The Pennsylvania State University

The Graduate School

Intercollege Graduate Degree Program in Plant Biology

EVOLUTIONARY AND FUNCTIONAL ANALYSIS

OF PECTINS AND CELLULOSE IN PLANT CELL WALLS

A Dissertation in

Plant Biology

by

Thomas Wayne McCarthy

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

December 2016
The dissertation of Thomas Wayne McCarthy was reviewed and approved* by the following:

Charles Anderson
Professor of Biology
Dissertation Advisor
Chair of Committee

Mark Guiltinan
Professor of Plant Molecular Biology

Michael J. Axtell
Professor of Biology

James Marden
Professor of Biology

Teh-hui Kao
Distinguished Professor of Biochemistry and Molecular Biology
Head of the Plant Biology Graduate Program

*Signatures are on file in the Graduate School
ABSTRACT

The plant cell wall is a complex and dynamic system of great importance both to plants and to humans. For plants, cell walls are necessary, among other things, for structure and support, for morphogenesis, for water and mineral management, and for protection. For humans they are sources of building materials, textiles, paper, fuel, and food. Improving our understanding of the plant cell wall is a precursor to being able to intelligently manipulate its properties for our own purposes.

The evolutionary history of the plant cell wall is of interest both academically and because of its relevance to the makeup of cell walls of modern economically important plants. Although the predecessor to land plants is widely agreed to have been a Charophycean green alga, the phylogeny of extant Charophycean green algae is not resolved, so they are of only limited use in researching plant terrestrialization. Bryophytes appear to be the earliest divergent group from the rest of the land plants, so comparisons between mosses and tracheophytic plants are currently our best option for learning about the embryophyte common ancestor.

In this work, phylogenetic analysis of 16 pectin-related gene families led to new insights about the history of plant cell walls. In contrast to the CESA family, which radiated in land plants after the divergence of bryophytes and tracheophytes, several of the pectin-related families showed clear evidence of having had multiple members in the land plant common ancestor. Additionally, two very large families,
the polygalacturonases and the pectin methylesterases, contained small monophyletic clades that suggest deeply conserved functions.

Six Arabidopsis genes from three deeply conserved clades in the polygalacturonase tree were chosen as targets for characterization. Attempts to isolate transcriptional knockouts using T-DNA insertional mutants were unsuccessful. One putative mutation in AT1G10640 was investigated heavily before being revealed to be an unusual T-DNA mutant but not a true knockout.

After exhausting the option of T-DNA mutants to produce knockouts of the genes of interest, the more recent molecular tools of artificial microRNAs and the CRISPR-Cas9 system were used to target the genes. The CRISPR-Cas9 system produced multiple mutant lines predicted to have early stop codons for five out of the six genes. These are expected to be a valuable resource for learning about the functions of polygalacturonases in eudicot cell walls, and gaining insight into their potential role in the embryophyte common ancestor.

Finally, I report on several experiments relating to the cell walls of Physcomitrella patens (Physcomitrella), a model moss with cell walls similar in composition to eudicot primary walls. A preliminary time course experiment suggests that protoplasts of Physcomitrella that have been released from their walls quickly generate a callose coating that is gradually strengthened by cellulose within 48 hours. Fluorescent Protein-labelled Cellulose Synthase (CESA) proteins in Physcomitrella move at similar speeds as those in Arabidopsis thaliana (Arabidopsis), but are much less densely spaced. Chemicals known to be cellulose
synthesis inhibitors in eudicots also appeared to slow *Physcomitrella* CESAs, but it is not clear that their mode of action is the same. A comparison of CESA protein sequences involved in secondary cell formation in Arabidopsis to chimeric proteins and heterologous CESAs from *Physcomitrella* suggested several amino acids as being important for proper assembly of the Cellulose Synthase Complex.
# TABLE OF CONTENTS

LIST OF FIGURES ................................................................................................. ix

LIST OF TABLES ................................................................................................. xi

ABBREVIATIONS ................................................................................................. xii

ACKNOWLEDGEMENTS ....................................................................................... xiii

Chapter 1 Introduction......................................................................................... 1

  The plant cell wall ......................................................................................... 2
    Makeup of the primary plant cell wall ....................................................... 3
    De novo production of cell walls during mitosis ...................................... 6
  Cellulose ........................................................................................................ 7
    Natural plant cellulose takes the form of cellulose microfibrils ............. 7
    Plant cellulose is produced at the plasma membrane by Cellulose Synthase Complexes ................................................................. 8
  Pectins .......................................................................................................... 9
    Pectin domains ....................................................................................... 9
    Pectin behavior in the wall .................................................................. 10
    Key pectin-related gene families ......................................................... 11
    Polygalacturonases ............................................................................. 12
  Evolutionary history of the plant cell wall ............................................ 13
    Cellulose Synthase evolution ............................................................... 17
    Evolutionary history of pectins .......................................................... 18
  References ................................................................................................ 19

Chapter 2 Phylogenetic analysis of pectin-related gene families in
Physcomitrella patens and nine other plant species yields evolutionary
insights into cell walls .................................................................................... 37

  Abstract ..................................................................................................... 37
  Background ................................................................................................. 38
  Results ....................................................................................................... 42
    Identification of pectin-related genes using PlantTribes 2.0 ............... 42
    Phylogenetic analysis of pectin-related gene families ....................... 44
    The GAUT superfamily contains at least five ancestral land plant
genomes ................................................................................................. 49
    Polygalacturonase and pectin methylesterase families are large and
deeply conserved ............................................................................. 50
Many pectin-related gene families appear to have had only one or two members in the common ancestor of land plants..............52
Discussion ..................................................................................................................................................52
Search and tree-building criteria for pectin-related genes.................................................................52
Several pectin-related gene families likely had multiple members in the common ancestor of mosses and tracheophytes..................54
Some pectin-related gene families were not detected in *Arabidopsis*.............................................55

*Arabidopsis* has an abundance of pectin-related genes, whereas grasses appear to have fewer pectin-related genes in some families........................................................................................................57
Conclusions .............................................................................................................................................57
Methods ....................................................................................................................................................58
Identification of pectin-related gene families.......................................................................................58
Phylogenetic analysis ..............................................................................................................................59
Author contributions .............................................................................................................................60
Acknowledgements ...............................................................................................................................61
References ................................................................................................................................................61

Chapter 3  Studying deeply conserved putative polygalacturonases ...............70

T-DNA insertion mutants of putative PGs did not disrupt expression.................71
Generating targeted transcriptional mutants of conserved putative PGs........73
CRISPR-Cas9 strategy for knocking out genes.................................................................................73
Choosing CRISPR-Cas9 protospacer sequences..............................................................................75
CRISPR-Cas9 induced mutation lines ...............................................................................................78
Targeting genes with artificial microRNAs......................................................................................82
Generation of transgenic plants and microorganisms for the characterization of the putative polygalacturonase AT1G10640........85
Discussion ..............................................................................................................................................88
Methods ................................................................................................................................................89
Plant material and growth conditions ..............................................................................................89
Assembling AT1G10640 complementation, promoter-GUS fusion reporter, and constitutive overexpression constructs..............................90
Assembling amiRNA and CRISPR-Cas9 constructs.........................................................................91
Plant transformation .............................................................................................................................93
Heterologous AT1G10640 expression ..............................................................................................93
Polygalacturonase activity assay ........................................................................................................95
Acknowledgements .............................................................................................................................96
References ..............................................................................................................................................96

Chapter 4  Insights into the cell wall of the moss *Physcomitrella patens* ..........99

Experimental background and motivation.........................................................................................100
LIST OF FIGURES

Figure 1-1: Tethered-network model of the primary cell wall ............................................. 5
Figure 1-2: Schematic structure of pectin domains ................................................................. 10
Figure 1-3: Simplified Archaeplastida phylogeny highlighting the occurrence of major wall components .................................................................................................................. 15
Figure 2-1: Summary of land plant phylogeny ................................................................. 42
Figure 2-2: GAUT family tree ............................................................................................... 46
Figure 2-3: Polygalacturonase family tree ............................................................................. 47
Figure 2-4: Pectin-methylesterase family tree ....................................................................... 48
Figure 2-5: GAUT superfamily tree ....................................................................................... 50
Figure 3-1: Polygalacturonase gene targets ............................................................................ 72
Figure 3-2: Schematic of Cas9/gRNA genome editing ............................................................ 74
Figure 3-3: CRISPR-Cas9 target sites in polygalacturonases .................................................. 77
Figure 3-4: Example CRISPR-Cas9 sequencing chromatogram ........................................... 78
Figure 3-5: Polygalacturonase amiRNA binding sites .......................................................... 84
Figure 4-1: CESA regions for chimeric protein design .......................................................... 103
Figure 4-2: Generation of de novo cell wall by Physcomitrella protoplasts ....................... 106
Figure 4-3: Physcomitrella protoplast regeneration time course ......................................... 107
Figure 4-4: GFP-PpCESA5 velocities ..................................................................................... 109
Figure 4-5: GFP-PpCESA5 at the cell surface ....................................................................... 109
Figure 4-6: Pooled GFP-PpCESA8 velocity experiments ....................................................... 110
Figure 4-7: GFP-PpCESA8 velocities after treatments .......................................................... 111
Figure 4-8: GFP-PpCESA8 at the cell surface ................................................................. 112
Figure 4-9: Part of the AtCESA7 alignment ................................................................. 114
Figure 2-S1: Xylogalacturonan xylosyltransferase family tree ..................................... 139
Figure 2-S2: Rhamnogalacturonan I arabinosyltransferase family tree ......................... 140
Figure 2-S3: Pectinmethylesterase inhibitor (PMEI) family tree .................................... 141
Figure 2-S4: Polygalacturonase inhibitor protein family tree ....................................... 142
Figure 2-S5: Pectate lyase-like (PLL) family tree ....................................................... 143
Figure 2-S6: Homogalacturonan methyltransferase family tree .................................... 144
Figure 2-S7: UDP-Glucuronic acid epimerase family tree ......................................... 145
Figure 2-S8: UDP-Rhamnose synthase family tree .................................................... 146
Figure 2-S9: Pectin acetyltransferase family tree ....................................................... 147
Figure 2-S10: Pectin acetyesterase family tree ........................................................... 148
Figure 2-S11: Rhamnogalacturonan II xylosyltransferase family tree ......................... 149
Figure 2-S12: β-1,4-Galactan β-1,4-Galactosyltransferase family tree ....................... 150
Figure 2-S13: GATL family tree ................................................................................. 151
LIST OF TABLES

Table 2-1: Representatives of pectin-related gene families in *Arabidopsis* and *Physcomitrella* ........................................................................................................................................ 44

Table 3-1: Non-null T-DNA mutant lines ................................................................................................................................................................................................. 71

Table 3-2: CRISPR-Cas9 mutant seed stocks .................................................................................................................................................................................. 79

Table 3-3: amiRNA mutant seed stocks ..................................................................................................................................................................................... 84

Table 3-4: AT1G10640 characterization mutant lines ....................................................................................................................................................... 85

Table 3-5: *Pichia* cultures expressing AT1G10640 show polygalacturonase activity ................................................................................................................................. 88

Table 4-1: Complementation of Arabidopsis secondary CESA knockouts .............. 104

Table 4-2: Arabidopsis secondary cell wall CESA residues of interest ................. 115
ABBREVIATIONS

PCW: Primary Cell Wall
NMR: Nuclear Magnetic Resonance
AFM: Atomic Force Microscopy
CESA: Cellulose Synthase
CSC: Cellulose Synthase Complex
GalA: Galacturonic Acid
HG: Homogalacturonan
RG-I: Rhamnogalacturonan I
RG-II Rhamnogalacturonan II
GAUT: Homogalacturonan Galacturonosyltransferase
PG: Polygalacturonase
PME: Pectin Methylesterase
CGA: Charophycean Green Alga
TC: Terminal Complex
CSR: Class-Specific Region
PCR: Plant Conserved Region, also Polymerase Chain Reaction
DCB: 2,6-dichlorobenzonitrile
GFP: Green Fluorescent Protein
VAEM: Variable Angle Epifluorescence Microscopy
ROI: Region Of Interest
XGA: Xylogalacturonan
GATL: GAUT-like
COMPP: Comprehensive Microarray Polymer Profiling
CDS: Coding Sequence
PAM: Protospacer Adjacent Motif
DSB: Double Stranded Break
InDel: Insertion or Deletion mutation
ACKNOWLEDGEMENTS

Charlie Anderson
  Mike Axtell
  Will Barnes
Magdalena Bezanilla
  Joseph Cho
  Mary Coleman
Claude dePamphilis
  Josh Der
  Jackie Doyle
William Ehlhardt
  Drew Fister
Mark Guiltinan
  Joseph Hill
Loren Honaas
Melissa Ishler
  Nikki Kapp
  Jim Marden
Siela Maximova
Christina McCarthy
  Michael McCarthy
  Daniel McClosky
  John McManus
Bastian Minkenberg
  Bill Murphy
  Seth Polydore
Alison Roberts
  Yue Rui
Lena Sheaffer
  Mai Tran
Feng Wang
  Shu-Zon Wu
Chaowen Xiao
Special thanks to Igor Andrzhievsky.
Thanks to Liz McCarthy and Laura Copley, without whom I couldn't have done this.

This work was supported as part of The Center for LignoCellulose Structure and Formation, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award # DE-SC0001090.
Universities are very familiar with bright, qualified school-leavers who arrive and then go into shock on finding that biology or physics isn't quite what they've been taught so far. "Yes, but you needed to understand that," they are told, "so that now we can tell you why it isn't exactly true." Discworld teachers know this, and use it to demonstrate why universities are truly storehouses of knowledge: students arrive from school confident that they know very nearly everything, and they leave years later certain that they know practically nothing. Where did the knowledge go in the meantime? Into the university, of course, where it is carefully dried and stored.

— Terry Pratchett, The Science of Discworld
Chapter 1

Introduction

The cell wall is key to the existence of plants as we know them. Plant cell walls provide structure and support both at the individual cell level and at the macroscopic level for the whole organism. They help regulate growth and water status within the plant, and they serve as a physical barrier against biotic and abiotic stressors. Humans use plant cell wall material for building supplies, paper, textiles, food, and fuel. More recently, plant cell walls have become significant to humans as both one of the biggest carbon sinks in the earth's ecosystem and as the most abundant potential source of renewable biofuels [1, 2].

Understanding the makeup of cell walls, their structure, and how they work is necessary if we want to improve our use of them. This is made challenging by the fact that the polysaccharides that make up the bulk of cell walls are complex and diverse, and unlike proteins, cannot be neatly linked to single encoding genes. Instead, the dynamic system that is the plant cell wall is actively built, modified, degraded, stretched, and rebuilt throughout much of the cell's life by a diverse group of enzymes present in vesicles, at the plasma membrane, and within the wall itself. Teasing out the functions of individual proteins and polysaccharides in the midst of a system that is both structurally (in that it has many moving parts) and functionally (in that it does many things simultaneously) complex is a daunting
endeavor, but is an integral part of understanding plant biology. This chapter provides a framework of background knowledge for the following research.

**The plant cell wall**

The plant cell wall is a complex fibrous structure surrounding the plasma membrane-bound plant protoplast [3–6]. Despite biologists’ love of classifying things, the plant cell wall evades easy classification: in some sense it is extracellular matrix that extends throughout the apoplastic space, and yet each plant cell is considered to include its own individual wall [3, 7, 8]. A layered structure, the plant cell wall is built from the outside in. The middle lamella is a pectin-rich region that provides adhesion between plant cells [3, 9]. The primary wall layer is internal to the middle lamella and composed mainly of three classes of polysaccharide, cellulose, hemicelluloses, and pectins, as well as glycosylated proteins [3, 5]. Some plant cells that have stopped expanding also produce a secondary wall layer that contains a fifth class of polymer, lignin [3, 10].

The plant cell wall serves as a defense against biotic (e.g. herbivory, infection) and abiotic (e.g. physical trauma, water loss, radiation) stresses, and at the macroscopic scale it provides support and strength for the growing plant [3]. On the microscopic scale, the wall also provides structure and support to the cell, serving as a stiff boundary against which the turgor of the protoplasm pushes [3]. However, the polysaccharide matrix that forms the bulk of the cell wall is not completely unyielding: cell growth is achieved by the controlled extension of the
wall [3, 11]. Extensibility of the wall is tightly regulated by the addition and modification of its polysaccharide components, and plant cell size and shape are controlled by anisotropic yielding of the cell wall [3, 12]. As the regulator of shape and size at the cellular level, the plant cell wall is consequently also responsible for organism-scale growth and development [3, 13].

**Makeup of the primary plant cell wall**

Because regulation of wall properties is important to plant cell function, the makeup of the wall can vary widely between cell type, tissue type, and species. In dicotyledonous plants, the dry mass of the "typical" primary cell wall (PCW) consists of about 25% cellulose, 30% hemicelluloses, and 40% pectins, with the remainder being an assortment of proteins [3]. Cellulose is made up of chains of β-1,4-linked glucans that are assembled into tightly packed bundles called microfibrils. Cellulose microfibrils are stiff and tough relative to other PCW components and provide much of the wall’s tensile strength [3, 14]. Hemicelluloses are a diverse collection of long polysaccharides that often carry short side chains [3, 15, 16]. Uncoiled hemicelluloses are narrower and more flexible than cellulose microfibrils, and can bind to both cellulose and pectins [16, 17]. Pectins are acidic polysaccharides, most of which have a poly-galacturonic acid backbone that is capable of forming calcium ion-mediated ionic bonds between pectin chains [18, 19]. These calcium bridges are responsible for the gelling effect of pectins used in cooking, and also affect the yielding properties of the plant cell wall.
Our conception of the PCW is currently in flux as research continues to refine our understanding of the properties and behavior of cell wall components. In the past the PCW has been described as being similar to rebar-reinforced concrete or fiberglass, with a stiff cellulose microfibril network bound together by a coating of hemicelluloses and embedded in, but not substantially interacting with, a pectic gel matrix (Figure 1-1) [4, 5, 16, 20]. However, recent experiments have accrued contradictions to this "tethered-network" model. The long-held assumption that xyloglucan was vital for PCW integrity and growth in eudicots has been shaken by the isolation of viable, apparently xyloglucan-free Arabidopsis mutants [21, 22]. Rather than being heavily coated with xyloglucan (the most common hemicellulose in eudicot PCWs), Nuclear Magnetic Resonance (NMR) data suggest that cellulose microfibrils have relatively few close interactions with xyloglucans, despite their high affinity in vitro [23–25]. On the other hand, pectin interactions with cellulose appear to be much more extensive than in vitro experiments would predict, with perhaps as much as 50% of cellulose microfibril surface in close proximity to pectin domains [17, 25–27].
The amount of the various polymers is shown based approximately on their ratio to the amount of cellulose. The amount of cellulose shown was reduced relative to a living cell for clarity. Because of the exaggerated distance between microfibrils, the hemicellulose cross-links [shown in dark orange (xyloglucan, XG) or light orange (glucoronoarabinoxylan, GAX)] are abnormally extended. Figure reused with permission from [4].

Current evidence suggests a model in which cellulose, hemicelluloses, and pectins are all part of the same network [17]. Based on Atomic Force Microscopy (AFM), NMR, and extractability data, cellulose microfibrils seem to interact physically with hemicelluloses, pectins, and other cellulose microfibrils [17, 25, 27–
Hemicelluloses and pectins also appear to be covalently bound at least some of the time, either during or soon after hemicellulose synthesis [32–35].

**De novo production of cell walls during mitosis**

Plant cell division is predicated on the production of a new cell wall between the daughter cells [36]. This developing cell wall, called the cell plate, is built in stages within a complex of cytoskeleton and vesicles called the phragmoplast [37, 38]. One of the polysaccharides to be incorporated into the cell plate is callose, a (1,3)-β-glucan [39, 40]. It appears early in the development of the phragmoplast and is present throughout much of cell plate development, but is removed from the completed new wall [41].

Pectins and hemicelluloses are delivered to the developing cell plate in vesicles, much as they are delivered to the plasma membrane during the growth of "normal" cell wall [42, 43]. Despite its importance as a load-bearing polysaccharide in external cell walls, cellulose is a relatively minor component of the developing cell plate [40, 44]. During cell plate development, fluorescent protein-labeled Cellulose Synthase (CESA) proteins are trafficked away from the plasma membrane, and fluorescent CESA delivery events almost exclusively occur at the cell plate until mitosis is complete [44].
Cellulose

Cellulose is of interest to humans both because of its integral role in the cell walls of plants, our ultimate source of food and oxygen, and because of its uses separate from the rest of the cell wall. Cellulose is the primary component of paper and cotton products, and is of large and growing interest for the production of biofuels. Despite our interest and a long history of research, our understanding of cellulose is still imperfect.

Natural plant cellulose takes the form of cellulose microfibrils

In the plant cell wall, cellulose takes the form of semi-crystalline microfibrils composed of β-1, 4-linked glucan chains held together by hydrogen bonds and van der Waals forces [3, 17]. Despite its biological and economic importance, much about cellulose and its biosynthesis remains uncertain. Dr. Daniel Cosgrove expresses this especially well in his 2014 review article "Re-constructing our models of cellulose and primary cell wall assembly" when he says "The [glucan] chains are packed into an ordered microfibril of indefinite length and uncertain cross sectional area and shape [17]." Recent papers suggest that plant PCW cellulose microfibrils contain 18 to 24 glucan chains and range from 2.2 to 5.3 nm in diameter [17, 30, 45–48]. Cellulose microfibrils can bundle into thicker ribbons sometimes called "macrofibrils", reaching widths of up to 56 nm [28, 45]. Macrofibrils often branch and condense to form reticulating networks of cellulose visible through AFM.
as distinct layers in the wall [28]. Because of its length and high tensile strength, cellulose organization anisotropically constrains wall yielding and influences the direction of cell expansion [3, 6, 11, 12].

**Plant cellulose is produced at the plasma membrane by Cellulose Synthase Complexes**

Unlike other PCW polysaccharides, which are assembled within the Golgi network and delivered to the cell wall via vesicles, cellulose microfibrils are produced at the plasma membrane and extruded into the apoplast by an enzymatic assembly called the Cellulose Synthase Complex (CSC). In land plants the CSC is visible by freeze-fracture electron microscopy as a 6-membered ring of particles (a "rosette") at the plasma membrane [49, 50]. The catalytic proteins responsible for the production of individual glucan chains in plants are a family of cellulose synthases (CESAs), of which there are ten in Arabidopsis [50–55].

In Arabidopsis PCWs cellulose is produced by CSCs containing the Cellulose Synthase subunits AtCESA1, AtCESA3, and AtCESA6, while secondary cell wall CSCs comprise AtCESA4, AtCESA7, and AtCESA8 [56, 57]. *AtCESA2, AtCESA5*, and *AtCESA9* are members of the same monophyletic clade as, and are partially redundant with, *AtCESA6* [57–59]. Experiments suggest that both primary and secondary CESAs exist in 1:1:1 stoichiometries within their respective CSCs, and the most recent proposed models of the rosette generally suggest a six-membered ring of heterotrimerers (although Vandavasi *et al.* induced the formation of AtCESA1 catalytic
domain homotrimers in vitro, Timmers et al. demonstrate that AtCESA7 and AtCESA8 do not homodimerize in vivo or in vitro) [60–65].

**Pectins**

Pectins make up approximately one third of the primary plant cell walls in eudicots [66–69]. Originally classified by extractability from the cell wall by hot water and calcium chelators, "pectin" is now a collective term for several complex polysaccharides characterized by the presence in high amounts of the negatively-charged sugar galacturonic acid (GalA) [16, 70]. Despite its importance, the complexity of pectin synthesis and modification means that unanswered questions about its many functions remain.

**Pectin domains**

Pectin consists primarily of homogalacturonan (HG), a linear polysaccharide of α-1,4-linked GalA (Figure 1-2) [68]. The GalA residues of HG can be methylesterified at the O-6 position and acetylated at O-2 and O-3 [71, 72]. The O-3 position can also be occupied by xylose, giving rise to the xylogalacturonan domain, or by apirose, forming the apiogalacturonan domain [73–75]. Decoration of HG with four complex and well-conserved side chains results in the pectic domain rhamnogalacturonan-II (RG-II) [76–78]. The final pectin domain is called rhamnogalacturonan-I (RG-I) and consists of an alternating rhamnose and GalA
backbone with branching side chains containing arabinose and/or galactose residues attached to the rhamnose residues [68]. Despite, or perhaps because of, their complexity, pectins are almost universally conserved in land plants and are also present in some algae [79].

Figure 1-2: Schematic structure of pectin domains

Four pectin domains are shown; apiogalacturonan is not portrayed. Domain relative amounts are not to scale, and the presence of all four domains within one macromolecule has not been experimentally confirmed. Kdo, 3-Deoxy- D-manno-2-octulosonic acid; DHA, 3-deoxy-D-lyxo-2-heptulosaric acid. Figure reused with permission from [67].

Pectin behavior in the wall

The gelling properties of pectin come from the ability of pairs of HG molecules to be bound together by ionic bonds between the carboxylate groups of GalA and ionic calcium present in the wall [18, 19]. The ability of HG to form Ca²⁺ bridges is influenced by the methylesterification status of its carboxylate groups,
and regulation of methylesterification is an important factor in controlling the properties of the cell wall [72]. Pectins can also stiffen the wall through borate diol ester bonds between the apiose residues of RG-II side chains [80–82], and prevention of borate crosslinking of RG-II causes dwarfism [81].

Pectin modification can stiffen, loosen, or degrade the cell wall, which can affect such diverse aspects of plant growth as root elongation, fruit ripening, pollen abscission, organ initiation, pollen tube extension, seed germination, and cell-cell adhesion [9, 83–92]. Pectins are also involved in regulating hydration status and porosity of the wall, and have been implicated in stress signaling [66, 68, 93–95]. Methylesterification status of pectin can directly influence wall stiffness and hygroscopic properties, but it may also affect the deposition of cellulose, and pectins are implicated in wall remodeling during cellular expansion [96–98].

**Key pectin-related gene families**

Pectin is the most structurally complex group of polysaccharides in the plant cell wall and is estimated to require at least 67 different enzyme activities for synthesis [66, 67, 70]. Not all of these enzymes have been identified at this point, but three particular gene families especially important for the production and modification of pectin are the homogalacturonan galacturonosyltransferases (GAUTs), the pectin methylesterases (PMEs), and the polygalacturonases (PGs) [99–101]. The GAUT family is responsible for the production of the HG backbone, and
disruption of specific GAUT genes can cause changes in wall composition, dwarfism and sterility, or presumed embryo lethality [70, 102, 103].

The methylesterification status of the HG backbone influences its ability to form calcium cross-links, but it also affects the susceptibility of pectin to degradation by both native and foreign enzymes [72, 104]. PMEs can remove the methyl groups from the HG backbone in either a progressive, block-wise fashion, leaving extended chains of carboxylate groups, or in a random fashion that makes the polymer prone to degradation but not to forming calcium bridges [105, 106]. Since the activity of different PMEs can either stiffen or loosen the cell wall, it is not surprising that interference in their functions cause a wide range of phenotypes, including weakening of the pollen tube, problems with pollen dehiscence, stunted root growth, dwarfism, and reduced stem strength [107–111].

**Polygalacturonases**

PGs are a class of enzymes belonging to the glycoside hydrolase family 28 [112]. Polygalacturonases hydrolyze the HG backbone of pectins, creating shorter pectin chains [100]. This decreases the stiffness of the wall and can affect its yielding properties. PGs are classified as endo- or exo-polygalacturonases based on whether they cleave within an HG chain or remove a single sugar from the end [100, 106, 113–115]. The PG gene family is huge and widespread, with members among the animal, fungal, and bacterial kingdoms as well as in Plantae [112, 116, 117].
Several functions of PGs have been noted in plants, but given the large size of the gene family, the full extent of their role is still only poorly understood [100, 106].

PG activity has long been known to be involved in fruit ripening [84, 118–123]. One of the first transgenic fruits to reach the consumer market was a tomato with reduced polygalacturonase activity [124]. Unsurprisingly, given their weakening effect on the cell wall, PGs have been shown to be important to cell separation, active in the abscission zones of fruit, floral organs, and seedpods, and necessary for pollen separation [85, 125–128]. They are also involved in anther and pollen development and pollen tube elongation [129, 130].

PGs have also recently been shown to be involved in cell enlargement and elongation, and to function in organ patterning [131, 132]. Given that PG activity weakens the pectin matrix as it cleaves homogalacturonan, it is not surprising that experimental treatments with exogenous PG induce long-term irreversible wall extension (creep) [22]. The large size of the PG family suggests that additional unknown functions in plant growth and development may remain.

**Evolutionary history of the plant cell wall**

It is generally agreed that modern land plants evolved from a Charophycean Green Alga (CGA) ancestor, and many scientists have suggested that the move to land is intimately tied up with the development of the plant cell wall [133–135]. Studying the evolutionary history of cell walls is especially challenging for several reasons. Unlike genes or proteins, the polysaccharides of the wall do not change
slowly over time or carry a record of their history in their sequence. A lot of what we know about the history of the wall has come from looking for the presence of specific polysaccharides in various extant plants and algae (Figure 1-3) [136–138]. This is complicated by the fact that we do not yet have precise methods for detecting every polysaccharide, and also because there is evidence of convergent evolution producing similar wall polysaccharides using unrelated genes [138, 139]. On top of this, there is evidence of gene loss in some families complicating our interpretation of phylogenetic data, and some works suggest that the prevalence of certain wall polysaccharides across kingdoms (e.g. cellulose, which is present in proteobacteria, cyanobacteria, red algae, green algae, brown algae, embryophytes, oomycetes, and tunicates) is the result of ancient horizontal gene transfer [140].
Figure 1-3: Simplified Archaeplastida phylogeny highlighting the occurrence of major wall components

The identification of specific wall components within lineages is symbolized as shown in the key. Genes responsible for cellulose and hemicellulose biosynthesis are indicated in boxes: CesA is the ancestral form of cellulose synthases, CslA/C is a single gene that is most similar to the land plant CslA and CslC gene families, and CesA* represents members of the cellulose synthase family whose proteins assemble into rosette terminal complexes. Adapted from [137].

As polysaccharide detection techniques have advanced and more plants and algae have been subjects of study, our models of wall evolution have had to adapt to surprises such as the discovery of wall components that had previously been thought absent. In some cases this has been the result of convergent evolution: 1-3, 1-4 mixed linkage glucans (MLGs), formerly thought to only be present in the walls of Poales, have now been found in Equisetum, the lycophyte Selaginella moellendorfii, and charophycean algae [136, 139, 141, 142]. However, Selaginella does not have orthologs of the Cellulose Synthase-like (Csl) CslF and CslH genes
implicated in the production of MLGs in grasses, suggesting convergent evolution [136, 143, 144]. Even more surprising, lignin-like material, long believed to solely be present in tracheophytic plants, was found in a red alga that last shared a common ancestor with land plants over 400 million years ago [138, 145]. Combine these convergent evolution red herrings with the fact that polysaccharides previously thought to be absent are now being detected with more sensitive methods, and it is clear that we need to be careful when interpreting polysaccharide presence data while reconstructing the history of plant cell walls.

Given that genes are the units of heritability, it is unsurprising that genetic analysis is the default method for investigating evolutionary history. Unfortunately, as previously stated, study of the cell wall is complicated by the fact that there is no simple one gene -> one protein -> one wall polysaccharide correlation. It is estimated that as much as 15% of the Arabidopsis genome (around 4100 genes) consists of cell wall-related genes [146]. With only a fraction of those genes identified in any way beyond simple automated domain annotation, the current scope of phylogenetic analysis in relation to cell walls is limited [147, 148]. Progress in reverse genetics is impeded by the problem of confirming a given gene's function in the face of the adaptability of the wall to perturbation and the high rate of redundancy among many wall-related genes.
**Cellulose Synthase evolution**

The wall-related genes that have received the most study are the CESA superfamily. CESA genes are widespread among bacteria and algae as well as land plants [140, 149]. In bacteria and algae they usually assemble into linear Terminal Complexes (TCs) at the plasma membrane, whereas in land plants they form a six-membered hexagonal TC commonly known as a rosette [140, 150–152]. Bacterial and land plant CESAs share significant homology of their catalytic regions, and a conserved set of amino acid residues referred to as the D-D-D-QXXRW motif [54]. However, land plant CESAs have additional domains not seen in bacterial CESAs. Near the N-terminal region is a putative Zn-binding domain, possibly involved in protein-protein interactions [153, 154]. Near the center of the protein, between the second and third putative trans-membrane helices are two other land plant-specific regions. The first shows high variability between intra-species CESAs but is highly conserved among inter-species orthologs and is called the Class-Specific Region (CSR), while the second is deeply conserved among land plants and is called the Plant Conserved Region (PCR) [59, 155].

The radiation of CESA genes in land plants did not appear to occur until after the divergence of the bryophytes from the embryophytes; phylogenetic trees suggest a single CESA in the common ancestor [156]. Surprisingly, phylogenetic trees also suggest that the divergence of primary and secondary wall CESAs preceded the development of hetero-oligomeric CSCs [59, 157]. While it may seem unparsimonious to suppose that primary and secondary wall CSCs independently
reached a three-member equilibrium, the evolutionary forces that drive a multi-
subunit complex toward a hetero-oligomeric state have been demonstrated in the V-
ATPase complex of yeast [157, 158].

**Evolutionary history of pectins**

The existence of pectin-like materials in cell walls extends back far beyond
the evolution of land plants. HG is present in all land plants as well as both
Charophycean and Chlorophycean green algae [159]. RG-I is also present in all land
plants and is suspected to be present in some Charophycean algae [138, 160].
Xylogalacturonan and RG-II appear to have evolved later: the former is not detected
in bryophytes, and the latter is hinted at in bryophytes by the presence of certain
sugars but, if present, is so only in very small quantities [77, 138, 161]. To date,
apiogalacturonan has only been detected in aquatic monocots [66, 75, 162]. There is
no evidence of RG-II in algae, but the rare KDO sugar that is a component of RG-II
side chains has been found in an alga, suggesting that at least some of biosynthetic
machinery necessary for production of RG-II predates the move to land [138, 163].
The complex structure of RG-II side chains is very highly conserved, but small
variations have been found in seedless plants [68, 77, 140]. It has been suggested
that increasing amounts of RG-II in the wall correlated with the evolution of an
upright growth structure, and that the wall-stiffening borate dimers of RG-II were
an essential part of this development [77].
Although pectins are integral parts of the plant cell wall, their complexity makes the study of all their relevant biosynthetic and modifying genes a daunting task. While we know that the pectin-related gene families of polygalacturonases and pectin methylesterases are large and underwent rapid divergence in tracheophytes, many of the genes required for the synthesis of pectin domains remain undiscovered, and much of the evolutionary history of pectins remains unknown. [116, 164–170].

References

8. Liu Z, Persson S, Sanchez-Rodriguez C: At the border: the plasma


Zabotina OA, Hahn MG, Burgert I, Pauly M, Raikhel N V, Keegstra K: **Disrupting two Arabidopsis thaliana xylosyltransferase genes results in plants deficient in xyloglucan, a major primary cell wall component.** *Plant Cell* 2008, **20**:1519–1537.

22. Park YB, Cosgrove DJ: **Changes in cell wall biomechanical properties in the xyloglucan-deficient xxt1/xxt2 mutant of Arabidopsis.** *Plant Physiol* 2012, **158**:465–75.


29. Park YB, Cosgrove DJ: **Xyloglucan and its interactions with other**


31. Park YB, Cosgrove DJ: **A Revised Architecture of Primary Cell Walls Based on Biomechanical Changes Induced by Substrate-Specific Endoglucanases.** *Plant Physiol* 2012, **158**:1933–1943.

32. Thompson JE, Fry SC: **Evidence for covalent linkage between xyloglucan and acidic pectins in suspension-cultured rose cells.** *Planta* 2000, **211**:275–86.


34. Popper ZA, Fry SC: **Xyloglucan-pectin linkages are formed intraprotoplastically, contribute to wall-assembly, and remain stable in the cell wall.** *Planta* 2008, **227**:781–794.


58. Persson S, Paredez AR, Carroll A, Palsdottir H, Doblin MS, Poindexter PD, Khitrov N, Auer M, Somerville CR: Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in

60. Wang J, Howles PA, Cork AH, Birch RJ, Williamson RE: **Chimeric proteins suggest that the catalytic and/or C-terminal domains give CesA1 and CesA3 access to their specific sites in the cellulose synthase of primary walls.** *Plant Physiol* 2006, 142:685–95.


70. Caffall KH, Mohnen D: **The structure, function, and biosynthesis of plant cell wall pectic polysaccharides.** *Carbohydr Res* 2009, **344**:1879–900.


75. Ovodova RG, Golovchenko V V, Shashkov AS, Popov S V, Ovodov YS: **Structural studies and physiological activity of lemnan, a pectin from Lemna minor L.** *Bioorganicheskaya Khimiya* 2000, **26**:743–751.

76. Pérez S, Rodríguez-Carvajal MA, Doco T: **A complex plant cell wall polysaccharide: rhamnogalacturonan II. A structure in quest of a function.** *Biochimie* 2003, **85**:109–21.


86. Peaucelle A, Braybrook SA, Le Guillou L, Bron E, Kuhlemeier C, Höfte H: Pectin-induced changes in cell wall mechanics underlie organ initiation


89. Parre E, Geitmann A: Pectin and the role of the physical properties of the cell wall in pollen tube growth of *Solanum chacoense*. *Planta* 2005, **220**:582–92.


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


Chapter 2

Phylogenetic analysis of pectin-related gene families in *Physcomitrella patens* and nine other plant species yields evolutionary insights into cell walls

Thomas W McCarthy, Joshua P Der, Loren A Honaas, Claude W dePamphilis and Charles T Anderson

Abstract

Background

Pectins are acidic sugar-containing polysaccharides that are universally conserved components of the primary cell walls of plants and modulate both tip and diffuse cell growth. However, many of their specific functions and the evolution of the genes responsible for producing and modifying them are incompletely understood. The moss *Physcomitrella patens* is emerging as a powerful model system for the study of plant cell walls. To identify deeply conserved pectin-related genes in *Physcomitrella*, we generated phylogenetic trees for 16 pectin-related gene families using sequences from ten plant genomes and analyzed the evolutionary relationships within these families.

Results

Contrary to our initial hypothesis that a single ancestral gene was present for each pectin-related gene family in the common ancestor of land plants, five of the 16 gene families, including homogalacturonan galacturonosyltransferases, polygalacturonases, pectin methylsterases, homogalacturonan methyltransferases, and pectate lyase-like proteins, show evidence of multiple members in the early land plant that gave rise to the mosses and vascular plants. Seven of the gene...
families, the UDP-rhamnose synthases, UDP-glucuronic acid epimerases, homogalacturonan galacturonosyltransferase-like proteins, β-1,4-galactan β-1,4-galactosyltransferases, rhamnogalacturonan-II xylosyltransferases, and pectin acetylesterases appear to have had a single member in the common ancestor of land plants. We detected no Physcomitrella members in the xylogalacturonan xylosyltransferase, rhamnogalacturonan-II arabinosyltransferase, pectin methylesterase inhibitor, or polygalacturonase inhibitor protein families.

**Conclusions**

Several gene families related to the production and modification of pectins in plants appear to have multiple members that are conserved as far back as the common ancestor of mosses and vascular plants. The presence of multiple members of these families even before the divergence of other important cell wall-related genes, such as cellulose synthases, suggests a more complex role than previously suspected for pectins in the evolution of land plants. The presence of relatively small pectin-related gene families in Physcomitrella as compared to Arabidopsis makes it an attractive target for analysis of the functions of pectins in cell walls. In contrast, the absence of genes in Physcomitrella for some families suggests that certain pectin modifications, such as homogalacturonan xylosylation, arose later during land plant evolution.

**Background**

Pectins make up approximately one third of the dry mass of primary cell walls in eudicots, affecting both water dynamics and the mechanical behavior of the wall [1]. Pectins consist of four domains: homogalacturonan (HG), xylogalacturonan (XGA), rhamnogalacturonan I (RG-I), and rhamnogalacturonan II (RG-II) [2]. Homogalacturonan makes up the majority of the pectic component of the cell wall
and also serves as the backbone of XGA and RG-II. Xylogalacturonan is made up of HG with attached xylose side-groups, whereas RG-II has four complex and distinct side-chains [3]. Rhamnogalacturonan I has side-chains containing galactose and arabinose, but its backbone consists of alternating rhamnose and galacturonic acid. These complex polysaccharides are almost universally conserved in land plants and are also present in some algae [4], although structural diversity in pectins is present between some species. For instance, there is evidence for RG-II in all land plant species analyzed to date [3,5] but its side chains are not perfectly conserved [6], and the side chains of RG-I vary among species [1]. Additionally, XGA has not been detected in Physcomitrella patens [7].

Pectins are important determinants of wall remodeling during cellular growth [8]. Pairs of HG molecules can be bound together by Ca$^{2+}$ bridges, stiffening the wall [9], and RG-II side-chains dimerize via borate diol ester bonds [10]. A decreased ability to form RG-II dimers leads to dwarfism [11]. Modifications to pectin can enhance or prevent these interactions and thus affect the properties of the wall as a whole: for example, alterations in wall stiffness mediated by pectin methylation have been implicated in organ primordium initiation and cell elongation [8,12]. Pectins also appear to be essential for normal cell-cell adhesion, since some pectin methylation-defective mutants lack tissue cohesion [13,14].

The complex structures of pectins require a large suite of biosynthetic genes, many of which are inferred only by the biochemical reactions required to synthesize the many linkages in pectins [15,16]. Nevertheless, many pectin-related genes have
been identified, and modification of their expression can have serious effects on the development and growth of mutant plants [17-20]. Pectins play an especially important role in the tip growth of pollen tubes, with methylation status regulating the yielding properties of the tip and side walls [21,22], but this system does not allow for easy genetic manipulation. *Physcomitrella patens*, the model moss [23], represents an attractive experimental system for the genetic and molecular analysis of pectins in the walls of tip-growing cells. Its primary growth form is a mass of protonemal filaments that extend exclusively via tip growth and might therefore rely heavily on pectins for normal development [24,25]. Genes in the *Physcomitrella* genome [26] can be modified directly using high-efficiency homologous recombination [27], which, combined with the dominant haploid generation of this moss, makes it ideal for genetic modification and analysis. As a moss, *Physcomitrella* is also likely to resemble an early stage in the transition of plants from aquatic to terrestrial life, giving us a clearer view of the cell wall architectures and physiology that made this transition possible.

As diverse plant genomes are sequenced, there are new opportunities to study gene families in an evolutionary context. The PlantTribes 2.0 database [28] is an objective gene family classification that can be used to investigate gene family composition and phylogeny on a global scale. By using the complete inferred protein sequences from ten diverse plant genomes (seven angiosperms plus the lycophyte *Selaginella moellendorffii*, the moss *Physcomitrella*, and the chlorophyte *Chlamydomonas reinhardtii*; see Figure 2-1), orthologous gene clusters
(orthogroups) were identified that represent deeply conserved, but often narrowly defined gene families. Orthogroups were constructed using OrthoMCL [29], resulting in gene clusters that typically align well across their length and have a conserved domain structure [30]. Leveraging the PlantTribes 2.0 classification is a conservative approach to identify gene family members from sequenced genomes, avoiding false positive hits that may be identified using less structured search algorithms (e.g. BLAST). To assess the complexity of the pectin biosynthetic and modification machinery in *Physcomitrella* and to investigate the evolutionary history of pectin-related gene families in land plants, we performed an orthogroup-based phylogenetic study of 16 gene families associated with pectin production and modification and mapped the relationships of these genes among terrestrial plant species with sequenced genomes. These analyses reveal that the *Physcomitrella* genome contains at least one member in most of the families analyzed and that the total number of pectin-related gene family members in *Physcomitrella* is much lower than that in *Arabidopsis*. Analysis of these families not only identified gene members in *Physcomitrella*, it also reveals that several pectin-related gene families had multiple members in the land-plant common ancestor.
Results

Identification of pectin-related genes using PlantTribes 2.0

We used the set of genes in *Arabidopsis* belonging to 16 pectin-related gene families identified in the literature (Supplemental File Table 2S1) to select orthogroups in the PlantTribes 2.0 database for in-depth phylogenetic analysis (Supplemental File Table 2S2) [28]. The number of genes from each species in each
family is displayed in Supplemental File Table 2S3. We found at least one

*Physcomitrella* gene in 12 of the 16 families examined (Table 2-1). Notably, no

*Physcomitrella* members of the xylogalacturonan xylosyltransferase (Figure 2-S1),
rhamnogalacturonan-I arabinosyltransferases (Figure 2-S2), pectin methylesterase
inhibitor (Figure 2-S3), or polygalacturonase inhibitor protein (Figure 2-S4) families
were detected. There were fewer *Physcomitrella* members in most of the pectin-
related gene families than in *Arabidopsis*, with the exception of the UDP-rhamnose
synthase (four *Arabidopsis*, six *Physcomitrella*), β-1,4-Galactan β-1,4-
Galactosyltransferase (three *Arabidopsis*, four *Physcomitrella*), and UDP-glucuronic
acid (UDP-GlcA) epimerase (five *Arabidopsis*, nine *Physcomitrella*) families.
Table 2-1: Representatives of pectin-related gene families in *Arabidopsis* and *Physcomitrella*

<table>
<thead>
<tr>
<th>Pectin-related gene family</th>
<th><em>Arabidopsis</em> genes</th>
<th><em>Physcomitrella</em> genes</th>
<th>Putative minimum # of family members in common ancestor</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDP-Rhamnose synthases</td>
<td>4</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>UDP-Glucuronic acid epimerases</td>
<td>5</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Galacturonosyltransferases (GAUTs)</td>
<td>15</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>GAUT-like proteins (GATLs)</td>
<td>10</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>β-1,4-Galactan β-1,4-Galactosyltransferase</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Rhamnogalacturonan II xylosyltransferases</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Rhamnogalacturonan I arabinosyltransferases</td>
<td>2</td>
<td>0</td>
<td>ND</td>
</tr>
<tr>
<td>Xylogalacturonan xylosyltransferases</td>
<td>2</td>
<td>0</td>
<td>ND</td>
</tr>
<tr>
<td>Homogalacturonan methyl-transferases</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Pectin methylesterases</td>
<td>66</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>Pectin methylesterase inhibitors (PMEIs)</td>
<td>2</td>
<td>0</td>
<td>ND</td>
</tr>
<tr>
<td>Polygalacturonases</td>
<td>67</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Polygalacturonase Inhibitor Proteins (PGIPs)</td>
<td>2</td>
<td>0</td>
<td>ND</td>
</tr>
<tr>
<td>Pectate lyase-like proteins</td>
<td>26</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Pectin acetylesterases</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pectin acetyltransferases</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Totals</td>
<td>229</td>
<td>69</td>
<td>24</td>
</tr>
</tbody>
</table>

Sixteen gene families were analyzed. For each gene family, the number under the species with the larger number of genes is highlighted in bold. In most cases there were more Arabidopsis members than *Physcomitrella* members. ND (not determined); phylogenetic ambiguity prevents an accurate estimation of ancestral gene number at this time.

**Phylogenetic analysis of pectin-related gene families**

Our identification of pectin-related genes in ten diverse plant species (Figure 2-1) provided an opportunity to examine their phylogenetic patterns [31]. To analyze the evolutionary relationships between gene family members, we aligned the sequences from the PlantTribes 2.0 search results for each family using the
MUSCLE algorithm [32] followed by manual curation, and constructed maximum likelihood trees from these alignments using RAxML [33]. Where possible, we also included a homologous gene from a green alga to root the trees. We tested the hypothesis that each pectin-related gene family would trace back to a single ancestral gene in the common ancestor of land plants, with any *Physcomitrella* genes forming a clade sister to all other land plants. Surprisingly, this was the case for only seven of the 16 families examined (Table 2-1). Five of the trees have multiple well-supported land plant-wide clades (Figures 3-2, 3-3, 3-4, 3-S5, 3-S6). Each clade is evidence for a separate ancestral gene in the early land plant ancestor of the terrestrial species examined. These trees and their implications are explored below.
Three well-supported clades that suggest ancestral GAUTs are highlighted (blue, pink, and green clouds), and an unresolved polytomy near the root of the tree is indicated in light grey. The green and pink clades, as well as the polytomy, contain monocot, eudicot, Selaginella, and Physcomitrella members, whereas the blue clade does not have any Physcomitrella members. The algal root gene from *Spirogyra pratensis* falls within the polytomy.
Figure 2-3: Polygalacturonase family tree.

Four monophyletic clades contain monocot, eudicot, Selaginella, and *Physcomitrella* genes. The tree contains two large polytomies, indicated in light grey and labeled “A” and “B”. Polytomy B contains unresolved *Physcomitrella* and Selaginella members. The algal root gene is from *C. reinhardtii*, a chlorophytic alga.
Figure 2-4: Pectin-methylesterase family tree.

Two large polytomies, labeled “A” and “B” and shown in light grey, indicate poor resolution of some of this family’s lineages. Four monophyletic clades contain members from the monocots, eudicots, Selaginella, and Physcomitrella. One of these clades (blue cloud) consists of polytomy B and a smaller clade of Physcomitrella and Selaginella genes. Additional moss and tracheophyte genes remain poorly resolved in polytomy A. The algal root (from P. margaritaceum) is within one of the polytomies.
The *GAUT* superfamily contains at least five ancestral land plant genes

The GAUT superfamily consists of the GAUT and the distantly-related GAUT-like (GATL) families [34]. Some Galacturonosyltransferases (GAUTs) are responsible for constructing HG and use UDP-galacturonic acid (UDP-GalA) as a substrate [34]. In *Arabidopsis*, mutations in GAUTs cause phenotypes ranging from changes in sugar composition of the wall to severe dwarfism to apparent lethality [34-37]. In our analysis, the GAUT family tree contains three large well-resolved clades, as well as an unresolved polytomy (Figure 2-2). Genes from *Physcomitrella* and tracheophytes are present in two of these clades and within the polytomy from which the root algal gene is not resolved. The third of these clades includes genes from *Selaginella*, monocots, and eudicots but no *Physcomitrella* genes. This tree suggests a minimum of four ancestral GAUTs in the earliest land plant.

The roles of the GATL proteins are not all clearly established: some of them have been implicated in pectin production, while at least one seems to be involved in xylan synthesis [37,38]. When we generated an alignment and phylogenetic tree of the entire superfamily (Figure 2-5), the GATL family (grey cloud) appeared as a well-resolved but distant clade derived from within the GAUT family that also contains representatives from all of the land plant species queried.
Figure 2-5: GAUT superfamily tree.

In this tree, phylogenetic distance is indicated by branch length. The GATL gene family (yellow cloud) is well-supported as being derived from within the GAUTs; due to a polytomy in the GATL family, clade relationships within this family are not well resolved. The distance of the GATLs from the GAUTs suggests an ancient divergence, but the position of the algal root supports the hypothesis that the GATLs descended from the GAUTs rather than diverging from a common ancestor. Scale bar, 0.7 substitutions/site.

Polygalacturonase and pectin methylesterase families are large and deeply conserved

Whereas GAUTs build the HG backbone of pectins, polygalacturonases (PGs) hydrolyze it, weakening the pectin matrix and potentially loosening the wall [39]. In eudicots, PGs are important in cell expansion and also in abscission and fruit softening [40]. The PG family is very large in Arabidopsis, with over 65 known members. Our phylogenetic analysis for these genes resulted in two large unresolved polytomies, each containing several monophyletic groups, four of which
contain representatives from mosses, lycophytes, monocots, and eudicots (Figure 2-3). Although the placement of several of the *Physcomitrella* genes is unresolved, the gene tree suggests a minimum of five genes in the common ancestor.

Like the PGs, the pectin methylesterase (PME) family is very large in *Arabidopsis* [41]. Galacturonic acid residues in the HG backbones of pectins often have attached methyl ester groups at the C6 position that can prevent interactions with other HG chains as well as pectin-modifying enzymes. Thus, the amount and pattern of methylation can affect wall dynamics in several ways. PMEs remove methyl groups from pectin, rendering it more prone to degradation by hydrolytic enzymes as well as to calcium cross-linking, potentially either weakening or stiffening the wall. This is complicated by the tendency of different PMEs to remove methyl groups in random or block-wise patterns: lone de-methylated GalAs make the polymer prone to enzyme degradation, whereas consecutive exposed carboxylate groups favor calcium-bridging [42]. Like the PGs, the PME gene tree we generated has two large polytomies and two smaller resolved clades (Figure 2-4). Unlike the PG tree, the algal root is a member of one of the polytomies. Within this polytomy are two well-supported land plant-wide monophyletic clades. Resolved from this polytomy is a third land plant-wide clade. Several *Physcomitrella* and *Selaginella* genes are in a clade that is sister to the second polytomy, which consists entirely of angiosperms. This tree suggests that a minimum of five PMEs existed in the common ancestor of the species examined.
Many pectin-related gene families appear to have had only one or two members in the common ancestor of land plants

Like the polygalacturonases, pectate lyase-like proteins cleave the HG backbone of pectins (Figure 2-S5) [43]. Homogalacturonan methyltransferases are responsible for methylating newly synthesized HG (Figure 2-S6) [13]. Both of these family trees indicate the existence of multiple members in the common ancestor by having multiple supported clades with members from every division of the plant lineage. The final seven of the family trees have *Physcomitrella* genes grouped sister to other land plants, indicating a single ancestral gene prior to the divergence of *Physcomitrella* and the tracheophytes: the UDP-GlcA epimerases, the UDP-rhamnose synthases, the pectin acetylerase, the pectin acetyltransferases, the RG-II xylosyltransferases, the β-1,4-galactan β-1,4-galactosyltransferases, and the GATLs (Figures 3-S7 to 3-S13). These families are listed as having one supported common ancestral gene in Table 2-1. The UDP-GlcA epimerase, UDP-rhamnose synthase, β-1,4-galactan β-1,4-galactosyltransferase, and GATL families all likely expanded in *Physcomitrella* after its divergence from the tracheophytes.

**Discussion**

**Search and tree-building criteria for pectin-related genes**

We adopted a relatively stringent set of criteria to identify putative orthologs of *Arabidopsis* pectin-related genes in *Physcomitrella* and other plant species, and
used these genes to build phylogenetic trees of pectin-related gene families. Rather than simply using database searches and overall sequence similarity to identify homologous genes, we leveraged the network of global gene relationships in the PlantTribes 2.0 database to identify clusters of orthologous genes (orthogroups) from the other species for analysis. Using BLAST to identify putative gene orthologs is a common practice, but increases the number of false positive sequences obtained because hits may only share high similarity in a small portion of the gene (i.e. a conserved domain), but may not be closely related and align poorly across the full length of the sequence. In contrast to BLAST-based methods, the use of PlantTribes 2.0 orthogroups increases the probability of identifying genes within the same evolutionary lineage, thus reflecting the history of these gene families more accurately. In some cases our search method detected fewer *Physcomitrella* members than other analyses of these families [39,44,45]. In all these cases the researchers used shared protein domains or sequence homology to identify their genes of interest. The search method we used was intended to identify high-confidence candidate genes for further experimental analysis that are more likely to share conserved functions within other model systems. We therefore employed a higher stringency approach at the cost of missing more distantly related homologs.

Although our trees largely agree with previously published phylogenies for some pectin-related gene families [35,39,44-48], the larger number of species we used improved our ability to resolve gene family topologies and to detect basal branchpoints that have been obscured in analyses using genome data from fewer
species [35,39,45-48]. An exception to this is the work of Wang et al., which identified PMEs and PMEIs in the same land plant species we examined, as well as *Amborella trichopoda* [44]. Wang et al. searched for conserved PME and PMEI protein domains and identified 35 putative *Physcomitrella* PMEs as compared with our ten. They also produced a large PMEI tree that included a putative *Physcomitrella* member. In contrast to our approach, their domain-based approach likely resulted in the detection of distantly related genes not included in our results.

**Several pectin-related gene families likely had multiple members in the common ancestor of mosses and tracheophytes**

The topologies of the trees we generated provide clues to the evolutionary relationships between known pectin-related genes and their orthologs in other species. This allows us to hypothesize about the state of the gene families in the last common ancestor of *Physcomitrella* and vascular plants. In seven of the families we analyzed, the paralogs in *Physcomitrella* are sister to all other genes in vascular plants. On the other hand, several of the families (GAUTs, HG methyltransferases, PMEs, PGs, pectate lyase-like proteins) each appear to have had multiple members in the common ancestor of land plants. Our analyses suggest that the suite of genes for the production, modification, and degradation of pectins had already diversified prior to the radiation of land plants. This contrasts with the cellulose synthase gene family (CESA), which likely contained a single gene in the ancestor of land plants and subsequently diversified after the divergence of mosses and vascular plants.
Multiple members of a gene family often have different expression patterns, allowing for tissue-specific regulation of the associated activity; for example, PpCESA5 is required only for gametophore development, implying that other PpCESAs produce cellulose in protonemal tissue [50]. Intriguingly, others have hypothesized that pectin synthesis and modification might originally have been central in wall production and modulation, with the importance of cellulose arising later [51]. There is also evidence for further diversification of these families before the flowering plant divergence in the form of angiosperm-wide clades in the GAUTs, PMEs, PGs, pectate lyase-like proteins, UDP-glucuronic acid epimerases, UDP-rhamnose synthases, and pectin acetylesterases.

Some pectin-related gene families were not detected in Physcomitrella

Since orthogroups in the PlantTribes 2.0 database generally represent narrowly defined gene lineages that typically align well across the whole length of the gene, we are confident that distantly related genes have been excluded from our analyses. However, it is possible that we failed to detect highly divergent members of some of these gene families. Nevertheless, most of the searches yielded at least one Physcomitrella gene per family. This was not true of the XGA xylosyltransferases, the RG I arabinosyltransferases, the PGIPs, and the PMEIs. It is not surprising that XGA xylosyltransferases were not detected in Physcomitrella given that a previous study using comprehensive microarray polymer profiling (COMPP) did not detect XGA in Physcomitrella cell walls [7]. On the other hand, α(1-
5)-arabinans characteristic of RG I were detected in the pectic fraction of *Physcomitrella* walls, which combined with the failure to detect *Physcomitrella* orthologs of *AtARAD* genes in this study and others [48] raises the possibility of the existence of other arabinan-arabinosyltransferases that are only distantly related to the currently known genes.

Although there are not any studies suggesting PGIPs are absent in *Physcomitrella*, we also did not detect any PGIP genes in *Selaginella*, suggesting that this gene family may have evolved after the divergence of lycophytes and euphyllophytes. PGIPs are thought to play a role in pathogen defense by preventing foreign PGs from degrading the plant cell wall [52], and it is interesting that none were detected in either our representative moss or lycophyte, given that *Physcomitrella* and other mosses are susceptible to fungal pathogens [53]. The PMEI tree we generated only contains genes from *Arabidopsis* and *Medicago truncatula*, and might not adequately represent the diversity in this gene family. This might be due to insufficient numbers of query genes to allow for the detection of all the family members, or because coding sequence information for some of the species might have been incomplete. Importantly, the *Arabidopsis* query genes were both contained within one orthogroup. Genome data for additional plant species and/or future improvements in genome annotations could potentially overcome this limitation.
Arabidopsis has an abundance of pectin-related genes, whereas grasses appear to have fewer pectin-related genes in some families

In nine of the 16 families analyzed, Arabidopsis had more members than any of the other species (Supplemental File Table3S3). This might be the result of the more extensive annotation of the Arabidopsis genome as compared to other species in the database, or the unique genome duplication histories of the species analyzed [30]. We see a general trend of more pectin-related genes in the eudicots than in the monocots and more in the monocots than in the more basal species such as Physcomitrella and Selaginella. This may reflect the lower levels of pectin in the walls of grasses compared to other flowering plants [54], as well as the relatively high abundance of other acidic polymers such as glucuronoarabinoxylans in grasses [55]. Further phylogenetic analyses of non-commelinid monocots, which have Type I cell walls [56], might be informative in determining the relationship between the elaboration of pectin-related gene families and the abundance of pectins in the cell wall.

Conclusions

Pectins play a key role in the cell walls of plants. We analyzed 16 gene families involved in the production, modification, and degradation of pectins in nine land plant species. Our analysis indicates that although many of these families appear to trace back to a single gene in the last common ancestor to the mosses and the vascular plants, several of the major families involved in pectin regulation likely
contained multiple genes. We did not detect *Physcomitrella* (or *Selaginella*) genes in four of the studied families, providing some evidence that they might have evolved after the divergence of seed plants from the lycophytes. This study has allowed us to identify *Physcomitrella* orthologs related to known pectin-related genes in *Arabidopsis* for in-depth experimental analysis. Our results also shed light on the evolutionary history of pectin biosynthesis and modification, suggesting that pectins may have played an important role in the transition from an aquatic to a terrestrial environment.

**Methods**

**Identification of pectin-related gene families**

We compiled a list of *Arabidopsis* genes with known and predicted pectin-related functions using TAIR and Uniprot annotations, as well as relevant literature (Supplemental File Table 3S1) [1,34,41,52,57-63]. In total, we used 108 genes from *Arabidopsis* to identify putative pectin-related gene families in the PlantTribes 2.0 database [64]. PlantTribes 2.0 is an objective gene family classification of protein coding genes from ten sequenced green plant genomes that have been clustered into orthogroups (putatively monophyletic gene lineages) using OrthoMCL [28]. Orthogroups containing pectin-related genes from *Arabidopsis* were extracted for phylogenetic analysis. This approach enabled us to include additional homologous genes from *Arabidopsis* not annotated with pectin-related gene functions. In some
cases, the pectin-related query genes from *Arabidopsis* did not belong to an orthogroup (i.e., they were singletons). The closest *Physcomitrella* gene to these singleton *Arabidopsis* genes were identified via TBLASTX and added to the family alignment. Because PlantTribes 2.0 includes the *Physcomitrella patens* version 1.1 gene annotations from Phytozome [65], we used a nucleotide BLAST+ search of a local database of *Physcomitrella patens* version 1.6 annotated coding sequences to identify the current gene annotations for ease of reference (Supplemental File Table3S2, which includes all of the genes used in this paper). Although PlantTribes 2.0 does include the chlorophyte alga *Chlamydomonas reinhardtii*, many of the gene families still lacked a non-land plant outgroup. To enhance the possibility of rooting our trees using an outgroup, we also included homologous transcript sequences from three additional green algae (*Nitella hyalina*, *Penium margaritaceum*, and *Spirogyra pratensis*) where possible [66]. We searched each transcriptome separately using coding sequences from *Physcomitrella* using TBLASTX with an E-value cutoff of $10^{-10}$. Full-length coding sequences were identified for the GAUT, pectin methyl esterase, UDP-rhamnose synthase, rhamnogalacturonan I arabinosyltransferase, and rhamnogalacturonan II xylosyltransferase families.

**Phylogenetic analysis**

Sequences for each family were aligned by translation in Geneious using MUSCLE (default parameters) [32], manually curated, and saved as relaxed Phylip files (Supplemental Files). In some cases this required removing non-homologous
genes and gene fragments from poorly annotated genomes. To generate trees (Supplemental Files), maximum likelihood phylogenetic analysis was performed using RAxML [33] with the following parameters: rapid bootstrap analysis and search for best-scoring maximum likelihood tree in one run, GTRGAMMA model of nucleotide evolution, random seed 12345, 1000 bootstrap replicates. Nodes with less than 50% bootstrap support were collapsed using TreeCollapserCL4 [67] and were visualized using FigTree [68]. Figures were manually edited for readability using Adobe Illustrator.

Author contributions

TWM contributed to experimental design, collected query sequences, performed the database searches, identified algal roots, performed the sequence alignments, ran the phylogenetic analyses, prepared the figures, and participated in drafting the manuscript. JPD contributed to experimental design, assisted in sequence alignment, and participated in drafting the manuscript. LAH contributed to experimental design, assisted in sequence alignment, and participated in drafting the manuscript. CWD contributed to experimental design and participated in drafting the manuscript. CTA contributed to experimental design and participated in drafting the manuscript. All authors read and approved the final manuscript.
Acknowledgements

Thanks to William Ehlhardt, William Murphy, and Jacqueline Doyle for help in building scripts to streamline database searches, and to Eric Wafula for bioinformatic assistance. Phylogenetic analysis was supported as part of The Center for LignoCellulose Structure and Formation, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award # DE-SC0001090 (TWM and CTA), and development of the PlantTribes 2.0 database was supported by NSF Plant Genome grant #0922742 (JPD, LAH, and CWD).

References


light on the evolution of increasingly complex organs in plants. *New Phytol* 2013, **197**:1353–1365.


**Arabidopsis thaliana** is a β-1,4-galactan β-1,4-galactosyltransferase. *Plant Cell* 2012, **24**:5024–5036.


64. **PlantTribes 2.0 Database.** http://fgp.bio.psu.edu/tribedb/10_genomes/index.pl.


68. *FigTree*. http://tree.bio.ed.ac.uk/software/figtree/.
Chapter 3

Studying deeply conserved putative polygalacturonases

Phylogenetic trees in Chapter 2 demonstrated that some of the pectin-related gene families had radiated before the divergence of land plants. However, this is not the full extent of evolutionary history we can infer. The expansion of the polygalacturonase gene family is attributed in part to whole-genome duplication events and subsequent neo- or sub-functionalization, but the family also contains embryophyte-wide monophyletic clades that do not show the same degree of radiation [1]. These small land plant-wide clades in the polygalacturonase family demonstrate deep conservation of certain genes, perhaps indicating PG functions that are key to plant growth and development. Intriguing, however, is the lack of radiation within these clades. Perhaps these genes’ functions are so specific that mutations altering transcription pattern or enzyme activity, responsible for neo- and sub-functionalization of genes in other clades, offer no fitness advantage to the plant. An Arabidopsis member of one of these small embryophyte-wide clades, PGX1 (AT3G26610), has already been characterized and demonstrated to be non-redundantly involved in plant growth and cell expansion [2]. Characterizing other deeply conserved putative PGs may reveal new functions for pectins and PG activity in plant development.
**T-DNA insertion mutants of putative PGs did not disrupt expression**

Six Arabidopsis polygalacturonase genes from three deeply conserved clades were chosen as subjects for analysis (Figure 3-1). Twenty mutant lines with T-DNA insertions in or near the coding sequences of the genes were found using The Arabidopsis Information Resource (TAIR) and T-DNA Express: Arabidopsis Gene Mapping Tool, and ordered from T-DNA library collections at the Arabidopsis Biological Resource Center (Table 3-1) [3, 4]. Homozygous mutant plants were identified by the presence of a PCR product specific to the T-DNA insertion site, and the absence of a PCR product specific to the wild-type version of the site. The homozygous mutant plants were tested for disruption of mRNA production by RT-PCR, but ultimately none of these lines appeared to be transcriptional null mutants (Table 3-1).

**Table 3-1: Non-null T-DNA mutant lines**

<table>
<thead>
<tr>
<th>Arabidopsis gene</th>
<th>ABRC Accession</th>
<th>Arabidopsis gene</th>
<th>ABRC Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G10640</td>
<td>SAIL_11_F04</td>
<td>AT1G19170</td>
<td>SALK_135999C</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>SALK_052933c</td>
<td>AT1G19170</td>
<td>SALK_100717</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>SALK_126876</td>
<td>AT1G60590</td>
<td>SAIL_594_D01</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>SALKseq_124078.4</td>
<td>AT1G60590</td>
<td>WiscDsLox447A07</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>CS858518</td>
<td>AT3G42950</td>
<td>SALK_016021C</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>SALK_053023C</td>
<td>AT3G42950</td>
<td>WiscDsLox361C05</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>SALK_053271</td>
<td>AT3G57790</td>
<td>SALK_072207C</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>SALK_126874</td>
<td>AT3G57790</td>
<td>SALK_116808C</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>SALK_085847</td>
<td>AT5G14650</td>
<td>CS404981</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>SALK_100709</td>
<td>AT5G14650</td>
<td>SALK_145451</td>
</tr>
</tbody>
</table>

Mutant lines were ordered from the Arabidopsis Biological Resource Center. None of these were transcriptional knockouts.
Arabidopsis members from three of the small, deeply conserved clades were chosen for further study (red arrows): At1g10640, At1g19170, At1g60590, At3g42950, At3g57790, At5g14650.

One of the mutant lines, SALKseq_124078.4, a putative insertion into the first exon of AT1G10640, had the surprising phenotype of producing no progeny that
tested as homozygous mutants: every single plant showed a PCR product derived from the wild-type version of the gene. We initially thought that this was a homozygous lethal mutant of AT1G10640, but eventually evidence suggested that the T-DNA insertion event in this line had caused a partial genome duplication and rearrangement [5]. All the plants of this T-DNA line had functional copies of the AT1G10640 gene.

**Generating targeted transcriptional mutants of conserved putative PGs**

Using T-DNA insertion libraries to try to find knockout mutants of the polygalacturonase genes of interest proved to be ineffective, but that does not change the fact that these polygalacturonases are deeply conserved and worth studying. Gene knockouts are still one of the most effective tools for learning about the function of a protein within a plant, so we decided to use newer molecular biology approaches to generate our own mutant lines.

**CRISPR-Cas9 strategy for knocking out genes**

The CRISPR-Cas9 system is a promising new technology for directed genome editing, and the limits of its uses have probably still not been found [6]. The system combines the bacterial endonuclease Cas9 with a synthetic guide RNA (gRNA) that contains a 20-nucleotide "protospacer" region that determines the target of the Cas9 protein's nuclease action. The Cas9/gRNA complex binds to DNA that matches the
protospacer sequence, and, importantly, has a "Protospacer-Adjacent Motif" (PAM), 5'-NGG-3' directly downstream of the protospacer. The Cas9 protein creates a double-stranded break (DSB) in the targeted DNA between the third and fourth base pairs upstream of the PAM (Figure 3-2) [6, 7]. This DSB is usually repaired by the cell's Non-Homologous End Joining repair machinery, which is error-prone and often causes base insertion or deletion mutations ("InDels") [8]. An InDel that is not a multiple of three bases within a gene coding sequence (CDS) will cause a frame-shift mutation that changes the coding sequence in the downstream portion of the gene, often introducing an early stop codon. This makes the CRISPR-Cas9 system extremely well suited to generating targeted knockout lines of genes of interest.

![Figure 3-2: Schematic of Cas9/gRNA genome editing.](image)

Cas9 is directed to its DNA target by base pairing between the gRNA and DNA. A PAM motif downstream of the gRNA-binding region is required for Cas9 recognition and cleavage. Cas9/gRNA cuts both strands of the target DNA, triggering endogenous DSB repair. In a knockout experiment, the DSB is repaired via the error-prone NHEJ pathway, which introduces an InDel at the DSB site that knocks out gene function. Figure adapted from [6].
The requirement for CRISPR-Cas9 activity to have a PAM limits the choice of targets somewhat, but NGG is not an uncommon sequence, and a bioinformatic analysis showed that seven out of eight sequenced plant genomes had specific protospacer targets next to the NGG PAM for >85% of their annotated transcriptional units [9]. In combination with rapid advances in Cas9 variants and orthologs that have alternate PAM sequences, the ability to specifically target virtually any gene is quickly becoming a reality [6].

Choosing CRISPR-Cas9 protospacer sequences

The CRISPR-PLANT web app (http://www.genome.arizona.edu/crispr/) allows the user to search several plant species, by gene ID or chromosome location, for viable CRISPR-Cas9 protospacer sequences. It returns a list of potential sequences, classified by their potential for off-target activity and with a count of the minimum number of mismatches necessary to bind to another spacer sequence in the genome. Four mismatches or more from the most similar potential recognition site is considered highly specific (three mismatches are sufficient if they are within 10 bases of the PAM). All six polygalacturonases of interest had several highly specific potential protospacer sequences within their coding sequences, but minimizing the risk of off-target Cas9 activity is not the only factor in choosing a target. An ideal CRISPR-Cas9 protospacer sequence for knocking out a gene will be specific (unlikely to have off-target activity), effective (likely to cause a mutation that renders the gene completely non-functional), and cause easily detectable
mutations. Because one Cas9-gRNA pairing can lead to multiple mutant alleles of a target gene, only one protospacer was chosen for each polygalacturonase of interest. The high-specificity protospacer sequences provided by CRISPR-PLANT were mapped onto the target gene sequence so that a protospacer that would lead to disruption in the first exon could be selected (Figure 3-3). This maximizes the potential specificity and effectiveness, but to provide a quick way to screen for mutations, protospacer sequences were chosen that would cause a disruption to a known restriction enzyme recognition site (Supplemental File Cas9_sequences.csv).
**Figure 3-3: CRISPR-Cas9 target sites in polygalacturonases**

**A**) The protospacer sequence targeting each polygalacturonase was within the first exon, making any frame shift mutations likely to totally disrupt gene function. **B**) InDels at the Cas9 cut site, three bp upstream of the PAM sequence, will disrupt a restriction enzyme site, speeding identification of mutant plants. **C**) Frame shift mutations caused by errors in Non-Homologous End Joining render the downstream gene nonsense and often introduce an early stop codon, effectively knocking out the gene.
**CRISPR-Cas9 induced mutation lines**

Arabidopsis plants were transformed with CRISPR-Cas9 constructs as described in Methods. The region surrounding the CRISPR-Cas9 target site was amplified from gDNA extracted from antibiotic-resistant T1 plants. PCR products were sequenced at the Penn State Sequencing Facility to assay the presence of a disruptive mutation. The chromatogram results of one such reaction are seen in Figure 3-4.

![Image](image-url)

**Figure 3-4: Example CRISPR-Cas9 sequencing chromatogram**

The sudden appearance of secondary peaks after the predicted Cas9 cut site (between 301 and 302) suggests that PCR amplified both a wild-type and a mutant version of the region. In this case it appears that NHEJ repair introduced a single G insertion into AT3G42950.

Mutations were detectable by the sudden appearance of secondary peaks at (or sometimes in the case of deletions, near) the predicted Cas9 cut site. While in
some cases InDels were apparent from obvious double peaks in the sequencing chromatogram, in other transformed plant samples no mutation was apparent. Sequencing results indicate that at least 45 out of 95 plants tested had Cas9-induced mutations, and 18 of these mutations introduce stop codons into the first exon of the targeted PG gene. All but one of the mutant plants also retained a wild-type copy of the gene. Interestingly, none of the tested plants transformed with the construct targeting AT1G60590 showed any sign of mutation near the predicted Cas9 cut site. Nevertheless, multiple putative null mutations for five out of the six target genes were isolated, and plenty of T1 seeds remain if additional independent lines are desired. The T2 seed lines from each transformed plant and their specific mutation, if known, are listed in Table 3-2.

Table 3-2: CRISPR-Cas9 mutant seed stocks

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Parent Mutation</th>
<th>Generation</th>
<th>Seed stock database number</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G10640</td>
<td>Chimeric</td>
<td>T2</td>
<td>CTA6436</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>Unknown</td>
<td>T1</td>
<td>CTA6437</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>1 A insertion</td>
<td>T1</td>
<td>CTA6438</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>Unknown</td>
<td>T2</td>
<td>CTA6440</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>Chimeric</td>
<td>T2</td>
<td>CTA6441</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>None</td>
<td>T2</td>
<td>CTA6442</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>Unknown</td>
<td>T2</td>
<td>CTA6443</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>None</td>
<td>T2</td>
<td>CTA6444</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>None</td>
<td>T2</td>
<td>CTA6445</td>
</tr>
<tr>
<td>Gene Name</td>
<td>Mutation Type</td>
<td>Validation Code</td>
<td>Sample Code</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------</td>
<td>-----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>None</td>
<td>T2</td>
<td>CTA6446</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>None</td>
<td>T2</td>
<td>CTA6447</td>
</tr>
<tr>
<td>AT1G10640</td>
<td><strong>1 A insertion</strong></td>
<td>T2</td>
<td>CTA6448</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>None</td>
<td>T2</td>
<td>CTA6449</td>
</tr>
<tr>
<td>AT1G10640</td>
<td><strong>1 G insertion</strong></td>
<td>T2</td>
<td>CTA6450</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>None</td>
<td>T2</td>
<td>CTA6451</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6483</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6484</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6485</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6486</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6487</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6488</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6489</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6490</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6491</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6492</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6493</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6494</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6495</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6496</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6497</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>None</td>
<td>T2</td>
<td>CTA6498</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>Unknown</td>
<td>T2</td>
<td>CTA6452</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>Unknown</td>
<td>T2</td>
<td>CTA6453</td>
</tr>
<tr>
<td>AT1G60590</td>
<td><strong>1 T insertion</strong></td>
<td>T2</td>
<td>CTA6454</td>
</tr>
<tr>
<td>AT1G60590</td>
<td><strong>1 T insertion</strong></td>
<td>T2</td>
<td>CTA6455</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>None</td>
<td>T2</td>
<td>CTA6456</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>Unknown</td>
<td>T2</td>
<td>CTA6457</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>Chimeric</td>
<td>T2</td>
<td>CTA6458</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>InDel</td>
<td>T2</td>
<td>CTA6459</td>
</tr>
<tr>
<td>AT1G60590</td>
<td><strong>1 A insertion</strong></td>
<td>T2</td>
<td>CTA6460</td>
</tr>
<tr>
<td>AT1G60590</td>
<td><strong>1 A insertion</strong></td>
<td>T2</td>
<td>CTA6461</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>Chimeric</td>
<td>T2</td>
<td>CTA6462</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>Chimeric</td>
<td>T2</td>
<td>CTA6463</td>
</tr>
<tr>
<td>AT1G60590</td>
<td><strong>1 G insertion</strong></td>
<td>T2</td>
<td>CTA6464</td>
</tr>
<tr>
<td>AT1G60590</td>
<td><strong>1 C insertion</strong></td>
<td>T2</td>
<td>CTA6465</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>Chimeric</td>
<td>T2</td>
<td>CTA6466</td>
</tr>
<tr>
<td>Gene</td>
<td>Mutation Type</td>
<td>Identify</td>
<td>Barcode</td>
</tr>
<tr>
<td>--------</td>
<td>---------------</td>
<td>----------</td>
<td>---------</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>1 A insertion</td>
<td>T2</td>
<td>CTA6467</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>1 G insertion</td>
<td>T2</td>
<td>CTA6468</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>1 base deletion</td>
<td>T2</td>
<td>CTA6469</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>InDel</td>
<td>T2</td>
<td>CTA6470</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>InDel</td>
<td>T2</td>
<td>CTA6471</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>Chimeric</td>
<td>T2</td>
<td>CTA6472</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>InDel</td>
<td>T2</td>
<td>CTA6473</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>Chimeric</td>
<td>T2</td>
<td>CTA6474</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>InDel</td>
<td>T2</td>
<td>CTA6475</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>InDel</td>
<td>T2</td>
<td>CTA6476</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>None</td>
<td>T2</td>
<td>CTA6477</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>InDel</td>
<td>T2</td>
<td>CTA6478</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>Chimeric</td>
<td>T2</td>
<td>CTA6479</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>1 G insertion</td>
<td>T2</td>
<td>CTA6480</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>InDel</td>
<td>T2</td>
<td>CTA6481</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>None</td>
<td>T2</td>
<td>CTA6482</td>
</tr>
<tr>
<td>AT3G57790</td>
<td>InDel</td>
<td>T2</td>
<td>CTA6420</td>
</tr>
<tr>
<td>AT3G57790</td>
<td>Unknown</td>
<td>T2</td>
<td>CTA6421</td>
</tr>
<tr>
<td>AT3G57790</td>
<td>Unknown</td>
<td>T2</td>
<td>CTA6422</td>
</tr>
<tr>
<td>AT3G57790</td>
<td>None</td>
<td>T2</td>
<td>CTA6423</td>
</tr>
<tr>
<td>AT5G14650</td>
<td>Large deletion</td>
<td>T2</td>
<td>CTA6430</td>
</tr>
<tr>
<td>AT5G14650</td>
<td>1 A insertion</td>
<td>T2</td>
<td>CTA6431</td>
</tr>
<tr>
<td>AT5G14650</td>
<td>Double InDel</td>
<td>T2</td>
<td>CTA6388</td>
</tr>
<tr>
<td>AT5G14650</td>
<td>None</td>
<td>T2</td>
<td>CTA6433</td>
</tr>
<tr>
<td>AT5G14650</td>
<td>None</td>
<td>T2</td>
<td>CTA6434</td>
</tr>
<tr>
<td>AT5G14650</td>
<td>None</td>
<td>T2</td>
<td>CTA6435</td>
</tr>
</tbody>
</table>
Targeting genes with artificial microRNAs

In many cases the complete disruption of gene activity that an early stop codon created by the CRISPR-Cas9 system causes is ideal for learning about the gene's role in the plant. However, in the case of genes essential to early development or growth, a homozygous knockout is fatal and can only reveal the earliest stage in the plant's development at which the gene becomes necessary. This is not very illuminating about the gene's role later in the plant's life, and may also obfuscate the specific activity of the protein encoded by the gene. In such cases, it can be more
useful to have a mutant line in which expression of the gene-of-interest is highly reduced but still present, allowing the plants to survive past infancy and show more informative phenotypes than "embryo death".

Artificial microRNAs (amiRNAs) are short sequences of RNA that serve as targeting guides for RNA-interference proteins such as DICER to destroy messenger RNA (mRNA) containing regions complementary to the amiRNA. Transformed Arabidopsis lines containing amiRNA constructs designed to target a specific gene often vary in the expression level of the targeted gene, potentially allowing the isolation of viable mutants of essential genes that show informative phenotypes [10]. Because the polygalacturonases being studied are deeply conserved and therefore more likely than average to be essential genes, amiRNA constructs were designed as a complementary addition to the CRISPR-Cas9 approach.

The James Carrington lab’s Plant Small RNA Maker Suite (P-SAMS) web app (http://p-sams.carringtonlab.org) was used to automatically design amiRNAs to target each of the polygalacturonases-of-interest with high specificity [11]. Three amiRNA constructs were designed and assembled for each gene (Supplemental File amiRNA_sequences.csv). The tool also offers the ability to design amiRNAs that will target multiple closely related Arabidopsis genes, and this was used to design an amiRNA against each of the closely related pairs At1g10640/At1g60590 and At3g42950/At1g19170. If these gene pairs are partially or wholly redundant, any mutant line targeting only one of the genes might not show a phenotype. Knocking down both at the same time improves the odds of seeing a meaningful effect. Figure
3-5 shows the amiRNA recognition sites on the targeted polygalacturonases. Cloning and transformation of the amiRNA constructs is described in the methods section, and Table 3-3 contains the location and database number of the T1 seeds, which have not yet been characterized.

<table>
<thead>
<tr>
<th>Target gene</th>
<th>Insert</th>
<th>Generation</th>
<th>Seed stock database number</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT1G10640</td>
<td>amiRNA1</td>
<td>T1</td>
<td>CTA6417</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>amiRNA2</td>
<td>T1</td>
<td>CTA6414</td>
</tr>
<tr>
<td>AT1G10640</td>
<td>amiRNA3</td>
<td>T1</td>
<td>CTA6413</td>
</tr>
<tr>
<td>AT1G19170</td>
<td>amiRNA3</td>
<td>T1</td>
<td>CTA6405</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>amiRNA1</td>
<td>T1</td>
<td>CTA6411</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>amiRNA2</td>
<td>T1</td>
<td>CTA6418</td>
</tr>
<tr>
<td>AT1G60590</td>
<td>amiRNA3</td>
<td>T1</td>
<td>CTA6408</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>amiRNA1</td>
<td>T1</td>
<td>CTA6406</td>
</tr>
<tr>
<td>AT3G42950</td>
<td>amiRNA2</td>
<td>T1</td>
<td>CTA6416</td>
</tr>
<tr>
<td>AT3G57790</td>
<td>amiRNA2</td>
<td>T1</td>
<td>CTA6407</td>
</tr>
<tr>
<td>AT3G57790</td>
<td>amiRNA3</td>
<td>T1</td>
<td>CTA6410</td>
</tr>
</tbody>
</table>

Figure 3-5: Polygalacturonase amiRNA binding sites
Unlike CRISPR-Cas9 target sites, amiRNAs do not have to be targeted to the first exon, as any binding to the mRNA should eventually lead to its complete digestion.
Generation of transgenic plants and microorganisms for the characterization of the putative polygalacturonase AT1G10640

Because we initially believed we had a genuine AT1G10640-disrupting mutant that exhibited an embryo-lethal phenotype, additional transgenic material was prepared in anticipation of characterizing AT1G10640. A complementation construct containing the wild-type gene, a promoter region:GUS reporter fusion construct, and a constitutive overexpression construct were prepared. The promoter:GUS and the overexpression constructs were transformed into Arabidopsis plants of the Columbia ecotype. Fourteen independent T2 promoter:GUS lines, 10 independent T2 overexpression lines, and 37 T3 lines descended from three T2 overexpression lines have been harvested and entered into the seed stock database (Table 3-4).

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Generation</th>
<th>Seed stock database number</th>
<th>Parent seed stock</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6015</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6016</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6017</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6018</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6037</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6038</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6039</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6040</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6041</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6042</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6043</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6044</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6045</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Promoter:GUS</td>
<td>T2</td>
<td>CTA6047</td>
<td>CTA5758</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T2</td>
<td>CTA5987</td>
<td>CTA5759</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T2</td>
<td>CTA5988</td>
<td>CTA5759</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T2</td>
<td>CTA5989</td>
<td>CTA5759</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T2</td>
<td>CTA5990</td>
<td>CTA5759</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T2</td>
<td>CTA6019</td>
<td>CTA5759</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T2</td>
<td>CTA6020</td>
<td>CTA5759</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T2</td>
<td>CTA6021</td>
<td>CTA5759</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T2</td>
<td>CTA6022</td>
<td>CTA5759</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T2</td>
<td>CTA6086</td>
<td>CTA5759</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T2</td>
<td>CTA6087</td>
<td>CTA5759</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6287</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6288</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6289</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6290</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6291</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6292</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6293</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6294</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6295</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6296</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6297</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6298</td>
<td>CTA6021</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6302</td>
<td>CTA6020</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6303</td>
<td>CTA6020</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6304</td>
<td>CTA6020</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6305</td>
<td>CTA6020</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6306</td>
<td>CTA6020</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6307</td>
<td>CTA6020</td>
</tr>
<tr>
<td>Overexpression</td>
<td>T3</td>
<td>CTA6308</td>
<td>CTA6020</td>
</tr>
</tbody>
</table>
We also worked toward heterologous expression of the AT1G10640 protein so that polygalacturonase activity could be confirmed. Attempts to produce detectable AT1G10640 in *E. coli* were unsuccessful, so we decided to try to use the yeast *Pichia pastoris* as our expression system. As a eukaryote, we hypothesized that protein-folding conditions for AT1G10640 might be better in *Pichia* than in *E. coli*, and several fungal polygalacturonases have been successfully expressed in *Pichia* [12–14]. Constructs for expression in *Pichia* were prepared and transformed into the wild-type X-33 *Pichia* ecotype and the slow-growing Mut<sup>s</sup> ecotype KM71H. Other researchers in our lab have attempted to heterologously express polygalacturonases and faced difficulties with obtaining soluble protein [2]. In keeping with this, we were not able to detect the His-tagged protein in supernatant from *Pichia* liquid.
cultures or in the supernatant of lysed cell pellets using a Western blot, and those solutions did not exhibit polygalacturonase activity. However, at least one of the lines of transformed *Pichia* (pPICZα A + truncated AT1G10640 CDS in the X-33 strain) demonstrates polygalacturonic acid (PGA) degrading activity in a slurry of disrupted *Pichia* cells (Table 3-5). This crude detection method leaves much to be desired, but it provides preliminary evidence that AT1G10640 is a genuine polygalacturonase, and that the *Pichia pastoris* system can be used to produce it.

| Table 3-5: *Pichia* cultures expressing AT1G10640 show polygalacturonase activity |
|-------------------------------|---------|---------|---------|---------|
| Construct                      | *Pichia* strain | 266 nm Absorbance | 266 nm Absorbance | Active/Boiled Ratio |
| pPICZα A + truncated AT1G10640 CDS | X-33     | 1.155   | 0.448   | 2.58    |
| pPICZα A empty                 | X-33     | 0.845   | 0.875   | 0.97    |
| pPICZα A + truncated AT1G10640 CDS | KM71H    | 0.6     | 0.484   | 1.24    |
| pPICZα A empty                 | KM71H    | 0.818   | 0.743   | 1.10    |
| pPICZ B + AT1G10640 CDS        | KM71H    | 0.839   | 0.492   | 1.71    |
| pPICZ B empty                  | KM71H    | 0.777   | 0.605   | 1.28    |

An increase in absorbance at 266 nm (the peak corresponding to 2-cyanoacetamide labeling of PGA reducing ends) was seen in PGA solutions that were 25% *Pichia* cell slurry by volume. None of the solutions that were only 10% *Pichia* cell slurry showed activity that was detectably different from the boiled sample.

**Discussion**

While CRISPR-Cas9 biotechnology has amazing potential, it does have limitations. T1 lines with mutations in both alleles of targeted genes are unfortunately more rare in Arabidopsis than they are in some of the animal cell systems in which it has been tested [8]. Transformed Arabidopsis plants heterozygous for knockout mutations were identified for five out of the six targeted
genes, including mutants of AT1G10640. Almost all of the first generation CRISPR-Cas9 mutant plants still carry a wild-type copy of the targeted gene, but self-crossing these plants should produce several independent lines of homozygous knockout mutants. Additionally, amiRNA lines were prepared in case the CRISPR-Cas9 system was ineffective (as in the case of AT1G60590) or the knockout mutant turned out to be lethal. Collectively, these plant lines should be a valuable resource for characterizing the function of several deeply conserved putative polygalacturonases.

Methods

Plant material and growth conditions

All Arabidopsis plants were of the Columbia ecotype. Seeds were germinated and grown on pH 5.6 ½MS plates consisting of 2.2 g/L Murashige and Skoog salts (Caisson Labs), 0.6 g/L MES (Research Organics), 1% w/v sucrose, and 0.8% w/v agar-agar (Research Organics) in a chamber kept at 22°C with 24-hour light. Seedlings were transplanted from plates to soil supplemented with MiracleGro fertilizer and grown in a growth room at 22°C under long-day conditions (16 hr light/8 hr dark).
Assembling AT1G10640 complementation, promoter-GUS fusion reporter, and constitutive overexpression constructs

All three of these constructs were assembled using the Gateway cloning system. In each case, the insert sequence was amplified by PCR using OneTaq polymerase (New England Biolabs) and then cloned into the pCR8 plasmid using the pCR8/GW/TOPO TA cloning system (Life Technologies), creating an entry clone.

The TOPO cloning reaction was transformed into chemically competent Top10 E. coli by heat shocking at 42°C for 45 seconds. Plasmids were purified from several transformed colonies using a Plasmid Mini Kit (Omega Bio-tek) and sequenced to confirm insertion direction and lack of mutation. The insert was then cloned into the appropriate destination vector via LR Gateway reaction to create the final expression plasmid.

To create the complementation construct, wild-type Arabidopsis genomic DNA (gDNA) was extracted using the Edwards method and used as a template to amplify the region starting 2000 bp upstream of the At1g10640 coding sequence (CDS) start site and ending right before the stop codon [15]. This PCR product was recombined into the pMDC110 destination vector to generate AT1G10640<sub>pro</sub>:AT1G10640-GFP [16].

To create the promoter-GUS fusion reporter construct, the 2000 bp region upstream of the AT1G10640 CDS start site was amplified from Arabidopsis gDNA and recombined into the pMDC162 destination vector to generate AT1G10640<sub>pro</sub>:GUS [16].
To create the constitutive overexpression construct, RNA was extracted from wild-type Arabidopsis using a Plant RNA Kit (Omega Bio-Tek). Samples were treated with RNase-free DNase I (NEB) on a column to remove genomic DNA. The purified RNA was used as a template to make cDNA using qScript cDNA SuperMix (Quanta Biosciences), and the cDNA was used as a PCR template to amplify the CDS of AT1G10640, which was recombined into the pEarleyGate104 destination vector to generate 35S:YFP-AT1G10640CDS [17].

**Assembling amiRNA and CRISPR-Cas9 constructs**

For each designed amiRNA, the P-SAMS program also provides a pair of overlapping oligonucleotide sequences that can be annealed to create double-stranded DNA fragments with 5' overhangs ("sticky ends") suitable for ligating into plasmids also designed in the Carrington lab [11]. These oligonucleotides were ordered from Integrated DNA Technologies. The pro-amiRNA oligonucleotides were annealed as previously described [11].

The plasmid combined with the annealed oligonucleotides to create the amiRNA constructs was pMDC32B-AtMIR390a-B/c (Addgene #51776). The Type IIS restriction enzyme Bsal does not cleave within its recognition site, which means that different Bsal restriction sites can create different hanging ends. In pMDC32B-AtMIR390a-B/c Bsal cleaves two sites, creating two different 5' overhangs (5'-TACA-3' and 5'-CATT-3') complementary to the sticky ends of the annealed oligonucleotides and allowing directional insertion. In theory the Bsal digestion and
the T4 ligation can be performed in one reaction, but in practice we found that it worked better to digest and gel purify the plasmid separately before ligation. The ligation reaction was set up as follows:

- Digested pMDC32B-AtmiR390a-B/c (87.7 ug/µL) 1 µL
- Diluted annealed oligonucleotides (0.15 µM) 1 µL
- 10x T4 DNA ligase buffer 2 µL
- T4 DNA ligase (400 U/µL) 1 µL
- Sterile milliQ water 15 µL

Total volume 20 µL

5 minutes at 22°C
8 hours at 16°C
10 minutes at 22°C
11 minutes at 65°C
Hold at 4°C

The plasmid used to create the CRISPR-Cas9 constructs was pKSE401 (Addgene #62202) [7]. Like the plasmid used for the amiRNA constructs, pKSE401 has two BsaI restriction sites that release a linear backbone with sticky ends that can be ligated to annealed oligonucleotides. In this case the oligonucleotides were simply the 20-nucleotide spacer sequence of the guide RNA and its complement, with the four-base overhangs necessary to create sticky ends. The "forward" oligonucleotides had a 5' ATTG addition, and the "reverse" oligonucleotides had a 5' AAAC addition. These were ordered and annealed as described above and ligated to the pKSE401 backbone.

The ligation products were transformed into chemically competent Top10 E. coli, purified, sequenced to confirm that the ligation occurred correctly, and then transformed into chemically competent GV3101 Agrobacterium tumefaciens.
Glycerol stocks of the *Agrobacterium* cultures used to transform Arabidopsis were stored at -80°C.

The oligonucleotide sequences used to prepare the CRISPR-Cas9 and amiRNA constructs are listed in Supplemental Files Cas9_sequences.csv and amiRNA_sequences.csv.

**Plant transformation**

Adult Arabidopsis plants that had bolted and produced at least two inflorescence meristems were transformed with *Agrobacterium tumefaciens* strain GV3101 using the floral dip method [18]. Positive transformants were selected on ½MS plates containing 5 μM MSO, 50 μg/mL kanamycin, or 25 μg/mL hygromycin (Omega Scientific).

**Heterologous AT1G10640 expression**

After unsuccessfully attempting to produce heterologously expressed AT1G10640 in *E. coli* as was previously described for the polygalacturonase PGX1, we hypothesized that we might have more success in a eukaryotic system [2]. We ordered the EasySelect™ *Pichia* Expression Kit (ThermoFisher Scientific) and designed two expression constructs, one for intracellular heterologous expression and one for secreted expression. The intracellular expression construct was created by inserting the full AT1G10640 CDS (amplified from cDNA) into the pPICZ B
plasmid Multiple Cloning Site using the restriction enzymes XhoI and NotI. The
secretion expression construct was created using the restriction enzymes KpnI and
SacII to combine the plasmid pPICZα A with a truncated version of the AT1G10640
CDS that excluded a 21 amino acid signal peptide sequence that was predicted by
the SignalP 4.0 webtool to be post-translationally cleaved [19]. Insertion sequences
were confirmed at the Penn State Sequencing Facility.

Constructs were transformed by heat shock into chemically competent cells
of *Picha* strains X-33 (a wild-type strain) and KM71H (a mutant strain that grows
slowly in methanol media) as described in the EasySelect™ manual. Positive
transformants were selected on YPDS medium plates with 100 μg/mL Zeocin.
Individual transformed colonies were used to inoculate 25 mL liquid cultures of
BMGY medium in 250 mL baffled flasks to grow overnight in a shaking incubator at
28°C, 250 rpm. The cultures were harvested by centrifugation at 1500 g for 5
minutes. Supernatant was poured off and the cell pellet was resuspended in BMMY
medium to induce expression. Transformed KM71H strains were resuspended in 20
mL of BMMY medium in 250 mL baffled flasks, while X-33 strains were resuspended
in 90 mL of BMMY medium in 1 L baffled flasks. These cultures were grown for 8
days, with sterile methanol added every day to 0.5% by volume. On day 8 the
cultures were harvested by centrifugation in 50 mL Falcon tubes at 1500 g for 5
minutes. The supernatants were transferred to separate tubes, and pellets and
supernatants were frozen in liquid nitrogen and stored at -80°C.
Polygalacturonase activity assay

*Pichia* cell pellets were resuspended in 1/3 pellet volume of Breaking Buffer (described in the EasySelect™ manual). From each suspension, 0.5 mL was transferred to a 2.0 mL microcentrifuge tube. Approximately 0.5 mL of 0.5 mm acid-washed glass beads were added to each tube. Cells were lysed by alternating between vortexing at max speed for 30 seconds and placing on ice for 30 seconds. This was repeated 10 times. In contrast to collection of highly soluble protein, cell slurry was pipetted out of the tube without centrifuging first. Each slurry sample was divided into two microcentrifuge tubes, one of which was placed in boiling water for ten minutes. Boiled cell samples were then vortexed at max speed until protein aggregates were resuspended in liquid, and then briefly centrifuged to collect the sample at the bottom of the tube.

Polygalacturonase activity assays were based on [20]. For each sample, two polygalacturonic acid (PGA) digestions were performed. One reaction contained 20 μL of cell slurry and 180 μL of PGA solution (37.5 mM Na-acetate pH 4.4, 0.2% w/v PGA) and the other contained 50 μL of cell slurry and 150 μL of PGA solution. These were incubated at 30°C for 2 hours. After incubation, 1 mL of cold borate buffer (pH 9.0) and 200 μL of 1% 2-cyanoacetamide solution were added to each sample, and the tubes were placed in boiling water for 10 minutes. After cooling to room temperature, absorbance values of each sample were measured using a quartz cuvette in a NanoDrop 2000c spectrophotometer. The absorbance peak of 2-
cyanoacetamide-labelled reducing ends of PGA has been previously reported at 274 nm, but the peak on our instrument for un-digested PGA was 266 nm [20].

Acknowledgements

Thanks to Chaowen Xiao and Mark Signs for help with experimental design. Thanks to Bastian Minkenberg for providing advice on CRISPR-Cas9 experimental design. Thanks to Seth Polydore for providing pMDC32B-AtmiR390a-B/c and advice on amiRNA experimental design. Thanks to Igor Andzhievsky for help with cloning and plant transformation.

References


Chapter 4

Insights into the cell wall of the moss *Physcomitrella patens*

The cell walls of *Physcomitrella patens* contain many of the same components as the primary cell walls of eudicots: cellulose, callose, xyloglucan, xylan, mannan, and most if not all of the pectin domains [1, 2]. There is no evidence for the presence of xylogalacturonan or mixed-linkage glucans, and the presence of RG-II has not been detected, although the release of boron and certain sugars by digestion suggest the presence of an RG-II-like polymer [3]. *Physcomitrella* lacks lignin, a hallmark of vascular plants, and thus it does not form canonical secondary cell walls. The filamentous protonemal cells do not have a cuticle, whereas the leafy gametophore does have a waxy external barrier [4, 5]. As a plant with a simple body plan and a well-developed molecular toolkit, *Physcomitrella* is a rapidly developing research system for cell walls [1, 2, 6–9]. Furthermore, the common ancestor to bryophytes and tracheophytes was also the common ancestor to all extant embryophytes. Comparisons between moss cell walls and those of "higher" plants is one of our best avenues for insights into the evolutionary history of the plant cell wall during terrestrialization. In collaboration with other researchers, I performed a variety of experiments related to *Physcomitrella* cell wall synthesis and assembly. The role of each collaborator is described at the end of this chapter.
Experimental background and motivation

De novo cell wall formation in *Physcomitrella* protoplasts

Researchers in the past have used microscopy and biochemical methods to analyze the cell wall material regenerated by protoplasts from various plant species [10–14]. Complete wall regeneration from protoplasts is possible in several plant species [14–17], but is complicated by the need for liquid suspension or embedding in alginate. *Physcomitrella* protoplasts are capable of regenerating walls and growing normally on plates of agar-based medium, making them far more convenient to work with than protoplasts from other plants. We were interested in whether polysaccharide-specific dyes could be used to track the regeneration of cell wall material by protoplasts over time and the order in which materials were incorporated into the new wall. While our plans to use this for Fluorescence-Activated Cell Sorting have not yet borne fruit, an initial test of the idea showed promise.

*Physcomitrella* CESA dynamics

Because Cellulose Synthase Complex (CSC) motility along the plasma membrane is generally believed to be driven by the extrusion of synthesized cellulose, CSC movement is often used as a proxy for measuring the rate of cellulose synthesis [18, 19]. Fluorescent protein-labeled CSCs can be tracked with
fluorescence microscopy, and changes in CSC dynamics are now widely used in analysis of mutations and chemicals that affect the function and synthesis of the cell wall [18–20]. In Arabidopsis, the drugs 2,6-dichlorobenzonitrile (DCB) and isoxaben are both cellulose synthesis inhibitors, but their modes of action are not the same [18, 21]. DCB appears to inhibit CSC motility and cause hyperaccumulation of Cellulose Synthases (CESAs) at the plasma membrane, whereas isoxaben causes the rapid disappearance of CSCs from the membrane [18, 21]. Point mutations in multiple Arabidopsis CESAs have been shown to confer resistance to the effects of isoxaben, suggesting that isoxaben interacts specifically with the CSC [22, 23]. Given the evolutionary distance between Physcomitrella and Arabidopsis CESAs, we hypothesized that chemicals that inhibit cellulose synthesis in Arabidopsis might have different effects in Physcomitrella. As part of an effort to compare the function of cellulose in the walls of Physcomitrella to those of embryophytes, we collaborated with the Roberts Lab and the Bezanilla Lab to examine the dynamics of Green Fluorescent Protein (GFP)-labeled Physcomitrella patens CESAs at the plasma membrane.

**Arabidopsis secondary CESA knockout rescue by Physcomitrella and chimeric CESAs**

In Arabidopsis primary cell walls, cellulose is produced by CSCs containing the Cellulose Synthase subunits AtCESA1, AtCESA3, and AtCESA6 (AtCESA2, AtCESA5, and AtCESA9 are partially redundant with AtCESA6), whereas secondary
cell wall CSCs comprise AtCESA4, AtCESA7, and AtCESA8 [20, 25, 26]. However, it has been demonstrated that primary and secondary CESAs can interact with each other in vitro, and furthermore, some can form functional CSCs consisting of primary and secondary CESAs together in mutant plants [27, 28]. Analysis of cellulose synthase sequences has led to recognition of regions conserved within CESA clades and across all the cellulose synthases, and has identified amino acid residues that are potentially significant to catalytic activity and CSC assembly [29]. Wang et al. performed experiments with chimeric primary CESAs to elucidate the function of various CESA domains, but plant CESA structures have not yet been experimentally determined and CSC assembly is still largely a mystery [30].

In order to learn more about the regions of CESA proteins responsible for CSC assembly and protein-protein interaction, Joseph Hill, of Ming Tien's research group at Penn State, prepared chimeric protein constructs consisting of different regions of the secondary cell wall Arabidopsis CESAs. The chimeric proteins were designed by dividing the AtCESA sequences into three regions: the N-terminus, containing the beginning of the protein sequence to the end of the second putative trans-membrane domain; the Central Domain, consisting of the region between the second and third putative trans-membrane domains; and the C-terminus, spanning from the beginning of the third putative trans-membrane domain to the end of the protein (Figure 4-1).
Permutations of these regions from each pair of secondary wall AtCESAs were combined to create chimeric protein sequences. Notation for the chimeric proteins takes the form of three digits, signifying from which AtCESA each region of the chimera came. The first digit describes the N-terminus, the second digit describes the Central Domain, and the third digit describes the C-terminus. Thus, Chimera 484 consists of the AtCESA4 N-terminus, the AtCESA8 central domain, and the AtCESA4 C-terminus.

These chimeric protein-coding sequences, along with the coding sequences of the Physcomitrella CESAs, were combined with each of the Arabidopsis secondary wall CESA promoters and transformed into their respective knockout lines, which display growth and vascular defects [27]. Some of the chimeric proteins and some of the Physcomitrella CESAs were able to complement the knockout phenotypes (Table
4-1). CESA sequences were aligned and compared in order to make inferences about the amino acid residues involved in clade-specific CESA functions.

Table 4-1: Complementation of Arabidopsis secondary CESA knockouts

<table>
<thead>
<tr>
<th>AtCesA4 knockout is rescued by:</th>
<th>AtCesA7 knockout is rescued by:</th>
<th>AtCesA8 knockout is rescued by:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chimera 744</td>
<td>Chimera 747</td>
<td>Chimera 488</td>
</tr>
<tr>
<td>AtCesA6</td>
<td>PpCesA7</td>
<td>Chimera 887</td>
</tr>
<tr>
<td>PpCesA5</td>
<td>PpCesA10</td>
<td>Chimera 788</td>
</tr>
<tr>
<td>PpCesA8</td>
<td></td>
<td>Chimera 787</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AtCesA1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpCesA7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PpCesA10</td>
</tr>
</tbody>
</table>

In most cases, lack of complementation cannot be taken as definitive evidence that the missing chimeric protein is incapable of complementing the knockout. Correct folding and expression of the chimeric proteins can only be inferred in the cases in which complementation was achieved.

Results

Protoplast regeneration time course

A time course experiment of regenerating *Physcomitrella* protoplasts was performed using the cell wall dyes Calcofluor White, which binds to both callose and cellulose, and Pontamine Fast Scarlet S4B, which binds specifically to cellulose, to assess the reappearance of regenerating wall [31, 32]. Protoplasts from chloronemal tissue were released from their existing walls with Driselase, a wall-degrading enzyme mixture, and distributed on several plates containing cellophane over solid regeneration medium. Every 12 hours the cells were rinsed off of a plate and resuspended in PRML liquid medium (suitable for protecting *Physcomitrella*
protoplasts from osmotic damage) containing 0.001% Calcofluor White and 0.01% S4B [8]. The cells were allowed to stain for 30 minutes and were then examined on a confocal microscope. Immediately after digestion, the protoplasts showed no fluorescence other than the autofluorescence of the chloroplasts stimulated by the 561 nm laser used to image S4B. The appearance of Calcofluor White staining was fast, showing up within two hours of the protoplasts leaving the enzyme solution. This agrees with previous analyses of the recovering Physcomitrella protoplast, where the authors saw fibrous material on the protoplast surface within one hour [33]. By the 12-hour time point, both Calcofluor and S4B staining were detectable, and the following time points saw an increase in incidence of S4B staining as well as cell division and elongation (Figure 4-2). This progression is shown in Figure 4-3.
Figure 4-2: Generation of de novo cell wall by *Physcomitrella* protoplasts

Twelve hours after release from its previous cell wall, a protoplast had a layer of fibrous material that stained with Calcofluor White but not with S4B, suggesting that it contains callose but not substantial amounts of cellulose. By 60 hours after protoplasting, most of the cells had walls that primarily stained with S4B, and many had divided and begun tip growth. Scale bars 10 microns.
Figure 4-3: Physcomitrella protoplast regeneration time course

Calcofluor White fluorescence can be detected at the cell periphery as quickly as two hours after protoplast release. Significant S4B-binding appears to first occur between 12 and 24 hours of recovery. The frequency of dividing cells increases steadily after 36 hours.

GFP-labeled Physcomitrella CESA velocities

Mai Tran, of Alison Roberts’ research group at the University of Rhode Island, produced Physcomitrella lines containing GFP-labeled PpCESA5 and PpCESA8 in their respective knockout backgrounds. Because of my experience with fluorescence microscopy of Physcomitrella tissue, we agreed to collaborate on imaging the GFP-PpCESAs in vivo. Unfortunately, spinning disk confocal microscopy turned out to not be effective for collecting high-quality time-lapses of Physcomitrella CESAs, which appear to be far less common at the plasma membrane than GFP-AtCESAs in Arabidopsis [20, 24]. We collaborated with Magdalena Bezanilla and Shu-Zon Wu of the Bezanilla research group at the University of Massachusetts, who use Variable Angle Epifluorescence Microscopy (VAEM) to produce high-quality images of Physcomitrella colonies. I visited their lab to collect initial time-lapse images and
show Dr. Wu how to find the GFP-PpCESA8-containing CSCs. After my visit she collected additional time-lapses and sent them to me for processing.

\textit{PpCESA5}

\emph{Physcomitrella patens} CESA5 is the only \textit{PpCESA} known to have a distinct phenotype (inhibition of gametophore development) in the single knockout mutant \cite{34, 35}. Measurement of fluorescent particle dynamics at the plasma membrane showed an average speed of 262 nm/min with a standard deviation of 80.0 nm/min (Figure 4-4). As in reports of FP-labeled Arabidopsis CESAs, treatment with DCB lowered CSC velocity to nearly zero \cite{21}. However, unlike in Arabidopsis, there was no hyperaccumulation of CSCs at the membrane: CSC density was not detectably different after treatment. Also in contrast to Arabidopsis, isoxaben treatment did not cause the disappearance of labeled CESAs from the plasma membrane. Like DCB, isoxaben caused GFP-PpCESA5 velocity to drop to near zero but left fluorescent CSC particle density the same (Figure 4-5).
Figure 4-4: GFP-PpCESA5 velocities

Velocity distribution of CSCs containing GFP-PpCESA5. Mean and standard deviation in nm/min were: 262, 80.0 (Control); 18.3, 37.2 (20 μM DCB); and 36.3, 66.5 (20 μM Isoxaben).

Figure 4-5: GFP-PpCESA5 at the cell surface

A) Control protonemal cell treated with 0.05% ethanol. Mean CSC density 0.484, s.d. 0.106 (particles/μm²), N=8 time-lapses. B) Protonemal cell treated with 20 μM DCB. Mean CSC density 0.512, s.d. 0.0927 (particles/μm²), N=8 time-lapses. C) Protonemal cell treated with 20 μM isoxaben. Mean CSC density 0.442, s.d. 0.115 (particles/μm²), N=8 time-lapses. Scale bars = 5 μm.

PpCESA8

Time-lapses of GFP-PpCESA8 were taken on two separate days. Velocities, with the control sample data from each day pooled, are shown in Figure 4-6A. The
problem with combining CSC velocity data like this is shown in Figure 4-6B: changes in imaging conditions can have a profound effect on CSC velocity. This phenomenon has been reported before: a temperature change of 8°C on the microscope stage (from 21°C to 29°C, a not unreasonable range during a long session in laser light) was found to quadruple CSC speed [36]. The clear difference in populations between Day 1 and Day 2 (Welch’s t-test p-value <2.2x10^{-16}) demonstrates the importance of comparing CSC velocity data to a control group from as similar conditions as possible, and the unfeasibility of comparing CSC velocity data between experiments, let alone between species.

![Figure 4-6: Pooled GFP-PpCESA8 velocity experiments](image)

**A)** GFP-PpCESA8 velocities measured over two days. **B)** The control group samples between two days are overlapping but distinct populations.

Separating the data by the day they were collected allows for more meaningful inferences (Figure 4-7). Treatment with 20 μM DCB and 20 μM isoxaben affected GFP-PpCESA8-containing CSC velocities in much the same way as it affected GFP-PpCESA5, although the reduction in speed was not as complete: unlike the GFP-
PpCESA5 line, the majority of particles in the inhibitor treatments were still moving faster than 25 nm/min. Surprisingly, increasing the concentration of DCB to 40 μM did not magnify its effect on CSC velocity: average velocity was higher in the 40 μM DCB treatment than in the 20 μM treatment. It is difficult to draw any conclusion from this because the average velocity of particles in the control treatment was also higher on the day that the 40 μM DCB data were collected.

![Figure 4-7: GFP-PpCESA8 velocities after treatments](image)

**A)** Day 1 treatments. Mean and standard deviation in nm/min were: 253, 79.0 (Control); 47.32, 38.3 (20 μM DCB); and 42.6, 40.4 (20 μM Isoxaben). **B)** Day 2 treatments. Mean and standard deviation in nm/min were: 364, 55.4 (Control) and 114, 106 (40 μM DCB).

Unlike in the experiments with GFP-PpCESA5, isoxaben treatment did appear to have some effect the density of GFP-PpCESA8-containing particles at the plasma membrane, in that fluorescent CSCs were harder to find and measure, possibly suggesting that the presence of isoxaben was increasing the rate of CSC endocytosis, albeit to a lesser degree than in Arabidopsis. Unfortunately, the number of cells imaged was too low to give statistical analysis of density sufficient power (Figure 4-
Additionally, we were forced to measure particle density in the GFP-PpCESA8 images in a way that excluded more particles from the counts than in GFP-PpCESA5, so comparisons in density between them are not meaningful (see Methods section).

**Figure 4-8: GFP-PpCESA8 at the cell surface**

**A)** Control protonemal cell treated with 0.05% ethanol. Mean CSC density 0.201, s.d. 0.0912 (particles/μm²), N=11 time-lapses. Density measurements from control data were pooled because no statistical difference between them was detected. **B)** Protonemal cell treated with 20 μM DCB. Mean CSC density 0.149, s.d. 0.117 (particles/μm²), N=5 time-lapses. **C)** Protonemal cell treated with 20 μM isoxaben. Mean CSC density 0.0691, s.d. 0.930 (particles/μm²), N=5 time-lapses. **D)** Protonemal cell treated with 40 μM DCB. Mean CSC density 0.243, s.d. 0.231 (particles/μm²), N=3 time-lapses. Scale bars = 5 μm.
Alignments between CESA protein sequences

The rescuing CESA protein sequences were aligned with the sequence of the knocked out Arabidopsis CESA (Supplemental Files). In most cases, a lack of rescue by a particular chimeric protein is not very informative because we cannot confirm whether that particular chimeric protein can fold into a viable cellulose synthase. However, we can infer that a chimeric construct is viable if it can rescue at least one knockout mutant, and draw some conclusions by comparing its sequence to the CESA it fails to complement.

The knockout of AtCESA7 was rescued by the 747 chimeric CESA construct, but not by the 787 construct that was able to rescue the AtCESA8 knockout. This suggests that residues shared by the AtCESA7 and AtCESA4 Central Domains, but missing or different in the AtCESA8 Central Domain, are necessary for the function of AtCESA7. The sequences of AtCESA7, Chimera 747, PpCESA7, and PpCESA10 (all of which rescue the AtCESA7 knockout) were aligned with Chimera 787 (Supplemental Files). Amino acid residues were identified as being potentially important for AtCESA7 function when they were perfectly conserved among all of the complementary sequences but different in Chimera 787 (Figure 4-9). Once identified, the most likely candidates were selected based on differences in polarity and hydrophobicity between the AtCESA7 residue and the AtCESA8 alternative (Table 4-2). The chimeric 744 which rescues the AtCESA4 knockout also failed to rescue the AtCESA7 knockout, and was used to identify key residues in the C-terminus of AtCESA7.
Potentially important amino acids are discovered by comparing the protein sequences that can rescue a knockout mutant with a sequence that cannot. Residues that are the same in all of the complementary sequences but differed in the non-complementary chimera are marked by orange boxes.

Similarly, the AtCESA4 knockout is rescued by Chimera 744 but not Chimera 747, suggesting that some necessary function of AtCESA4 is conserved in its C-terminus. The AtCESA8 knockout is rescued by Chimera 787 and several other chimeric proteins, all of which contain the AtCESA8 Central Domain, but not by Chimera 747. Potentially important residues in the C-terminus of AtCESA4 and the Central Domain of AtCESA8 were identified in the manner described above, using the sequence of Chimera 747 for comparison (Table 4-2).
### Table 4-2: Arabidopsis secondary cell wall CESA residues of interest

<table>
<thead>
<tr>
<th>ATCESA4</th>
<th>ATCESA7</th>
<th>ATCESA8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residue of Interest</td>
<td>Chimera 747</td>
<td>Residue of Interest</td>
</tr>
<tr>
<td>G848</td>
<td>D871</td>
<td>I282</td>
</tr>
<tr>
<td>N857</td>
<td>T880</td>
<td>L298</td>
</tr>
<tr>
<td>V900</td>
<td>I923</td>
<td>P316</td>
</tr>
<tr>
<td>F908</td>
<td>V931</td>
<td>S361</td>
</tr>
<tr>
<td>V914</td>
<td>I937</td>
<td>S416</td>
</tr>
<tr>
<td>A987</td>
<td>S1008</td>
<td>E373</td>
</tr>
<tr>
<td>I1017</td>
<td>V1038</td>
<td>T410</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E421</td>
</tr>
<tr>
<td></td>
<td></td>
<td>V427</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P451</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M504</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K549</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y553</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N566</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I579</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C597</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P625</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K655</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M659</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N663</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K791</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E796</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L814</td>
</tr>
</tbody>
</table>

Amino acids that are completely conserved in all sequences that can rescue the knockout and the alternate residues present in the non-complementary chimeric sequence. Bolded residues are of special interest because substitution for the non-complementary form would cause a change in charge or polarity.
Discussion

**Glucan-labeling fluorescent dyes can be used to track the regeneration of Physcomitrella protoplast cell walls**

While biochemical analyses of regenerating protoplast cell walls of various plant species tend to report the presence of β-linked glucans, different researchers have reported different identities of the polysaccharide making up the recovered wall in their experimental system [13]. Cellulose, callose, non-microfibrillar 1,4-linked glucan, and mixtures of 1,3-linked and 1,4-linked glucans have all been reported in various plant species [10, 13–15, 37–39]. In this experiment, Calcofluor White fluorescence was seen over 12 hours before the presence of S4B, suggesting that callose is the first glucan to be delivered to the nascent *Physcomitrella* wall. Callose is associated both with wound-response and the generation of a new cell wall during cytokinesis, so it is not clear if the early presence of callose is related to the trauma of protoplasting or is necessary for the production of the new cell wall [9, 40]. The growing presence of S4B labeling after 24 hours of recovery indicates that cellulose is also deposited in the new wall, and the increasing frequency of cell division and elongation suggests that the protoplast is able to regenerate a wall that is sufficiently similar to the original to allow for normal cell growth. This time-lapse experiment was performed to gather preliminary data supporting the viability of labeling recovering protoplasts with fluorescent dyes. The eventual goal was to use Fluorescence Activated Cell Sorting experiments to track the presence of various cell wall components in the regenerating wall over time. Collaborators in the Roberts
Lab recently published similar experiments using immunolabeling that agree with this experiment’s detection of cellulose and callose in the walls of 24-hour post-protoplasting cells [9].

*Physcomitrella* CESA proteins do not respond to cellulose synthesis inhibitors in the same way as Arabidopsis CESAs

The data from the GFP-PpCESA imaging experiments suggest that while the cellulose synthesis inhibitors DCB and isoxaben do affect *Physcomitrella* CSC dynamics, they may not have the same effects as in Arabidopsis. *Physcomitrella* CSCs may be reduced at the cell membrane in the presence of isoxaben, but if so, not to the extent of complete loss. Both DCB and isoxaben significantly slow the speed of CSCs in *Physcomitrella*, but at least in the case of GFP-PpCESA8-containing CSCs, significant movement still occurs. The fact that GFP-PpCESA5 and GFP-PpCESA8 appear to respond differently to these inhibitors may be relevant to the open question of the nature of *Physcomitrella* CESA assembly into CSCs. Unlike in Arabidopsis, it has not yet been established whether *Physcomitrella* CSCs require multiple CESA isoforms to assemble and function [35]. It is possible that PpCESA5 and PpCESA8 are never part of the same CSC assembly, and that the difference in degree of slowing in response to the inhibitors reflects differences in their activity at the plasma membrane.

The next steps for this project are to collect more time lapses to provide more statistical power, and to apply a range of cellulose synthase inhibitor
concentrations to establish a dose-response curve. The Roberts Lab is also analyzing *Physcomitrella* colony macroscopic response to the application of the inhibitors.

**CESA complementation experiments suggest key amino acid residues involved in clade-specific protein behavior**

Complementing Arabidopsis CESA knockouts with chimeric proteins and CESA s from other plant species offers us a powerful method for teasing out key residues necessary for specific function and CSC assembly in each CESA. It is somewhat surprising that *Physcomitrella* CESA proteins are able to rescue Arabidopsis knockout mutants, given their evolutionary distance and the fact that none of them are direct orthologs of any of the embryophyte CESAs [7]. The apparent redundancy of most *Physcomitrella* CESA s makes characterizing them individually more difficult, but we may be able to draw insights about their functional relationships with each other from their patterns of complementation in other species [35]. Hopefully future experiments will include attempts to complement Arabidopsis primary CESA mutants as well as CESA mutants in other tracheophyte species with *Physcomitrella* CESAs.

A limitation to this experiment is the necessity of inferring conclusions from negative data. As complementation experiments with CESAs from other species are added to the literature, support for some of the residues of interest being functionally conserved will grow. Other residues may be eliminated from the list when they are not present in new complementary sequences. Some of the strongest
support for our conclusions could come from experiments using orthologous CESAs from additional plant species, including some that have different residues at the putatively conserved locations.

With just the complementation data we have now we were able to generate a short list of amino acids in AtCESA4, AtCESA7, and AtCESA8 that would make good candidates for targeted mutagenesis experiments. An idealized experiment would show that changing just one particular amino acid would prevent an Arabidopsis CESA from rescuing its knockout mutant, but it is likely that multiple residues are involved in CESA interaction and CSC formation and that any single change might not be enough to completely destabilize the structure. Nevertheless, these results are a promising step on the path to more complete models of CESA function. As 3D models of CESA structure improve, we hope to discover correlations between the charged residues of interest with each other and with charged residues on other CESAs that are predicted to be near each other in the CSC [41, 42]. One proposed future experiment to definitively establish the role of some of these residues in CESA interactions is swapping two oppositely charged residues between interacting CESAs and demonstrating that the modified proteins can only form functional CSCs with each other.
Methods

Plant materials and growth

*Physcomitrella patens* colonies were grown in petri dishes on cellophane disks over BCDAT medium in a chamber at 22°C under 24-hour light [8]. The Gransden 2009 *Physcomitrella* ecotype was used for protoplast regeneration experiments. The GFP-PpCESA5 and GFP-PpCESA8 lines provided by the Roberts Lab were in knockout backgrounds of PpCESA5 and PpCESA8, respectively.

*Physcomitrella protoplast regeneration time course*

*Physcomitrella patens* protoplasts were generated from protonemal tissue grown on BCDAT medium as previously described [8]. Protoplasts were resuspended in PRML liquid media at a concentration of approximately $2 \times 10^4$ protoplasts/mL as measured by hemocytometer. One mL of suspension was added to each of ten petri plates containing cellophane disks over PRMB solid regeneration medium.

At time 0 and every 12 hr thereafter, the protoplasts on one of the plates were rinsed off with 4 mL of 8.5% mannitol solution and transferred into a 15 mL tube using a 1 mL pipette tip with the opening widened by cutting off the end with a razor blade. The protoplasts were gently pelleted by centrifugation for 7 minutes at 30 g. The supernatant was poured off and the protoplasts were resuspended in 900
μL of staining solution (PRML with 0.01% S4B and 0.001% calcofluor white). Time point 2.5 hours is the result of examining the protoplasts from time point zero 2.5 hours after resuspending in staining solution.

In the two hours following resuspension of the protoplasts at each time point, several slides were prepared by applying approximately 40 μL of the suspension between four dabs of vacuum grease on a glass microscope slide. A cover slip was gently applied over the drop and the grease to avoid crushing the recovering cells. Slides were searched in brightfield mode for healthy-looking cells (spherical, even distribution of chloroplasts, cytoplasmic streaming). At later time points, elongating and divided cells were also counted as healthy. Once found, healthy cells were examined for the blue fluorescence indicating Calcofluor White binding and the red fluorescence of S4B using a Zeiss Axio Observer microscope attached to a Yokogawa CSU-X1 spinning disk head with a 100× 1.4 numerical aperture immersion oil objective. All images were captured using the AxioVision 4.8 software (Carl Zeiss). A 405 nm excitation laser and a 425-475 nm emission filter were used to detect Calcofluor white; a 561 nm excitation laser and a 581-654 nm emission filter were used for S4B.

**Physcomitrella patens GFP-CESA time-lapses**

*Physcomitrella* colonies were grown from protoplasts for seven or eight days on cellophane disks over PRML medium [8]. To maintain the flat, evenly spread shapes of the colonies for imaging, agar pads were prepared on slides by placing
100 mL of molten 1% agar in Hoagland’s solution onto the surface of a slide, then quickly covering it with a second slide. After the agar hardened, the upper slide was removed and the excess agar trimmed with a razor blade. A square of cellophane surrounding one or more Physcomitrella colonies was cut and lifted away from the plate with a razor blade, and placed colony-side down on the agar pad. The cellophane was pulled parallel to the slide surface, leaving the colonies in place on the agar pad. Twenty to 30 microliters of Hoagland’s solution was added on top of the colonies, and then gently covered with a cover slip [43]. Excess liquid was absorbed with a Kimwipe, and the edges of the cover slip were sealed with Valap (1:1:1 parts of Vaseline, lanoline and paraffin) using a hot glass pipette.

After a slide was prepared, it was promptly moved to the microscope. A Nikon Eclipse Ti microscope with a 100X 1.49 NA TIRF objective and Andor DU-897 EMCCD camera was used to capture images every two seconds to create time-lapse videos of GFP-tagged PpCESA5 and PpCESA8 at the plasma membrane. Although Variable Angle Epifluorescence Microscopy imaging allowed for much clearer images of Physcomitrella CSCs than spinning disk confocal, they were not distinct enough for particle tracking software to be able to automatically process them.

**Treatment with cellulose synthesis inhibitors**

Ethanol was used as a solvent for DCB and isoxaben with stock concentrations of 40 mM. *Physcomitrella* colonies were saturated with 100 μL of 20 μM DCB, 40 μM DCB, or 20 μM isoxaben in PNO3 liquid medium. For negative
controls, colonies were treated with 100 μL of 0.05% ethanol in PNO3 liquid medium. Slides were prepared as above after 20 to 25 minutes of exposure to inhibitors.

**Measuring CSC velocities**

Time-lapse files were opened in Fiji, image-processing software based on ImageJ [44]. Image contrast was normalized to improve visibility of particles. CSC velocity was measured by identifying each particle of the expected CSC size and brightness within the first slice of a time lapse (time 0 seconds), measuring its position, tracking it until it was no longer visible, and measuring its last position. The displacement of the particle and the time necessary for travel were calculated to yield velocity measurements for each particle. Particles that did not persist for at least one minute were ignored, as were particles whose size, speed, or brightness identified them as CESA-containing vesicles rather than CSCs. This process of identifying particles that appeared to be CSCs was repeated at slices corresponding to increments of 2.5 minutes, and particles that were identified were traced backwards and forwards through the time lapse so that the beginning and end points would correspond to the first and last appearance of the identified particles. This was done to minimize bias in the experimenter toward particles that moved at “normal” CSC speeds. Raw measurement data are available in Supplemental Files.
Measuring CSC density

To estimate CSC density in the GFP-PpCESA5 protonemal cells, time-lapse files were opened in Fiji and a Region of Interest (ROI) was selected with the freehand tool and its area measured with the measure function of Fiji. The Particle Detector plugin was used to detect fluorescent CSCs with the following settings: 2 pixel radius, 0 cutoff, 1.9 percentile. The Particle Analysis Point Picker tool was used to select each particle within the ROI to acquire a count. Density was estimated as the number of particles detected by the Particle Detector plugin divided by the area of the ROI.

Unfortunately, the noise levels of the GFP-PpCESA8 were too high for the Particle Detector plugin to be able to reliably select CSC particles. Area of the cell was again determined using the freehand tool, but CSC count had to be approximated by the number of measured velocities from within the first slice of each time-lapse. This undercounted the number of CSCs because it ignored any that did not persist for 60 seconds or whose paths could not be tracked during velocity measurements. Raw measurement data are available in Supplemental Files.

Aligning CESA sequences

Protein sequences were aligned in Geneious version 7.6 using the MUSCLE alignment tool and manually curated [45, 46]. Potentially important residues were discovered and classified by manual examination of the alignments.
Collaborator Roles

Mai Tran produced the GFP-labeled PpCESA lines. Shu-Zon Wu captured the time-lapses of GFP-PpCESAs. Mai Tran, Thomas Wayne McCarthy, Magdalena Bezanilla, Shu-Zon Wu, and Alison Roberts contributed to the intellectual design of the imaging experiments. Thomas Wayne McCarthy processed the GFP-PpCESA time-lapses and analyzed the data.

Joseph Hill produced the chimeric CESA constructs and the complementation lines and analyzed the plants to identify successful complementation. Joseph Hill provided the protein sequences for analysis. Thomas Wayne McCarthy, Joseph Hill, and Charles T. Anderson contributed to the intellectual design of the alignment experiments. Thomas Wayne McCarthy aligned the sequences and detected the amino acid residues of interest.

References


30. Wang J, Howles PA, Cork AH, Birch RJ, Williamson RE: **Chimeric proteins suggest that the catalytic and/or C-terminal domains give CesA1 and CesA3 access to their specific sites in the cellulose synthase of primary walls.** *Plant Physiol* 2006, 142:685–95.


32. Anderson CT, Carroll A, Akhmetova L, Somerville CR: **Real-time imaging of cellulose reorientation during cell wall expansion in Arabidopsis roots.**


Chapter 5

Concluding thoughts

Where did pectins originate?

Chapter 2 contains the first published phylogenetic analyses of most of the known pectin-related gene families. Other researchers have published papers on the evolution of the GAUT, PG, and PME families, but analyzing all the known pectin-related gene families at one time, and including members from a moss and a lycophyte, allowed us to see patterns in the evolution of pectin-related genes in land plants [1–15]. We established that the divergence of several important families predated the divergence of mosses from the tracheophytes, but this raises new questions about the history of these families. Harholt et al. speculate that the land plant common ancestor was among the Charophycean Green Algae and that the cell wall evolved in response to terrestrialization pressures before the development of the Embryophytes [16]. Whether this is true or not, our results indicate that the history of pectin-related gene evolution predates the divergence of bryophytes, and the origins of the GAUT, PG, and PME genes remain unknown. Databases like PlantTribes are being updated as new plant genome sequences are published, and as more algal genomes become available the re-analysis of pectin-related gene families may reveal new facts about pectin evolutionary history and the development of its functions in ancient cell walls.
Did CSC assembly evolve in parallel in mosses and tracheophytes?

It has been previously established that *Physcomitrella* CESAs are evolutionarily very distant from those of Arabidopsis, radiating before the events leading to primary and secondary cell wall CESAs and the three clades within each [17, 18]. Despite this evolutionary distance, PpCESAs can rescue Arabidopsis secondary CESA knockout plants. Little is known about *Physcomitrella* CSC assembly, although the lack of detectable phenotype in most PpCESA knockout mutants suggests that if heteromeric *Physcomitrella* CSCs exist, there is more redundancy between PpCESAs than AtCESAs. The fact that each secondary AtCESA knockout could be rescued by two PpCESAs lends some credence to this hypothesis, but it does not answer the question of why *Physcomitrella* CESAs are able to rescue the mutants at all. This question is closely tied to the nature of CSC assembly, and our collaboration with Joseph Hill will contribute to understanding both the evolutionary history of CESAs and the nature of their extant interactions.

Why are some putative polygalacturonases so deeply conserved?

The following section is a description of experimental plans to use the transgenic material that we generated to study previously uncharacterized putative polygalacturonases. The deep conservation of these genes makes them good candidates for the discovery of new polygalacturonase functions.
Generating homozygous Cas9-induced mutants

Almost all of the first generation CRISPR-Cas9 mutant plants still carry a wild-type copy of the targeted gene, but if the germ-line cells of the T1 plants are heterozygous for the detected mutations, their progeny should include homozygous mutants. Table 4-2 shows lines with bolded mutations predicted to knock out protein production: these seed lines should be the first priority to be grown on kanamycin-containing ½ MS plates. Although Cas9-induced mutations are stable without the presence of the T-DNA insertion, selecting for its continued presence maximizes the probability of recovering plants without wild-type versions of the targeted gene. If "non-transgenic" lines lacking the CRISPR-Cas9 T-DNA insertion are desired, seeds can be germinated on non-selective plates and the presence or absence of the insert can be detected with PCR. Plants can be quickly screened for mutation homozygosity by using PCR to amplify the region surrounding the Cas9 site and digesting the PCR product with the appropriate restriction enzyme (Figure 4-3). In cases where the PCR product is not digested, homozygosity and the nature of the mutation can be confirmed by sequencing the PCR product. Unlike in T-DNA insertional mutants, RT-PCR will not be informative about the effectiveness of Cas9-induced mutations to knock out production of the targeted protein, because the mutant mRNA will still be present.

Given that the gene targets are very deeply conserved, mutant phenotypes seem likely. The current literature on Cas9-induced mutation has put the weight of evidence on having multiple independent mutant lines showing the same
phenotype, but if additional confirmation is desired the mutant plants can be rescued with a complementation construct using the promoter region and CDS of the target gene.

**Using amiRNA lines for essential genes**

It would not be surprising if one or more of the deeply conserved polygalacturonases is essential to plant development. If so, isolating homozygous knockout mutants will be unlikely. Dissecting the siliques and seeds of heterozygous plants would be informative as to the stage at which the presence of the conserved putative polygalacturonase is necessary for continued development, but such experiments will not be informative as to the function of the protein later in the plant’s development. Many of the amiRNA lines are likely to cause only partial silencing, and therefore will have the potential to survive to viability while still exhibiting phenotypes from lowered gene expression. Unlike the CRISPR-Cas9 mutants, amiRNA interference should be a dominant trait and therefore phenotypes may be visible in the T1 plants. However, selecting several T1 plants for each amiRNA construct and performing quantitative RT-PCR on T2 lines in order to find lines with relatively consistent reductions in mRNA between sibling plants will be necessary in order to quantify any detected phenotypes.
Characterizing AT1G10640

The Cas9-induced early stop-codon mutants should provide good information on the function of AT1G10640, either through phenotypes in the adult plant, or embryo lethality confirming that this deeply conserved polygalacturonase is essential. If AT1G10640 is an essential gene, the amiRNA lines will be used to assess the effect of lowered levels of the AT1G10640 protein on the plant. A complementation construct containing the genomic sequence of AT1G10640 and the 2000 bp promoter region upstream of the CDS start site has been prepared, and can be used if necessary to rescue AT1G10640 knockout mutants, but only if the CRISPR-Cas9 cassette has been crossed out of the line.

Promoter:GUS and putative overexpression lines have been prepared for AT1G10640, but data have not yet been collected on them. Genuine overexpression of AT1G10640 will be confirmed by quantitative RT-PCR in multiple lines in order to demonstrate that any detected phenotypes are the result of AT1G10640 overexpression and not, e.g., disruption of an unrelated gene by T-DNA insertion. Given that AT1G10640 is in the same clade as PGX1, the overexpression and knockout lines will be studied in much the same way that PGX1 was, looking for changes in cell and seedling elongation as well as changes in wall composition [19]. The promoter:GUS lines will be used to visualize the transcription pattern of AT1G10640.

Data on the heterologous expression of AT1G10640 in Pichia pastoris are only preliminary, but they suggest that at least one of the expression constructs led
to the production of insoluble protein with polygalacturonase activity. After replicating this result for confirmation, the next step will be to attempt to solubilize the protein so that its biochemical activity can be assayed. Because our attempts to Western blot *Pichia* proteins have been unsuccessful, we do not yet know if SDS solubilizes AT1G10640. If so, purification of the hexahistidine-tagged protein on a Nickel column and renaturation may allow for characterization of purified protein [20]. If SDS is not viable, other denaturants such as urea or guanidinium hydrochloride will be tried. Beyond confirming pectin-degrading activity using 2-cyanoacetamide to detect an increase in labeled reducing ends, the endo- or exo-polygalacturonase activity of AT1G10640 will be determined by High Performance Liquid Chromatography of the digested polygalacturonic acid [21]. To our knowledge no specific plant PG has been characterized as an endo- or exo-polygalacturonase.

References


3. Yin Y, Chen H, Hahn MG, Mohnen D, Xu Y: *Evolution and function of the plant cell wall synthesis-related glycosyltransferase family 8*. *Plant


Appendix A

Supplemental Figures

Figure 2-S1: Xylogalacturonan xylosyltransferase family tree.

Physcomitrella and Selaginella genes were not detected in this family.
**Figure 2-S2: Rhamnogalacturonan I arabinosyltransferase family tree.**

This tree contains no *Physcomitrella* members and two algal members, one from *Penium margaritaceum* and one from *Nitella hyalina.*
Figure 2-S3: Pectinmethylesterase inhibitor (PMEI) family tree.

This tree contains only *Arabidopsis* and *Medicago trunculata* members and likely does not represent the whole family.
Physcomitrella and Selaginella genes were not detected in this family. Monocot and eudicot family members are contained in separate clades that are well-resolved from each other.
**Figure 2-55: Pectate lyase-like (PLL) family tree.**

A small land plant-wide clade is resolved from the rest of the tree, indicating at least two genes in the common ancestor of land plants.
Figure 2-S6: Homogalacturonan methyltransferase family tree.

This tree consists of three monophyletic clades, two of which are land plant-wide. An algal root with reasonably homology was not detected for this gene family, preventing the determination of whether two or three ancestral genes were present in the common ancestor of land plants.
Figure 2-S7: UDP-Glucuronic acid epimerase family tree.

This family appears to be land plant-wide and is rooted by a gene from *C. reinhardtii*. However, the grouping of all the *Physcomitrella* genes into one monophyletic clade implies that there was only one family member in the common ancestor.
Not only is this family land plant-wide, it includes members from the algae *C. reinhardtii*, *Spirogyra pratensis*, and *Penium margaritaceum*, but the grouping of all the *Physcomitrella* genes into one monophyletic clade implies that there was only one family member in the common ancestor.
Figure 2-S9: Pectin acetyltransferase family tree.

This family appears to be land plant-wide and is rooted by a gene from *C. reinhardtii*. The grouping of all the *Physcomitrella* genes into one monophyletic clade implies that there was only one family member in the common ancestor.
Figure 2-S10: Pectin acetylesterase family tree.

This family contains only one *Physcomitrella* and no *Selaginella* members.
Figure 2-S11: Rhamnogalacturonan II xylosyltransferase family tree.

This family appears to be land plant-wide, with one member in the common ancestor of land plants. The algal root gene is from *Nitella hyalina*.
Figure 2-S12: β-1,4-Galactan β-1,4-Galactosyltransferase family tree.

This tree has no algal root. The *Physcomitrella* genes are grouped together in a well-supported clade separate from other species. There is no evidence for more than one gene in the common ancestor.
Figure 2-S13: GATL family tree.

This tree is poorly resolved, with no root and large polytomies. The *Physcomitrella* genes group together in one well-supported clade.
Appendix B

Supplemental File List

Table2S1.xls: Query Arabidopsis genes. A list of all the Arabidopsis genes used as queries to the PlantTribes 2.0 database and the sources for collecting them.

Table2S2.csv: Pectin-related genes. This table contains all of the genes examined in this study.

Table2S3.csv: Species distribution by family. Plant Tribes 2.0 species list, with the number of pectin-related genes found in each.

galactangalactosyltransferasefamilyalignment.phy: Raw β-1,4-galactan β-1,4-galactosyltransferase Phylip alignment.

GATLfamilyalignment.phy: Raw GATL Phylip alignment.

GAUTfamilyalignment.phy: Raw GAUT Phylip alignment.

GAUTsuperfamilyalignment.phy: Raw GAUT superfamily Phylip alignment.

homogalacturonanmethyltransferasefamilyalignment.phy: Raw homogalacturonan methyltransferase Phylip alignment.

pectatelyaselikefamilyalignment.phy: Raw pectate lyase-like Phylip alignment.

pectinacetylerasefamilyalignment.phy: Raw pectin acetylerase Phylip alignment.

pectinacetyltransferasefamilyalignment.phy: Raw pectin acetyltransferase Phylip alignment.

PGIPfamilyalignment.phy: Raw polygalacturonase inhibitor protein Phylip alignment.

PMEfamilyalignment.phy: Raw pectin methylesterase Phylip alignment.

PMEIfamilyalignment.phy: Raw pectin methylesterase inhibitor Phylip alignment.

polygalacturonasefamilyalignment.phy: Raw polygalacturonase Phylip alignment.
RGLarabinosyltransferasefamilyalignment.phy: Raw rhamnogalacturonan I arabinosyltransferase Phylip alignment.
RGIxylosyltransferasefamilyalignment.phy: Raw rhamnogalacturonan II xylosyltransferase Phylip alignment.
UDPGlcAepimerasefamilyalignment.phy: Raw UDP-glucuronic acid epimerase Phylip alignment.
xylogalacturonanxylosyltransferasefamilyalignment.phy: Raw xylogalacturonan xylosyltransferase Phylip alignment.
galactangalactosyltransferase.tree: Raw β-1,4-galactan β-1,4-galactosyltransferase Newick tree.
GATL.tree: Raw GATL Newick tree.
GAUT_superfamily.tree: Raw GAUT superfamily Newick tree.
GAUT.tree: Raw GAUT Newick tree.
homogalacturonanmethyltransferase.tree: Raw homogalacturonan methyltransferase Newick tree.
pectatelyaselylike.tree: Raw pectate lyase-like Newick tree.
pectinacetylenesterase.tree: Raw pectin acetylenesterase Newick tree.
pectinacetyltransferase.tree: Raw pectin acetyltransferase Newick tree.
PGIP.tree: Raw polygalacturonase inhibitor protein Newick tree.
PME.tree: Raw pectin methylesterase Newick tree.
PMEI.tree: Raw pectin methylesterase inhibitor Newick tree.
polygalacturonase.tree: Raw polygalacturonase Newick tree.
RGLarabinosyltransferase.tree: Raw rhamnogalacturonan I arabinosyltransferase Newick tree.
RGIxylosyltransferase.tree: Raw rhamnogalacturonan II xylosyltransferase Newick tree.
UDPGlcAepimerase.tree: Raw UDP-glucuronic acid epimerase Newick tree.
xylogalacturonanxylosyltransferase.tree : Raw xylogalacturonan xylosyltransferase Newick tree.

Cas9_sequences.csv : Spreadsheet of sequence data related to CRISPR-Cas9 experiments.

amiRNA_sequences.csv : Spreadsheet of sequence data related to amiRNA experiments.

PpCESA5_velocities_20150918.csv : Raw velocity data of GFP-labeled PpCESA5.

PpCESA8_velocities_20151007.csv : Day 2 raw velocity data of GFP-labeled PpCESA8.

PpCESA8_velocities_20151008.csv : Day 1 raw velocity data of GFP-labeled PpCESA8.

PpCESA8_velocities_20151007and20151008.csv : Combination of Day 1 and Day 2 raw velocity data of GFP-labeled PpCESA8.

PpCESA8_velocities_20151007and20151008_controls.csv : Combination of Day 1 and Day 2 raw velocity data of GFP-labeled PpCESA8 control treatments.

PpCESA8_densities.csv : Density measurements from PpCESA8 time lapses.

PpCESA5_densities.csv : Density measurements from PpCESA5 time lapses.

AtCESA4_full_complementation_alignment.phy : Extended Phylip protein alignment of AtCESA4 and sequences that effectively rescue the AtCESA4 knockout.

AtCESA4_including_noncomplementation_by_747_alignment.phy : Extended Phylip protein alignment of AtCESA4 and sequences that effectively rescue the AtCESA4 knockout, and the 747 chimeric sequence that did not rescue.

AtCESA7_full_complementation_alignment.phy : Extended Phylip protein alignment of AtCESA4 and sequences that effectively rescue the AtCESA7 knockout.

AtCESA7_including_noncomplementation_by_744_alignment.phy : Extended Phylip protein alignment of AtCESA7 and sequences that effectively rescue the AtCESA7 knockout, and the 744 chimeric sequence that did not rescue.

AtCESA7_including_noncomplementation_by_787_alignment.phy : Extended Phylip protein alignment of AtCESA7 and sequences that effectively rescue the AtCESA7 knockout, and the 787 chimeric sequence that did not rescue.
AtCESA8_full_complementation_alignment.phy: Extended Phylip protein alignment of AtCESA8 and sequences that effectively rescue the AtCESA8 knockout.

AtCESA8_including_noncomplementation_by_747_alignment.phy: Extended Phylip protein alignment of AtCESA8 and sequences that effectively rescue the AtCESA8 knockout, and the 747 chimeric sequence that did not rescue.
VITA

Thomas Wayne McCarthy

Education
- **Pennsylvania State University**  State College, PA
  PhD in Plant Biology  Anticipated graduation December 2016
- **Purdue University**  Colleges of Science, Agriculture  West Lafayette, IN
  B.S. in Biology  Dec 2009
  B.S. in Plant Biology  Dec 2009

Publications and Presentations

Research Experience
- Feb 2010 to July 2011  **Wilker Chemistry Group**  West Lafayette, IN
  Research Technician
- May 2008 to Dec 2009  **Wise Plant Pathology Lab**  West Lafayette, IN
  Lab and Field Assistant
- Jan 2008 to May 2008  **Banks Plant Biology Lab**  West Lafayette, IN
  Undergraduate Researcher

Awards received
- **Penn State University**
  University Graduate Fellowship 2011-2012
- **Purdue University**
  Indiana Top Scholar Scholarhip 2005-2006
  Purdue Academic Success Scholarship 2005-2006

Science outreach
- Sept 2014 to present  **PA Junior Academy of Science**  State College, PA
  Judge
- June 2008 to Aug 2011  **iGEM**  State College, PA
  Graduate advisor