bacterial macrocin-*O*-methyltransferase (TylF) superfamily, which includes Mtf1 as a member. The prototype of this superfamily, TylF, is a bacterial *O*-methyltransferase that methylates a sugar residue on macrocin to produce tylosin, a macrolide antibiotic. The conserved domain search relies more on higher level features, such as likely helical and fold structures, and less on strict primary structure than does a BlastP search. Using this TylF domain to search the *Physcomitrella* predicted proteome resulted in three proteins annotated as AdoMet-dependent, family 3 *O*-methyltransferases. The three hits were PHYPADRAFT_15168, PHYPADRAFT_116394, and PHYPADRAFT_143295, all of which were added to the candidate list for study in this dissertation where they were designated as KO8, KO9, and KO10, respectively.

The final one added to the candidate list was PHYPADRAFT_115978 (alias Pp1s15_437V6), designated here as KO11. In reporting on *Arabidopsis* AtGXMT1, an *O*-methyltransferase responsible for the methylation of D-GlcU residues on glucuronoxylan to form 4-*O*-Me-GlcU, Urbanowicz et al. (2012) identified AtGXMT1 as a domain of unknown function family 579 protein and further identified Pp1s15_437V6

as the only member of the *Physcomitrella* predicted proteome belonging to this family. Because analyses of *Physcomitrella* cell walls resulted in no evidence of 4-*O*-Me-GlcU and in fact detected 3-*O*-Me-Rha as the only methylated sugar (Nothnagel & Nothnagel, 2007), PHYPADRAFT_115978 seemed to be a worthy candidate for investigation as the rhamnosyl 3-*O*-methyltransferase.

All 11 candidate genes and corresponding predicted proteins are listed in Table 2.1, together with their respective abbreviated names *KO1* to *KO11*. The updated physical locations of the genes in *Physcomitrella* genome v1.6 or v3.0 are also shown in Table 2.1, by chromosome if known otherwise by scaffold.

Gene locus and predicted models of candidate genes. Predicted gene models for all 11 candidate genes *KO1* through *KO11*are diagramed in Figure 2.1. The *KO1* gene is 8927 base pairs (bp) long, contains 22 exons and 21 introns, and results in a very long predicted protein of 1137 amino acids. The longest exon in *KO1* is 509 bp, while the shortest exon is 68 bp. The longest intron is 748 bp, and the shortest intron is 107 bp. The predicted transcript of *KO1* is approximately 3.6 kb.

A complete gene model for *KO2* is not available in the JGI (v1.1) genome. Sequence information on *KO2* is limited to an EST sequence, which consists of exons of length 40 bp and 422 bp, these two exons being interrupted by an 84 bp intron. The resulting predicted protein contains 154 amino acids.

The gene model of *KO3* is 8930 bp long, contains 22 exons and 21 introns, and encodes a predicted protein of 705 amino acids. The longest exon is 223 bp, while the longest intron is 1192 bp.

The gene model of *KO4* is 9109 bp long, consists of 23 exons and 22 introns, and encodes a predicted protein of 704 amino acids. The longest exon is 213 bp, while the shortest exon is only 2 bp. The first intron of *KO4* is 1021 bp, which is the longest intron of the gene.

A complete gene model for *KO5* is not available in the JGI (v1.1) genome. Sequence information on *KO5* is limited to EST sequences which are 2354 bp long, consist of a single exon with no introns, and encode a predicted polypeptide of 615 amino acids.

The gene model of *KO6* is 8209 bp long, consists of 16 exons, and encodes a predicted protein of 1489 amino acids.

The gene model of *KO7* is 3498 bp long, consists of 10 exons and 9 introns, and encodes a predicted protein of 632 amino acids. While the longest exon is 400 bp, the shortest exon is only 85 bp. The longest intron is 363 bp, and the shortest intron is 112 bp.

The predicted gene of *KO8* is 2087 bp long, consists of 4 exons and 3 introns, and encodes a predicted protein of 331 amino acids. The second exon is 407 bp, the longest exon in this gene. The shortest exon is the last exon at only 156 bp. The first intron of KO8 is 525 bp.

The gene model of *KO9* is 1363 bp long, consists of 5 exons and 4 introns, and encodes a predicted protein of 255 amino acids. The last exon is 294 bp, the longest exon of this gene. The last and the longest intron is 263 bp, and the shortest exon is 102 bp.

The gene model of *KO10* is 3020 bp long, consists of 9 exons and 8 introns, and encodes a predicted protein of 284 amino acids. The longest exon is the first and contains 264 bp, while the shortest exon is the last and contains only 42 bp. The longest intron is 696 bp, while the shortest intron is 115 bp.

The gene model for *KO11* has three exons in a gene of 1259 bp, resulting in a transcript of 690 bp.

The preliminary release of the Phytozome v3.0 genome for *Physcomitrella* (www.phytozome.net) includes chromosomal locations for some, but not all, of the 11 candidate genes *KO1* through *KO11*. Those genes without available chromosomal locations are shown by the number of genomic sequences in scaffolds. As judged from the identified locations of these 11candidate genes (.Table 2.1), none of the 11 genes are clustered.

Components and use of the knockout cassettes. The principal components of the moss transformation vector pTN80 are illustrated in Figure 2.2. Based on the backbone of pBluescript SK (+), vector pTN80 (Fig. 2.2A) contains a neomycin phosphotransferase *(npt)* gene between multiple cloning sites. The modified 35S promoter in front of the *npt* gene can drive expression in both *E. coli* and plants. To construct knockout cassettes in pTN80, left and right genomic fragment of each targeted gene were amplified by PCR using the primers listed in Table 2.2. The amplified genomic sequences were then cloned into the multiple cloning sites in the left and right flanking regions of the npt gene and its promoter (Fig. 2.2B). After linearization of each knockout cassette with BssHII (except *KO5*, see Methods) and PEG-facilitated introduction of the linearized cassette into the

Physcomitrella cytoplasm, these flanking genomic sequences from the targeted *Physcomitrella* gene promoted homologous recombination and integration of the linearized knockout cassette into a specific chromosomal location (Table 2.3) where it replaced a portion of the targeted wild-type *Physcomitrella* gene.

Mapping insertion of knockout cassettes in stable mutants surviving antibiotic selection. After three rounds of selection on medium containing G418 antibiotic, stable knockouts were obtained in which the *npt* gene, which confers G418 antibiotic resistance, was integrated into the genome. Knockout mutants of *ko1, 2, 3, 4, 5, 6, 7, 9, 10, and 11* were successfully generated by this method. When the *ko8* knockout cassette was used, no colonies survived after three rounds of selection on G418, indicating that knockout of the *KO8* gene was lethal.

Although homologous recombination (HR) is the major method used by *Physcomitrella* when repairing double strand DNA breaks (Kamisugi et al., 2005, 2006), repair by non-homologous end joining (NHEJ) cannot be completely ruled out *a priori*. Repair by HR leads to specific gene disruption by targeted insertion of knockout cassettes, but repair by NHEJ leads to illegitimate integration of knockout cassettes at random sites. To test whether the stable mutants surviving three rounds of antibiotic selection were specific knockouts of the targeted genes or instead were products of random insertion, genomic PCR was performed to check the integrity of the targeted gene.

A pair of gene-specific PCR primers was designed for each targeted gene and were tested in both wild-type and mutant plants. Figure 2.3A shows that, as expected, use

of the gene-specific primer pairs resulted in amplification of a genomic fragment of appropriate size when used with genomic DNA from wild-type plants. When used with genomic DNA from the mutant plants, the gene-specific primer pairs for *KO1, KO2, KO4, KO5, KO6, KO7, KO9, KO10,* and *KO11* did not result in amplification of a genomic fragment of the size observed with the wild-type plants. This observation is consistent with specific gene disruption by the NPT insertion and indicates that the desired knockout plants were obtained for these genes. In the case of *KO3*, however, PCR with the gene specific primers resulted in amplification of similar sized genomic fragments in both the wild-type and mutant plants.

An additional PCR experiment was performed to further investigate the *ko3* mutant plants. The left and right borders of the *ko3* knockout cassette were detected in these plants by using, for each border, one gene-specific PCR primer from outside of the KO cassette and one PCR primer from the NPT sequence (Fig. 2.3B). While this PCR confirmed that the *KO3* cassette had inserted into the *KO3* gene locus, the continuing detection of the wild type *KO3* gene fragment indicated that the culture was either diploid or a mixture of wild-type and knockout plants. The PEG-mediated transformation process can generate diploids or even higher polyploids by protoplast fusion during the PEG treatment (Hiwatashi & Hasebe, 2012). The observations on *ko3* (Fig. 2.3B) could be the result of the culture originating as the fusion of one wild-type protoplast with one HR knockout protoplast. Chimeric cultures that are mixtures of wild-type plants and knockout plants can result if too large of portions of colonies are lifted from plates and propagated during antibiotic selection (Hiwatashi & Hasebe, 2012).

Screening of KO mutants by analysis of glycosyl composition of AGPs. The hypothesis is that if one of the candidate *KO* genes does encode the rhamnosyl 3-*O*-methyltranferase responsible for the synthesis of 3-*O*-Me-Rha, then knockout of that gene will result in a reduced or zero level of 3-*O*-Me-Rha in AGPs from the knockout mutant plants. To investigate this hypothesis, AGPs were purified from wild-type and knockout mutants and analyzed for glycosyl composition by GC. Because some of the mutants grew slowly, the amounts of even gametophytic tissue available were quite limited. Thus, it was necessary to develop purification and assay procedures that could start with amounts of tissue in the range of 0.25 to 0.5 gfw, or even less. Obtaining high purity fractions of AGPs (or really any macromolecule) from such small starting specimens is challenging.

One particular challenge was incomplete removal of $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside from the final AGP preparation. Because each $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside molecule contains three Glc residues, the presence of even small amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside significantly skewed the measured glycosyl composition of the AGP preparation. The amount of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside was quite variable but proportionately larger for the smaller specimen sizes. Typically the measured glycosyl compositions of the AGP preparations included less than 20 mol% Glc, but in some cases, usually those where the starting mass of tissue was less than 0.25 gfw, the glycosyl composition included as much as 80 mol% Glc. The practical solution to this problem was to mathematically remove Glc from glycosyl compositions and renormalize the results for the other sugars. Because highly purified

AGPs in general (Nothnagel, 1997) and *Physcomitrella* AGPs in particular (Fu et al., 2007) contain only low amounts (commonly 1-2 mol%) of Glc residues, mathematical removal of Glc resulted in little distortion of the glycosyl composition.

Even with mathematical removal of Glc, the glycosyl compositions of the AGP preparations were somewhat variable. Tables 2.4-2.13 show that the average contents of 3-*O*-Me-Rha ranged from 1.41 mol% to 10.81 mol%. As these experiments were being done, it became evident that much of this variability arose from week-to-week and investigator-to-investigator variations in success with the microscale purification procedure. To deal with this variation, specimens were prepared and analyzed in batches with each batch consisting of one wild-type specimen and two to four knockout mutant specimens. Always pairing knockout specimens with a wild-type specimen in the same batch enabled use of the paired t-test in assessing the statistical significance of the results. A further step to reduce effects of contaminants in the final AGP preparation was to focus on the 3-*O*-Me-Rha/Rha ratio, rather than the absolute level of 3-*O*-Me-Rha, when assessing possible effects of a knockout. Such an approach is sensible in any case, since Rha resides are likely the substrate for the rhamnosyl 3-*O*-methyltransferase enzyme.

With the focus on the 3-*O*-Me-Rha/Rha ratio, a remaining concern was contamination by pectic polysaccharides. Pectic polysaccharides contain some amount of Rha residues, particularly so for the rhamnogalacturonan I subclass of pectic polysaccharides where the polymer backbone consists of alternating residues of GalU and Rha. Such pectic polysaccharides are common in a wide evolutionary range of plants (Nothnagel & Nothnagel, 2007). Pectic polysaccharide contamination in the AGP

preparation would thus skew the 3-*O*-Me-Rha/Rha ratio downward because of the extra Rha residues present in the pectic polysaccharides. Because highly purified AGPs in general (Nothnagel, 1997) and *Physcomitrella* AGPs in particular (Fu et al., 2007) contain only low amounts (commonly 0-5 mol%) of GalU residues, pectic polysaccharide contamination in the AGP preparation was easily recognized by high GalU content. For most of the AGP preparations from the *ko* mutants, the GalU content was low. In the relatively few instances where AGP preparations had high GalU content, the results from those specimens were generally excluded from the analysis. The captions for Tables 2.4 through 2.13 include information on how many specimens were so excluded. This approach was generally successful except for the *ko2* knockouts where all AGP specimens had high (10-65 mol%) GalU content, and thus they could not be excluded. This odd characteristic of the KO2 AGP specimens remains an intriguing mystery.

With these cautions and procedures, the 3-*O*-Me-Rha /Rha ratio in AGPs was adopted as an indicator of whether the knockout mutants showed disrupted function of rhamnosyl 3-*O*-methyltransferase.

The results presented in Tables 2.4 through 2.13 show that all of the preparations from both wild-type and knockout mutants showed glycosyl compositions that are typical of AGPs, as indicated by highest Gal content (39.34-60.69 mol%) and also high Ara (9.73-24.18 mol%) and GlcU (10.94-16.34 mol%) contents. Thus, the microscale AGP purification procedure was effective and showed that none of the knockouts exhibited grossly disturbed AGPs.

As regards the key 3-*O*-Me-Rha /Rha ratio, statistically significant reductions in the ratio were observed in *ko1* (P=0.0031, Table 2.4), *ko2* (P=0.0329, Table 2.5), *ko6* (P=0.0140, Table 2.9), and *ko9* (P=0.0027, Table 2.11) relative to the wild-type. No statistically significant effects on 3-*O*-Me-Rha /Rha ratios were observed in *ko3* (Table 2.6), *ko5* (Table 2.8), *ko7* (Table 2.10), *ko10* (Table 2.12), and *ko11* (Table 2.13) relative to the wild-type. Curiously, a statistically significant increase in the 3-*O*-Me-Rha /Rha ratio was observed in *ko4* (P=0.0328, Table 2.7) relative to the wild-type. These results are summarized in Table 2.14.

DISCUSSION

In AGPs in the moss *Physcomitrella patens*, up to 15 mol% of the sugar residues are 3-*O*-Me-Rha in terminal non-reducing positions on the glycan chains (Fu et al., 2007). This relatively high abundance of methylated residues makes *Physcomitrella* AGPs an attractive model system for studying the mechanism of synthesis of methylated sugar residues. The attractiveness of the system is further enhanced by the high frequency of homologous recombination in *Physcomitrella*, which enables the use of gene targeting by knockout as a powerful technique for study of gene function (Kamisugi et al., 2005; 2006; 2012). A typical knockout cassette usually contains a selection marker gene with promoter and terminator, plus flanking homologous genomic fragments (Fig.2.2). These flanking partial gene sequences serve as templates for integration of the knockout cassette into the genome by HR. The length of the genomic sequence in the cassette determines the recombination frequency.

Use of gene targeting by knockout as the principal approach in this study required identification of genes that might be candidates to encode the targeted rhamnosyl 3-*O*-methyltransferase. With no plant genes yet identified as encoding methyltransferases acting on sugar residues in cell wall polymers, little information to guide the selection of candidate genes was available at the start of this dissertation. The best starting point found in the literature was the *Mycobacterium smegmatis* gene Mtf1, which encodes a rhamnosyl *O*-methyltransferase that produces 3-*O*-Me-Rha (Patterson et al., 2000), the exact sugar residue found in *Physcomitrella* AGPs. Using the primary amino acid sequence of Mtf1 as the query, BlastP (Madden, 2002) and other searches against the

whole *Physcomitrella* predicted protein database at JGI (v1.1) resulted in seven candidate proteins. The sequence similarities of these best seven candidate proteins to Mtf1 were all very low, perhaps not a surprising result given the evolutionary distance between *Mycobacteria* and *Physcomitrella* and the notoriously low sequence similarities observed even between methyltransferases of the same subclass (Gomez García et al., 2010). A search for homologs based on the conserved domain of the TylF superfamily of methyltransferases, which includes Mtf1, resulted in three additional candidates in *Physcomitrella*. The conserved domain of the TylF superfamily typically includes a class I methyltransferase fold and a divalent metal-binding motif (Gomez García et al., 2010). The final candidate protein added to this study was KO11. During the course of this dissertation work, Urbanowicz et al. (2012) identified AtGXMT1, a domain of unknown function family 579 protein, as being the methyltransferase that produces 4-*O*-Me-GlcU in *Arabidopsis*, and they identified KO11 as the only domain of unknown function family 579 protein in *Physcomitrella*.

All of these candidate proteins, KO1 through KO11, were investigated for their genomic location and sequences (Table 2.1, Fig. 2.1). Complete gene sequences for *KO2* and *KO5* were not available in the JGI (v1.1) genome, these two candidates being known only by ESTs. Partial genomic sequences for all 11 genes were amplified and cloned into the moss transformation vector pTN80 to generate knockout cassettes. Stable knockouts of all these genes except KO8 were successfully selected on medium containing G418 antibiotic. None of the *ko8* mutants survived after three rounds of selection, indicating that *KO8* might be an essential gene for cell division and plant growth. Another oddity

arose with the *ko3* mutant for which the insertion of the knockout cassette was detected at its genomic locus, but PCR showed that this gene locus was not completely disrupted. These results indicated that the *ko3* culture was either diploid or a mixture of wild-type and knockout plants.

Analyses of the ten knockout mutants (all but KO8) that survived selection on medium containing G418 antibiotic involved measurement of the 3-*O*-Me-Rha/Rha ratio in AGPs purified from the mutants and from the wild-type for comparison. The hypothesis was that if one of the candidate *KO* genes does encode the rhamnosyl 3-*O*-methyltranferase responsible for the synthesis of 3-*O*-Me-Rha, then knockout of that gene would result in a reduced or zero level of 3-*O*-Me-Rha in AGPs from the knockout mutant plants. The results presented in Tables 2.4 through 2.13 are summarized in Table 2.14. Statistically significant reductions in the 3-*O*-Me-Rha/Rha ratio were observed in *ko1* (P=0.0031, Table 2.4), *ko2* (P=0.0329, Table 2.5), *ko6* (P=0.0140, Table 2.9), and *ko9* (P=0.0027, Table 2.11) relative to the wild-type. In considering which of these four genes should be further investigated in this dissertation, *KO2* was eliminated because all of the AGP preparations from this mutant had very high levels of GalU, indicating significant contamination by pectic polysaccharides. These pectic polysaccharides likely brought in Rha residues which might have been responsible for the lowering of the 3-*O*-Me-Rha/Rha ratio. Although this consistent presence of pectic polysaccharides with the *ko2* AGPs might be an interesting topic for further study, it was considered beyond the scope of this dissertation.

During the course of this dissertation, functional annotations for KO1, KO6, and KO9 appeared in databases and helped guide selection of the genes for further study. KO1 was annotated as lipid-A-disaccharide synthase, which synthesizes the lipid-A-disaccharide in bacteria. Lipid A has not generally been considered to be present in plant cells, although a pathway for synthesis of lipid A precursors in *Arabidopsis* has been recently reported (Li et al., 2011). Still, it is not certain that the annotation of KO1 as lipid-A-disaccharide synthase is correct. KO6 was annotated as peptidase, which discouraged its further study. KO9 was annotated as caffeoyl-CoA O-methyltransferase (CCoAMT), which is an *O*-methyltransferase in lignin monomer biosynthesis. The annotation of KO9 as an *O*-methyltransferase, albeit of different function, was encouraging.

The statistically very significant effects of the *ko1* and *ko9* knockouts on the 3-*O*-Me-Rha/Rha ratio in *Physcomitrella* AGPs (Tables 2.4 & 2.11), together with the recent annotations of these genes, led to selection of these genes for further study in this dissertation. To these two candidates was also added *KO11* for further study. Although the *ko11* knockouts did not show a statistically significant reduction in the 3-*O*-Me-Rha/Rha ratio in *Physcomitrella* AGPs (Table 2.13), the homology of this protein to the only protein thus far proven to be an *O*-methyltransferase working on a sugar residue in a plant cell wall polysaccharide (Urbanowicz et al., 2012) encouraged its further study.

REFERENCES

Abel S, Theologis T (1994) Transient transformation of *Arabidopsis* leaf protoplasts: a versatile experimental system to study gene expression. Plant J **5**: 421-427

Akiyama T, Tanaka K, Yamamoto S, Iseki S (1988) Blood-group active proteoglycan containing 3-O-methylrhamnose (acofriose) from young plants of *Osmunda japonica*. Carbohydr Res 178: 320-326

Allard B, Casadevall E (1990) Carbohydrate composition and characterization of sugars from the green microalga *Botryococcus braunii*. Phytochemistry 29: 1875-1878

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25**: 3389-3402

Anderson DMW, Munro AC (1969) The presence of 3-*O*-methylrhamnose in *Araucaria* resinous exudates. Phytochemistry 8: 633-634

Aono N, Sato Y, Nishiyama T (2003) Genomic DNA extraction. Moss.nibb.ac.jp

Ashton NW, Cove DJ (1977) The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants in the moss *Physcomitrella patens*. Molecular and General Genetics **154**: 87-95

Bacon JSD, Cheshire MV (1971) Apiose and mono-*O*-methyl sugars as minor constituents of the leaves of deciduous trees and various other species. Biochem J **124:** 555-562

Barsett H, Paulsen BS (1992) Separation, isolation and characterization of acidic polysaccharides from the inner bark of *Ulmus glabra* Huds. Carbohydr Polymers **17**: 137-144

Cove D (2005) The moss *Physcomitrella patens*. Annu Rev Genet **39**: 339-358

Cove DJ, Perroud P-F, Charron AJ, McDaniel SF, Kandelwal A, Quatrano RS (2009) The moss *Physcomitrella patens*: a novel model system for plant development and genomic studies. In: A Gann, D Crotty, eds, Emerging Model Organisms: A Laboratory Manual, Vol. 1. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 69-104

Fu H, Yadav MP, Nothnagel EA (2007) *Physcomitrella patens* arabinogalactan proteins contain abundant terminal 3-*O*-methyl-L-rhamnosyl residues not found in angiosperms. Planta **226**: 1511-1524

Girke T, Schmidt H, Zahringer U, Reski R, Heinz E (1998) Identification of a novel delta 6-acyl-group desaturase by targeted gene disruption in *Physcomitrella patens*. Plant J **15**: 39-48

Gomez García I, Stevenson CE, Usón I, Caren L. Freel Meyers CL, Walsh CT, Lawson DM (2010) The crystal structure of the novobiocin biosynthetic enzyme NovP: the first representative structure for the TylF *O*-methyltransferase superfamily. J Mol Biol **395**: 390-407

Gorrod ARN, Jones JKN (1954) The hemicelluloses of Scots pine (*Pinus sylvestis*) and black spruce (*Picea nigra*) woods. J Chem Soc 2522-2525

Grimsley NH, Ashton NW, Cove DJ (1977) The production of somatic hybrids by protoplast fusion in the moss *Physcomitrella patens*. Molecular and General Genetics **154**: 97-100

Hiwatashi Y (2012) Culture and storage of protonemata and gametophores. http://moss.nibb.ac.jp/

Hiwatashi Y, Hasebe M (2012) How to transform *Physcomitrella patens*: PEG-mediated transformation. http://moss.nibb.ac.jp/

Jeevarajah D, Patterson JH, Taig E, Sargeant T, McConville MJ, Billman-Jacobe H (2004) Methylation of GPLs in *Mycobacterium smegmatis* and *Mycobacterium avium*. J Bacteriol 186: 6792-6799

Kamisugi Y, Cuming AC, Cove DJ (2005) Parameters determining the efficiency of gene targeting in the moss *Physcomitrella patens*. Nucleic Acids Res **33**: e173

Kamisugi Y, Schaefer DG, Kozak J, Charlot F, Vrielynck N, Hola M, Angelis KJ, Cuming AC, Nogue F (2012) MRE11 and RAD50, but not NBS1, are essential for gene targeting in the moss *Physcomitrella patens*. Nucleic Acids Res **40**: 3496-3510

Kamisugi Y, Schlink K, Rensing SA, Schween G, Stackelberg M, Cuming AC, Reski R, Cove DJ (2006) The mechanism of gene targeting in *Physcomitrella patens*: homologous recombination, concatenation and multiple integration. Nucleic Acids Res **34**: 6205-6214

Kaplan M, Stephen AM, Vogt D (1966) 3-*O*-Methyl-L-rhamnose as a constituent of plant polysaccharide gums. S African Med J 40: 702 (single page article)

Kauss H, Hassid WZ (1967) Biosynthesis of the 4-O-methyl-D-glucuronic acid unit of hemicellulose B by transmethylation from S-adenosyl-L-methionine. J Biol Chem **242**:1680-1684

Kremer C, Pettolino F, Bacic A, Drinnan A (2004) Distribution of cell wall components in *Sphagnum* haline cells and in liverwort and hornwort elaters. Planta **219**: 1023-1035

Li C, Guan Z, Liu D, Raetz CRH (2011) Pathway for lipid A biosynthesis in Arabidopsis thaliana resembling that of Escherichia coli. Proc Natl Acad Sci USA **108**: 11387-11392

Madden T (2002) Chapter 16, The BLAST sequence analysis tool. In: J McEntyre, J Ostell, eds, The NCBI handbook [Internet]. National Center for Biotechnology Information, Bethesda, MD

Matsunaga T, Ishii T, Matsumoto S, Higuchi M, Darvill A, Albersheim P, O'Neill, MA (2004) Occurrence of the primary cell wall polysaccharide rhamnogalacturonan II in pteridophytes, lycophytes, and bryophytes. Implications for the evolution of vascular plants. Plant Physiol **134**: 339-351

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48: 443-453

Nothnagel EA (1997) Proteoglycans and related components in plant cells. Int Rev Cytol 174: 195-291

Nothnagel AL, Nothnagel EA (2007) Primary cell wall structure in the evolution of land plants. J Integr Plant Biol **49**: 1271-1278

Ogawa K, Yamaura M, Maruyama I (1994) Isolation and identification of 3-*O*-methyl-D-galactose as a constituent of a neutral polysaccharide of *Chlorella vulgaris*. Biosci Biotech Biochem **58**: 942-944

Ogawa K, Yamaura M, Maruyama I (1997) Isolation and identification of 2-*O*-methyl-L-rhamnose and 3-*O*-methyl-L-rhamnose as constituents of an acidic polysaccharide of *Chlorella vulgaris*. Biosci Biotech Biochem 61: 539-540

Patterson JH, McConville MJ, Haites RE, Coppel RL, Billman-Jacobe H (2000) Identification of a methyltransferase from *Mycobacterium smegmatis* involved in glycopeptidolipid synthesis. J Biol Chem **275**: 24900-24906

Percival E, Foyle RAJ (1979) The extracellular polysaccharides of *Porphyridium cruentum* and *Porphyridium aerugineum*. Carbohydr Res **72**: 165- 176

Popper ZA, Fry SC (2003) Primary cell wall composition of bryophytes and charophytes. Ann Bot 91: 1-12

Popper ZA, Sadler IH, Fry SC (2001) 3-*O*-methyl-D-galactose residues in lycophyte primary cell walls. Phytochemistry **57**: 711-719

Popper ZA, Sadler IH, Fry SC (2004) 3-*O*-Methylrhamnose in lower land plant primary cell walls. Biochem Syst Ecol 32: 279-289

Reski R (1998) Development genetics and molecular biology of mosses. Bot Acta **111**: 1-15

Reski R, Abel WO (1985) Induction of budding on chloronemata and caulonemata of the moss *Physcomitrella patens* using isopentenyladenine. Planta **165**: 354-358

Serpe MD, Nothnagel EA (1999) Arabinogalactan-proteins in the multiple domains of the plant cell surface. Adv Bot Res **30**: 207-289

Seifert G, Roberts K (2007) The biology of arabinogalactan proteins. Annu Rev Plant Biol **58**: 137-161

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

Schaefer D, Zryd JP, Knight CD, Cove DJ (1991) Stable transformation of the moss *Physcomitrella patens*. Molec Gen Genetics **226**: 418-424

Schaefer DG (2001) Gene targeting in *Physcomitrella patens*. Curr Opin Plant Biol **4**: 143-150

Springer GF, Takahashi T, Desai PR, Kolecki BJ (1965) Cross reactive human blood-group H(O) specific polysaccharide from *Sassafras albidum* and characterization of its hapten. Biochemistry **4**: 2099-2112

Turvey JR, Griffiths LM (1973) Mucilage from a fresh-water red alga of the genus *Batrachospermum*. Phytochemistry 12: 2901-2907

Urbanowicz BR, Pena MJ, Ratnaparkhe S, Avci U, Backe J, Street HF, Foston M, Li H, O'Neill MA, Ragauskas AJ, Darvill AG, Wyman C, Gilbert HJ, York WS (2012) 4-*O*-methylation of glucuronic acid in *Arabidopsis* glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. Proc Natl Acad Sci USA **109**: 14253-14258

Vidali L (2009) Rapid formin-mediated actin-filament elongation is essential for polarized plant cell growth. Proc Natl Acad Sci USA **106**: 13341-13346

Wood AJ, Oliver MJ, Cove DJ (2000) Bryophytes as model systems. Bryologist **103**: 128-133

TABLES AND FIGURES

Table 2.1. All candidate genes to encode rhamnosyl 3-*O*-methyltransferase.
List of all genes (named here as *KO1* to *KO11*) considered in this dissertation to be candidates to encode a rhamnosyl-3-*O*-methyltransferase, together with the corresponding predicted proteins. Protein ID numbers are referred from the JGI database (http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html). Since a complete genetic map of *P. patens* is not yet available, not all of the gene loci on chromosomes are available. The physical locations of each gene, either shown by chromosome if known otherwise by scaffold, are cited from the JGI database (www.phytozome.net). Sequences of all genes, each labeled with the number of genomic sequence, are located in different chromosomes or scaffolds in the whole genome.

Abbreviation	Protein ID	Physical locations
KO1	169825	Chromosome 08: 16943902 - 16948629
KO2	83743	Scaffold_116:1020026-1020571
KO3	134505	Chromosome 03: 23381202 - 23389110
KO4	153277	Scaffold_391: 85294 - 96668
KO5	163311	Scaffold_54:1106904-1109257
KO6	209593	Scaffold_55: 738519 - 752900
KO7	139716	Chromosome 26: 2315425 - 2320598
KO8	15168	Chromosome 01: 16951217 - 16954773
KO9	116394	Chromosome 23: 3829443 - 3831349
KO10	143295	Chromosome 04: 14012136 - 14016881
KO11	115978	Scaffold_15: 2475216 - 2476718

A) *KO1*

B) *KO2*

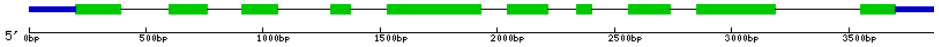
C) *KO3*

D) *KO4*

E) *KO5*

F) *KO6*

G) *KO7*



H) *KO8*

I) *KO9*

J) *KO10*

K) *KO11*

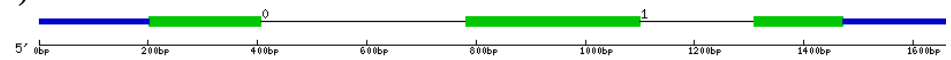
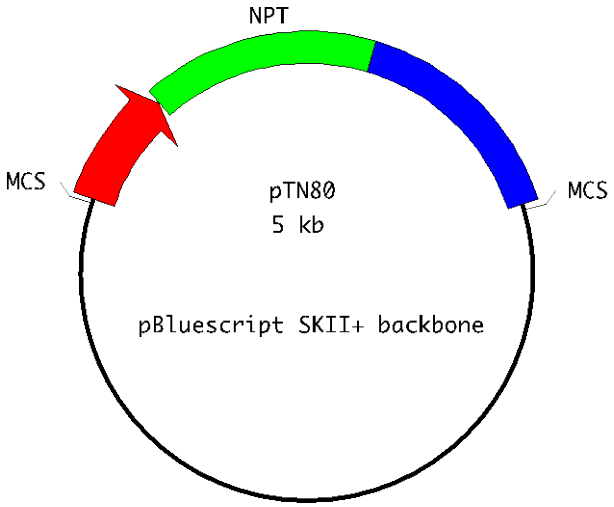


Figure 2.1. Predicted gene structures of all 11 candidate genes.
Predicted gene structures of all 11 *Phycomitrella* genes considered in the present work to be candidates to encode the rhamnosyl-3-*O*-methyltransferase. Green boxes indicate exons, grey lines indicate introns, and blue lines indicate UTR. The length of the whole genes, exons, and introns are drawn in proportion. *KO2* (B) and *KO5* (E) only have EST sequences available, so gene models drawn for these two genes are based on ESTs. The other nine gene structures (A,C,D,F,G,H,I,J,K) are based on gene sequences and transcript sequences. Reference genomic sequences and transcript sequences were downloaded from JGI Physcomitrella (http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html). The structure of each gene was drawn in a gene structure display server (http://gsds.cbi.pku.edu.cn)

A)



B)

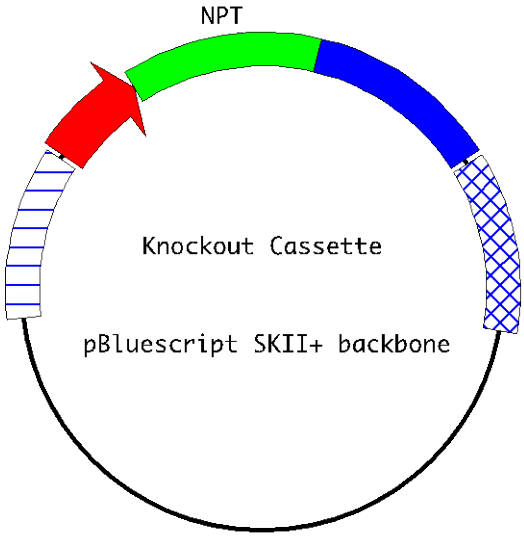


Figure 2.2. Scheme of generation of knockout cassettes.
Schematics of (A) the moss transformation vector pTN80 and (B) the knockout cassette. Knockout cassettes were generated on the backbone of pBluescript SK+. The NPT gene (green bar) was located between multiple cloning sites. The modified 35S promoter (red arrow) can be expressed in both *E. coli* and plants. Left and right genomic fragments (lines and grids) cloned into cassettes at positions flanking the *npt* gene were for promotion of homologous recombination.

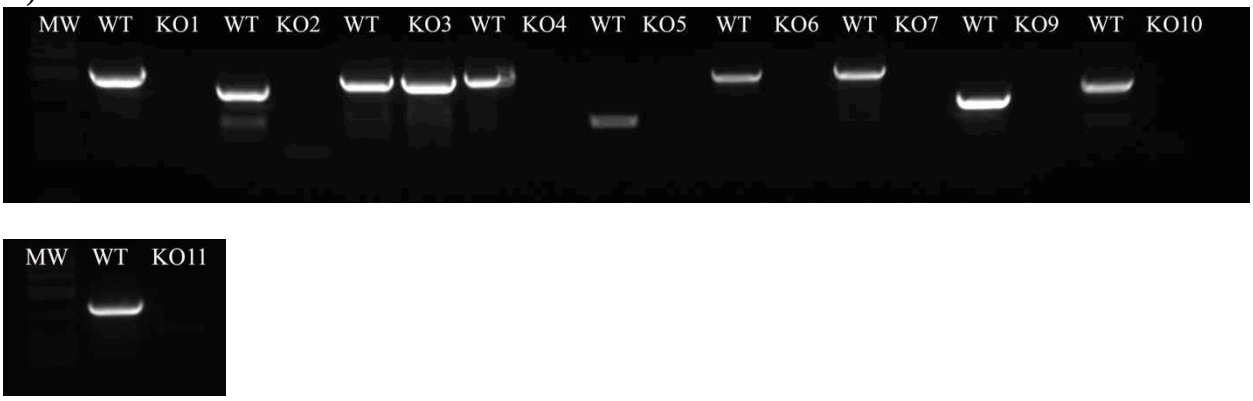
Table 2.2. Forward and reverse primers for generation knockout cassettes.
List of forward and reverse primers that were used in PCR to amplify left and right genomic fragments for the knockout cassettes. Knockout cassettes of each candidate gene were generated by cloning genomic sequences of each gene into the left and right sides of the *NPT* gene in the moss transformation vector pTN80 (shown in Figure 2.2). These flanking genomic sequences integrate into the *Physcomitrella* genome through homologous recombination. Protein ID corresponding to the *KO* number abbreviation is shown, corresponding to Table 2.1.

Genomic fragments in KO cassette	Protein ID	Forward primer	Reverse primer
Left of *KO1*	169825	TAGGGCCCATCGATCTGTTTTA	TACCGCGGCAACGTAGCATGT
Right of *KO1*	169825	TAACTAGTGTGGGATGTTCACCTTC	TAGTCGACGAACCAGACAAGGC
Left of *KO2*	83743	TAGGGCCCTCTTTGCTCTTGT	TGCTCCTATTCCCCAA
Right of *KO2*	83743	TAACTAGTGGGAGGAGGAGAAA	TAGAGCTCACAGATGCCTTCAGT
Left of *KO3*	134505	TAGGGCCCAGTGTCACCTGG	ACCGCGGATAATAGGCTCAA
Right of *KO3*	134505	TAGTCGACTAATGGAAACCGCTAT	TAGAGCTCATTTTGAAGAGTGCTT
Left of *KO4*	153277	AAGGGCCCTAATTTTAGAATT	CAGAATTGAGGAGATTGTAAC
Right of *KO4*	153277	TAACTAGTCAAGGCATGTCCTCTA	TAGAGCTCAACTCGACCACCCT
Left of *KO5*	163311	CCTGCTCGAGGCCAACGG	GGGGTACCTGTTTCCAAGTCAGGGATC
Right of *KO5*	163311	GAGCTCAGGAGTTATGCGGAAG	ACTAGTTGCGTCGGTTGTTTGC
Left of *KO6*	209593	TAGGGCCCCCCTCTATGTGC	TACCGCGGTCATACGGCGA
Right of *KO6*	209593	TAACTAGTCTCAGCAGCCATTCG	TAGTCGACGGGACATCCCAC
Left of *KO7*	139716	TAGGGCCCACAAACCCATCTA	TACCGCGGAAAGAACCCTTC
Right of *KO7*	139716	TAGTCGACTCTGGAGCTAAGACTGG	TAGAGCTCTGGGACGAATGCAC
KO8	15168	GGTTGGATGTGGAGCAGAGT	ATTCCCACCGTTGGTCTACA
Left of *KO9*	116394	TGGGCCCTAACCAGGCAAAGG	TTGCTGATGTCCAGCGCAAT
Right of *KO9*	116394	AAAAGGCGGGCGTTGC	GAGCTCCGGTTACGACCATTT
Left of *KO10*	143295	CAATGCACGGATCTTCAACA	TGAATTGGACATGGCGAGTA
Right of *KO10*	143295	CCACCATATAATTCCGCACA	GTGTAGCGTCGACAGCAGTG
Left of *KO11*	115978	TTAAAGCTTTGAGATAGTCTCTCTT	TTAGGTACCGATTGCACCTCTTCTA
Right of *KO11*	115978	TTAACTAGTTCTGTGCAATGGGATGT	TATCCGCGGTCCCATTGTATGCTT

Table 2.3. Genomic sequence numbers of left and right genomic fragments in each knockout cassette. Numbers shown in the columns of left and right genomic sequences indicate sequence locations of left and right genomic fragments of knockout cassettes in the whole *Physcomitrella* genome. The protein ID corresponding to each *KO* number abbreviation is shown, corresponding to Table 2.1.

Abbreviation	Protein ID	Left genomic sequence	Right genomic sequence
KO1	169825	561097::561721	561799::562423
KO2	83743	1020762::1020355	1020345::1019942
KO3	134505	597134::597765	597780::598384
KO4	153277	93783::94417	94446::95049
KO5	163311	1107326::1107764	1107823::1108223
KO6	209593	740182::739511	739459::738860
KO7	139716	302263::302909	303045::303602
KO8	15168	1407833::1406994	1406993::1405833
KO9	116394	126632::127524	127557::128388
KO10	143295	78362::77505	77473::76623
KO11	115978	2476653::2475790	2475713::2475958

A)



B)

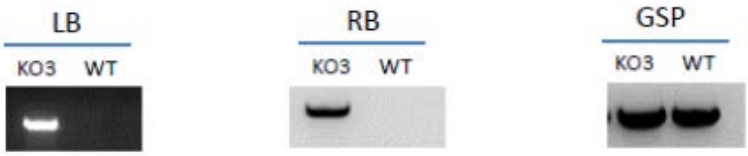


Figure 2.3. Genomic PCR to confirm stable knockouts of candidate genes.
(A) Genomic PCR to confirm stable knockout mutants of candidate genes. Knockout mutants were selected through three rounds of growth on antibiotic medium. Portions of genes considered to be candidates to encode rhamnosyl 3-*O*-methyltransferase were amplified by genomic PCR in wild type (WT) plants, but not in stable knockout mutants, suggesting native gene loci were disrupted by integration of the knockout cassette into the targeted genomic sequences. The *KO3* gene sequence was found to be amplified from both wild type and *ko3* mutants, indicating that the *KO3* gene locus was not completely disrupted. Gene specific primers for genomic PCR were, *KO1* forward- CGGGGAGTGTGCATATTTCT, *KO1* reverse- GGAGCCAAAGGTAGCACAAA; *KO2* forward- GAAGTGTGTGAGTGGAAGGTG, *KO2* reverse- CCCAATCCCAACGACAATAG; *KO3* forward- AGATTCTTGGAAACATTTGCG, *KO3* reverse- GAGAGAGCAACCTGCTCCAGTGCAT; *KO4* forward- TCGTCAATCGATGTTGGAAA, *KO4* reverse- ATTTTCGACCTGTCCGTATGC, *KO5* forward- ATTTTTACGCAAGGCCAATG, *KO5* reverse- AAACGGAACTGGTGACGAAC; *KO6* forward- TCCTTCGGAGGTATCGCTTA, *KO6* reverse- AAGTTGCACTATTGGCAGCA; *KO7* forward-

CAACGGAAACAATTCACACG, *KO7* reverse- CACCATTTCTTTGCGTTTCA; *KO9* forward- GGCTATTCGCTTCTTTGCAC, *KO9* reverse- GCCAGGAAGCTGTTCAACTC; *KO10* forward- TAATGCGCAATGAGGAGATG, *KO10* reverse- ATCGATGCAACGACAGATTG; *KO11* forward- TGCGCTTGCTCTGTTACATT, *KO11* reverse-TGCGCACAAAAACTCTTCAC. (B) The left and right borders of the *KO3* knockout cassette were detected by PCR using, for each border, one gene-specific primer from outside of the *KO* cassette and one primer from the *NPT* sequence. While this PCR confirmed that the *KO3* cassette had inserted into the *KO3* gene locus, the continuing detection of the wild type *KO3* gene fragment indicated that the culture was either diploid or a mixture of wild-type and knockout plants. The PCR primers used for this *KO3* left border (LB) detection were AGATTCTTGGAAACATTTGCG and AGTGGGATTGTGCGTCATCCC. The PCR primers used for this *KO3* right border (RB) detection were GACGACTAAACCTGGAGCCCA and GAGAGAGCAACCTGCTCCAGTGCAT. The primers used for the gene-specific primer (GSP) PCR were the same as used in (A).

Table 2.4. Glycosyl composition of AGPs in *ko1* and wild type.
Glycosyl compositions and 3-*O*-Me-Rha/Rha ratios of total soluble AGP fractions from wild type and *ko1* knockout plants of *Physcomitrella*. Glc contents, which were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations, were excluded prior to normalization of mole percents. Data are averages of n=12 trials. For each trial, a wild type sample was paired with a knockout sample. Four additional trials were excluded due to high pectic polysaccharide content, indicated by high (>30 mol%) GalU content.

Residue or ratio	Glycosyl composition (mol%)	
	Wild type	*ko1*
3-*O*-Me-Rha	4.52	5.08
Ara	17.30	16.39
Rha	1.25	1.94
Fuc	0.64	0.42
Xyl	5.11	4.20
GlcU	12.46	12.50
GalU	1.84	1.53
Man	2.50	2.18
Gal	54.39	55.37
3-*O*-Me-Rha/Rha	3.02	2.19

Paired t test on 3-*O*-Me-Rha/Rha ratios for n=12 trials resulted in a two-tailed P value of 0.0031, considered very significant.

Table 2.5. Glycosyl composition of AGPs in *ko2* and wild type.
Glycosyl compositions and 3-*O*-Me-Rha/Rha ratios of total soluble AGP fractions from wild type and *ko2* knockout plants of *Physcomitrella*. Glc contents, which were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations, were excluded prior to normalization of mole percents. Data are averages of n=8 trials. For each trial, a wild type sample was paired with a knockout sample. All trials showed high pectic polysaccharide content in the *ko2* samples, indicated by high (10 to 65 mol%) GalU content, so these trials could not be excluded.

Residue or ratio	Glycosyl composition (mol%)	
	Wild type	*ko2*
3-*O*-Me-Rha	1.58	1.41
Ara	15.87	9.73
Rha	1.07	1.85
Fuc	0.59	0.84
Xyl	2.44	3.72
GlcU	14.62	11.80
GalU	2.74	27.93
Man	2.54	3.38
Gal	58.57	39.34
3-*O*-Me-Rha/Rha	1.69	0.70

Paired t test on 3-*O*-Me-Rha/Rha ratios for n=8 trials resulted in a two-tailed P value of 0.0329, considered significant.

Table 2.6. Glycosyl composition of AGPs in *ko3* and wild type.
Glycosyl compositions and 3-*O*-Me-Rha/Rha ratios of total soluble AGP fractions from wild type and *ko3* knockout plants of *Physcomitrella*. Glc contents, which were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations, were excluded prior to normalization of mole percents. Data are averages of n=5 trials. For each trial, a wild type sample was paired with a knockout sample. One additional trial was excluded due to high pectic polysaccharide content, indicated by high (23 mol%) GalU content.

Residue or ratio	Glycosyl composition (mol%)	
	Wild type	*ko3*
3-*O*-Me-Rha	5.99	6.26
Ara	18.36	18.08
Rha	1.50	1.28
Fuc	0.38	0.43
Xyl	4.24	3.89
GlcU	13.02	14.20
GalU	1.29	1.53
Man	1.47	1.51
Gal	53.75	52.82
3-*O*-Me-Rha/Rha	3.98	5.19

Paired t test on 3-*O*-Me-Rha/Rha ratios for n=5 trials resulted in a two-tailed P value of 0.3239, considered not significant.

Table 2.7. Glycosyl composition of AGPs in *ko4* and wild type.
Glycosyl compositions and 3-*O*-Me-Rha/Rha ratios of total soluble AGP fractions from wild type and *ko4* knockout plants of *Physcomitrella*. Glc contents, which were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations, were excluded prior to normalization of mole percents. Data are averages of n=3 trials. For each trial, a wild type sample was paired with a knockout sample. No trials were excluded due to high pectic polysaccharide content.

Residue or ratio	Glycosyl composition (mol%)	
	Wild type	*ko4*
3-*O*-Me-Rha	5.83	5.89
Ara	20.35	20.40
Rha	1.78	1.42
Fuc	0.27	0.45
Xyl	3.02	3.38
GlcU	13.00	12.82
GalU	0.88	1.16
Man	1.43	1.62
Gal	53.44	52.87
3-*O*-Me-Rha/Rha	2.92	4.25

Paired t test on 3-*O*-Me-Rha/Rha ratios for n=3 trials resulted in a two-tailed P value of 0.0328, considered significant.

Table 2.8. Glycosyl composition of AGPs in *ko5* and wild type.
Glycosyl compositions and 3-*O*-Me-Rha/Rha ratios of total soluble AGP fractions from wild type and *ko5* knockout plants of *Physcomitrella*. Glc contents, which were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations, were excluded prior to normalization of mole percents. Data are the result of n=1 trial. For this trial, a wild type sample was paired with a knockout sample. No trials were excluded due to high pectic polysaccharide content.

Residue or ratio	Glycosyl composition (mol%)	
	Wild type	*ko5*
3-*O*-Me-Rha	1.76	1.47
Ara	13.96	9.77
Rha	1.16	1.31
Fuc	0.24	0.20
Xyl	4.02	8.56
GlcU	15.82	16.34
GalU	1.04	3.35
Man	1.32	1.52
Gal	60.69	57.48
3-*O*-Me-Rha/Rha	1.52	1.12

No statistical test was possible because only n=1 trial was available.

Table 2.9. Glycosyl composition of AGPs in *ko6* and wild type.
Glycosyl compositions and 3-*O*-Me-Rha/Rha ratios of total soluble AGP fractions from wild type and *ko6* knockout plants of *Physcomitrella*. Glc contents, which were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations, were excluded prior to normalization of mole percents. Data are averages of n=5 trials. For each trial, a wild type sample was paired with a knockout sample. No trials were excluded due to high pectic polysaccharide content.

Residue or ratio	Glycosyl composition (mol%)	
	Wild type	*ko6*
3-*O*-Me-Rha	8.09	3.77
Ara	21.69	17.04
Rha	2.07	1.36
Fuc	0.23	0.55
Xyl	2.77	4.12
GlcU	11.95	14.82
GalU	0.58	1.35
Man	1.18	2.29
Gal	51.45	54.68
3-*O*-Me-Rha/Rha	3.84	2.85

Paired t test on 3-*O*-Me-Rha/Rha ratios for n=5 trials resulted in a two-tailed P value of 0.0140, considered significant.

Table 2.10. Glycosyl composition of AGPs in *ko7* and wild type.
Glycosyl compositions and 3-*O*-Me-Rha/Rha ratios of total soluble AGP fractions from wild type and *ko7* knockout plants of *Physcomitrella*. Glc contents, which were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations, were excluded prior to normalization of mole percents. Data are averages of n=8 trials. For each trial, a wild type sample was paired with a knockout sample. One additional trial was excluded due to high pectic polysaccharide content, indicated by high (25 mol%) GalU content.

Residue or ratio	Glycosyl composition (mol%)	
	Wild type	*ko7*
3-*O*-Me-Rha	2.59	2.98
Ara	14.74	14.98
Rha	0.97	1.58
Fuc	0.78	0.49
Xyl	4.66	5.59
GlcU	15.20	13.47
GalU	7.56	5.43
Man	2.06	2.93
Gal	51.44	52.55
3-*O*-Me-Rha/Rha	2.58	1.83

Paired t test on 3-*O*-Me-Rha/Rha ratios for n=8 trials resulted in a two-tailed P value of 0.0515, considered not quite significant.

Table 2.11. Glycosyl composition of AGPs in *ko9* and wild type.
Glycosyl compositions and 3-*O*-Me-Rha/Rha ratios of total soluble AGP fractions from wild type and *ko9* knockout plants of *Physcomitrella*. Glc contents, which were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations, were excluded prior to normalization of mole percents. Data are averages of n=13 trials. For each trial, a wild type sample was paired with a knockout sample. No trials were excluded due to high pectic polysaccharide content.

Residue or ratio	Glycosyl composition (mol%)	
	Wild type	*ko9*
3-*O*-Me-Rha	6.02	4.08
Ara	19.99	16.38
Rha	1.49	1.57
Fuc	0.37	0.61
Xyl	3.44	4.81
GlcU	13.88	14.37
GalU	1.31	2.57
Man	1.96	2.68
Gal	51.54	53.01
3-*O*-Me-Rha/Rha	4.28	2.45

Paired t test on 3-*O*-Me-Rha/Rha ratios for n=13 trials resulted in a two-tailed P value of 0.0027, considered very significant.

Table 2.12. Glycosyl composition of AGPs in *ko10* and wild type.
Glycosyl compositions and 3-*O*-Me-Rha/Rha ratios of total soluble AGP fractions from wild type and *ko10* knockout plants of *Physcomitrella*. Glc contents, which were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations, were excluded prior to normalization of mole percents. Data are averages of n=2 trials. For each trial, a wild type sample was paired with a knockout sample. No trials were excluded due to high pectic polysaccharide content.

Residue or ratio	Glycosyl composition (mol%)	
	Wild type	*ko10*
3-*O*-Me-Rha	10.81	5.95
Ara	24.18	18.85
Rha	2.42	1.18
Fuc	0.25	0.55
Xyl	2.11	5.58
GlcU	10.94	15.90
GalU	0.43	1.40
Man	1.17	1.78
Gal	47.71	48.82
3-*O*-Me-Rha/Rha	4.47	5.05

Paired t test on 3-*O*-Me-Rha/Rha ratios for n=2 trials resulted in a two-tailed P value of 0.8134, considered not significant.

Table 2.13. Glycosyl composition of AGPs in *ko11* and wild type.
Glycosyl compositions and 3-*O*-Me-Rha/Rha ratios of total soluble AGP fractions from wild type and *ko11* knockout plants of *Physcomitrella*. Glc contents, which were primarily due to variable amounts of residual (β-D-Glc)$_3$ Yariv phenylglycoside in the preparations, were excluded prior to normalization of mole percents. Data are averages of n=4 trials. No trials were excluded due to high pectic polysaccharide content.

Residue or ratio	Glycosyl composition (mol%)	
	Wild type	*ko11*
3-*O*-Me-Rha	7.75	7.92
Ara	20.21	22.20
Rha	1.60	2.13
Fuc	0.37	0.58
Xyl	4.18	4.08
GlcU	12.06	15.43
GalU	0.99	4.58
Man	1.65	2.00
Gal	51.19	41.09
3-*O*-Me-Rha/Rha	4.83	4.15

Unpaired t test on 3-*O*-Me-Rha/Rha ratios for n=4 trials resulted in a two-tailed P value of 0.6823, considered not significant.

Table 2.14. Summary of all 3-*O*-Me-Rha/Rha ratios.
Effects of knockouts of 10 candidate genes on 3-*O*-Me-Rha/Rha content of *Physcomitrella* AGPs were summarized. See Tables 2.4-2.13 for complete results and statistical tests.

3-O-Me-Rha/Rha ratio			
Significantly reduced	Significantly reduced but high GalU	Significantly increased	No significant effect
ko9 (P=0.0027) *ko1* (P=0.0031) *ko6* (P=0.0140)	*ko2* (P=0.0329)	*ko4* (P=0.0328)	*ko3* *ko5* *ko7* *ko10* *ko11*

CHAPTER 3 – FUNCTIONAL ANALYSIS OF KO1 IN THE MOSS *PHYSCOMITRELLA PATENS* AND IN *NICOTIANA TABACUM* CV XANTHI

ABSTRACT

The gene *KO1* was identified in Chapter 2 of this dissertation as one of the candidates to encode the rhamnosyl 3-*O*-methyltransferase that synthesizes 3-*O*-Me-Rha in the moss *Physcomitrella patens*. To begin to characterize the function of *KO1*, the gene model for *KO1* was analyzed and revised. A phylogenic tree was constructed and indicated that KO1 was evolved from cyanobacteria and contained a LpxB domain characteristic of gram-negative bacteria. The *ko1* knockout mutant generated by HR in Chapter 2 was further examined by PCR and RT-PCR to confirm the placement of the knockout cassette. Phenotypic analysis showed that polarized tip growth was disrupted in the *ko1* knockout. Protonemal filaments grew in a curlier path in the knockout than in the wild-type. In the leafy gametophyte stage, the *ko1* knockout had fewer and shorter lateral rhizoids than did the wild-type. Although the primary screening showed that the 3-*O*-Me-Rha/Rha content ratio was statistically very significantly reduced in AGPs from the *Physcomitrella ko1* knockout mutant compared to AGPs from the *Physcomitrella* wild-type, transgenic expression of the *KO1* gene in *Nicotiana tabacum* cv Xanthi did not result in detectable 3-*O*-Me-Rha content in the AGPs of the transgenic tobacco plants.

INTRODUCTION

Arabinogalactan-proteins (AGPs) are a family of highly glycosylated hydroxyproline-rich glycoproteins (HRGPs) that are found bound to the plasma membrane, bound to the cell wall, and soluble in the cell wall space of probably all plant cells (Jauh & Lord, 1996; Serpe & Nothnagel 1999; Seifert & Roberts 2007). Most AGPs consist of a single core polypeptide to which are attached several or many galactosyl- and arabinosyl-rich glycans. Although many studies have focused on the core polypeptides of AGPs (Chen et al., 1994; Du et al., 1994; Schultz et al., 2002), less progress has been made on biosynthesis of the arabinogalactan glycans (Liang et al., 2010; Wu et al., 2010; Basu et al., 2013; Geshi et al., 2013). The present study is motivated by our earlier finding that the glycans on AGPs in the moss *Physcomitrella patens* contain 3-*O*-Me-Rha in terminal non-reducing positions amounting to as much as 15 mol%, a remarkably high level for a methylated sugar residue (Fu et al., 2007). Our bioinformatics searches identified several *Physcomitrella* genes that could be considered to be candidates to encode the rhamnosyl 3-*O*-methyltransferase involved in the synthesis of this 3-*O*-Me-Rha. Primary screening of knockout mutants of these candidate genes revealed that the *ko1* knockout mutant had statistically very significantly reduced 3-*O*-Me-Rha/Rha content ratio in its AGPs compared to the AGPs of the wild-type moss (data in Chapter 2).

The moss *Physcomitrella patens* is a valuable model organism for research into plant cell, developmental, and evolutionary biology (Cove & Knight, 1993; Vidali & Benzanilla, 2012). The *Physcomitrella patens* genome was first completely sequenced and assembled by Rensing et al. (2008) and continues to undergo further refinement. The whole genome of the moss is approximately 511 MB, about twice as large as the *Arabidopsis* genome, and consists of 27 chromosomes (n=27). Studies of the genome, EST sequences, and the transcriptome have revealed that *Physcomitrella patens* shares a high degree of homology with *Arabidopsis* (Nishiyama et al., 2003). A variety of genes that have been cloned from *Physcomitrella* are highly homologous to genes found in more evolved plants including angiosperms.

The dominant portion of the *Physcomitrella* life cycle, from the germination of spores to the fertilization of the egg, is the haploid gametophyte. Early in the gametophytic stage, protonema grow in two filamentous cell types, chloronema and caulonema. Both types of filaments are uniseriate, consisting of one row of end-to-end cells with occasional branching. The first filamentous cells that grow from germinated spores are chloronema, which contain many well-developed chloroplasts that make chloronemal tissue green. In contrast, the caulonemal cells, which usually begin to appear about 7 days after spore germination, have far fewer chloroplasts but grow faster (Schmiedel & Schnepf, 1980; Menand et al., 2007). Filamentous protonemal tissues later

form buds that differentiate into leafy gametophytes. Rhizoids, filamentous cells which seem to be most closely related to caulonemal cells, grow out from surface cells of leafy stems (Pressel et al., 2008; Jang & Dolan, 2011). Caulonemal cells and rhizoid cells seem to have similar cellular organization, but the rhizoid cells are pigmented reddish-brown. Growth of both protonemal cells and rhizoids is extremely polarized towards the cell apex (Lee et al., 2005; Jones & Dolan, 2012; Vidali & Bezanilla, 2012). Exocytosis in protonemal cells has been observed in the moss (Pressel et al., 2008).

This dissertation chapter is focused on functional characterization of the *Physcomitrella KO1* gene. Further study of the *KO1* gene expression profile through searches in an updated database have led to revision of the *KO1* gene structure. A phylogenetic tree was generated to elucidate evolutionary relationships between KO1 and its orthologs. Phenotypic studies revealed abnormal polarized tip growths, including a more curly protonemal growth pattern and fewer and shorter rhizoids, in the *ko1* knockout mutant compared to the wild-type moss. The *KO1* gene was then cloned to a binary vector for transgenic expression in tobacco. The glycosyl composition of the AGPs from the *KO1* transgenic tobacco plants was analyzed.

MATERIALS AND METHODS

Plant materials. *Physcomitrella patens* (Hedw.) Br. Eur. spores were kindly provided by Karen Schumaker (University of Arizona). Wild-type and stable knockout cultures of *Physcomitrella* in predominantly leafy gametophytic form were routinely cultured on KP medium, either liquid or agar-solidified, with transfers at approximately monthly intervals.. To maintain gametophytic cultures in predominantly protonemal form, tissue was cut to small pieces in a Polytron homogenizer and grown on BCDAT medium overlaid with cellophane (Sigma-Aldrich, St. Louis, MO, USA). Normal conditions in the culture room were 21°C and 16/8 h light/dark photoperiod with 40-60 µmol m^{-2} s^{-1} intensity during the light period. Formulations of the growth media were as follows.

KP medium. A modified Knop's medium prepared exactly as described by Fu et al. (2007).

Knop medium. 1.84 mM KH_2PO_4, 3.35 mM KCl, 1.0 mM $MgSO_4$, 6.09 mM $Ca(NO_3)_2$, 45 µM $FeSO_4$, KOH or HCl as needed to give pH 5.8. For solid medium, 0.8% (w/v) agar (Sigma, A1296) added. Medium sterilized by autoclaving (Reski & Abel, 1985).

BCDAT medium. 1 mM $MgSO_4$, 10 mM KNO_3, 45 µM $FeSO_4$, 1.8 mM, KH_2PO_4 (pH 6.5 adjusted with KOH), 1mM $CaCl_2$, 0.22 µM $CuSO_4$, 0.19 µM $ZnSO_4$, 10 µM

H_3BO_3, 0.1 µM, Na_2MoO_4, 2µM $MnCl_2$, 0.23 µM $CoCl_2$, and 0.17 µM KI, 5 mM diammonium tartrate, 0.8% (w/v) agar (Sigma, A1296). Medium sterilized by autoclaving (Ashton & Cove, 1977).

Plasmid constructions. Construction of the KO1 (Phypa_169825) knockout cassette, along with the other 10 knockout cassettes used in this dissertation, has been described in detail in Chapter 2. Because of its relevance to the revised *KO1* gene structure that will be presented in this chapter, however, the construction of the *KO1* knockout cassette will be briefly summarized here.

Construction of the *KO1* knockout cassette was based on the moss transformation vector pTN80 (gift from Dr. Mitsuyasu Hasebe, National Institute for Basic Biology, Japan, GI: 124377588) that contained a modified CaMV35S promoter-npt ii CDS-NOS terminator gene construct that confers resistance to G418 antibiotic. The modified 35S promoter can be expressed both in plants and *E.coli*. Two partial genomic sequences of *KO1* were cloned into the left and right multiple cloning sites of the pTN80 vector. Those homologous gene sequences were PCR amplified from *Physcomitrella patens* genomic DNA with exTaq (Takara Bio., Mountain View, CA, USA). After purification by gel electrophoresis, the PCR fragments were cloned into the pGEM T-easy vector (Promega Corp., Wisconsin, WI, USA). Positive colonies were selected by colony PCR and enzyme digestion and confirmed by sequencing (Genomics Core, UC Riverside). One 624 bp

genomic fragment was cloned from a pGEM T-easy vector into a flanking region of pTN80 using ApaI and EcoRI sites. Another 624 bp fragment genomic fragment was inserted from a pGEM T-easy vector into the other flanking region of pTN80 using SpeI and NotI sites. The ApaI, EcoRI, SepI, and NotI restriction endonucleases were from NEB Corp. (Ipswich, MA, USA). The PCR primers used to amplify the two 624 bp genomic sequences were TAGGGCCCATCGATCTGTTTTA, TACCGCGGCAACGTAGCATGT; and TAACTAGTGTGGGATGTTCACCTTC, TAGTCGACGAACCAGACAAGGC. Oligonucleotides were synthesized by IDT-DNA Corp. (Coralville, Iowa).

The cDNA clone pdp32i13 was purchased from the RIKEN center (Saitama, Japan, http://www.brc.riken.jp/lab/epd/Eng/). The open reading frame (ORF) of this cDNA was PCR amplified with exTaq using the primers TTACTCGAGAGAGTGAGGGCGTAC and TTAAAGCTTGGCAGAGCGCTCAAGT, and then cloned into the pART7 vector at XhoI and Hind III sites. The fragment between two NotI sites was digested from pART7 and inserted into pART27 for expression in tobacco. The pART7 and pART27 vectors (Gleave, 1992) were gifts from Dr. Martha L. Orozco (University of California, Riverside)

Protoplast transformation. The method for introducing the linearized *KO1* knockout cassette into *Physcomitrella* protoplasts by polyethylene glycol (PEG)

mediated transformation (Schaefer et al., 1991; Cove et al., 2009) has been described in detail in Chapter 2.

Knockout mutant selection. The method for selecting the *Physcomitrella ko1* knockout mutants by growth on BCDAT medium containing G418 antibiotic has been described in detail in Chapter 2.

Tobacco transformation. Transformation of *Nicotiana tabacum* cv Xanthi was performed by the Plant Transformation Research Center at the University of California, Riverside using the pART7/pART27 binary vector system (Gleave, 1992) and standard *Agrobacterium*-mediated methods (Duca et al., 2009). See a previous above section for insertion of the *KO1* ORF into pART7 and pART27.

DNA extraction from Physcomitrella *and analysis of stable ko1 knockout mutants*. A cetyltrimethylammonium bromide (CTAB)-based method was used for extraction of genomic DNA from *Physcomitrella* (Aono et al., 2012). Genomic DNA of the wild-type and the knockout mutants was extracted from one-month-old leafy gametophyte cultures. The gametophytes (approximately 0.2-0.4 gfw) were ground in liquid nitrogen in a mortar with a pestle. The ground, still frozen tissue powder was put into a 1.5 ml microcentrifuge tube with 400 µl of 2x CTAB buffer [2% (w/v) CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 8.0] and then incubated in a 60ºC water bath for 1 h.

The DNA was then extracted with an equal volume of chloroform: isoamylalcohol (25:1, v/v). After centrifugation (15,000 RPM, 10 min), the aqueous (upper) phase was collected and transferred to a clean microcentrifuge tube. An equal volume of isopropanol was then added to precipitate the DNA. After centrifugation (15,000 RPM, 10 min), the resulting supernatant was discarded and the pellet was washed with 70% (v/v) ethanol by centrifugation. The final pellet was air-dried and dissolved in 100 µl of TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA) containing 1µg/ml RNase A (Sigma-Aldrich, St. Louis, MO, USA).

For analysis of stable *ko1* knockout mutants, genomic DNA was used as the template with a PCR temperature program of 35 cycles of 94ºC for 1 min, 54 ºC for 1 min, and 72 ºC for 2 min with Promega 2X Green Master Mix (Promega Corp. Wisconsin, WI, USA). The primers designed to map the insertion junctions of knockout cassette were: P1-CGGGGAGTGTGCATATTTCT, P2-TTACGTCAGTGGAGATGTCAC, P3-ACGAGACGACTAAACCTGGA, and P4-GGAGCCAAAGGTAGCACAAA. Fig. 3.5 shows the locations of these primers on the KO1 gene.

RNA extraction from Physcomitrella *and RT-PCR for analysis of KO1 gene expression level.* Total RNA was extracted from *Physcomitrella* through use of the RNeasy Plant Mini Kit (Qiagen Inc, Valencia, CA, USA). One microgram of total RNA, treated with DNase I (Qiagen Inc, Valencia, CA, USA), was used to synthesize first

strand cDNA with Superscript III according to the manufacturer's protocol (Life Technologies, Carlsbad, CA, USA). Two microliters of first strand cDNA were then used as the template in RT-PCR to detect the *KO1* gene expression level. The RT-PCR primers for *KO1* were AGCGCCGGTAATTCAAAAG, and TCCAGAGCGTGTTGTCGTAG.

DNA and RNA extraction from tobacco and analysis of transgenic tobacco plants. DNA and RNA were extracted from leaves of transgenic tobacco using same methods as described above for *Physcomitrella* samples. The primers used in PCR and RT-PCR were Forward- TGCGCTTGCTCTGTTACATT, Reverse-TGCGCACAAAAACTCTTCAC

AGP extraction from transgenic tobacco plants. Leaves from young regenerated tobacco plants were processed for purification of total soluble AGPs and analysis of AGP glycosyl composition using the methods of Fu et al. (2007), modified to accommodate micro-scale samples of 0.25-0.5 gfw. Samples were homogenized by grinding in a mortar and pestle with liquid nitrogen, and then the frozen powder was added to 4.5 ml of a pH 7.5 buffer of 50 mM tris (hydroxymethyl) aminomethane-HCl, 10 mM KCl, 1 mM EDTA, 0.1 mM $MgCl_2$, 8% (w/v) sucrose, 1 mM phenylmethanesulfonyl fluoride, and 0.1% (v/v) plant protease inhibitor (Sigma-Aldrich, St. Louis, MO, USA). After stirring for 5 min, the homogenate was centrifuged at 35,600xg for 30 min at 4°C. The resulting supernatant was filtered through Whatman #1 paper and collected in a glass conical

centrifuge tube (Kimax 45200 with Teflon-lined screw cap). To each filtrate was added $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside, NaN_3 solution, and dry NaCl to give final concentrations of 60 µM, 0.02% (w/v), and 1%(w/v), respectively. The specimens were then incubated overnight at 4°C to allow formation and precipitation of the AGP-Yariv phenylglycoside complex. The following morning, the specimens were centrifuged at 1,700xg for 10 min at room temperature. The supernatants were discarded, and the pellets were extracted with 0.5-ml aliquots of distilled water. After mixing and centrifugation, the supernatants were poured off and saved. Extraction of the pellets with 0.5-ml aliquots of distilled water, and saving and pooling of the supernatants individually for each specimen, was continued until essentially all of the red color was in the pooled supernatants. To the pooled supernatant for each specimen was added NaN_3 solution and dry NaCl to give final concentrations of 0.02% (w/v) and 1% (w/v), respectively, and the specimens were then incubated overnight at 4°C to allow formation and precipitation of the AGP-Yariv phenylglycoside complex. This process was repeated for a total of three additions of dry NaCl. After the centrifugation after the third addition of NaCl, the supernatants were discarded, and the pellets were dissolved in dissolved in minimal (0.02 or 0.05 ml) dimethylsulfoxide, with sonication to achieve full solubility. Three volumes (0.06 or 0.15 ml) of acetone and 0.08 volumes (1.6 or 4 µl) of 1% (w/v) NaCl solution were then added to each specimen, the tubes were quickly closed with Teflon-lined caps,

vortexed, and incubated overnight at 4°C to allow precipitation of the AGPs while leaving the Yariv phenylglycoside in the supernatant. The following morning, the specimens were centrifuged at 1,700xg for 10 min at room temperature. The supernatants were discarded, and the pellets were again dissolved in minimal dimethylsulfoxide and then precipitated by addition of acetone and NaCl solution. This procedure for extraction of Yariv phenylglycoside was usually repeated for a total of three additions of dimethylsulfoxide, although for some very small specimens only one or two additions of dimethylsulfoxide were used. After the centrifugation after the final addition of dimethylsulfoxide, acetone, and NaCl solution, the supernatants were discarded, and the pellets were washed twice by addition of 0.3 ml of acetone and centrifugation. The pellets after the second acetone wash were dried and then dissolved in 0.2 ml of distilled water. The 0.2 ml solutions of purified AGPs were then split to two vials with 0.1 ml being processed for GC and the other 0.1 ml being stored frozen as back-up.

Analysis of glycosyl composition of AGPs. Glycosyl compositions of AGPs were determined by gas chromatography (GC) and gas chromatography-mass spectrometry (GC-MS) of trimethylsilyl (TMS) ether-methylglycoside derivatives of sugars after methanolysis. The methods were as described by Fu et al. (2007), with minor modifications. To each vial of AGP sample, 100 nmoles of *myo*-inositol were added as an internal standard. Another 4 ml vial was prepared to contain 100 nmoles of each of 10

sugar standards, including inositol. The samples were dried by evaporation under a stream of filtered air in a 40 ºC water bath and then placed in a 40 °C vacuum oven overnight for further drying. On the next afternoon, 400 µl of 1.5 M methanolic-HCl and 100 µl of methyl acetate were added to each vial. Teflon-lined screw caps were tightly secured to the vials, which were then placed in a 80°C dry heater block for overnight incubation. On the next morning, five drops of *t*-butanol were added to each vial, and the samples were blown dry with a stream of filtered air in a water bath at room temperature. To form TMS ether derivatives of the methyglycosides, 5 drops of Tri-Sil Reagent (Thermo-Fisher Scientific) were added to each vial. The Teflon-lined screw caps were secured on the vials, and the vials were manually rotated for 15 min to insure contact of the reagents with the methylglycosides on the walls of the vial. The samples were then evaporated just to dryness by blowing with filtered air in a water bath at room temperature. Each dried sample was dissolved in iso-octane (200 µl for sugar standards, 50 µl for AGP samples) just before injection of 1-4 µl into the GC or GC-MS.

Light microscopy of ko1 knockout and wild-type Physcomitrella. Protonemal filaments of live 7-day old cultures were observed and counted with a Leica MZIII Pursuit stereo microscope fitted with a 4MP RGB/gray SPOT camera. Lateral rhizoids of live 3-month-old gametophytes were photographed with a Nikon D5000 SLR camera

with a macro imaging lens. Both instruments were at the Core Microscopy Facility of the Center for Plant Cell Biology at UCR.

Construction of a phylogenetic tree for KO1. KO1 homologs were searched in the UniprotKB protein database (http://www.uniprot.org/) by BlastP. The resulting sequences of found homologs were downloaded and input to the ONE CLICK model at Phylogeny.fr (http://www.phylogeny.fr) for multiple sequence alignment and construction of phylogenetic tree (Dereeper et al., 2008).

RESULTS

Search of the Physcomitrella *transcriptome for the KO1 cDNA and revision of the KO1 gene model.* In the Joint Genome Institute (JGI) v1.1 *Physcomitrella* genome (http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html), which became available in 2008 near the start of this dissertation project, *KO1* was proposed to have a transcript of 3.4 kb length. Despite many efforts early in this dissertation project, however, a 3.4 kb transcript of *KO1* could not be cloned. Searches for a cDNA clone of *KO1* were also made in public *Physcomitrella* transcriptome databases. The RIKEN center (http://www.brc.riken.jp/lab/epd/Eng/) database contains cDNA libraries from moss protonema tissue, leafy gametophytes, and sporophytes. Searches of those cDNA libraries found several transcripts that matched portions of the predicted *KO1* transcript sequence (Fig. 3.1A). The COSMOSS (http://www.cosmoss.org/) transcriptome database for *Physcomitrella* was also searched and yielded two similar transcripts that overlapped with cDNA clone pdp32i13 that had been found in the RIKEN cDNA libraries (Fig. 3.1B). Clone pdp32i13 was then purchased from the RIKEN center and sequenced. Analysis of the full-length sequence of pdp32i13 showed that the *KO1* transcript was 2.1 kb long (Fig. 3.2A). The NCBI ORF finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) found six possible reading frames for this transcript. The results are shown in Figure 3.2B, where only the second 5’ to 3’ reading

frame was considered to be correct. This second reading frame contained 5' and 3' UTRs bordering a long coding region. Figure 3.2C shows the translated amino acid sequence that was also obtained from ORF finder. After alignment of the Figure 3.2A transcript sequence to the original JGI v1.1 genomic sequence for the *KO1* gene, a revised gene model for *KO1* was drawn (Fig. 3.3B). Compared to the original model that predicted the *KO1* gene to be 8.9 kb long and containing 22 exons (Fig. 3.3A), the revised model predicts the *KO1* gene to be only about 5 kb long and containing 11 exons (Fig. 3.3B), which encode a predicted protein of 536 amino acids (Fig. 3.2C).

Predicted KO1 protein contains a LpxB domain. The predicted KO1 protein sequence was analyzed at the NCBI conserved domain database website (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) which found the sequence to contain a LpxB domain. The LpxB domain is the characteristic conserved domain of lipid A disaccharide synthase. A BlastP search for KO1 orthologs was also performed at the UniprotKB protein database (http://www.uniprot.org/). Except for one ortholog found in a red alga, the other KO1 orthologs were LpxB or putative lipid A disaccharide synthase genes in bacteria (Table 3.1). The length of these KO1 orthologs ranged from 368 to 501 amino acids, while the identity to the KO1 protein ranged from 27% to 43%. Most of the E-values of these orthologs from the BlastP search were much smaller than 7E-8.

Amino acid sequences of the KO1 orthologs in Table 3.1 were downloaded from the UniprotKB protein database and uploaded to the ONE CLICK model at Phylogeny.fr (http://www.phylogeny.fr) for phylogenetic analysis (Castresana, 2000; Guindon & Gascuel, 2003; Edgar, 2004; Anisimova & Gascuel; 2006; Dereeper et al., 2008; 2010). MUSCLE alignment was utilized to draw a phylogenetic tree of KO1 and its orthologs (Fig. 3.4). The evolutionarily closest neighbor to KO1 (Phypa_169625) in the phylogenetic tree was a protein from *Galdieria sulphuraria*, a red algae.

Generation of stable KO1 knockout mutant. To study the function of *KO1*, stable knockout mutants of *KO1* gene were generated. A knockout cassette (Fig. 3.5A) consisting of a NPT selection marker and bordering *KO1* genomic sequences was transformed into moss protoplasts. After regeneration of growing protonema, stable knockout mutants were isolated through three rounds of selection on medium containing G418 antibiotic. PCR performed with primers that spanned from the *NPT* gene to genomic sequence outside the borders of the knockout cassette (Fig. 3.5A) showed that the expected knockout cassette-to-genome junctions were present in the *ko1* knockout mutant but absent in the wild-type (Fig. 3.5B). Similarly confirming, PCR conducted with primers for genomic sequences before and after the expected borders of the knockout cassette produced the expected amplified product with genomic DNA from wild-type plants but not with genomic DNA from *ko1* mutants (Fig. 3.5B). RT-PCR

performed with total RNA extracted from wild-type *Physcomitrella* and from the *ko1* showed that *KO1* mRNA was detectable in the wild-type but not in the *ko1* (Fig. 3.5C), indicating that the *KO1* gene was knocked out in mutant.

Curly protonema phenotype of the ko1. Protonemal tissue of *Physcomitrella patens* includes chloroplast-rich chloronema cells and fast-growing caulonema cells. These two types of cells expand by polarized tip growth. Wild type protonemal filaments were observed to extend outward from the center of the colonies with the apex of the filament growing relatively straight ahead (Fig. 3.6A-C). A few filaments in wild-type colonies extended outward with the apex exhibiting a slightly curly or wavering growth path. In the *ko1* colonies, however, most of the protonemal filaments exhibited the curly or wavering growth path (Fig. 3.6 D-E). To quantitate this phenotype, colonies with only straight filaments were counted as normal colonies. Colonies containing at least some curly filaments were counted as curly colonies. By these criteria, 83% of the total wild-type colonies were normal, and 17 % of the wild-type colonies were curly. In contrast, all of the *ko1* colonies were curly (Table 3.2). Within the 17% curly wild-type colonies, only 6% of the outgrowing filaments were curly, and the remaining 94% of the outgrowing filaments were relatively straight. Within the *ko1* colonies, all of which were counted as curly, 82% of the outgrowing filaments were curly (Table 3.3). Overall, 86% of the filaments in wild type colonies exhibited normal, relatively straight, tip growth. Only 19%

of the outgrowing filaments in *ko1* colonies exhibited straight growth, while 81% exhibited curly or wavering growth (Table 3.4).

Reduced lateral rhizoid growth phenotype of the ko1. Rhizoids of mosses are uniseriate filaments, consisting of one row of end-to-end cells with growth occurring at the tip of the apical cell. Rhizoids serve to anchor the moss but have no known uptake functions, unlike roots of more derived plants that both anchor the plant and function in uptake of water and minerals. Rhizoids commonly grow out laterally from surface cells of leafy stems (Pressel et al., 2008; Jang & Dolan, 2011; Jang et al., 2011). The *ko1* showed a reduced rhizoid growth phenotype. Lateral rhizoids were shorter in leafy gametophytes of the *ko1* than in the wild type (Fig. 3.7A). To quantitate this phenotype, the whole leafy stem was divided into upper, middle and lower segments (Fig. 3.7A), and the numbers and lengths of rhizoids in each segment were determined. Wild-type plants had more (Fig. 3.7B) and longer (Fig. 3.7C) rhizoids than did the *ko1*. For both the wild-type and the *ko1*, the largest number of rhizoids and the longest rhizoids occurred in the middle segment.

Transgenic expression of KO1 in Nicotiana tabacum cv Xanthi. To further test the function of *KO1*, this *Physcomitrella* gene was transgenically expressed in another species. Tobacco has been shown to synthesize arabinogalactan proteins with rhamnosyl residues in terminal non-reducing positions on the glycan chains (Tan et al., 2010), the

same positions in which 3-*O*-Me-Rha is found in *Physcomitrella* AGPs. Because tobacco is readily transformed, and because preliminary studies showed that AGPs purified from wild-type tobacco did not contain a detectable level of 3-*O*-Me-Rha, this plant was selected for the transgenic expression study.

The *KO1* cDNA was cloned into the binary vector pART27, where expression could be expected to be driven to high level by the 35S promoter, for transformation into tobacco. Following transformation, 30 plants regenerated and grew on Kanamycin selection medium. These 30 regenerated plants were tested for the presence of the KO1 gene by genomic PCR, which indicated that 14 of the 30 plants contained the *KO1* gene in the tobacco genome, although the PCR band for six of the 14 plants was rather weak (Fig. 3.8). Two of these 14 plants were not growing healthy in the growth chamber, were not successfully transferred to the green house, and thus were not further tested. RT-PCR was performed on the remaining 12 plants and showed that 10 of the transgenic tobacco plants had detectable *KO1* mRNA accumulation (Fig. 3.9). Although transgenic plants *KO1-5* and *KO1-16* showed detectable *KO1* mRNA accumulation, they were not successfully transferred into green house, so they were not further studied.

Total soluble AGPs were purified from eight of the transgenic tobacco plants that showed *KO1* mRNA accumulation, as well as from wild-type tobacco (negative control) and wild-type *Physcomitrella* (positive control) plants. The preparations from both the

wild-type and the transgenic *KO1* tobacco plants had glycosyl compositions that included high levels of Gal, Ara, and GlcU, as is typical of AGPs. The AGPs preparations from all of the tobacco plants also contained approximately 5 mol% of Rha residues. These Rha residues should have been good substrates for a rhamnosyl 3-*O*-methyltransferase enzyme, but no 3-O-Me-Rha was detected in AGPs from any of the eight transgenic *KO1* tobacco plants.

DISCUSSION

The *KO1* gene was predicted to have a very complicated gene structure in the JGI draft v1.1 *Physcomitrella* genome. The KO1 protein was annotated in the JGI database as having some similarity to bombesin-like peptide while also containing a LpxB domain, which is characteristic of gram-negative bacteria. Bombesin-like peptide has been proposed to be involved in a neuropeptide signaling pathway in animals. No evidence has been found, however, to suggest that such a signaling pathway exists in plants. As predicted in the JGI v1.1 *Physcomitrella* genome, the full-length *KO1* cDNA was about 3.4 kb long. Efforts in this dissertation project were unsuccessful in cloning a 3.4 kb transcript of *KO1*. The RIKEN center and the COSMOSS database both have moss transcriptome sequences available (Fig. 3.1). Analyses of these available ESTs and cDNAs led to the proposal of a revised gene model for *KO1* in this dissertation (Figs. 3.2, 3.3). The revised *KO1* gene model proposes a truncated version of the original JGI v1.1 *KO1* gene model, encoding a polypeptide largely matching the C-terminal half of the KO1 polypeptide originally predicted by JGI v1.1. The KO1 polypeptide predicted by the revised gene model still has similarity to a LpxB domain but no indication of involvement in a neuropeptide signaling pathway. The preliminary release of *Physcomitrella* genome v3.0 (www.phytozome.net) includes a gene model for *KO1* that is consistent with the revised *KO1* model developed in this dissertation project.

The KO1 polypeptide sequence predicted by the revised gene model (Fig. 3.2C) contains a LpxB domain, as recognized through analysis at the NCBI conserved domain database website. The LpxB domain is the characteristic conserved domain of lipid A disaccharide synthase. Lipid A is the hydrophobic part of lipopolysaccharides and serves to anchor the lipopolysaccharide to the outer membrane of gram-negative bacteria (Louis et al., 2009). Lipid A is also responsible for the endotoxicity of lipopolysaccharide to animal cells (Raetz & Whitfield, 2002) and can stimulate immunoreaction of animals to bacteria. Nine enzymes have been shown to participate in the biosynthesis of lipid A (Opiyo et al., 2010). Cloned in 1986, LpxB is one of these nine enzymes and is a glycosyltransferase called lipid A disaccharide synthase (Crowell et al., 1986). LpxB catalyzes the joining of UDP-2,3-bis(3-hydroxytetradecanoyl)glucosamine to 2,3-bis(3-hydroxytetradecanoyl)-β-D-glucosaminyl-1-phosphate to form 2,3-bis(3-hydroxytetradecanoyl)-D-glucosaminyl-1,6-β-D-2,3-bis(3-hydroxytetradecanoyl)-β-D-glucosaminyl 1-phosphate while releasing UDP as the second product of the reaction (Raetz & Whitfield, 2002; Metzger & Raetz, 2009).

Fully assembled lipid A has not been found in plants. Orthologs of five Lpx enzymes including LpxB, however, have been recently identified in *Arabidopsis* (Li et al., 2011). Mutant analysis of those *Lpx* genes suggested that *Arabidopsis* synthesizes at least some precursors of lipid A. It has been proposed that lipid A precursors may function in

signal transduction, since the lipid A receptor TLR4 of animal cells is orthologous to BRI1/BAK1 of *Arabidopsis* that induces a brassinosteriod signaling pathway (Edwards, 2005).

A BlastP search for KO1 orthologs performed in this study found numerous othologs that were LpxB or putative lipid A disaccharide synthase genes in bacteria (Table 3.1). One additional ortholog was found in a red alga, which was positioned evolutionarily closest to KO1 in a phylogenetic tree (Fig. 3.4). Interestingly, the KO1 orthologs found in the BlastP search did not include *E.coli* LpxB or any protein from more highly derived plants such as angiosperms. The orthologs of KO1 included cyanobacteria instead of *E.coli*, possibly indicating that KO1 evolved from cyanobacteria.

Physcomitrella patens is considered to be a good model organism for the study of polarized tip growth in plants (Jang et al., 2011), since its life cycle includes several phases of tip growth and since its high frequency of HR enables gene targeting as an experimental tool in the study of gene function (Kammerer & Cove, 1996; Schaefer & Zry, 1997). Protonema and rhizoids in *Physcomitrella* both exhibit polarized tip growth. The two best-studied forms of tip growth in higher plants occur in pollen tubes and root hairs, both involving tip growth of single cells. In *Physcomitrella*, tip growth in

chloronema, caulonema, and rhizoids occurs in uniseriate filaments, consisting of one row of end-to-end cells with growth occurring at the tip of the apical cell.

In the present study, the *ko1* of *Physcomitrella* exhibited several phenotypes, two of which involved tip growth in protonema (Fig. 3.6, Tables 3.2-3.4) and in rhizoids (Fig. 3.7). In protonema, the *ko1* showed a curly or wavering tip growth. In rhizoids, the *ko1* mutant showed reductions in both numbers and lengths of rhizoids. The mechanisms through which *KO1* influences these processes remain to be established. In general, transmission electron microscopy of tip growth in plants has shown that polarized tip cells contain many vesicles near the tip (Lancelle & Hepler, 1992; McCauley and Hepler, 1992; Derksen et al., 1995; Galway et al., 1997). This apical extension region, also called the clear zone, is rich in endoplasmic reticulum and, to a lesser extent, rich in Golgi and mitochondria. Golgi and mitochondria have also been observed to aggregate directly behind the clear zone (Lancelle & Hepler, 1992; Derksen et al., 1995; Cheung & Wu, 2008; Rounds et al., 2010; Furt et al., 2012). In *Physcomitrella*, mitochondria are abundant in the clear zone but are also distributed at lower density throughout the chloronema and caulonema cells (Furt et al., 2012). The importance of mitochondria in tip cells is generally assumed to be as a source of energy for exocytosis (Lovy-Wheeler et al., 2007). Li et al. (2011) found that LpxB and other Lpx proteins are targeted to mitochondria in *Arabidopsis*, Analysis at the TargetP website

(http://www.cbs.dtu.dk/services/TargetP/) suggests that *KO1* might be targeted to *Physcomitrella* mitochondria as well. Thus, one hypothesis regarding the mechanism of the *ko1* knockout phenotypes observed in tip growth might be that the knockout impairs mitochondria, thereby impairing the energy source for tip growth.

As shown in Chapter 2 of this dissertation, another phenotype of the *ko1* knockout mutant in *Physcomitrella* is a statistically very significant reduction in the 3-*O*-Me-Rha/Rha ratio in AGPs. Data presented in the present chapter show successful transgenic expression of *KO1* in tobacco to the level of mRNA accumulation (Figs. 3.8, 3.9). None of the transgenic tobacco plants that contained *KO1* mRNA, however, had a detectable level of 3-*O*-Me-Rha in their AGPs (Table 3.5). At least two alternative hypotheses seem consistent with these results on transgenic tobacco. One hypothesis would be that *KO1* does encode a rhamnosyl 3-*O*-methyltransferase, but the expressed mRNA does not lead to accumulation of the protein, or that the protein is incorrectly targeted, i.e., the protein accumulates in a subcellular location where it does have access to the tobacco AGPs. An alternative hypothesis would be that *KO1* does not encode a rhamnosyl 3-*O*-methyltransferase, and that the effect of the *ko1* knockout on 3-*O*-Me-Rha/Rha ratio in *Physcomitrella* AGPs occurs indirectly. It remains unclear whether knockout of a LpxB-related function of *KO1* could lead to such an indirect effect on 3-*O*-Me-Rha/Rha ratio in AGPs.

REFERENCES

Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branchs: A fast, accurate and powerful alternative. Syst Biol **55**: 539-552

Aono N, Sato Y, Nishiyama T (2012) Genomic DNA extraction. http://moss.nibb.ac.jp/

Ashton NW, Cove DJ (1977) The isolation and preliminary characterisation of auxotrophic and analogue resistant mutants in the moss *Physcomitrella patens*. Molecular and General Genetics **154**: 87-95

Basu D, Liang Y, Liu X, Himmeldirk K, Faik A, Kieliszewski M, Held M, Showalter AM (2013) Functional identification of a hydroxyprol-*O*-galactosyltransferase specific for arabinogalactan protein biosynthesis in *Arabidopsis*. J Biol Chem **288**: 10132-10143

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

Cheung AY, Wu HM (2008) Structural and signaling networks for the polar cell growth machinery in pollen tubes. Annu Rev Plant Biol **59**: 547-572

Chen CG, Pu ZY, Moritz RL, Simpson RJ, Bacic A, Clarke AE, Mau SL (1994) Molecular cloning of a gene encoding an arabinogalactan-protein from pear (*Pyrus communis*) cell suspension culture. Proc Natl Acad Sci USA **91**: 10305-10309

Cove DJ, Knight CD (1993) The moss *Physcomitrella patens*, a model system with potential for the study of plant reproduction. Plant Cell **5**: 1483-1488

Cove DJ, Perroud P-F, Charron AJ, McDaniel SF, Khandelwal A, Quatrano RS (2009) The moss *Physcomitrella patens*. A novel model system for plant development and genomic studies. In: Crotty DA, Gann A, editors. Emerging model organisms, a laboratory manual. Vol. I. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. pp 69-104

Crowell DN, Anderson MS, Raetz CR (1986) Molecular cloning of the genes for lipid A disaccharide synthase and UDP-N-acetylglucosamine acyltransferase in *Escherichia coli*. J Bacteriol **168**: 152-159

Dereeper A, Audic S, Claverie JM, Blanc G (2010) BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol **10**: 8

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acid Research **36**: W465-469

Derksen J, Rutten T, Lichtscheidl IK, Dewin AHN, Pierson ES, Rongen G (1995) Quantitative analysis of the distribution of organelles in tobacco pollen tubes: implications for exocytosis and endocytosis. Protoplasma **188**: 267-276

Du H, Simpson RJ, Moritz RL, Clarke AE, Bacic A (1994) Isolation of the protein backbone of an arabinogalactan-protein from the styles of *Nicotiana alata* and characterization of a corresponding cDNA. Plant Cell **6**:1643-1653

Duca M, Glijin A, Lupashcu V, Duca D, Orozco-Cardenas ML (2009) The expression of CsLFY and bar genes at transcription and translation levels in transgenic tobacco plants. Romanian Biotechnological Letters **14**: 4887-4892

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**: 1792-1797

Edwards D (2005) Regulation of signal transduction pathways by estrogen and progesterone. Annu Rev of Physiol **67**: 335-376

Fu H, Yadav MP, Nothnagel EA (2007) *Physcomitrella patens* arabinogalactan proteins contain abundant terminal 3-*O*-methyl-L-rhamnosyl residues not found in angiosperms. Planta **226**: 1511-1524

Furt F, Lemoi K, Tuzel E, Vidali L (2012) Quantitative analysis of organelle distribution and dynamics in *Physcomitrella patens* protonemal cells. BMC Plant Biol **12**: 70

Galway ME, Heckman JW, Schiefelbein JW (1997) Growth and ultrastructure of *Arabidopsis* root hairs: The *rhd3* mutation alters vacuole enlargement and tip growth. Planta **201**: 209-218

Geshi N, Johansen JN, Dilokpimol A, Rolland A, Belcram K, Verger S, Kotake T, Tsumuraya Y, Kaneko S, Tryfona T, Dupree P, Scheller HV, Hofte H, Mouille G (2013) A galactosyltransferase acting on arabinogalactan protein glycans is essential for embryo development in *Arabidopsis*. Plant J doi: 10.1111/tpj.12281

Gleave AP (1992) A versatile binary vector system with a T-DNA organizational structure conducive to efficient integration of cloned DNA into the plant genome. Plant Mol Biol **20**: 1203-1207

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52**: 696-704

Jauh GY, Lord EM (1996) Localization of pectins and arabinogalactan-proteins in lily (*Lilium longiflorum* L.) pollen tube and style, and their possible roles in pollination. Planta **199**: 251-261

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

Jang G, Yi K, Pires ND, Menand B, Dolan L (2011) RSL genes are sufficient for rhizoid system development in early diverging land plants. Development **138**: 2273-2281

Jones VAS, Dolan L (2012) The evolution of root hairs and rhizoids. Ann Bot **110:** 205-212

Kammerer W, Cove DJ (1996) Genetic analysis of the effects of re-transformation of transgenic lines of the moss *Physcomitrella patens*. Mol Gen Genet **250**: 380-382

Lancelle SA, Hepler PK (1992) Ultrastructure of freeze-substituted pollen tubes of *Lilium longiflorum.* Protoplasma **167**: 215-230

Lee KJD, Sakata Y, Mau SL, Pettolino F, Bacic A, Quatrano RS, Knight CD, Knox JP (2005) Arabinogalactan proteins are required for apical cell extension in the moss *Physcomitrella patens*. Plant Cell **17**: 3051– 3065

Li C, Guan Z, Liu D, Raetz CR (2011) Pathway for lipid A biosynthesis in *Arabidopsis thaliana* resembling that of *Escherichia coli*. Proc Natl Acad Sci USA **108**:11387-11392

Liang Y, Faik A, Kieliszewski M, Tan L, Xu W-L, Showalter AM (2010) Identification and characterization of in vitro galactosyltransferase activities involved in arabinogalactan-protein glycosylation in tobacco and *Arabidopsis*. Plant Physiol **154**: 632-642

Louis E, Metzger IV, Raetz CR (2009) Purification and characterization of the lipid A disaccharide synthase (LpxB) from *Escherichia coli:* a peripheral membrane protein. Biochemistry **48**: 11559-11571

Lovy-Wheeler A, Cárdenas L, Kunkel JG, Hepler PK (2007) Differential organelle movement on the actin cytoskeleton in lily pollen tubes. Cell Motil Cytoskeleton **64**: 217-232

Menand B, Calder G, Dolan L (2007) Both chloronemal and caulonemal cells expand by tip growth in the moss *Physcomitrella patens*. J Exp Bot 58: 1843-1849

Metzger LE 4th, Raetz CR (2009) An alternative route for UDP-diacylglucosamine hydrolysis in bacterial lipid A biosynthesis. Biochemistry 49: 6715-6726

McCauley MM, Hepler PK (1992) Cortical ultrastructure of freeze-substituted protonemata of the moss *Funaria hygrometrica.* Protoplasma **169**: 168-178

Nishiyama T, Fujita T, Shin-I T, Seki M, Nishide H, Uchiyama I, Kamiya A, Carninci P, Hayashizaki Y, Shinozaki K, Kohara Y, Hasebe M (2003) Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. Proc Natl Acad Sci **100**: 8007-8012

Opiyo S, Pardy RL, Moriyama H, Moriyama EN (2010) Evolution of the Kdo_2-lipid A biosynthesis in bacteria. BMC Evol Biol **10**: 362

Pressel S, Ligrone R, Duckett JG (2008) Cellular differentiation in moss protonemata: a morphological and experimental study. Annu Bot **102**: 227-245

Raetz CR, Whitfield C (2002) Lipopolysaccharide endotoxins. Annu Rev Biochem **71**: 635-700

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-I T, 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, Cuming 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

Reski R, Abel WO (1985): Induction of budding on chloronemata and caulonemata of the moss, *Physcomitrella patens*, using isopentenyladenine. Planta **165**: 354-358

Rounds CM, Hepler PK, Fuller SJ, Winship LJ (2010) Oscillatory growth in lily pollen tubes does not require aerobic energy metabolism. Plant Physiol **152**: 736-746

Schaefer DG, Zrÿd J, Knight CD, Cove DJ (1991) Stable transformation of the moss *Physcomitrella patens*. Mol Gen Genet **226**: 418-424

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

Schmiedel G, Schnepf E (1980) Polarity and growth of caulonema tip cells of the moss Funaria hygrometrica. Planta **147**: 405-413

Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YM, Bacic A (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. Plant Physiol **129**: 1448-1463

Serpe MD, Nothnagel EA (1999) Arabinogalactan-proteins in the multiple domains of the plant cell surface. Adv Bot Res **30**: 207-289

Seifert G, Roberts K (2007) The biology of arabinogalactan proteins. Annu Rev Plant Biol **58**: 137-161

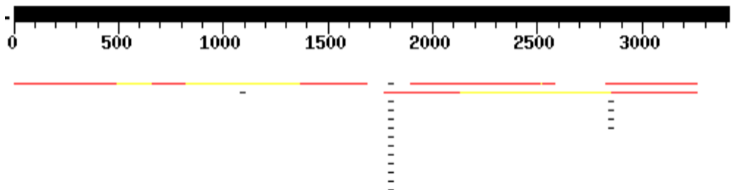
Tan L, Varnai P, Lamport DTA, Yuan C, Xu J, Qiu F, Kieliszewski MJ (2010) Plant *O*-hydroxyproline arabinogalactans are composed of repeating trigalactosyl subunits with short bifurcated side chains. J Biol Chem **285**: 24575-24583

Vidali L, Bezanilla M (2012) *Physcomitrella patens*: a model for tip cell growth and differentiation. Curr Opin Plant Biol **15**: 625-631

Wu Y, Williams M, Bernard S, Driouich A, Showalter AM, Faih A (2010) Functional identification of two nonredundant Arabidopsis α(1,2)fucosyltransferases specific to arabinogalactan proteins. J Biol Chem **285**: 13638-13645

TABLES AND FIGURES

A)



B)

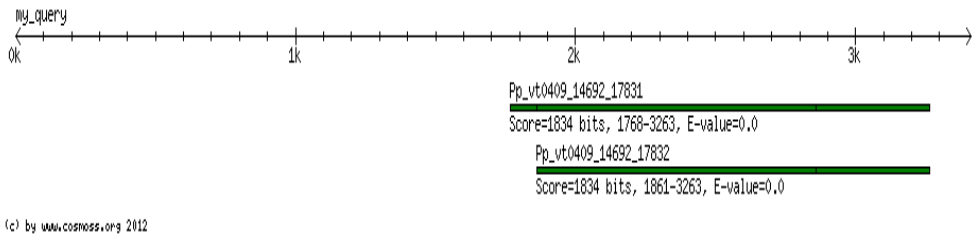


Figure 3.1. Search of cDNA of *KO1* in the moss transcriptome.
Results of searches for a *KO1* cDNA in public transcriptome databases for *Physcomitrella patens*. (A) The predicted full-length transcript sequence of the *KO1* gene was used as query to do a BlastN search of the *Physcomitrella* cDNA database at the RIKEN center (http://www.brc.riken.jp/lab/epd/blast/). The thick black line labeled with numbers 0 to 3000+ was the predicted *KO1* transcript. The red (higher quality match) and yellow (lower quality match) lines illustrated positions of different transcripts found in the RIKEN database. Black dashes indicated non-specific short nucleotide fragments found in the BlastN search. The lower red and yellow line showed the cDNA clone pdp32i13, which covered a portion of the predicted *KO1* gene model that included the site of the knockout cassette insertion used in this project. (B) The predicted full-length transcript sequence of the *KO1* gene was used as query to do a BlastN search of the

Physcomitrella transcriptome at the COSMOSS database (http://www.cosmoss.org/). The black arrow labeled with 0k to 3k+ showed the predicted full-length query sequence. The two bottom green lines indicated two cDNA clones found to match the predicted *KO1* transcript.

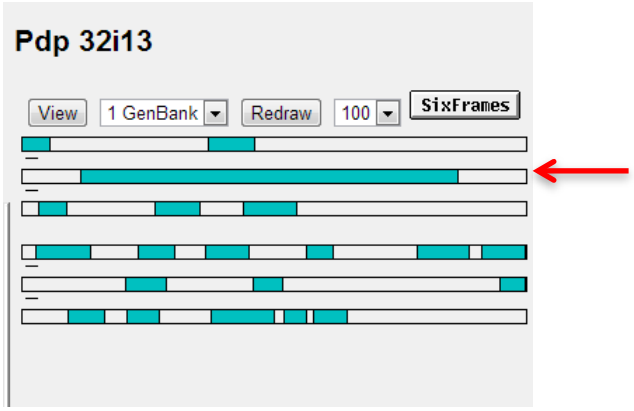
A)

>Pdp 32i13

AGCAACTCTATAGGGCGATTGGAGCTCCACCGCGGTGGCGGCCGCATAACTT
CGTATAGCATACATTATACGAAGTTATGGATCAGGCCAAATCGGCCGAGCTC
GAATTCGTCGAGAACCGACTGAACTCGCAATTCTCGAACCCTTTCGGAAACG
GAATCGACAATTTTCTCCTTGTAATCTACGTGACTCGGTGAATCACCCGCTGG
ATTGCGGTGATATAGCAATGCTGTGAGGAAAAGAGTGAGGGCGTACAATGGC
GGGTTCAATGTGCCCCGCGACTGTTTTGAGCGGGCCATCCTGCGCCCGGGTCT
CTTCTCCAAGCTGGAACGTTTTAAACAGGGCTTCCAAATGTTTCATTAGAGTT
TCGAAGTTTGCGCCATTGTTTTGTGGGGCGTTGAAGACGAGAAAACGGGATC
AAAGCACTATTTTCAGAGTGAGTGTCAAATCATCTCATATACGAGCGGAAGT
CAAGAATGAGCCAGTGGATGTAGTGATTCTCACGAATGGGCCCGGTGAGGTG
GCCACCTGGGTGAAGCCAGTGGTTGCAAGGTTGCGGCGGACAGCGGAGGAC
CATGCAATGGACATGCGCATCTCGGTTCTCTTGGCACCCTGCCCTCATGCCTC
TGGAAAAGAGCTCCAACTCCTTCAATCATTCGACGAGATTGATCGGTGCCAG
GGACCAGAAGAGTTTTACTCACTCCTACTGCTTGGTCGAACACGAAGTGGTT
GGGATTGGCGTAGGCGGGGAGTGTGCATATTTCTAGGCGGCGACCAATTCAA
CACGCTTATCCTGGGATGGAGGTTAGGTTACAAGACGGTAGTATACGCAGAA
GACGCAGCACGATGGCCGGGCTTTGTAGATTTATACATGCTACGTTCCGAAG
AGCTCGTGTTGGAAGTGCCGCAATGGGCTCGCAGCCGATGCCTGGTTGTTGG
AGACCTTTTCGTAGATGCAGTGGGATGTTCACCTTCATCAATCTCCTGGAGTT
TATCTGACAAACTGCAGTCTCGGGCGCAAGATAGGAATGTTGTCCCAATTGTT
GGTCTTCTGCCAGGTTCCAAGGATGCAAAATTGGCCATTGGCGTCCCCTATTT
TATGGCGGTCGCAGATCATTTACATCAACTTTTGCAGGGCCAGGTGCGGTTTG
TGCTACCTTTGGCTCCCACTGTTACCGTGACTGAGCTTGAACGCTTTGCCGAT
GCAGCTTCCAACCCCCTCATATCACGTTTTCACTGGTCTTCTGGTCACTTCGTT
GGGAAGCAATCTGTGAAAGTAGAGAAAATCTCAAAATTGCGGCAAGAAGAA
CCTGAGGAAGGGTCAACTTGGAATTCATCGGAGTTACTCAAAGTTGAAGAAC
TGGGGCAGCTAGTGACGGAGGGAGGGGTGGTCATAGATATTCAGCAGCAATT
CCCGCCATATTCTCTTTATCAAGGCTGCTCCGTCTGTCTTACAACAGTGGGTA
CAAATACGGCAGAGCTAGGGACATTAGGTGTCCCCATGCTTGTGGTATTGCC
GACTTATTTTCTTGAGACTTTCAGGGGAGCGACAGGAGGAATACTTGGCCTG
CTATCAGCAATCCCCGGTCAGGCTGGTGCAGCGATGGCTCACTTTGTCAACCT

TTTCATACTCAAAACAGCCGGGTTTATCTCGTGGCCCAACAGGTGGGCTGGA
GAGACAATTGTTCCAGAGCTTATTGGTGAGATCGATCCTCAGGAAGTAGCAG
CTTTAGCTGCTGAATATCTTCAATCACCAGACCGACTCCAAGCAATGCACGA
ACGCCTTCTTGACCTCCAAATGCCGCGGTCTGTTTCCAAAGGAGGCGCAGCA
CATTCCATCGCAGTCGCTGTTCAGCAATTATTATAGTATGTAGCGTAGAACCT
TCCTTCTTGCCTACCCAGTATTTGTGCAGATACATGAATGGCAAGTTTACTTG
AGCGCTCTGCCTAAAGAATTACTGACATCAATGTTATTCAGGTCGACATCTTT
TCTGCTTTGGAAGAAGTTGATGTGACTCATTGTTTATAATTCGGTCAGCAAAG
CCACTTTTTATCCTAAAAAAAAAAAAAAAACTCTCCAGGGTTGGATCCGGCC
ATAAGGGCCTGATCCTTCGAGGGGGGGCCCGGTACCAGCTTTGTCCCTTAGT
GAGGTAATGAGT

B)



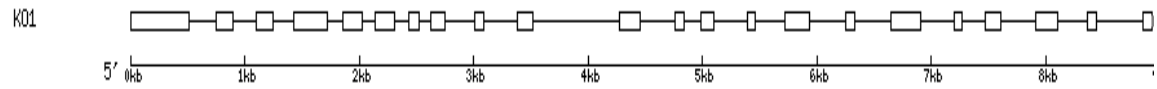
C)

MAGSMCPATVLSGPSCARVSSPSWNVLNRASKCFIRVSKFAPLFCGALKTRKRD
QSTIFRVSVKSSHIRAEVKNEPVDVVILTNGPGEVATWVKPVVARLRRTAEDHA

MDMRISVLLAPCPHASGKELQLLQSFDEIDRCQGPEEFYSLLLLGRTRSGWDWR
RRGVCIFLGGDQFNTLILGWRLGYKTVVYAEDAARWPGFVDLYMLRSEELVLE
VPQWARSRCLVVGDLFVDAVGCSPSSISWSLSDKLQSRAQDRNVVPIVGLLPGS
KDAKLAIGVPYFMAVADHLHQLLQGQVRFVLPLAPTVTVTELERFADAASNPLI
SRFHWSSGHFVGKQSVKVEKISKLRQEEPEEGSTWNSSELLKVEELGQLVTEGG
VVIDIQQQFPPYSLYQGCSVCLTTVGTNTAELGTLGVPMLVVLPTYFLETFRGAT
GGILGLLSAIPGQAGAAMAHFVNLFILKTAGFISWPNRWAGETIVPELIGEIDPQE
VAALAAEYLQSPDRLQAMHERLLDLQMPRSVSKGGAAHSIAVAVQQLL

Figure 3.2. Transcript, amino acid sequence and open reading frame of revised *KO1*. (A) Sequence of cDNA clone pdp32i13, which was purchased from the RIKEN center (Saitama, Japan) and sequenced in the UCR Genomics Core Facility (genomics.ucr.edu). This cDNA clone is now regarded as the actual expressed transcript of the *KO1* gene (Phypa_169825) and included the site of the knockout cassette insertion used in this project. The full-length transcript was 2159 bp long. (B) *KO1* open reading frames as predicted by the NCBI ORF finder. The full-length sequence of the *KO1* transcript shown in Fig. 3.2A was put into NCBI ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) to find the correct open reading frame. All six possible reading frames found by the program are shown. The second reading frame, which was the longest of the six and included 5' and 3' UTRs, was selected as the *KO1* ORF. (C) Amino acid sequence translated from the *KO1* open reading frame identified in Fig. 3.2B. Analysis of the amino acid sequence by CDD in NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) showed that KO1 contained a LpxB domain. Neither signal peptides nor transmembrane domains were detected in this sequence by signalP (http://www.cbs.dtu.dk/services/SignalP/) and TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) web servers

A)



B)

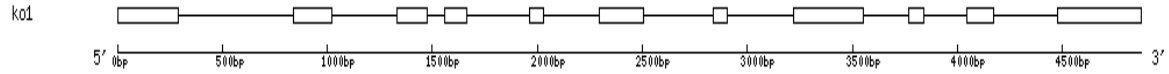


Figure 3.3. Original and revised gene models for *KO1*.
(A) Original JGI v1.1 predicted gene model of *KO1*as drawn in a gene structure display server (gsds.pku.edu.cn). For this gene model, the predicted transcript sequence and reference genomic sequence of *KO1* were downloaded from the *Physcomitrella* draft genome v1.1 in the JGI database (http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html). The predicted *KO1* gene was about 8.9 kb long, containing 22 exons. (B) Revised expressed gene model of *KO1*. The transcript sequence was obtained from cDNA clone pdp32i13 (Fig. 3.2A). The reference genomic sequence was downloaded from the *Physcomitrella* draft genome v1.1 in the JGI database. The revised gene model was drawn in a gene structure display server (gsds.pku.edu.cn). The revised *KO1* gene was about 5 kb long and contained 11 exons.

Table 3.1. BlastP search of KO1orthologs.

Orthologs of the *Physcomitrella* KO1 protein. A BlastP search using the predicted KO1 polypeptide sequence (Fig. 3.2C) as query at the UniprotKB protein database (http://www.uniprot.org) resulted in many hits. The Entry column shows the entry name as it appears in the UniprotKB database. Gene names, protein names, and the length (amino acids) of the proteins are alsol isted, along with the percent identity with KO1 and the E-values from the BlastP search. Most of the found orthologs were in microorganisms, the exception being L8LAA6_9CYAN from a red alga.

Entry	Gene names	Protein names	Organism	Length	Identity (%)	E-value
K9S995_9CYAN	GEI7407_2066	Putative lipid-A-disaccharide synthase	Acaryochloris marina (strain MBIC 11017)	431	43	9.00E-108
K9W0J6_9CYAN	Cri9333_2400	Putative lipid-A-disaccharide synthase	Anaerobaculum hydrogeniformans ATCC BAA-1850	444	42	2.00E-105
B7K7Y4_CYAP7	PCC7424_2769	Putative lipid-A-disaccharide synthase	Arthrospira maxima CS-328	416	39	4.00E-105
K9TX59_9CYAN	Chro_1639	Putative lipid-A-disaccharide synthase	Arthrospira platensis C1	446	40	6.00E-103
I4I4L9_MICAE	MICAH_560002	Putative lipid-A-disaccharide synthase	Chamaesiphon minutus PCC 6605	412	39	1.00E-101
D8FW25_9CYAN	OSCI_950014	Putative lipid-A-disaccharide synthase	Chroococcidiopsis thermalis PCC 7203	435	40	3.00E-101
K9XXI6_STAC7	Sta7437_2857	Putative lipid-A-disaccharide synthase	Crinalium epipsammum PCC 9333	418	39	4.00E-101
K9UGF7_9CHRO	Cha6605_3190	Lipid A disaccharide synthetase	Cyanothece sp. (strain PCC 7424)	420	39	1.00E-99
L8LAA6_9CYAN	Lep6406DRAFT_00012580	Lipid A disaccharide synthetase	Galdieria sulphuraria (Red alga)	425	41	2.00E-95
D4TMH9_9NOST	CRD_00139	Cyanobacteria-specific lipid-A-disaccharide synthase	Geitlerinema sp. PCC 7407	413	37	2.00E-94
K6CU44_SPIPL	APPUASWS_09150	Putative lipid-A-disaccharide synthase	Jonquetella anthropi E3_33 E1	426	37	3.00E-93
H1WJS2_9CYAN	ARTHRO_690138	Putative lipid-A-disaccharide synthase, LpxB	Leptolyngbya sp. PCC 6406	424	37	7.00E-90
K9Q4S8_9CYAN	Lepto7376_3964	Lipid A disaccharide synthetase	Meiothermus ruber	415	36	2.00E-88
B8HRZ7_CYAP4	Cyan7425_0261	Putative lipid-A-disaccharide synthase	Microcoleus vaginatus FGP-2	450	35	6.00E-88
B0CEQ6_ACAM1	lpxB AM1_3166	Lipid-A-disaccharide synthase, putative	Microcystis aeruginosa	427	40	3.00E-82
A2C6G6_PROM3	P9303_03231	Cyanobacteria-specific lipid A disaccharide synthase	Nostoc punctiforme	451	36	5.00E-78
G4FNX8_9SYNE	Syn8016DRAFT_2330	Lipid A disaccharide synthetase-like protein	Oscillatoria sp. PCC 6506	428	36	8.00E-75
M2X6U7_GALSU	Gasu_06440	Lipid-A-disaccharide synthase	Prochlorococcus marinus	501	33	2.00E-70
D3PNW5_MEIRD	Mrub_2763 K649_08710	Lipid A disaccharide synthetase	Raphidiopsis brookii D9	389	27	4.00E-20
D3L346_9BACT	HMPREF1705_01453	Cyanobacteria-specific lipid A disaccharide synthase	Synechococcus sp. WH 8016	368	29	7.00E-08

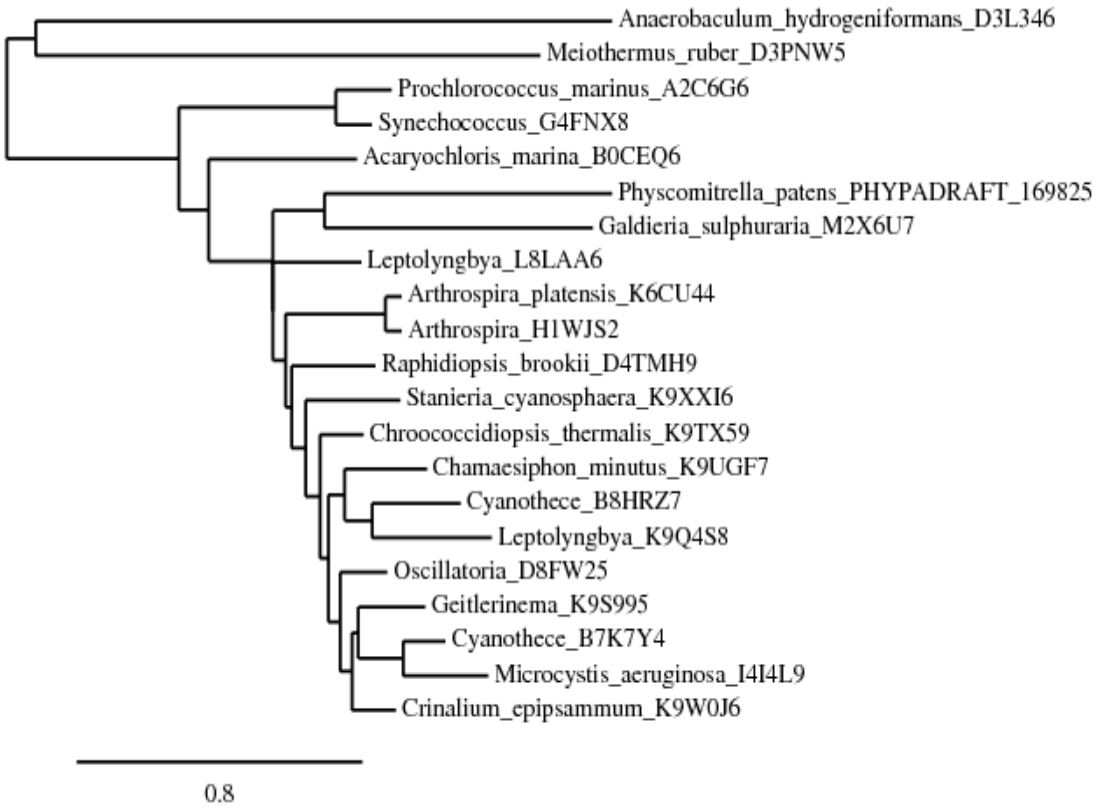


Figure 3.4. Phylogenetic tree of KO1.
Based on KO1 and the orthologous proteins listed in Table 3.1, MUSCLE alignment and the ONE CLICK model at www.phylogeny.fr were used to draw the tree. The bar with the number 0.8 indicates tree branch length, which usually represents changes in the branch. KO1 appears in the tree as PHYPADRAFT_169825 and was evolutionarily closest to M2X6U7 from *Galdieria sulphuraria*, a red alga.

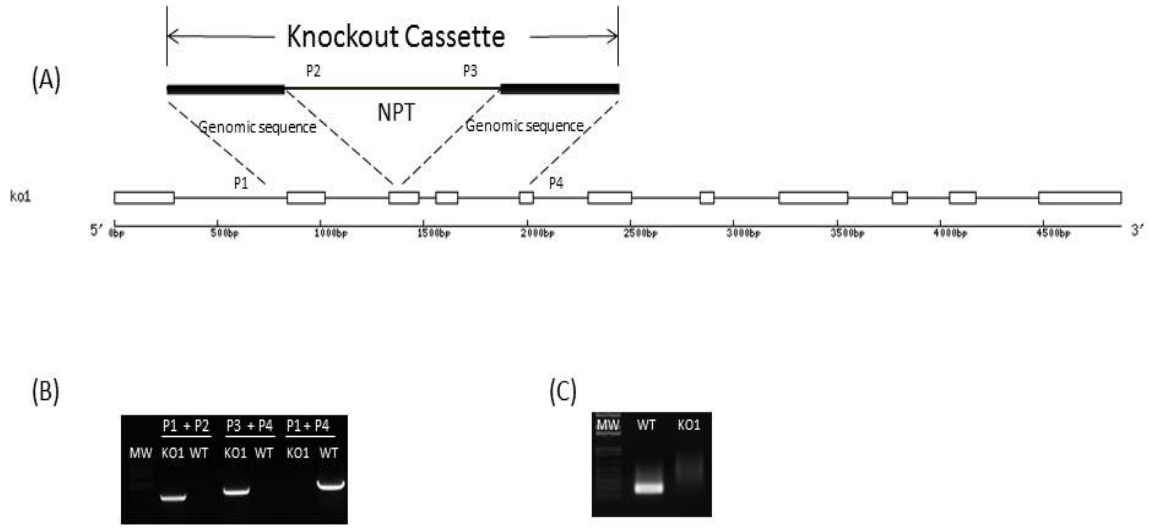


Figure 3.5. Molecular analysis of *ko1* knockout mutant.
(A) Scheme to show the anticipated knockout cassette insertion into the *KO1* genomic sequence. The knockout cassette consisted of a *npt* gene for selection and bordering *KO1* genomic sequences for HR. Primers P1 and P4 were located as shown on the *Physcomitrella* genome. Primers P2 and P3 were located on the *npt* portion of the knockout cassette. (B) PCR analysis using primers to map the insertion junctions of the knockout cassette integrated into the *Physcomitrella* genome. The left insertion junction was mapped with primers P1 and P2, and the right insertion junction was mapped with primers P3 and P4. The anticipated portion of the *KO1* genomic sequence was amplified from wild-type (WT) genomic DNA with primers P1 and P4, but not with genomic DNA from the *ko1* knockout mutant. (C) RT-PCR to test *KO1* gene expression. RT-PCR product was amplified with total RNA from wild-type *Physcomitrella* but not with total RNA from the *ko1* knockout mutant.

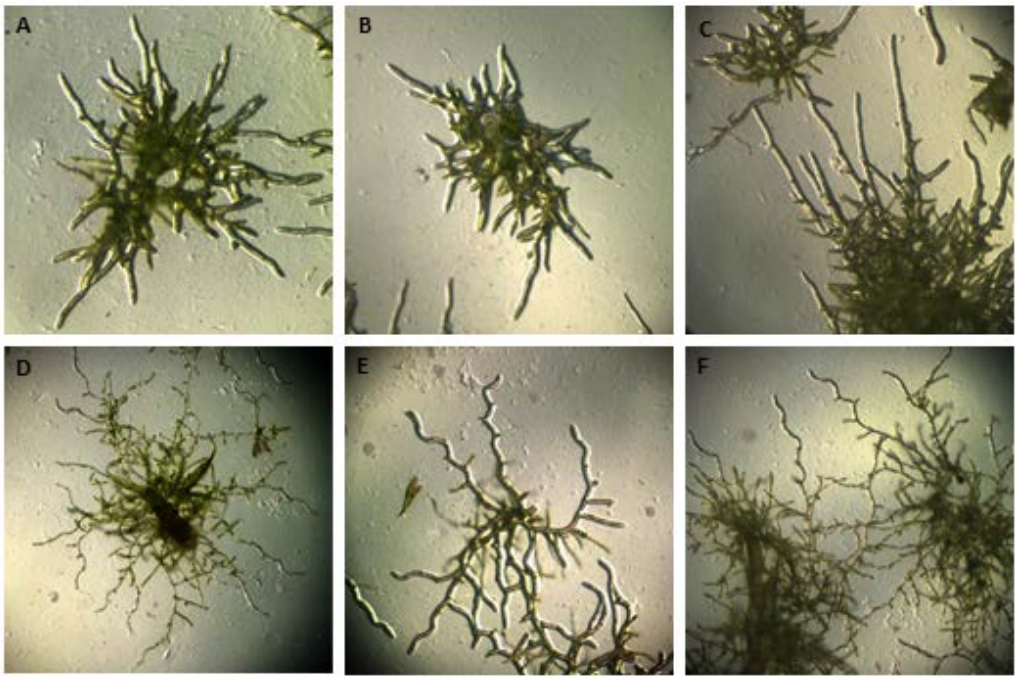


Figure 3.6. Protonema filamentous growth in *ko1* and wild type. Filaments of wild-type protonema tissue grew relatively straight by polarized tip growth (A-C). Polarized tip growth was disrupted in the *ko1* mutant as indicated by the appearance of curly filaments (D-E). See Tables 3.2-3.4 for quantitation of these effects.

Table 3.2. Abundances of normal colonies and abnormal curly colonies of protonema tissue growth in wild type plants and *ko1* knockout mutants.
See Figure 3.6 for illustrations of these colonies. Results in table are based on counts of 50 colonies in each of three Petri dishes.

	Wild type (%)	KO1 mutant (%)
Normal colony	83±2	0±0
Curly colony	17±2	100±0

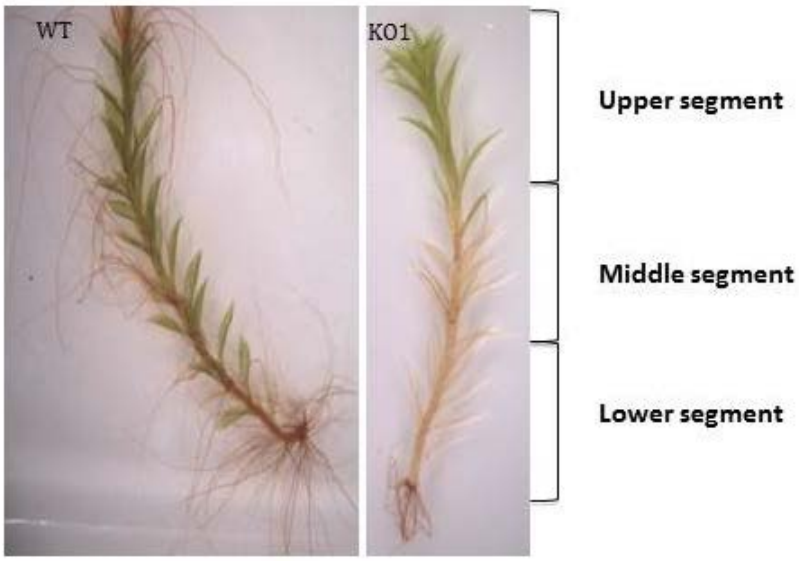
Table 3.3. Abundances of normal filaments and curly filaments in abnormal curly colonies reported in Table 3.2 for wild type and *ko1* mutant cultures. Within abnormal curly colonies, not all filaments are curly. Rather, only some filaments are curly, while the others are normal, i.e., relatively straight. Results in table are based on counts of 50 colonies in each of three Petri dishes.

	Wild type (%)	*ko1* (%)
Normal filaments	94±2	18±3
Curly filaments	6±2	82±3

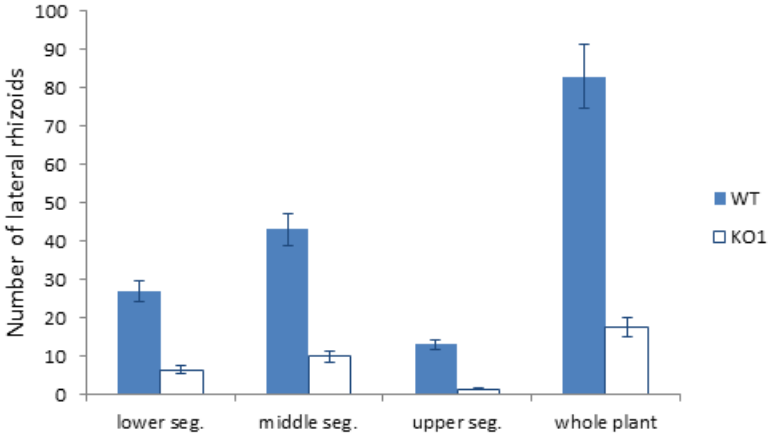
Table 3.4. Total abundances of normal and curly filaments in *ko1* and wild type.
Protonema filaments were counted from cultures reported in Table 3.2. Total normal filaments were counted irrespective of whether they occurred in normal colonies or in curly colonies. Similarly, total curly filaments were counted irrespective of whether they occurred in normal colonies or in curly colonies. Results in table are based on counts of 50 colonies in each of three Petri dishes.

	Wild type (%)	*ko1* (%)
Normal filaments	86±5	19±6
Curly filaments	14±5	81±6

(A)


WT
KO1
Upper segment
Middle segment
Lower segment

(B)


Number of lateral rhizoids
100
90
80
70
60
50
40
30
20
10
0
lower seg.
middle seg.
upper seg.
whole plant
WT
KO1

(C)

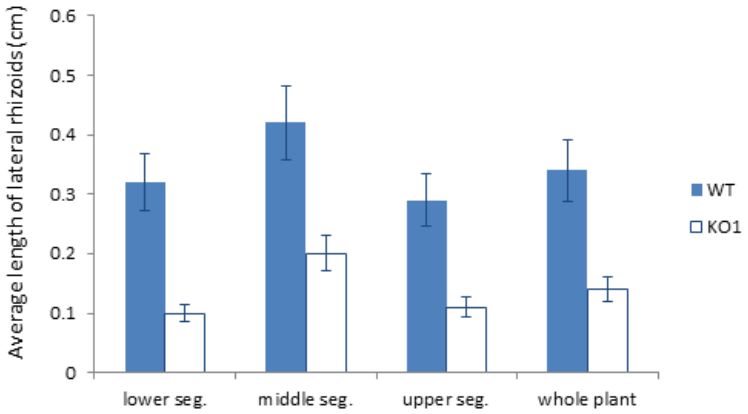


Figure 3.7. Length and density of lateral rhizoids in gametophytes of *ko1* and wild type.
(A) Each gametophyte was divided into three segments of equal length. (B) Number of lateral rhizoids in the three segments and in the whole plant. Lower seg: lower segments; middle seg: middle segments, upper seg: upper segments. The number of lateral rhizoids in each segment was counted and normalized with regard to the length of each segment. The number of lateral rhizoids for the whole plant is the sum of numbers in three segments. (C) Average length (cm) of lateral rhizoids in different segments and in whole plant. Counts and measurements were made in three replicates with seven different plants counted in each replicate for a total of 21 wild type plants and 21 *ko1* plants.

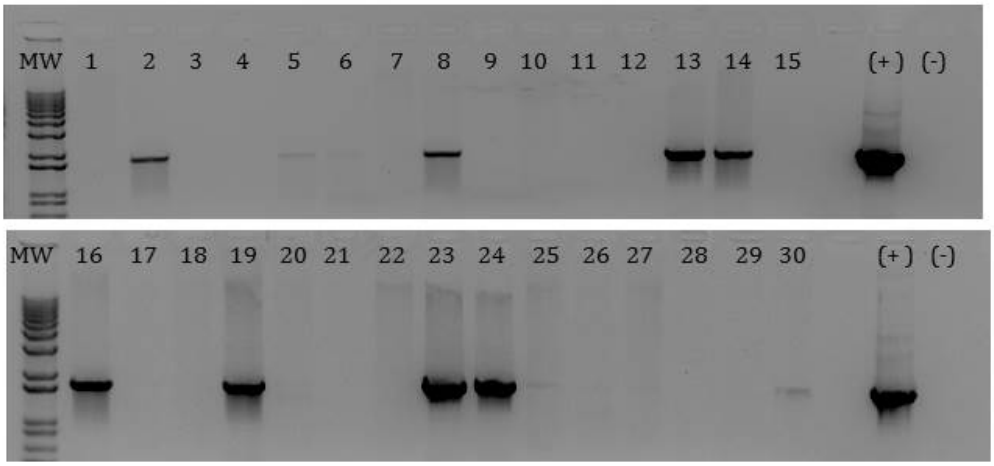


Figure 3.8. Genomic PCR of transgenic *KO1* tobacco plants.
Genomic PCR to test for the presence of the *KO1* transgene in regenerated transgenic tobacco plants. In each panel, the left lane shows DNA molecular weight markers (1kb plus DNA ladder, No. 10787018, Life Technologies). The transgenic plant number is shown above each lane. The (+) lane is a positive control using the expression plasmid as the template. The (–) lane is a negative control using wild-type genomic DNA as the template. Among the 30 regenerated transgenic plants grown on Kanamycin selection medium, 14 plants (#2, 5, 6, 8, 13, 14, 16, 19, 23, 24, 25, 26, 27, 30) showed a detectable PCR band indicating that the KO1 gene had inserted into the tobacco genome. In six plants (#5, 6, 25, 26, 27, 30), the PCR band was weak.

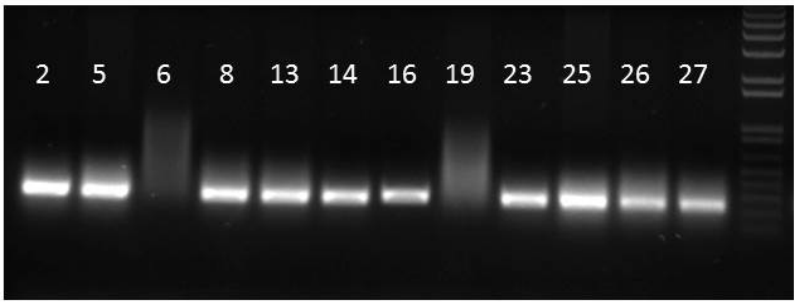


Figure 3.9. RT-PCR of transgenic *KO1* tobacco plants.
RT-PCR to test for *KO1* gene expression (presence of *KO1* mRNA) in transgenic tobacco plants. The transgenic plant number is shown above each lane. Transgenic plants #24 and 30 were not healthy in the growth chamber, were not successfully transferred to the greenhouse, and thus were not tested in RT-PCR. Among the other 12 transgenic lines showing a band in genomic PCR (Fig. 3.8), only transgenic plants #6 and 19 showed no detectable *KO1* gene expression. The other 10 plants, including #2, 5, 8, 13, 14, 16, 23, 25, 26, and 27 were shown to have KO1 gene expression. The far right lane is 1kb plus DNA ladder (No. 10787018, Life Technologies).

Table 3.5. Glycosyl composition of total soluble AGPs from transgenic *KO1* tobacco plants.
Glycosyl compositions of total soluble AGPs from wild-type Physcomitrella (positive control), wild-type tobacco (negative control), and transgenic *KO1* tobacco plants. No 3-O-Me-Rha was detected in AGPs from any of the transgenic *KO1* tobacco plants. Glc contents were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations.

	Glycosyl composition (mol%)									
	Physco	Tobacco plants								
Residue	Wild type	Wild type	KO1-2	KO1-8	KO1-13	KO1-14	KO1-23	KO1-25	KO1-26	KO1-27
3-O-Me Rha	3.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ara	15.58	30.93	29.27	27.89	23.58	23.88	13.27	27.34	32.16	24.82
Rha	1.29	4.89	5.00	5.73	5.82	4.97	4.18	5.74	5.67	5.78
Fuc	0.26	0.09	0.10	0.24	0.09	0.09	0.09	0.14	0.11	0.12
Xyl	2.71	2.09	5.01	5.47	6.30	5.25	13.04	5.68	4.97	6.38
GlcU	11.73	8.23	6.44	8.84	8.04	8.62	6.46	7.26	8.03	8.23
GalU	2.28	0.67	1.89	4.14	2.67	2.62	11.10	2.73	1.87	4.93
Man	1.63	0.92	0.97	1.07	1.18	1.60	2.58	0.81	0.96	1.29
Gal	46.92	45.96	41.06	41.28	46.12	42.74	30.17	44.39	41.15	41.40
Glc	13.85	6.24	10.24	5.34	6.19	10.23	19.10	5.92	5.09	7.04

CHAPTER 4 - FUNCTIONAL ANALYSIS OF KO9 IN THE MOSS *PHYSCOMITRELLA PATENS* AND IN *NICOTIANA TABACUM CV XANTHI*

ABSTRACT

The gene *KO9* was identified in Chapter 2 of this dissertation as one of the candidates to encode the rhamnosyl 3-*O*-methyltransferase that synthesizes 3-*O*-Me-Rha in the moss *Physcomitrella patens*. The initial identification of *KO9* as a candidate was based on a bioinformatics search that found the predicted KO9 protein to contain the conserved domain of the TylF superfamily of *O*-methyltransferases. This *KO9* candidate was selected for further study in this dissertation because the 3-*O*-Me-Rha/Rha ratio in arabinogalactan proteins (AGPs) from the *Physcomitrella ko9* knockout mutant was found to be statistically very significantly reduced compared to the ratio in AGPs from the wild type. The present studies find that KO9 has strong sequence similarity to higher plant caffeoyl-CoA O-methyltransferases (CCoAOMTs) of lignin biosynthesis. To gain evidence on whether the KO9 protein functions as rhamnosyl 3-*O*-methyltransferase or as CCoAOMT, properties of *Physcomitrella ko9* knockout mutants and *KO9* transgenic tobacco plants were examined. The amount of lignin-like content in the cell wall of *ko9* moss knockouts was not significantly different from the amount of lignin-like content in the cell wall of wild-type moss. Transgenic expression of the *KO9* gene in *Nicotiana tabacum* cv Xanthi did not result in detectable 3-*O*-Me-Rha content in the AGPs of the transgenic tobacco plants. Thus, the function of *KO9* in *Physcomitrella* remains uncertain.

INTRODUCTION

Methyltransferases (MT) transfer methyl groups to different substrates from methyl donors. The most common methyl donor, by far, is *S*-adenosyl-methionine (AdoMet), which is believed to exist in all living organisms. AdoMet is formed through the joining of methionine and ATP by methionine adenosyltransferase. When the methyl group is transferred from AdoMet to an acceptor, the left-over part of AdoMet is the by-product *S*-adenosyl-L-homocysteine. Eventually, *S*-adenosyl-L-homocysteine is recycled via conversion to homocysteine and then to back to methionine and finally to AdoMet (Luka et al., 2009). AdoMet is initially synthesized in the cytosol (Wallsgrove et al., 1983; Ravanel et al., 1998) and then transferred to different subcellular locations for methylation of different substrates. For example, Ibar and Orellana (2007) found that AdoMet was transported from the cytosol into the Golgi apparatus for the methylesterification of pectin homogalactouronan. AdoMet-dependent MTs usually contain an AdoMet binding domain plus a substrate binding domain.

The methyl group from AdoMet is the methyl source for various acceptor substrates such as nucleic acids, proteins, lipids, lignin subunits, carbohydrates, and other small metabolites (Cantoni, 1975; Kende, 1993; Fontecave et al., 2004; Luka et al., 2009). When the acceptor is an oxygen (as in a hydroxyl group on a sugar) and an *O*-methyl ether is formed, then this kind of methyltransferase is called an *O*-methyltransferase. Mtf1 is an *O*-methyltransferase that methylates a single rhamnosyl residue in the glycopeptidolipids (GPL) of *Mycobacterium smegmatis* cell wall (Patterson et al., 2000).

Sequence analysis revealed that Mtf1 shares a high degree of homology with methyltransferases of other *Mycobacterium* species, especially in the motifs in the conserved domain. Mutants of *mtf1* in *Mycobacterium* do not have the ability to methylate the rhamnosyl residue of GPL, which results in 90% inhibition of GPL synthesis. Catechol *O*-methyltransferase (COMT) was the first structurally characterized AdoMet-dependent MT and plays important roles in the nervous system and in Parkinson's disease (Vidgren et al., 1994). The crystal structure of COMT showed the protein to contain a core fold of seven β-sheets linked by α-helices.

Lignin, next to cellulose, is the second most abundant polymer in some plant cell walls (Battle et al., 2000). Lignin is mainly deposited in the secondary cell wall of all vascular plants, especially in the tracheary woody plants. As a phenolic polymer, lignin is deposited in xylem treachery elements of vascular bundles that function in water transport (Zhong & Ye, 2009). Lignin adds rigidity and thus structural support to the cell walls of tracheids, vessel elements, fibers, and various other differentiated cells. Lignin also functions in plant-pathogen interactions, protecting plant cell wall polysaccharides from degradation by enzymes secreted by invading microorganisms. Lignification is commonly found at the site of pathogen invasion (Lange et al., 1995).

Due to its importance relative to biofuel production, the biosynthetic pathway of lignin has been widely investigated in woody plants and angiosperms in the past ten to fifteen years (Boerjan et al., 2003). The three most abundant monomers of lignin are the *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (G) residues, these being derived from

p-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol, respectively (Weng et al., 2008). Woody angiosperm lignin is composed of mainly G and S residues, with only few H residues. Monocot lignin contains closer to equal amounts of G, S, and H residues, while gymnosperm lignin consists of mostly G residues, some H residues, but generally no S residues. Synthesis of lignin monomers begins with phenylalanine. Eight core enzymes are involved in the monolignol biosynthetic pathway (Weng & Chapple, 2010), including phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-hydroxycinnamoyl-CoA ligase (4CL), hydroxycinnamoyl transferase (HCT), p-coumaroyl shikimate 3'-hydroxylase (C3'H), caffeoyl-CoA O-methyltransferase (CCoAOMT), (hydroxy)cinnamoyl-CoA reductase (CCR) and (hydroxy)cinnamyl alcohol dehydrogenase (CAD). Lignin monomers are transported across the plasma membrane and then polymerized in the cell wall through the action of peroxidases.

As one of the eight core enzymes of the monolignol biosynthetic pathway, CCoAOMT is a well-characterized AdoMet-dependent MT that occurs in various plant species as a family of enzymes. Acting within the monolignol biosynthetic pathway, CCoAOMT catalyzes the first methylation reaction, which involves transfer of a methyl group to an aromatic 3-hydroxyl group (Weng & Chapple, 2010). CCoAOMT is a homolog of mammalian catechol-*O*-methyl transferase (COMT), containing a catalytic divalent cation binding site and an AdoMet-binding domain (Vidgren et al., 1994; Ferrer et al., 2005). Several studies agreed that downregulation of CCoAOMT activity resulted in reduced lignin content (Meyermans et al., 2000; Zhong et al., 2000; Guo et al., 2001) but differed in other details. Downregulation of CCoAOMT led to a relative increase in H

residues in poplar (Meyermans et al., 2000; Zhong et al., 2000) but not in alfalfa (Guo et al., 2001). Likewise, downregulation of CCoAOMT led to a decreased G/S ratio in some cases (Meyermans et al., 2000; Guo et al., 2001) but to comparable reductions in both G and S units in another case (Zhong et al., 2000).

The moss *Physcomitrella patens* is an emerging model organism. Its genome is already completely sequenced by the Joint Gene Institute (JGI). The availability of a sequenced genome is a particular advantage for molecular genetic studies using *Physcomitrella* as a model organism (Rensing et al., 2008). *Physcomitrella* has 27 chromosomes (n=27) and, according to current information, a genome size of 511 MB, about twice that of *Arabidopsis*. Current hypotheses on evolution hold that Bryophytes and flowering plants share a common ancestor, with Bryophytes diverging approximately 450 Mya from the main line of evolution leading to flowering plants. Recent EST sequence data and transcriptome data reveal that *Physcomitrella* and *Arabidopsis* share a high degree of homology (Reski et al., 1998). More than 66% of *Arabidopsis* proteins have homologs in *Physcomitrella*. A variety of genes that have been cloned from *Physcomitrella* are remarkably homologous to higher plant genes (Reski, 1998). High efficiency of homologous recombination (Schaefer & Zryd, 1997; Girke et al., 1998) enables gene targeting as a practical technique to study gene function in *Physcomitrella*.

The gene *KO9* was identified in Chapter 2 of this dissertation as one of the candidates to encode the rhamnosyl 3-*O*-methyltransferase that synthesizes 3-*O*-Me-Rha in *Physcomitrella*. The initial identification of *KO9* as a candidate was based on a

bioinformatics search that found the predicted KO9 protein to contain the conserved domain of the TylF superfamily of *O*-methyltransferases. This *KO9* candidate was selected for further study in this dissertation because the 3-*O*-Me-Rha/Rha ratio in arabinogalactan proteins (AGPs) from the *Physcomitrella ko9* knockout mutant was found to be statistically very significantly reduced compared to the ratio in AGPs from the wild type. Results presented in this chapter show that KO9 has high homology with angiosperm CCoAOMTs. To gain evidence on whether the KO9 protein functions as rhamnosyl 3-*O*-methyltransferase or as CCoAOMT, properties of *Physcomitrella ko9* knockout mutants and *KO9* transgenic tobacco plants were analyzed.

MATERIALS AND METHODS

Nucleic acid extraction. DNA and RNA of moss leafy gametophytes and tobacco leaves were extracted using the protocols described in detail in Chapter 3.

Plasmid construction and generation of mutant plants. Construction of the *KO9* knockout cassette, its insertion into *Physcomitrella* protoplasts, and selection of the resulting *KO* knockout mutants have been described in detail in Chapter 2. For transgenic expression of *KO9* in *Nicotiana tabacum* cv Xanthi, the cDNA clone pdp74952 was purchased from the RIKEN center (Saitama, Japan, http://www.brc.riken.jp/lab/epd/Eng/). The open reading frame (ORF) of this cDNA was PCR amplified with exTaq (Takara Bio, Mountain View, CA, USA) using the primers [TGCTCTAGAAGCTGTTTACCACTAT] and [TTACCCGGGGTTAAAACAACTTTCG], and then cloned into the binary vector pBI121 (Komori et al., 2007) with XbaI and SmaI (NEB, Ipswich, MA, USA). Aside from this use of the pBI121 binary vector in place of the pART7/pART27 binary vector, the details of generation of the transgenic tobacco plants are the same as in Chapter 3.

PCR and RT-PCR: PCR and RT-PCR programs were the same as used in Chapter 3. The PCR primers used for testing the *ko9* knockout cassette insertion were P1-TTATGTGGGATAGTATCAAGG, P2-CATCCCTTACGTCAGTGGAGA, P3-GAGGCCGTTAGGGAAAAGATG, and P4-CATGTTTAAGACGCATCAACGAAAT. The RT-PCR primers for testing the mRNA level in *ko9* knockout mutants of moss were Forward-AGCGCCGGTAATTCAAAAG, Reverse-GCCAGGAAGCTGTTCAACTC. Genomic PCR and RT-PCR primers for testing transgenic tobacco plants were Forward-CGTCGAAGAGGTGGAAGAAG, Reverse-TCCAGAGCGTGTTGTCGTAG.

Moss culture. Moss cultures were grown in same media and conditions as described in detail in Chapter 2.

Analysis of glycosyl composition of tobacco AGPs. Extraction of AGPs from transgenic tobacco plants and analysis of the glycosyl composition of the AGPs by gas chromatography (GC) and gas chromatography-mass spectrometry (GC-MS) of trimethylsilylether derivatives of methylglycosides were done exactly as described in Chapter 3.

Bioinformatics searches and generation of a phylogenetic tree. BlastP searches for orthologs and generation of a phylogenetic tree were done using the same procedures and internet resources as described in Chapter 3.

Determination of lignin content by the Klason assay. Moss leafy gametophytes were ground to fine powders in liquid nitrogen with a mortar and pestle. These fine powders were extracted in 80% (v/v) ethanol until the extract remained clear. Pellets were collected by centrifugation at 3000 rpm. After extraction, the pellets were dried overnight in an 80°C oven to yield the crude cell wall fraction (alcohol-insoluble residue). A 150 mg aliquot of the dried cell wall powder was mixed with 3 ml of 72% (w/w) H_2SO_4, and the cell wall material was allowed to swell at room temperature for 0.5 h. The suspension of wall material was then diluted with distilled water to 4% H_2SO_4 and autoclaved for 1 h at 121°C to hydrolyze all glycosidic linkages. The resulting black solid residue was filtered through a preweighed Whatman GF/A glass fiber filter paper (70 mm). The filtrate still remaining on the glass fiber filter was washed with distilled water,

dried at 90°C overnight, and then weighed. The mass of dried residue on the filter divided by the 150 mg starting mass of cell wall powder gave the lignin content expressed as w/w percentage (Weng et al., 2010).

Determination of lignin content by the acetyl bromide assay. A cell wall fraction was purified from leafy gametophytes of *Physcomitrella* wild type and *ko9* mutants, according to the methods described by Nothnagel & Nothnagel (2007). In brief, the leafy gametophytes were blended in a homogenizing buffer, filtered, and centrifuged to remove buffer-soluble components. The resulting crude cell wall pellets were then extracted with a sequence of solvents to remove impurities, with centrifugation after each extraction to recover the cell wall pellets. The sequential extractions were done with phenol-acetic acid-water to remove proteins, 70% (v/v) ethanol to remove traces of phenol-acetic acid-water, 90% (v/v) dimethylsulfoxide to remove starch, 70% (v/v) ethanol to remove traces of dimethylsulfoxide, 2:1 chloroform:methanol (v/v) to remove lipids and other organic-soluble components, and finally with acetone to remove traces of chloroform:methanol. After the final centrifugation and removal of the supernatants, the remaining acetone was removed from the pellets by overnight incubation in a vacuum desiccator. The purified cell wall pellets were then transferred to a different vacuum desiccator where they were dried over P_2O_5 to remove any residual traces of water.

Total lignin contents of the cell wall fractions were determined by the acetyl-bromide assay (Fry, 1988; Hatfield et al., 1999; Fukushima & Hatfield, 2001). Dry, purified cell wall (0.5-2 mg) was weighed into 2 ml conical glass vials with Teflon-lined

screw caps, and 0.5 ml of 1:3 (v/v) acetyl bromide/acetic acid was carefully added with an oven-dried glass pipet. (Caution: Acetyl bromide is a fuming liquid that reacts violently with water and causes severe skin burns and eye damage. This chemical must be used only in a fume hood, and appropriate personal protective equipment must be worn by the user.) The vials were tightly capped and placed in a dry heating block at 50°C for 4 h to solubilize cell wall material. After centrifugation, oven-dried glass pipets were used to transfer 0.4 ml of the resulting supernatant from each vial into a clean glass tube where it was mixed with 2 ml of acetic acid and 0.4 ml of 2 M NaOH. After that, 0.2 ml of 1.5 M hydroxylamine hydrochloride and another 1.0 ml of acetic acid were added and mixed. The absorbance at 280 nm of the final solution was then measured against a blank prepared in the same manner but with no cell wall material. Lignin content was calculated according to the conversion that lignin at 10 μg/ml in the final solution gives absorbance of 0.24 at 280 nm (Fry, 1988).

Analysis of variance and the Tukey-Kramer posttest of significance were performed on the acetyl bromide assay results using InStat (version 2.0, Graphpad Software, San Diego, CA).

RESULTS

Physcomitrella KO9 has high sequence similarity to CCoAOMTs in other plants. KO9 is annotated as a family 3, *O*-methyltransferase in the JGI *Physcomitrella patens* v1.1 database. A BlastP search revealed that *Physcomitrella* KO9 shared high degrees of similarity with CCoAOMT, a key enzyme in monolignol biosynthesis, from a variety of seedless plants, gymnosperms, and angiosperms (Table 4.1). Orthologs of KO9 were found in spike moss, coniferals, angiosperm monocots, and angiosperm dicots. Identity between KO9 and these orthologs ranged from 58% to 69%. The E-values of these orthologs are also shown in Table 4.1 and ranged from E-83 to E-118. A phylogenetic tree was built in the PHYLIP server using the ONE-CLICK model (Fig. 4.1). From this phylogenetic tree, evolutionary relationships between CCoAOMTs can be inferred. The Bryophyte *Physcomitrella* KO9 was positioned in the tree closest to orthologous CCoAOMTs from *Selaginella moellendorffii* (common name spikemoss, but it is not a moss, rather a Lycophyte). Based on this phylogenetic tree, KO9 was also evolutionarily related to gymnosperm CCoAOMTs including those in conifer and pine, and to angiosperm CCoAOMTs, such as those in rice, maize, *Arabidopsis* and tobacco.

Generation of stable KO9 knockout mutants. *KO9* is a gene with five exons and four introns. As described in detail in Chapter 2, a knockout cassette for *KO9* was constructed and consisted of a NPT selection marker and *KO9* genomic sequences for the promotion of homologous recombination (Fig. 4.2A). The knockout cassette was introduced into moss protoplasts for the generation of stable knockout mutants. Seven lines of *ko9* knockouts, arising from separate transformation events, survived through

three rounds of selection on medium containing G418 antibiotic. PCR analysis showed that the NPT knockout cassette had been integrated into the *Physcomitrella* genome in all seven knockout lines (Fig. 4.2B). PCR primers P1+P4 amplified the endogenous *KO9* gene in the wild-type moss but not in any of the seven *ko9* knockout mutants. Primers P1+P2 and P3+P4 amplified the left and right insertion junctions, respectively, between the knockout cassette and the moss genome in all seven *ko9* knockouts but not in wild-type *Physcomitrella* (Fig. 4.2B).

Knockout of KO9 does not alter lignin-like abundance in Physcomitrella *cell walls*. In the analysis of glycosyl composition of AGPs purified from the *ko9* knockouts, the 3-*O*-Me-Rha/Rha ratio was statistically very significantly decreased in *ko9* knockout mutants compared to the wild-type (Chapter 2). Because the phylogenetic analysis showed that KO9 shared high degrees of homology with CCoAOMTs from a variety of plants (Table 4.1, Fig. 4.1), the lignin or lignin-like contents of the cell wall of the *Physcomitrella* KO9 knockout mutants was measured and compared with that of the wild-type to test whether KO9 functions as a CCoAOMT in moss.

Two different methods of lignin analysis were used in these experiments. Klason lignin analysis started with alcohol insoluble residue, a relatively crude cell wall preparation. The strategy of the Klason assay is to hydrolyze all glycosidic bonds in the cell wall, thus releasing all of the carbohydrate from the cell wall. The residual mass of the hydrolyzed walls is interpreted as lignin. Applied to cell walls from leafy gametophytes from four *ko9* knockout lines, the Klason assay found lignin contents

ranging from 10.9% to 13.8% (w/w) (Table 4.2). The lignin content similarly determined for cell walls from the wild type was 12.7% (w/w), indicating that lignin content did not significantly differ between the *ko9* knockouts and the wild type. The Klason assay of lignin is subject to interference from proteins and other molecules that might not be completely cleaved and removed by treatment with 4% H_2SO_4 for 1 h at 121°C. Because many cytoplasmic proteins precipitate in 80% (v/v) ethanol, the alcohol insoluble residue used as the crude cell wall fraction for the Klason assay likely contained many proteins unrelated to the cell wall.

To avoid these potential concerns with the Klason assay, a different assay for lignin was also used. The strategy of the acetyl bromide assay is to thoroughly acetylate all cell wall polymers, thereby rendering them soluble in acetic acid. The 280 nm absorbance of the solubilized walls is then measured and attributed to absorbance by the phenolic groups of lignin. This acetyl bromide assay of lignin is subject to interference from aromatic residues of proteins, nucleic acids, and any other non-lignin molecules that might be present and have 280-nm absorbance. Thus, it is essential that a highly purified cell wall fraction be used when the acetyl bromide assay is used determine lignin content. In this experiment, highly purified cell walls from the *Physcomitrella* wild-type, *ko9* knockouts, and a *ko1* knockout were assayed. The lignin-like content of the wild-type cell walls was 145 mg lignin/g cell wall (Table 4.3A), slightly less than the 184 mg lignin/g cell wall reported for *Physcomitrella* by Espineira et al. (2011) and slightly more than the 12.7% (=127 mg lignin/g cell wall) found by the Klason assay (Table 4.2). Six *ko9* knockout lines were grouped into two pools, *ko9*-127 and *ko9*-345, for which the acetyl

bromide assay detected 135 and 145 mg lignin/g cell wall, respectively (Table 4.3A), neither of which was statistically different from the result for the wild-type (Table 4.3B). Highly purified cell walls from *Physcomitrella ko1* knockout mutant were included in this experiment as an out group, and the acetyl bromide assay surprisingly detected only 117 mg lignin/g cell wall for these *ko1* samples, an amount statistically significantly less than in either the wild type or the *ko9* mutant.

Transgenic expression of KO9 in Nicotiana tabacum cv Xanthi. To further analyze the function of *KO9* gene, *KO9* was expressed in *Nicotiana tabacum* cv Xanthi. Transgenic tobacco plants were regenerated and selected on medium containing Kanamycin antibiotic and then moved to soil in pots in a greenhouse. PCR using as the template genomic DNA from transgenic *KO9* plants showed a detectable PCR band in 32 plants, indicating that the *KO9* gene had inserted into the genomes of these 32 plants (Fig. 4.3). RT-PCR bands, indicating the presence of KO9 mRNA, were observed for 12 transgenic plants (Fig. 4.4).

Transgenic KO9 tobacco plants do not produce AGPs containing 3-O-*Me-Rha*. All 12 transgenic tobacco plants showing evidence of *KO9* mRNA were tested for the presence of 3-*O*-Me-Rha in AGPs. Leaf tissue was homogenized, AGPs were purified from the homogenate, and then AGPs were methanolyzed to simple methyl glycosides. The methylglycosides were then derivatized to trimethylsilyl ethers and analyzed by GC and GC-MS. The GC-MS results were closely examined for the presence of a m/z 146 ion at the expected retention times for the derivatives of 3-*O*-Me-Rha. The m/z 146 ion

arises from a fragment consisting of two side-by-side C atoms of the sugar backbone (usually C2 and C3, or C3 and C4), one of these C atoms carrying a methyl ether and the other carrying a trimethylsilyl ether. A strong m/z 146 ion is diagnostic for a methylated sugar, since this ion is very weak or absent from the mass spectra of unmethylated sugars. A very, very weak m/z 146 ion was observed in four of the 12 specimens from transgenic *KO9* plants (results not shown). Very close examination of specimens from wild-type tobacco, however, sometimes also showed a similar very, very weak m/z 146 ion. Further investigation was undertaken by hydrolyzing the tobacco AGPs to simple sugars which were then derivatized to alditol acetates. Careful analysis of the alditol acetates by GC-MS revealed the presence of a very, very weak 3-*O*-methyl-pentose in the AGPs from both wild-type and *KO9* transgenic tobacco plants (results not shown). When derivatized as a trimethylsilyl ether of a methylglycoside, this 3-*O*-methyl-pentose would produce a m/z 146 ion fragment, just the same as does 3-*O*-Me-Rha. The very, very weak m/z 146 ion observed in the four transgenic *KO9* plants was no doubt due to this 3-*O*-methyl-pentose, which seems to have not been previously detected in tobacco. With this 3-*O*-methyl-pentose issue resolved, the glycosyl compositions presented in Table 4.4 show that none of the 12 transgenic *KO9* tobacco plants produced leaf AGPs having detectable 3-*O*-Me-Rha content. Flowers, stems, and roots of two of the *KO9* transgenic plants (*ko9*-2 and *ko9*-46) were also sampled, but AGPs from these organs also lacked detectable 3-*O*-Me-Rha (results not shown).

DISCUSSION

Lignification is an important feature that seems to have originated at the time that plants emerged from aquatic environments and adapted to terrestrial environments (Kenrick & Crane, 1997). Comparative genomics studies have revealed that the complete monolignol biosynthesis pathway first appeared in moss as represented by *Physcomitrella* (Xu et al., 2009). Nevertheless, differences of opinion exist as to whether mosses actually make polymeric lignin (Vanholme et al., 2010). Weng and Chapple (2010) argue that bryophytes do not synthesize insoluble polymeric lignin in their cell walls but instead use the lignin monomers as building blocks for lignans, which are soluble dimers in the protoplasm. Weng and Chapple (2010) further argue that mosses make lignans, not to strengthen the cell wall, but rather to accumulate them in the vacuoles of the cells to protect against damaging UV-B radiation from the direct sun to which early terrestrial plants were exposed when they moved out of their aquatic environment and onto the land. Besides lignans, other soluble phenylpropanoids such as flavonoids have also been proposed to protect against UV irradiation (Basile et al., 1999; Umezawa, 2003).

While *Physcomitrella* seems to have all eight of the enzymes required to make the three lignin monomer types H, G, and S, chemical analysis of lignin monomer composition has revealed that *Physcomitrella* contains only the H building blocks and none of the G or S blocks that carry one or two methyl ether groups (Espiñeira et al., 2011).

In the present chapter, phylogenetic analysis revealed that *Physcomitrella* KO9 is an ortholog of plant CCoAOMTs, one of the eight important enzymes in lignin monomer biosynthesis (Table 4.1, Fig. 4.1). The *Physcomitrella ko9* knockouts generated in Chapter 2 were thus revisited here to test for possible *KO9* function relative to lignin monomer biosynthesis. Two different lignin assays were performed on cell wall preparations from *ko9* knockout moss, however, and neither assay showed statistical differences relative to the wild-type (Tables 4.2, 4.3). These results appear to be in contrast to results from poplar and alfalfa plants, where downregulation of CCoAOMT activity resulted in reduced lignin content (Meyermans et al., 2000; Zhong et al., 2000; Guo et al., 2001). One possible reason for this contrast might be that *KO9* does not encode a CCoAOMT in *Physcomitrella*. Another possible reason for the contrast might be that *Physcomitrella* does not make polymeric lignin, so the Klason and acetyl bromide assays are measuring the lignan dimers or other lignin monomer adducts.

In this regard, the sequential extractions used to purify the cell wall fraction from *Physcomitrella* for the acetyl bromide assay are noteworthy (see Methods). If lignan dimers in mosses and other primitive plants are principally accumulated in soluble form in the vacuole, as Weng and Chapple (2010) suggest, then it seems highly likely that these lignans would have been washed out of the cell wall fraction by the extractions with phenol-acetic acid-water, 70% (v/v) ethanol, 90% (v/v) dimethylsulfoxide, 2:1 chloroform:methanol (v/v), and acetone. Even after these extractions, however, the purified cell wall fraction of wild-type *Physcomitrella* still showed 145 mg lignin/g cell wall by the acetyl bromide assay (Table 4.3A), slightly less than the 184 mg lignin/g cell

wall reported for *Physcomitrella* by Espiñeira et al. (2011). This lignin-like content found for *Physcomitrella* by Espiñeira et al. (2011) was actually higher than the 60 to 100 mg lignin/g cell wall levels found by the same authors for some lycophytes and ferns that have vasculature and make polymeric lignin. Clearly, more remains to be learned about why *Physcomitrella* makes lignin monomers, and how so much supposedly non-polymeric lignin like material associates so tightly with the moss cell wall.

In Chapter 2, analysis of the *Physcomitrella ko9* knockout mutants showed that the 3-*O*-Me-Rha/Rha content ratio in AGPs was significantly decreased in *ko9* mutants compared to the wild type, indicating that *KO9* might be involved in methylation of terminal rhamnosyl residues in the arabinogalactans. Similar to the approach taken with *KO1* in Chapter 3, the hypothesis that KO9 is a rhamnosyl 3-*O*-methyltransferase was tested in this study by transgenic expression of *KO9* in tobacco. Since tobacco AGPs contain terminal, unmethylated rhamnosyl residues on the glycan chains (Liang et al., 2010), tobacco AGPs might be a good substrate for the rhamnosyl 3-*O*-methyltransferase from *Physcomitrella*. Although 12 transgenic tobacco plants generated in the present study accumulated some level of *KO9* mRNA (Figs. 4.3, 4.4), none of the transgenic tobacco plants had a detectable level of 3-*O*-Me-Rha in their AGPs (Table 4.4). As in Chapter 3, at least two alternative hypotheses seem consistent with these results on transgenic tobacco. One hypothesis would be that *KO9* does encode a rhamnosyl 3-*O*-methyltransferase, but the expressed mRNA does not lead to accumulation of the protein, or that the protein is incorrectly targeted, i.e., the protein accumulates in a subcellular location where it does have access to the tobacco AGPs. An alternative hypothesis would

be that *KO9* does not encode a rhamnosyl 3-*O*-methyltransferase, and that the effect of the *ko9* knockout on 3-*O*-Me-Rha/Rha ratio in *Physcomitrella* AGPs occurs indirectly.

REFERENCES

Basile A, Giordano S, Lopez-Saez JA, Cobianchi RC (1999) Antibacterial activity of pure flavonoids isolated from mosses. Phytochemistry **52**: 1479-1482

Battle M, Bender ML, Tans PP,White JW, Ellis JT, Conway T, Francey RJ (2000) Global carbon sinks and their variability inferred from atmospheric O_2 and δ13C. Science **287**: 2467–2470

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol **54**: 519-546

Cantoni GL (1975) Biological methylation: selected aspects. Annu Rev Biochem **44**: 435-451

Espiñeira JM, Novo Uzal E, Gomez Ros LV, Carrion JS, Merino F, Ros Barcel A, Pomar F (2011) Distribution of lignin monomers and the evolution of lignification among lower plants. Plant Biol **13**: 59-68

Ferrer JL, Zubieta C, Dixon RA, Noel JP (2005) Crystal structures of alfalfa caffeoyl coenzyme a 3-O-methyltransferase. Plant Physiol **137**: 1009–1017

Fontecave M, Atta M, Mulliez E (2004) *S*-adenosylmethionine: nothing goes to waste. Trends Biochem Sci **29**: 243-249

Fry SC (1988) The growing plant cell wall: Chemical and metabolic analysis. Longman Scientific and Technical, Harlow, UK

Fukushima RS, Hatfield RD (2001) Extraction and isolation of lignin for utilization as a standard to determine lignin concentration using the acetyl bromide spectrophotometric method. J Agric Food Chem **49**: 3133-3139

Girke T, Schmidt H, Zahringer U, Reski R, Heinz E (1998) Identification of a novel delta 6-acyl-group desaturase by targeted gene disruption in *Physcomitrella patens*. Plant J **15**: 39-48

Guo D, Chen F, Inoue K, Blount JW, Dixon RA (2001) Downregulation of caffeic acid 3-*O*-methyltransferase and caffeoyl CoA 3-*O*-methyltransferase in transgenic alfalfa impacts on lignin structure and implications for the biosynthesis of G and S lignin. Plant Cell **13**: 73-88

Hatfield RD, Grabber J, Ralph J, Brei K (1999) Using the acetyl bromide assay to determine lignin concentrations in herbaceous plants: Some cautionary notes. J Agric Food Chem **47**: 628-632

Ibar C, Orellana A (2007) The import of *S*-adenosylmethionine into the Golgi apparatus is required for the methylation of homogalacturonan. Plant Physiol **145**: 504-512

Kende H (1993) Ethylene biosynthesis. Annu Rev Plant Physiol Plant Mol Biol **44**: 283-307

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

Komori T, Imayama T, Kato N, Ishida Y, Ueki J, Komari T (2007) Current status of binary vectors and superbinary vectors. Plant Physiol **145**: 1155-1160

Lange BM, Lapierre C, Sandermann H (1995) Elicitor induced spruce stress lignin. Structural similarity to early developmental lignins. Plant Physiol **108**: 1277-1287

Liang Y, Faik A, Kieliszewski M, Tan L, Xu WL, Showalter AM (2010) Identification and characterization of in vitro galactosyltransferase activities involved in arabinogalactan-protein glycosylation in tobacco and *Arabidopsis*. Plant Physiol **154**: 632-642

Luka Z, Mudd SH, Wagner C (2009) Glycine *N*-methyltransferase and regulation of *S*-adenosylmethionine levels. J Biol Chem **284**: 22507-22511

Meyermans H, Morreel K, Lapierre C, Pollet B, De Bruyn A, Busson R, Herdewijn P, Devreese B, Van Beeumen J, Marita JM, Ralph J, Chen C, Burggraeve B, Van Montagu M, Messens E, Boerjan W (2000) Modifications in lignin and accumulation of phenolic glucosides in poplar xylem upon down-regulation of caffeoyl-coenzyme A *O*-methyltransferase, an enzyme involved in lignin biosynthesis. J Biol Chem **275**: 36899-36909

Nothnagel AL, Nothnagel EA (2007) Primary cell wall structure in the evolution of land plants. J Integr Plant Biol **49**: 1271-1278

Patterson J, McConville M, Haites R, Coppel R, Billman-Jacobe H (2000) Identification of a methyltransferase from *Mycobacterium smegmatis* involved in glycopeptidolipid synthesis. J Biol Chem **275**: 24900-24906

Ravanel S, Gakière B, Job D, Douce R (1998) The specific features of methionine biosynthesis and metabolism in plants. Proc Natl Acad Sci U S A **95**: 7805-7812

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-I T, 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, Cuming 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

Reski R, Reynolds S, Wehe M, Kleber-Janke T, Kruse S (1998) Moss (*Physcomitrella patens*) expressed sequence tags include several sequences which are novel for plants. Bot Acta **111**: 145-151

Reski R (1998) Development genetics and molecular biology of mosses. Bot Acta **111**: 1-15

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

Umezawa T (2003) Diversity in lignan biosynthesis. Phytochemistry Reviews **2**: 371-390

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. Plant Physiol **153**: 895-905

Vidgren J, Svensson LA, Liljas A (1994) Crystal structure of catechol *O*-methyltransferase. Nature **368**: 354–358

Wallsgrove RM, Lea PJ, Miflin BJ (1983) Intracellular localization of aspartate kinase and the enzymes of threonine and methionine biosynthesis in green leaves. Plant Physiol **71**: 780-784

Weng JK, Li X, Stout J, Chapple C (2008) Independent origins of syringyl lignin in vascular plants. Proc Natl Acad Sci USA **105**: 7887-7892

Weng JK, Chapple C (2010) The origin and evolution of lignin biosynthesis. New Phytol **187**: 273-285

Weng JK, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C (2010) Convergent

evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte *Selaginella* and flowering plants. Plant Cell **22**: 1033-1045

Xu Z, Zhang D, Hu J, Zhou X, Ye X, Reichel KL, Stewart NR, Syrenne RD, Yang X, Gao P, Shi W, Doeppke C, Sykes RW, Burris JN, Bozell JJ, Cheng ZM, Hayes DG, Labbe N, Davis M, Stewart CN Jr, Yuan JS (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. BMC Bioinformatics **10**: S3

Zhong R, Morrison WH, Himmelsbach DS, Poole FL, Ye ZH (2000) Essential role of caffeoyl coenzyme A *O*-methyltransferase in lignin biosynthesis in woody poplar plants. Plant Physiol **124**: 563-578

Zhong R, Ye ZH (2009) Secondary cell walls. In: Encyclopedia of Life Sciences. John Wiley Sons, Ltd: Chichester DOI: 10.1002/9780470015902.a0021256

TABLES AND FIGURES

Table 4.1. BlastP search of KO9 orthologs. A BlastP search of the EBI Uniprot protein database (www.uniprot.org) found many orthologs of *Physcomitrella* KO9, which are listed in this table along with their species origins. The Entry column in the table shows the entry name of each homolog in the Uniprot database. Blast scores and the lengths of the proteins are also listed. Identities between KO9 and each orthologous protein, as well as the E-value of each match, are shown in the table. All of the KO9 orthologs shown were annotated as CCoAOMTs in different plant species.

Entry	Length	Identity	Score	E-value	Organism
O49499	259	63%	852	3.0×10-111	Arabidopsis thaliana (Mouse-ear cress)
A0SDG2	249	64%	879	1.0×10-115	Chamaecyparis formosensis
A0S6W3	278	58%	842	2.0×10-109	Chamaecyparis obtusa var. formosana
K7YEU4	255	63%	878	3.0×10-115	Cunninghamia lanceolata (China fir) (Pinus lanceolata)
Q9SWB8	247	66%	874	8.0×10-115	Eucalyptus globulus (Blue gum)
E4W6N3	250	69%	897	3.0×10-118	Gossypium hirsutum (upland cotton)
G8XQW1	247	68%	890	3.0×10-117	Gossypium hirsutum (Upland cotton)
Q40313	247	65%	865	9.0×10-114	Medicago sativa (Alfalfa)
Q42945	247	65%	862	5.0×10-113	Nicotiana tabacum (Common tobacco)
O04899	240	64%	861	6.0×10-113	Nicotiana tabacum (Common tobacco)
O24144	239	64%	856	3.0×10-112	Nicotiana tabacum (Common tobacco)
O24151	242	64%	856	4.0×10-112	Nicotiana tabacum (Common tobacco)
O24149	242	65%	856	4.0×10-112	Nicotiana tabacum (Common tobacco)
O24150	242	65%	855	5.0×10-112	Nicotiana tabacum (Common tobacco)
Q9XJ19	260	64%	845	3.0×10-110	Oryza sativa subsp. japonica (Rice)
I3RV46	179	66%	657	5.0×10-83	Oryza sativa subsp. japonica (Rice)
Q1JR82	259	64%	866	2.0×10-113	Picea abies (Norway spruce) (Picea excelsa)
Q9ZTT5	259	67%	861	1.0×10-112	Pinus taeda (Loblolly pine)
Q43095	247	65%	870	3.0×10-114	Populus tremuloides (Quaking aspen)
O65922	247	64%	860	1.0×10-112	Populus trichocarpa (Western balsam poplar)
O65862	247	65%	869	5.0×10-114	Populus trichocarpa (Western balsam poplar)
D8RYY7	258	62%	873	2.0×10-114	Selaginella moellendorffii (Spikemoss)
D8SV37	258	62%	870	5.0×10-114	Selaginella moellendorffii (Spikemoss)
D8RU62	278	64%	851	7.0×10-114	Selaginella moellendorffii (Spikemoss)
D8QWD3	291	62%	846	6.0×10-110	Selaginella moellendorffii (Spikemoss)
Q8H9B6	242	66%	873	2.0×10-114	Solanum tuberosum (Potato)
A0S6W2	249	64%	877	3.0×10-115	Taiwania cryptomerioides
Q43237	242	68%	884	2.0×10-116	Vitis vinifera (Grape)
Q9XGD6	258	63%	839	2.0×10-109	Zea mays (Maize)
Q9XGD5	264	63%	825	4.0×10-107	Zea mays (Maize)

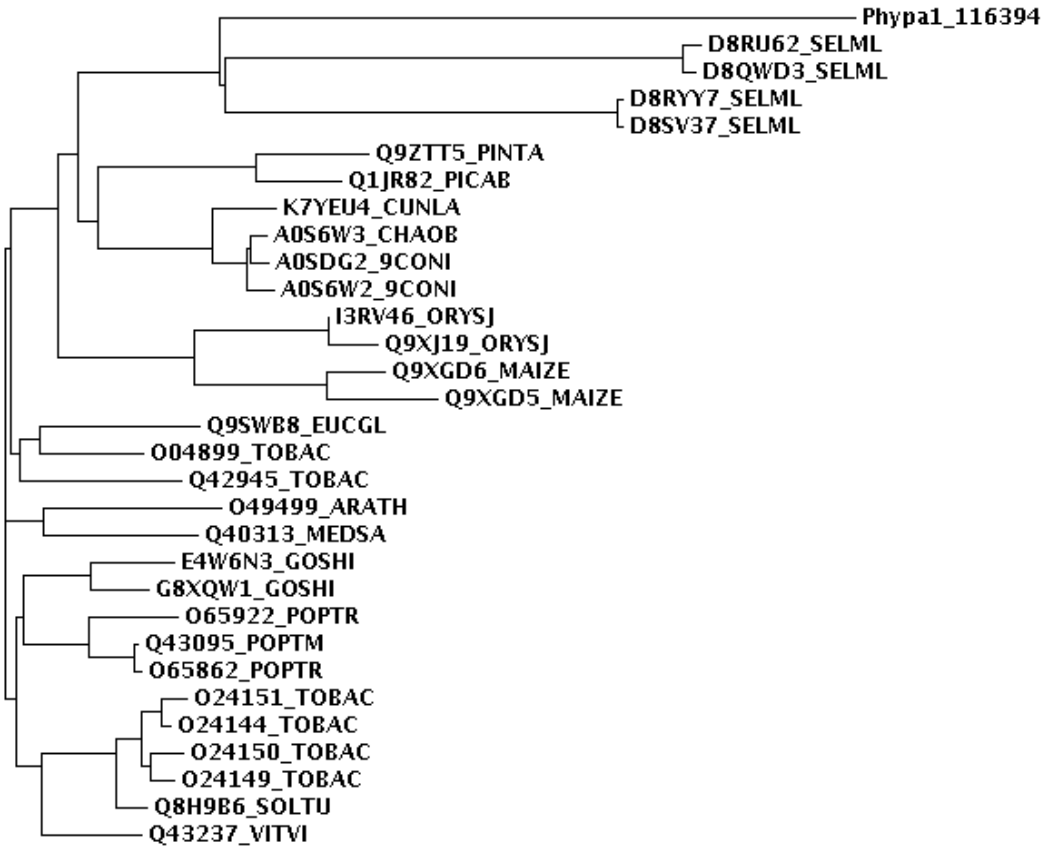


Figure 4.1. Phylogenetic tree of KO9 and its orthologs. The amino acid sequences of the KO9 orthologs listed in Table 4.1, together with the *Physcomitrella* KO9 amino acid sequence, were used to do a MUSCLE alignment for generation of a phylogenetic tree by the ONE CLICK model in the webserver of www.phylogeny.fr. KO9 is shown in the tree as Phypa1_116394. This phylogenetic tree indicates evolutionary relationships of orthologous CCoAOMTs from seedless land plants to gymnosperms (coniferales) and angiosperms (monocots and eudicots). The Bryophyte *Physcomitrella* KO9 was positioned in the tree closest to orthologous CCoAOMTs from *Selaginella moellendorffii* (common name spikemoss, but it is not a moss, rather a Lycophyte).

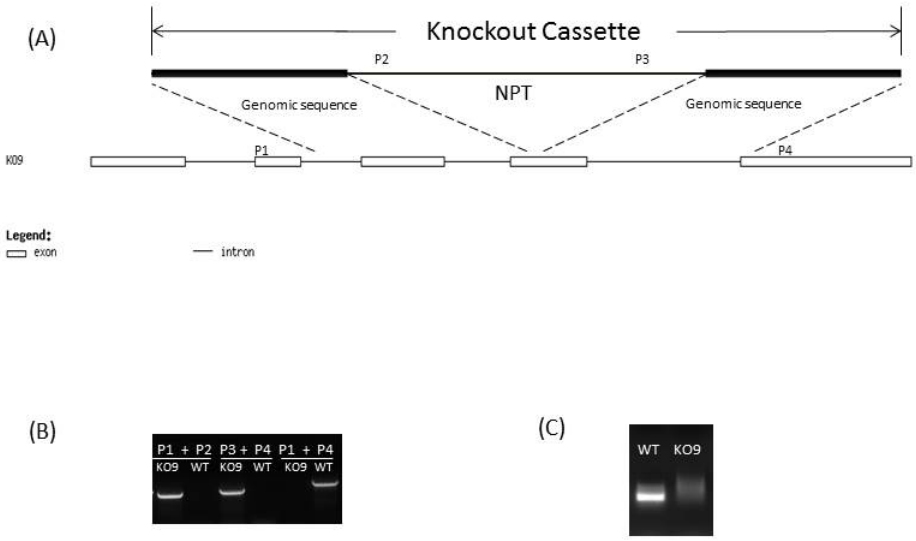


Figure 4.2. Molecular analysis of *ko9* knockout mutant.
Generation and analysis of the *Physcomitrella ko9* knockout mutant. Scheme to show insertion of the knockout cassette into the *KO9* genomic sequence. The *KO9* gene has five exons and four introns. The NPT insertion is in the fourth exon. Mapping of the left and right insertion junctions of the knockout cassette was performed by PCR with primer pairs P1+P2, and P3+P4. Seven stable knockout lines were obtained, and all seven showed positive PCR bands. PCR with primers P1+P4 was performed to amplify a portion of the endogenous *KO9* gene. PCR with DNA from wild type (WT) plants as the template amplified a portion of the *KO9* gene of anticipated size, but PCR with DNA from *ko9* mutants as the template did not amplify a sequence of that size.

Table 4.2. Klason lignin contents of *Physcomitrella* wild type and *ko9*. A 150 mg aliquot of crude cell wall fraction (alcohol-insoluble residue) was used in the Klason assay of lignin content, and the results are dry weight of lignin per dry weight of crude cell wall fraction, expressed as a percent. Data are means and standard deviations from n=2 measurements on the wild-type and each of four separate *ko9* knockout lines. No significant difference of Klason lignin content between wild type and *ko9* knockout mutants was found.

Knockout lines	Klason lignin content (w/w)
Wild type	12.7%±0.9%
ko9-2	11.6%±1.2%
ko9-3	13.8%±1.0%
ko9-5	11.2%±2.1%
ko9-7	10.9%±1.7%

Table 4.3. Analysis of lignin-like content of *ko9* and wild type by acetyl bromide assay.

The assay was applied to dry, purified cell wall fractions from leafy gametophyte cultures of wild type and *ko9* and *ko1* knockout mutants of *Physcomitrella*. (A) Lignin-like content (mean and standard error of the mean, SEM) for several cultures. *ko9*-127 designates a pool composed of plants from the *ko9*-1, *ko9*-2, and *ko9*-7 knockout lines. *ko9*-345 designates a pool composed of plants from the *ko9*-3, *ko9*-4, and *ko9*-5 knockout lines. The *ko1* knockout was from Chapter 3. (B) Analysis of variance for results presented in (A).

A)

Plant type	Number of trials	A_{280}/mg cell wall		mg lignin[a] g cell wall
		Mean	SEM	
Wild-type	16	0.698	0.017	145
KO9-127	16	0.646	0.022	135
KO9-345	14	0.695	0.025	145
KO1	26	0.562	0.020	117

[a]Calculated using A_{280}=0.24 for 10 μg lignin/ml (Fry, 1988)

B)

Comparison	P value
Wild-type vs KO1	P<0.001 Extremely significant
Wild-type vs KO9-127	P>0.05 Not significant
Wild-type vs KO9-345	P>0.05 Not significant
KO1 vs KO9-127	P<0.05 Significant
KO1 vs KO9-345	P<0.001 Extremely significant
KO9-127 vs KO9-345	P>0.05 Not significant

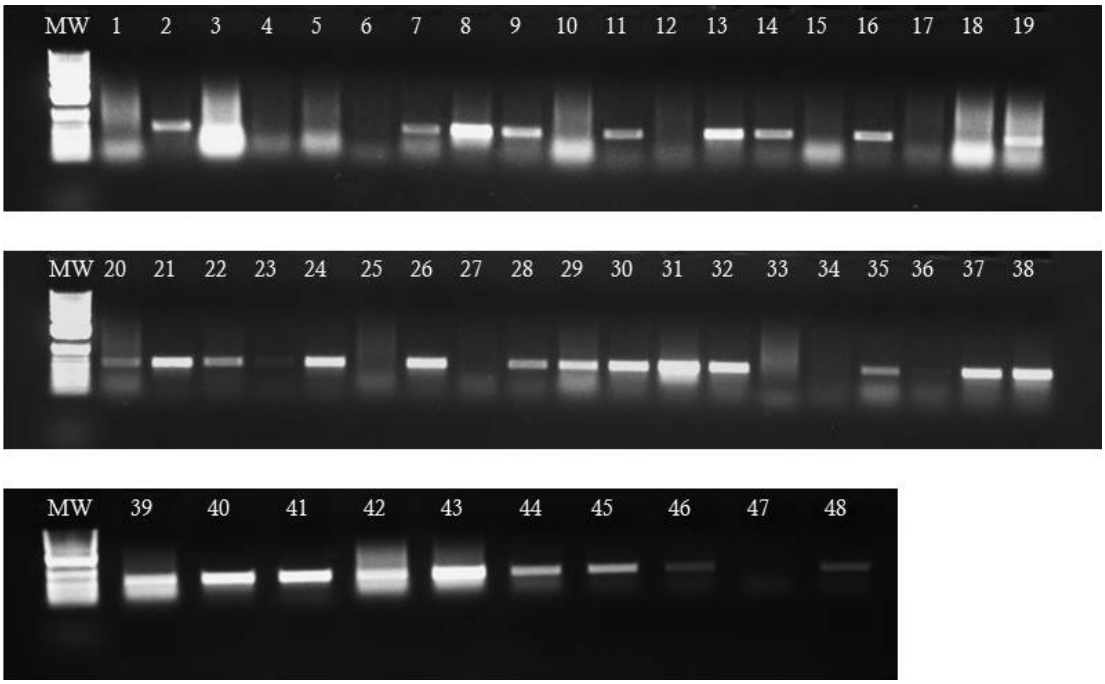


Figure 4.3. Genomic PCR of transgenic *KO9* tobacco plants.
Genomic PCR to test for the presence of the *KO9* gene in regenerated transgenic tobacco plants. In each panel, the left lane shows DNA molecular weight markers (1kb plus DNA ladder, Life Technologies). The transgenic plant number is shown above each lane. Among 48 regenerated transgenic plants, 32 plants (#2, 7, 8, 9, 11, 13, 14, 16, 19, 20, 21, 22, 23, 24, 26, 28, 29, 30, 31, 32, 35, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48) showed a detectable PCR band indicating that the *KO9* gene had inserted into the tobacco genome. In three plants (#23, 46, 48), the PCR band was weak.

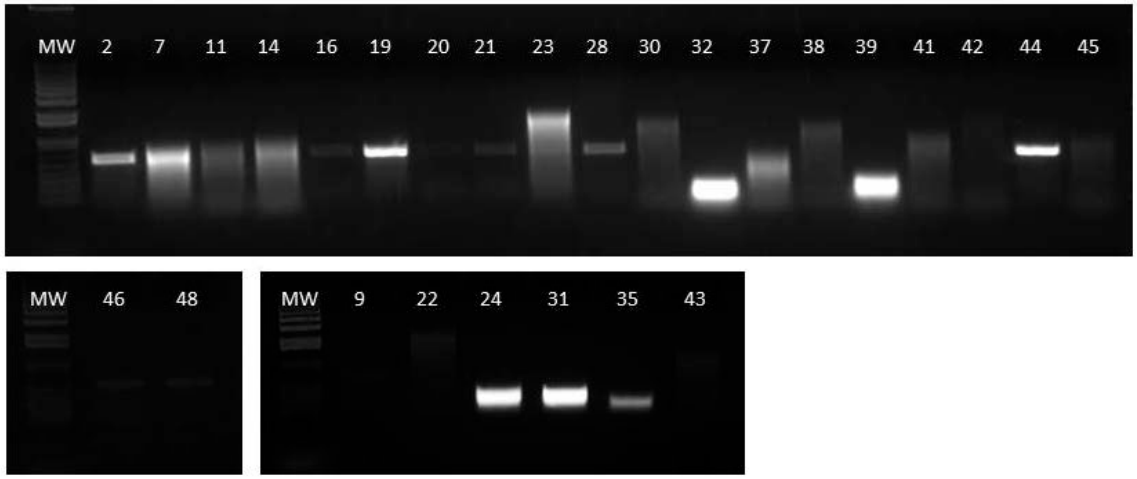


Figure 4.4. RT-PCR of transgenic *KO9* tobacco plants.
RT-PCR to test for *KO9* gene expression (presence of *KO9* mRNA) in transgenic tobacco plants. The transgenic plant number is shown above each lane. Of the 32 successful transgenic lines shown in Fig. 4.3, 27 lines were tested by RT-PCR. Transgenic plants #8, 13, 26, 29, and 40 were not healthy in the growth chamber, were not successfully transferred to the greenhouse, and thus were not tested in RT-PCR. Twelve transgenic plants, including #2, 7, 9, 14, 16, 19, 20, 21, 28, 44, 46, and 48 were observed to have *KO9* gene expression, although for plants #9, 16, 20, 21, 46, and 48 the bands were only weakly detectable at the luminometer and did not convincingly appear in the digital image. The far left lane in each panel is 1kb plus DNA ladder (No. 10787018, Life Technologies).

Table 4.4. Glycosyl compositions of total soluble AGPs from transgenic *KO9* tobacco plants. Wild-type *Physcomitrella* was used as positive control, wild-type tobacco was used as negative control. No 3-*O*-Me-Rha was detected in AGPs from any of the transgenic *KO9* tobacco plants. Glc contents were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations

	Glycosyl composition (mol%)													
	Physco	Tobacco plants												
Residue	Wild type	Wild type	KO9-2	KO9-7	KO9-9	K09-14	KO9-16	KO9-19	KO9-20	KO9-21	KO9-28	KO9-44	KO9-46	KO9-48
3-O-Me Rha	3.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ara	15.58	30.93	24.90	20.33	24.41	27.98	27.78	25.71	25.72	19.61	21.33	28.05	15.97	28.63
Rha	1.29	4.89	4.92	5.36	4.59	4.56	5.09	4.63	4.15	3.51	4.34	4.48	5.07	4.22
Fuc	0.26	0.09	0.11	0.21	0.13	0.20	0.21	0.22	0.10	0.08	0.10	0.12	0.08	0.10
Xyl	2.71	2.09	4.61	3.06	4.99	3.08	3.43	3.52	2.73	3.58	3.28	3.89	3.25	3.10
GlcU	11.73	8.23	10.75	14.20	10.09	11.89	11.84	11.08	7.93	8.52	9.55	8.96	15.78	9.31
GalU	2.28	0.67	2.23	2.18	3.13	1.79	1.91	2.92	1.08	1.85	1.50	1.05	3.24	1.48
Man	1.63	0.92	0.62	0.63	0.72	0.75	0.57	0.63	0.52	0.48	0.53	0.80	1.05	0.97
Gal	46.92	45.96	45.04	48.52	45.82	44.56	43.62	45.89	54.48	59.37	57.05	44.31	48.70	46.68
Glc	13.85	6.24	6.80	5.49	6.12	5.18	5.55	5.39	3.30	2.98	2.32	8.33	6.82	5.51

CHAPTER 5 - HETEROLOGOUS GENE EXPRESSION OF KO11 IN *NICOTIANA TABACUM* CV XANTHI

ABSTRACT

Arabidopsis AtGXMT1 is a domain of unknown function (DUF) family 579 protein that methylates GlcU side residues on a 1,4-linked β-D-xylan backbone hemicellulose to form 4-*O*-Me-GlcU residues. KO11 has been identified as the only predicted *Physcomitrella* protein to contain the same DUF 579 domain as the AtGXMT1 glucuronosyl 4-*O* methyltransferase. No 4-*O*-Me-GlcU has been found in *Physcomitrella*, however, and 3-*O*-Me-Rha is the only methylated sugar found thus far in the moss cell wall. Hence, KO11 might be the rhamnosyl 3-*O*-methyltransferase. To test the hypothesis that KO11 functions as a rhamnosyl 3-*O*-methyltransferase, the *KO11* gene was transgenically expressed in *Nicotiana tabacum* cv Xanthi, which produces arabinogalactan proteins with glycan chains having unmethylated rhamnosyl residues in terminal non-reducing positions. Genomic PCR and RT-PCR showed that eight of the transgenic tobacco plants exhibited *KO11* gene expression. Purification of arabinogalactan proteins from these plants and analysis of their glycosyl compositions revealed that none of the transgenic plants produced arabinogalactan proteins containing 3-*O*-Me-Rha. These results suggest that KO11 may not be a rhamnosyl 3-*O*-methyltransferase, or that KO11 may not be functional in tobacco.

INTRODUCTION

Plant *O*-methyltransferases transfer methyl groups to oxygen atoms on a variety of molecules such as lignin subunits, flavonoids, and carbohydrates to form methyl ethers or methyl esters. The focus of this dissertation is on *O*-methyltransferases that form methyl ethers. Some *O*-methyltransferases that function in lignin biosynthesis, such as caffeic acid *O*-methyltransferase (COMT) and caffeoyl-CoA *O*-methyltransferase (CCoAOMT), have been characterized (Chen et al., 2000; Zhong et al., 2000; Guo et al., 2001). Few *O*-methyltransferases that function in synthesis of plant cell wall carbohydrates have been described. Pectin methyltransferase (PMT), which forms a methyl ester at C6 of GalU resides in pectic polysaccharides, has been biochemically characterized at the enzyme activity level and has been localized in the Golgi apparatus (Vannier et al., 1992; Goubet et al., 1998; Baydoun et al., 1999; Ishikawa, 2000). Through mutational analyses, TSD2 and QUA2, Golgi-localized proteins with a putative methyltransferase domain, have been identified as very good candidates for a PMT in *Arabidopsis* (Krupkova et al., 2007; Mouille et al., 2007).

Methyltransferases involved in synthesis of 4-*O*-Me-D-GlcU residues of hemicelluloses have been characterized at the level of enzyme activity (Baydoun et al. 1989) and have been localized in the Golgi apparatus (Vannier et al., 1992; Baydoun et al., 1999). Recently, Urbanowicz et al. (2012) identified an *Arabidopsis* gene that encodes glucuronoxylan methyltransferase (AtGXMT1), an *O*-methyltransferase that methylates GlcU side residues on a 1,4-linked β-D-xylan backbone hemicellulose to form 4-*O*-Me-GlcU residues. The discovery of this gene marks the first identification of a gene

that encodes an *O*-methyltransferase that forms an *O*-methyl ether on a sugar residue in a plant cell wall polymer. The AtGXMT1 protein contains a conserved DUF579 domain that had not been previously linked to *O*-methylation of polysaccharides. Enzyme activity assays with recombinately expressed AtGXMT1 (less its transmembrane domain) revealed that the enzyme would transfer methyl groups from *S*-adenosyl-methionine (AdoMet) to substrate GlcU residues already incorporated into the glucuronoxylan, but would not transfer methyl groups to monomers such as UDP-GlcU, or free GlcU. Rather unexpectedly, this AtGXMT1 was found to require Co^{2+}, rather than other divalent metal ions such as Mg^{2+}, Ca^{2+}, or Zn^{2+}, for enzyme activity.

The moss *Physcomitrella patens* is becoming an increasingly popular model organism for studies of plant evolution, development and physiology. High efficiency homologous recombination (HR) that enables gene targeting is a key feature of *Physcomitrella* which, when combined with other advantages such as a completely sequenced genome, short life cycle, easy culture and maintenance, makes this moss a powerful model system (Schaefer & Zryd, 1997; Girke et al., 1998).

Phylogenetic analysis of AtGXMT1 revealed that Pp1s15_437V6 (alias PHYPADRAFT_115978, designated here as KO11) is the only *Physcomitrella* ortholog of AtGXMT1 (Urbanowicz et al., 2012). These two proteins both contain a conserved DUF579 conserved domain, which has been identified as being involved in the biosynthesis of xylan and the methylation of GlcU of xylan (Urbanowicz et al., 2012; Lee et al., 2012). Because analyses of *Physcomitrella* cell walls resulted in no evidence of 4-*O*-Me-GlcU and in fact detected 3-*O*-Me-Rha as the only methylated sugar (Nothnagel &

Nothnagel, 2007), KO11 is worthy of investigation as a candidate to be the rhamnosyl 3-*O*-methyltransferase.

Some glycans in tobacco AGPs share a similar structure with some glycans in *Physcomitrella* AGPs, both having non-reducing terminal rhamnosyl residues (Fu et al., 2007; Tan et al., 2010). In the moss most of these terminal rhamnosyl residues are methylated as 3-*O*-Me-Rha, while in tobacco these terminal rhamnosyl residues are unmethylated. This similar structure might make tobacco AGPs a good substrate for the *Physcomitrella* rhamnosyl 3-*O*-methyltransferase *in vivo*.

In this study, *Physcomitrella KO11* was subcloned to a binary vector pBI121 that was used to transgenically express *KO11* in tobacco. AGPs were purified from leaves of the transgenic tobacco plants and analyzed for 3-*O*-Me-Rha content.

MATERIALS AND METHODS

Generation of transgenic tobacco. For transgenic expression of *KO11* in tobacco, the cDNA clone pdp67167 was purchased from the RIKEN center (Saitama, Japan, http://www.brc.riken.jp/lab/epd/Eng/). The open reading frame (ORF) of this cDNA was PCR amplified with exTaq (Takara Bio, Mountain View, CA, USA) using the primers TCATCTAGACCGCAAGTATAGGCCA and TATCCCGGGTCCCATTGTATGCTT, and then subcloned into the binary vector pBI121 (Komori et al., 2007; gift from Dr. Martha L. Orozco-Cardenas, UCR) at XbaI and SmaI sites (NEB, Ipswich, MA, USA). Transformation of this vector into *Nicotiana tabacum* cv Xanthi was done at the UCR Plant Transformation Research Center using a standard *Agrobacterium*-mediated method (Duca et al., 2009).

DNA and RNA extraction. Young leaves from regenerated transgenic plants were ground to a fine powder in liquid nitrogen with a mortar and pestle. DNA was extracted from the frozen powder by the method of Murray & Thompson (1980). Total RNA was extracted using a plant RNA purification buffer (Life Technologies, Carlsbad, CA, USA) according to manufacturer's protocol.

PCR and RT-PCR. The following temperature program was used for PCR: 94 °C 3 min, then 30 cycles of 94 °C 30s, Tm-2 °C 45s, 72 °C min/1kb, final extension 5 min. For RT-PCR, 1 μg of total RNA treated with DNase I (Life Technologies, Carlsbad, CA, USA) was used as the template with the one-step RT-PCR kit (Life Technologies, Carlsbad, CA, USA) according to the manufacturer's protocol.

AGP extraction from transgenic tobacco plants. Tobacco leaves were processed for purification of total soluble AGPs and analysis of AGP glycosyl composition using the methods of Fu et al. (2007), modified to accommodate micro-scale samples of 0.25-0.5 gfw. Full details are provided in Chapter 3.

Analysis of glycosyl composition of AGPs. Glycosyl compositions of AGPs were determined by gas chromatography (GC) and gas chromatography-mass spectrometry (GC-MS) of trimethylsilyl (TMS) ether-methylglycoside derivatives of sugars after methanolysis (Fu et al., 2007). Full details are provided in Chapter 3.

RESULTS

Transgenic *KO11* tobacco plants were generated at the UCR Plant Transformation Research Center. To test for the presence of the *KO11* gene in the genome of the transgenic tobacco plants, genomic PCR was performed (Fig. 5.1). Of the 30 plants that regenerated on selection medium containing antibiotic, 22 plants showed a band in genomic PCR with gene-specific primers. From these 22 plants, 19 plants were successfully transferred to the greenhouse and subsequently analyzed by RT-PCR to test for the presence of *KO11* mRNA. RT-PCR revealed that eight transgenic plants showed gene expression of *KO11* (Fig. 5.2).

Of the eight transgenic plants showing the presence of *KO11* mRNA, seven were available in the greenhouse for harvest of leaves and purification of AGPs from the leaves. Analysis of the glycosyl compositions of AGPs by GC and GC-MS showed that none of the transgenic *KO11* plants had any detectable 3-*O*-Me-Rha in their AGPs (Table 5.1).

DISCUSSION

Physcomitrella KO11 has been identified as the ortholog of *Arabidopsis* glucuronosyl 4-*O*-methyltransferase AtGXMT1 (Urbanowicz et al., 2012). Hemicellulosic xylan in dicots consists of a backbone of β-1, 4-linked xylose with single-residue side chains of glucuronic acid (GlcU), 4-*O*-methylglucuronic acid (4-*O*-Me-GlcU) or arabinose. Among these side residues, 4-*O*-Me-GlcU is particularly abundant in the secondary wall of eudicotyledons leading to the presence of 4-*O*-Me-glucuronoxylan as a major hemicellulose of the secondary cell wall (Pauly et al., 1999). AtGXMT1 transfers methyl groups from *S*- adenosylmethionine to GlcU residues of hemicellulosic xylan in the secondary cell wall of *Arabidopsis* (Urbanowicz et al., 2012). Because *Physcomitrella* KO11 has the same conserved DUF579 domain as AtGXMT1, and because *Physcomitrella* has only 3-O-Me-Rha and no 4-O-Me-GlcU in its cell wall (Nothnagel & Nothnagel, 2007), it seemed possible that KO11 might be the rhamnosyl 3-*O*-methyltransferase.

Although data presented in Chapter 2 of this dissertation showed that the *ko11* knockout mutant of *Physcomitrella* produced AGPs with a 3-*O*-Me-Rha/Rha content ratio that was not different from the wild-type, a second test of the function of *KO11* was carried out by expressing *KO11* in transgenic tobacco. Since tobacco AGPs contain terminal, unmethylated rhamnosyl residues on the glycan chains (Tan et al., 2010), tobacco AGPs might be a good substrate for the rhamnosyl 3-*O*-methyltransferase from *Physcomitrella*. Although eight transgenic tobacco plants generated in the present study

accumulated some level of *KO11* mRNA (Figs. 5.1, 5.2) and AGPs were purified from seven of these plants, none of them had a detectable level of 3-*O*-Me-Rha in their AGPs (Table 5.1).

Thus, the hypothesis that KO11 might be a rhamnosyl 3-*O*-methyltransferase in *Physcomitrella* has been subjected to two different tests. As judged by both lack of effect on 3-*O*-Me-Rha/Rha content ratio in AGPs of *ko11* knockout mutants in *Physcomitrella* and by absence of any detectable 3-*O*-Me-Rha in AGPs from transgenic *KO11* tobacco, the hypothesis appears to be false, i.e., KO11 does not appear to be a rhamnosyl 3-*O*-methyltransferase.

REFERENCES

Baydoun EA, Waldron KW, Brett CT (1989) The interaction of xylosyltransferase and glucuronyltransferase involved in glucuronoxylan synthesis in pea (Pisum sativum) epicotyls. Biochem J **257**: 853-858

Baydoun AR, Wileman SM, Wheeler-Jones CP, Marber MS, Mann GE, Pearson JD, Closs EI (1999) Transmembrane signalling mechanisms regulating expression of cationic amino acid transporters and inducible nitric oxide synthase in rat vascular smooth muscle cells. Biochem J 344 Pt 1:265-272

Chen C, Meyermans H, Burggraeve B, De Rycke RM, Inoue K, De Vleesschauwer V, Steenackers M, Van Montagu MC, Engler GJ, Boerjan WA (2000) Cell-specific and conditional expression of caffeoyl-coenzyme A-3-*O*-methyltransferase in poplar. Plant Physiol **123**: 853-868

Duca M, Glijin A, Lupashcu V, Duca D, Orozco-Cardenas ML (2009) The expression of CsLFY and bar genes at transcription and translation levels in transgenic tobacco plants. Romanian Biotechnological Letters **14**: 4887-4892

Fu H, Yadav MP, Nothnagel EA (2007) *Physcomitrella patens* arabinogalactan proteins contain abundant terminal 3-*O*-methyl-L-rhamnosyl residues not found in angiosperms. Planta **226**: 1511-1524

Girke T, Schmidt H, Zahringer U, Reski R, Heinz E (1998) Identification of a novel delta 6-acyl-group desaturase by targeted gene disruption in *Physcomitrella patens*. Plant J **15**: 39-48

Goubet F, Council LN, Mohnen D (1998) Solubilization and partial characterization of homogalacturonan-methyltransferase from microsomal membranes of suspension-cultured tobacco cells. Plant Physiol **116**: 337-347

Guo D, Chen F, Inoue K, Blount JW, Dixon RA (2001) Downregulation of caffeic acid 3-*O*-methyltransferase and caffeoyl CoA 3-*O*-methyltransferase in transgenic alfalfa. Plant Cell **13**: 73-88

Ishikawa M, Kuroyama H, Takeuchi Y, Tsumuraya Y (2000) Characterization of pectin methyltransferase from soybean hypocotyls. Planta **210**:782-791

Komori T, Imayama T, Kato N, Ishida Y, Ueki J, Komari T (2007) Current status of binary vectors and superbinary vectors. Plant Physiol **145**: 1155-1160

Krupkova E, Immerzeel P, Pauly M, Schmülling T (2007) The *TUMOROUS SHOOT DEVELOPMENT2* gene of *Arabidopsis* encoding a putative methyltransferase is required for cell adhesion and co-ordinated plant development. Plant J **50**:735-750

Lee C, Teng Q, Zhong R, Yuan Y, Haghighat M, Ye ZH (2012) Three *Arabidopsis* DUF579 domain-containing GXM proteins are methyltransferases catalyzing 4-*O* methylation of glucuronic acid on xylan. Plant Cell Physiol **53**: 1934-1949

Mouille G, Ralet MC, Cavelier C, Eland C, Effroy D, Hématy K, McCartney L, Truong HN, Gaudon V, Thibault JF, Marchant A, Höfte H (2007) Homogalacturonan synthesis in *Arabidopsis thaliana* requires a Golgi-localized protein with a putative methyltransferase domain. Plant J **50**: 605-614

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research **8**: 4321-4325

Nothnagel AL, Nothnagel EA (2007) Primary cell wall structure in the evolution of land plants. J Integr Plant Biol **49**: 1271-1278

Pauly M, Andersen LN, Kauppinen S, Kofod LV, York WS, Albersheim P, Darvill A (1999) A xyloglucan-specific endo-β-1,4-glucanase from *Aspergillus aculeatus*: expression cloning in yeast, purification and characterization of the recombinant enzyme. Glycobiology **9**: 93-100

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

Tan L, Varnai P, Lamport DTA, Yuan C, Xu J, Qiu F, Kieliszewski MJ (2010) Plant *O*-hydroxyproline arabinogalactans are composed of repeating trigalactosyl subunits with short bifurcated side chains. J Biol Chem **285**: 24575-24583

Urbanowicz BR, Peña MJ, Ratnaparkhe S, Avci U, Backe J, Steet HF, Foston M, Li H, O'Neill MA, Ragauskas AJ, Darvill AG, Wyman C, Gilbert HJ, York WS (2012) 4-O-methylation of glucuronic acid in *Arabidopsis* glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. Proc Natl Acad Sci U S A **109**: 14253-14258

Vannier MP, Thoiron B, Morvan C, Demarty M (1992) Localization of methyltransferase activities throughout the endomembrane system of flax (*Linum usitatissimum* L) hypocotyls. Biochem J **286**: 863-868

Zhong R, Morrison WH 3rd, Himmelsbach DS, Poole FL 2nd, Ye ZH (2000) Essential role of caffeoyl coenzyme A *O*-methyltransferase in lignin biosynthesis in woody poplar plants. Plant Physiol **124**: 563-578

TABLES AND FIGURES

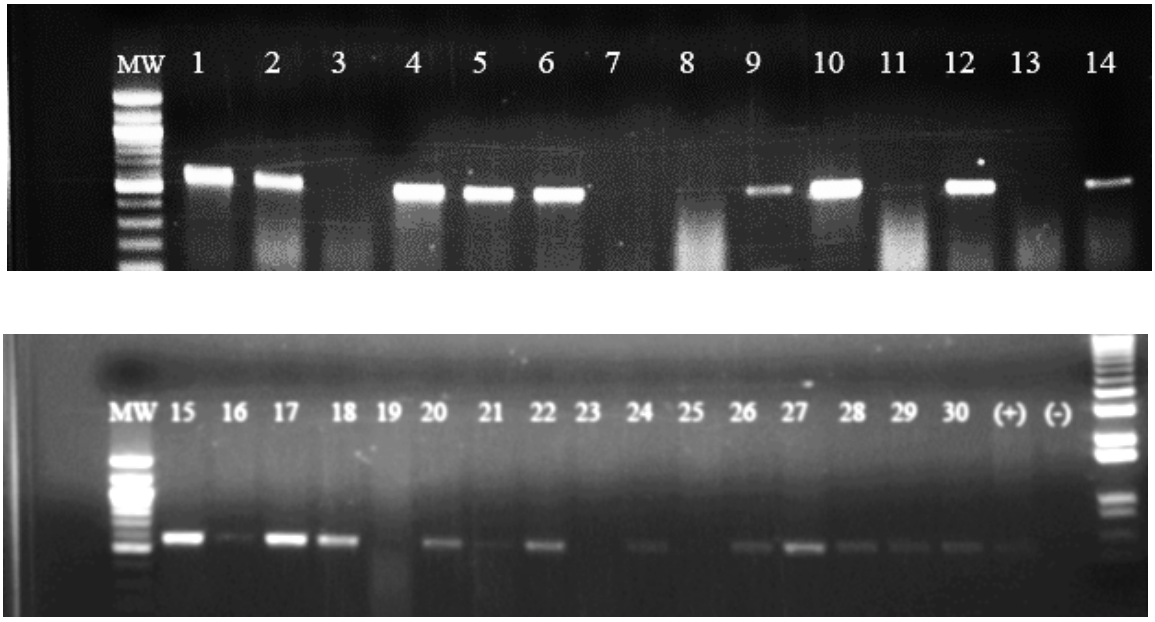


Figure 5.1. Genomic PCR of transgenic *KO11* tobacco plants. PCR was performed with gene specific primers of *KO11* to test for the presence of the *KO11* gene in the tobacco genome. Thirty transgenic plants were tested, and the number of each plant, from 1 to 30, appears above the respective lane in the agarose gel. The first lane (MW) at the left edges of both the upper and lower gels was 1kb DNA ladder (NEB, Ipswich, MA, USA). The far right lane of the lower gel is also a 1 kb plus DNA ladder (Life Technologies, Carlsbed, CA, USA). A gene-specific band was amplified from genomic DNA of 22 transgenic plants (#1, 2, 4, 5, 6, 9, 10, 12, 14, 15, 16, 17, 18, 20, 21, 22, 24, 26, 27, 28, 29, 30). In five plants (#16, 21, 24, 29, 30), the PCR band was weak. Primers used in genomic PCR were TGCGCTTGCTCTGTTACATT and TGCGCACAAAAACTCTTCAC.

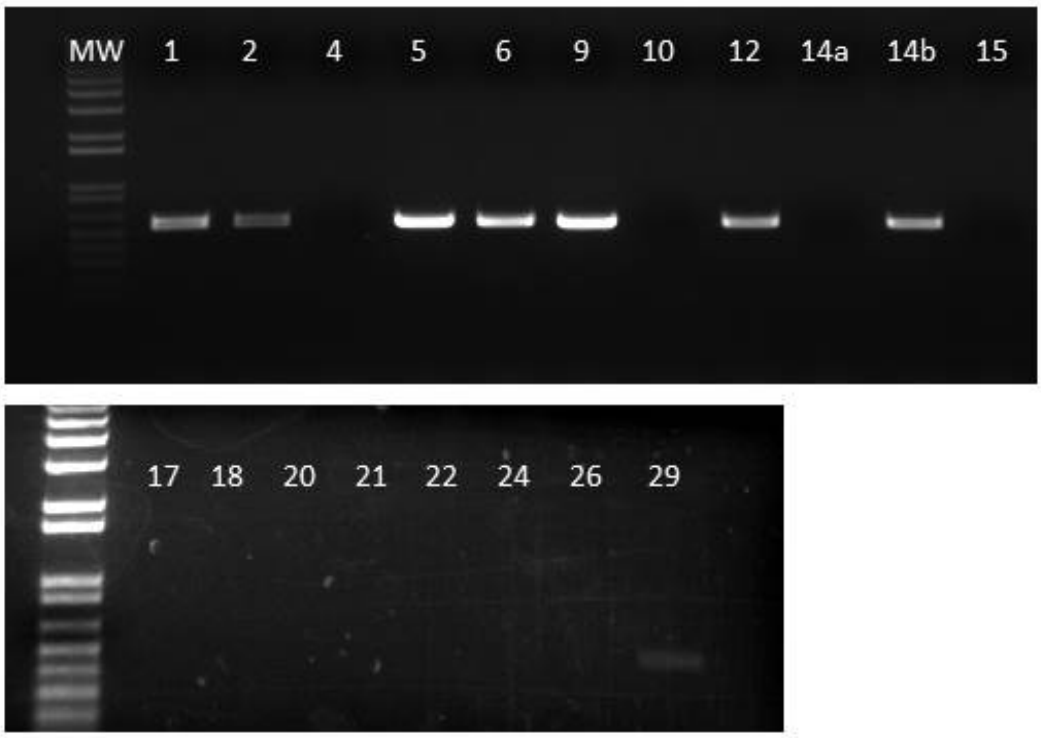


Figure 5.2. RT-PCR of transgenic *KO11* plants. From each transgenic *KO11* plant, 1μg of extracted total RNA was used as the template in RT-PCR to test for the presence of *KO11* mRNA. Primers used were TGCGCTTGCTCTGTTACATT and TGCGCACAAAAACTCTTCAC. The first lane (MW) at the left edges of both the upper and lower gels were 1 kb plus DNA ladder (Life Technologies, Carlsbed, CA, USA). A RT-PCR band, indicating gene expression of *KO11*, was detected in 8 transgenic plants (#1, 2, 5, 6, 9, 12, 14b, 26), although for plant #26 the band was only weakly detectable at the luminometer and did not convincingly appear in the digital image.. Two transgenic plants were found labeled as #14 in the greenhouse and so were relabeled as 14a and 14b. Transgenic plants #16, 27, 28, and 30 were not growing well in the growth chamber and were not successfully transferred to the greenhouse.

Table 5.1. Glycosyl compositions of total soluble AGPs from transgenic *KO11* tobacco plants. Wild-type *Physcomitrella* was used as positive control, wild-type tobacco was used as negative control. Eight *KO11* tobacco plants, seven were confirmed to have *KO11* gene expression in Figure 5.2. One tobacco plant, KO11-7, did not exhibit *KO11* in the genome (Fig. 5.1) and was included as a blind negative control. No 3-*O*-Me-Rha was detected in AGPs from any of the transgenic *KO11* tobacco plants. Glc contents were primarily due to variable amounts of residual $(\beta\text{-D-Glc})_3$ Yariv phenylglycoside in the preparations.

	Glycosyl composition (mol%)									
	Physco	Tobacco plants								
Residue	Wild type	Wild type	KO11-2	KO11-5	KO11-6	KO11-7	KO11-9	KO11-12	KO11-14	KO11-26
3-O-Me Rha	3.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ara	15.58	30.93	27.74	26.80	22.99	27.86	25.49	25.44	15.81	18.90
Rha	1.29	4.89	4.82	4.44	4.68	4.79	4.58	5.61	3.82	4.41
Fuc	0.26	0.09	0.19	0.11	0.17	0.28	0.13	0.11	0.12	0.12
Xyl	2.71	2.09	2.96	3.63	4.23	1.37	3.23	4.37	3.81	3.86
GlcU	11.73	8.23	9.17	8.82	10.32	10.28	10.43	11.36	10.09	11.25
GalU	2.28	0.67	0.82	1.37	2.77	1.30	1.57	1.51	5.01	3.41
Man	1.63	0.92	0.87	0.63	1.32	0.92	0.76	0.99	1.33	1.41
Gal	46.92	45.96	44.96	43.51	42.04	44.60	47.43	45.64	39.18	46.21
Glc	13.85	6.24	8.46	10.69	11.48	8.60	6.38	4.97	20.83	10.43

CHAPTER 6 - CONCLUSIONS

Summary of findings in the dissertation. Arabinogalactan proteins (AGPs) in the moss *Physcomitrella patens* have been found to contain up to 15 mol% of 3-*O*-methyl-L-rhamnosyl (3-*O*-Me-Rha) residues in the glycan chains. This level of methylated sugar residues is among the highest known for plant cell wall polymers, attracting our interest to identify the rhamnosyl 3-*O*-methyltransferase. No plant rhamnosyl 3-O-methyltransferases have been identified so far. We used *Mycobacterium* Mtf1, a proven rhamnosyl 3-O-methyltransferase, as query to search for candidates to be rhamnosyl 3-O-methyltransferase in *Physcomitrella*. Eleven *Physcomitrella* genes were selected as candidates to encode the rhamnosyl 3-*O*-methyltransferase. Some of these candidate genes were selected via BlastP and other searches of the *Physcomitrella* predicted protein database using Mtf1 as query. Other candidates were selected through a conserved domain search focused on TylF, the prototypic member of a superfamily of *O*-methyltransferases including Mtf1. One candidate was selected on the basis of having being identified as the only *Physcomitrella* homolog of *Arabidopsis* AtGXMT1, a recently identified glucuronosyl-4-*O*-methyltransferase gene that is thus far the only gene shown to encode an *O*-methyl ether transferase working on plant cell wall polysaccharides.

Stable knockouts of nine of the 11 candidate genes in *Physcomitrella* were obtained. The 3-*O*-Me-Rha /Rha ratio in AGPs was adopted as an indicator of whether the knockout mutants showed disrupted function of rhamnosyl 3-O-methyltransferase.

Analysis of glycosyl composition of AGPs purified from knockout mutants showed that some of the knockout lines had reduced 3-*O*-Me-Rha/Rha ratio, a result consistent with the predicted effect of disabling the rhamnosyl 3-*O*-methyltransferase. The statistically very significant effects of the *ko1* and *ko9* knockouts on the 3-*O*-Me-Rha/Rha ratio, together with the recent annotations of these genes, led to selection of the *KO1* and *KO9* genes for further study in this dissertation. To these two candidates was also added *KO11* for further study. Although the *ko11* knockouts did not show a statistically significant reduction in the 3-*O*-Me-Rha/Rha ratio in *Physcomitrella* AGPs, the homology of the KO11 protein to AtGXMT1, the only protein thus far proven to be an *O*-methyltransferase working on a sugar residue in a plant cell wall polysaccharide, encouraged its further study.

Further study of the gene *KO1* gene and its transcripts led to a revision of the gene model. A phylogenic tree was constructed and indicated that KO1 was evolved from cyanobacteria and contained a LpxB domain, which is characteristic of lipid A synthesis in gram-negative bacteria. Phenotypic analysis of the *Physcomitrella ko1* knockout mutant showed that polarized tip growth was disrupted. Protonemal filaments grew in a curlier path in the *ko1* knockout than in the wild type. In the leafy gametophyte stage, the *ko1* knockout had fewer and shorter lateral rhizoids than the wild type. Although the primary screening showed that the 3-*O*-Me-Rha/Rha content ratio was statistically very significantly reduced in AGPs from the *Physcomitrella ko1* knockout mutant compared to AGPs from the *Physcomitrella* wild-type, transgenic expression of the *KO1* gene in

Nicotiana tabacum cv Xanthi did not result in detectable 3-*O*-Me-Rha content in the AGPs of the transgenic tobacco plants.

The initial identification of *KO9* as a candidate gene was based on finding it through a conserved domain search focused on TylF, the prototypic member of a superfamily of *O*-methyltransferases including Mtf1. Further study revealed that the predicted KO9 protein has strong sequence similarity to higher plant caffeoyl-CoA O-methyltransferases (CCoAOMTs) of lignin biosynthesis. Because of this similarity, the lignin-like contents of *Phycomitrella ko9* and *ko1* knockout mutants and wild type were measured. The lignin-like content of *ko9* was not significantly different from the wild type, but surprisingly the lignin-like content of *ko1* was statistically very significantly less than that of the wild-type. Although the primary screening showed that the 3-*O*-Me-Rha/Rha content ratio was statistically very significantly reduced in AGPs from the *Physcomitrella ko9* knockout mutant compared to AGPs from the *Physcomitrella* wild-type, transgenic expression of the *KO9* gene in *Nicotiana tabacum* cv Xanthi did not result in detectable 3-*O*-Me-Rha content in the AGPs of the transgenic tobacco plants.

Arabidopsis AtGXMT1 is an *O*-methyltransferase that transfers methyl groups to GlcU residues on glucuronoxylan hemicellulose to form 4-*O*-Me-GlcU. AtGXMT1 contains a conserved DUF579 domain, and KO11 has been identified in the literature as being the only predicted *Physcomitrella* protein to also contain a DUF579 domain. No 4-*O*-Me-GlcU has been found in *Physcomitrella*, however, and in fact 3-*O*-Me-Rha is thus far the only methylated sugar detected in the moss cell wall. The *ko11 Physcomitrella*

knockout exhibited no effect on 3-O-Me-Rha/Rha content of AGPs. To further test whether KO11 could be the rhamnosyl 3-*O*-methyltransferase, the *KO11* gene was transgenically expressed in tobacco. Glycosyl composition analysis detected no 3-*O*-Me-Rha in AGPs purified from the transgenic *KO11* tobacco.

Thus, it remains uncertain whether any of the studied candidate genes encodes rhamnosyl 3-*O*-methyltransferase in *Physcomitrella*. As knockouts in *Physcomitrella*, ko1 and ko9 exhibited statistically very significant reductions in the 3-*O*-Me-Rha/Rha content ratio of AGPs, yet neither these two knockouts nor any of the other knockouts showed complete loss of 3-*O*-Me-Rha. These observations leave open the possibilities that rhamnosyl 3-*O*-methyltransferase might be encoded by a multi-gene family or that the candidate proteins might influence 3-*O*-Me-Rha content only through indirect mechanisms. Likewise, the observed failure of transgenic *KO1*, *KO9,* and *KO11* tobacco plants to produce any detectable 3-*O*-Me-Rha in the tobacco AGPs leaves open at least two alternative hypothesis. One hypothesis would be that one of these three genes does encode a rhamnosyl 3-*O*-methyltransferase, but the expressed mRNA does not lead to accumulation of the protein, or that the protein is incorrectly targeted, i.e., the protein accumulates in a subcellular location where it does have access to the tobacco AGPs. An alternative hypothesis would be that none of these three genes encodes a rhamnosyl 3-*O*-methyltransferase, and that the effects of the *ko1* and *ko9* knockouts on the 3-*O*-Me-Rha/Rha content ratio in *Physcomitrella* AGPs occur indirectly.

Recommended future research directions in this field. For continuing research in this field, further bioinformatics searches should receive high priority. New information in the form of the preliminary release of the Phytozome (www.phytozome.net) v3.0 *Physcomitrella* genome has become available since the start of this project. Improved bioinformatics tools, such as those of the Conserved Domain Database at NCBI (http://www.ncbi.nlm.nih.gov/cdd), have also become available. After more certain candidate genes are selected through this new information and tools, *Physcomitrella* knockout mutants of the candidate genes could be generated by the same methods used in this dissertation. If analysis of the new knockouts suggests that a multiple-gene family is involved, then RNAi would be a better choice to knock-down multiple genes at once. Expression the new candidate genes in tobacco would still be an important tool and might be extended to other species of plants that produce AGPs containing terminal rhamnosyl residues.

KO1 is an interesting gene with the *Physcomitrella ko1* knockout showing several phenotypic effects including abnormal growth in protonema and rhizoids, reduced 3-*O*-Me-Rha/Rha content ratio in AGPs, and reduced lignin-like content in cell walls. A pressing issue here is to examine the *ko1* knockouts by Southern blotting to check whether there is a single insertion in the moss genome, or whether multiple insertions of the knockout cassette occurred and thereby caused the several, seemingly unrelated, phenotypic effects. Complementation of the *ko1* knockout would likewise be valuable in this regard. If polarized tip growth is directly related to *KO1* gene function, then subcellular localization studies of the KO1 protein should be performed to check the

mitochondrial localization predicted by TargetP. Exocytosis might be related to *KO1* gene function in polarized tip growth, so studies of exocytosis in the *ko1* knockout and wild type would be valuable. With regards to the finding that the predicted KO1 protein contains a LpxB domain, biochemical analyses should be performed to determine whether lipid-A precursors accumulate in the *Physcomitrella ko1* knockout mutant.