IDENTIFICATION OF A BAHD ACYLTRANSFERASE INVOLVED IN THE ACETYLATION OF MONOLIGNOLS IN KENAF (*Hibiscus cannabinus*)

By

Sasha Annabel Peers

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Plant Biology

2012

ABSTRACT

IDENTIFICATION OF A BAHD ACYLTRANSFERASE INVOLVED IN THE ACETYLATION OF MONOLIGNOLS IN KENAF (*Hibiscus cannabinus*)

By

Sasha Annabel Peers

Plant cell walls are the most abundant renewable resource on Earth and have been proposed as a major feedstock for biofuel production. Lignin is a component of cell walls that provides support and structural integrity to vascular tissues. These properties are also responsible for lignin being the major obstacle for the processing of plant biomass to biofuels. Lignin structure can be altered by acylation of lignin monomer units, namely, *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol. The significance of this phenomenon has not been established primarily because the enzyme(s) responsible for the acylation of monolignols have not been identified. This study sought to identify the enzyme(s) responsible for acetylation in phloem fiber lignin in kenaf (*Hibiscus cannabinus*). Kenaf is a non-woody plant whose lignin has a high syringyl content and extensive acetylation of syringyl units. This study reports on the identification of an enzyme (AMT; acetyl monolignol transferase) that catalyzes the synthesis of sinapyl acetate using a synthetic gene for a BAHD acyltransferase present in kenaf. AMT had a high activity with sinapyl alcohol and very little activity with coniferyl alcohol, which is the expected selectivity given the acetylation found in kenaf lignin. AMT was transformed into arabidopsis and seedlings are being screened for AMT expression. The significance of lignin acylation is an intriguing unanswered question in plant cell wall biology. Elucidating the enzymes involved in monolignol acylation will provide a basis for uncovering their functional role in lignin structure and whether changes to acylation levels in plants will prompt facile processing of biomass into biofuels.

This thesis is dedicated to my beloved family and friends.
I am forever grateful for your love and support.

ACKNOWLEDGMENTS

The author would like to thank the following people: Dr. Curtis Wilkerson for his guidance, understanding, and financial support, Drs. Federica Brandizzi, Brad Day, and Amy Iezzoni for their advice and understanding, Doug Whitten and Saunia Withers for their assistance on experimental techniques, the Wilkerson lab for their positive energy and good humor, the Iezzoni lab (and individuals previously listed above) for volunteering their time and effort during the cherry pollination experiments, Dr. John Ralph and his lab members for providing the chemically synthesized substrates used in the HPLC experiments, and lastly, the MSU Plant Biology Department and the Great Lakes Bioenergy Research Center for their financial support.

PREFACE

This thesis is presented in two chapters because it describes two areas of plant biotechnology – fuel and food. First, "Identification of a BAHD Acyltransferase Involved in the Acetylation of Monolignols in Kenaf (*Hibiscus cannabinus*)" was presented as the central project due to the complete story it gave from start to finish. In short, lignin is a major obstacle in the conversion of cellulose to fermentable sugars for ethanol production. In Chapter One, I present an enzyme (acetyl monolignol transferase; AMT) that is involved in monolignol acylation in kenaf, a modification that modification may improve sugar yields from bioenergy crops. Thus, the purpose of this study was to identify and biochemically characterize AMT and overexpress AMT in arabidopsis for an analysis of lignin structure and cell wall digestibility. Result from this study will elucidate on the implications of altering lignin acylation in plants and reveal whether lignin acylation can be utilized for facile processing of biomass into biofuels. Chapter Two, titled "Characterization of the Sweet Cherry Pollination System", was my initial project when I began graduate school that involved an analysis of gametophytic self-incompatibility (GSI) in sweet cherry in order to understand the genetic and molecular basis of this phenomenon. However, I strongly pursued the AMT project because the pollination experiments were not successful for the first two years and factors such as unexpected weather conditions and harvest frequency made my time and resources very limited. Nevertheless, I was able to successfully perform the laboratory pollinations, which resulted in an established laboratory pollination protocol and an in-depth proteome analysis of cherry styles across different experimental conditions. These studies have not been described in sweet cherry and as such, will provide a foundation for the study of GSI in *Prunus* and further support the development of effective breeding programs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Identification of a BAHD Acyltransferase Involved in the Acetylation of Monolignols in Kenaf (*Hibiscus cannabinus)*

1.1 INTRODUCTION

In recent years, plant cell walls have captured significant attention as being the most abundant renewable energy resource on Earth. It has been reported that cell walls represent 70% of the 170-200 x 10^9 tons of worldwide biomass production in land plants (1). Yet, humans only utilize 2% of this cell wall biomass for the production of heat energy, timber, paper, and textiles. Due to the considerable availability of plant biomass, plant cell walls have been proposed as a major feedstock for the production of biofuels.

All plant cells contain a primary wall composed of polysaccharide-rich cellulose, hemicelluloses, and pectin, and some also form a thickened secondary cell wall composed of cellulose and hemicelluloses (1). During the formation of the secondary cell wall, phenylpropanoid compounds are polymerized in the wall to form lignin. Lignin is a complex phenolic polymer that provides structural integrity and support to water-conducting tissues. Arising from its complex structure, lignin confers a hydrophobic and impenetrable barrier to land plants, which not only supports vascular tissue but also aids in pathogen defense (2). It is because of these characteristics that lignin has aided in the evolutionary adaptation of plants from aquatic to terrestrial environments. However, due to its strength and prevalence in the cell wall, lignin presents a major obstacle for the processing of plant biomass to biofuels.

In order to design more effective cell wall processing techniques, it is important to understand the processes involved in the formation of cell wall components. All components of the cell wall vary across plant species (1). Similarly, lignin amount and composition vary across plant species and even among cell types (3). The classical mechanism for the biosynthesis of lignin monomer units, also known as monolignols, involves a series of reactions that transform the amino acid phenylalanine into one of three monolignols- hydroxyphenyl (H), guaiacyl (G) or

syringyl (S) (Figure 1.1) (2). The phenylpropanoid pathway begins with the deamination of phenylalanine to cinnamic acid by phenylalanine ammonia lyase (PAL), which is the first committed step in phenylpropanoid biosynthesis (2). Cinnamic acid is then hydroxylated at the 4′-position of the aromatic ring by cinnamate 4-hydroxylase (C4H). The resulting *p*-coumaric acid molecule is then activated (e.g. S-CoA) by 4-coumarate:CoA ligase (4CL), a key enzyme that catalyzes the synthesis of *p*-coumaroyl CoA. At this point, monolignol biosynthesis begins by the transesterification of *p*-coumaroyl CoA to a shikimic/quinic acid by hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase (HCT) and subsequent hydroxylation of the aromatic ring by *p*-coumaroyl shikimate 3′-hydroxylase (C3H). Shikimic/quinic acid is removed by HCT to generate caffeoyl CoA. Caffeoyl CoA then undergoes successive O-methylation reactions by caffeic acid O-methyltransferase (COMT) and caffeoyl CoA O-methyltransferase (CCoAOMT), followed by hydroxylation via ferulate 5-hydroxylase (F5H). Monolignols are then produced by reduction of hydroxycinnamyl CoA to hydroxycinnamaldehyde, followed by conversion to an alcohol by cinnamate alcohol dehydrogenase (CAD).

Monolignols are synthesized in the cytoplasm and transported across the plasma membrane to the apoplast via a poorly understood transport system. It has been suggested that monolignol transport is an ATP-driven process and as such, it may involve ATP-binding cassette (ABC) type transporters (4). However, the mechanism for monolignol transport remains unclear. Once in the apoplast, monolignols undergo dehydrogenative polymerization, which begins by the oxidation of monolignols by peroxidases and laccases to form monolignol radicals. Then, the monolignol radicals are incorporated into the lignin polymer via radical coupling, wherein bond formation between any two radicals is promoted by the single-electron oxidation of the monolignols, resulting in formation of lignin throughout the cell wall (2,3).

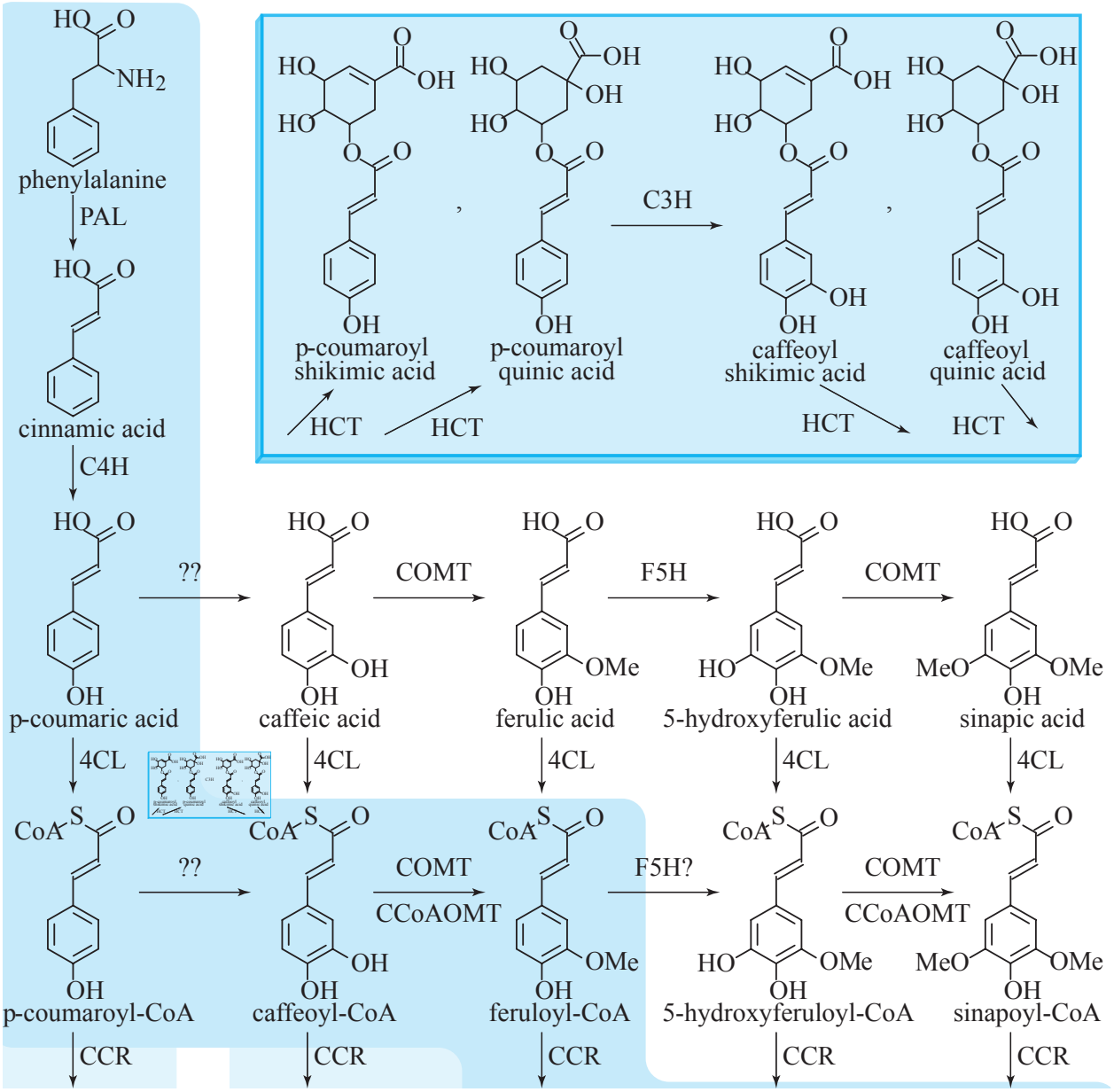
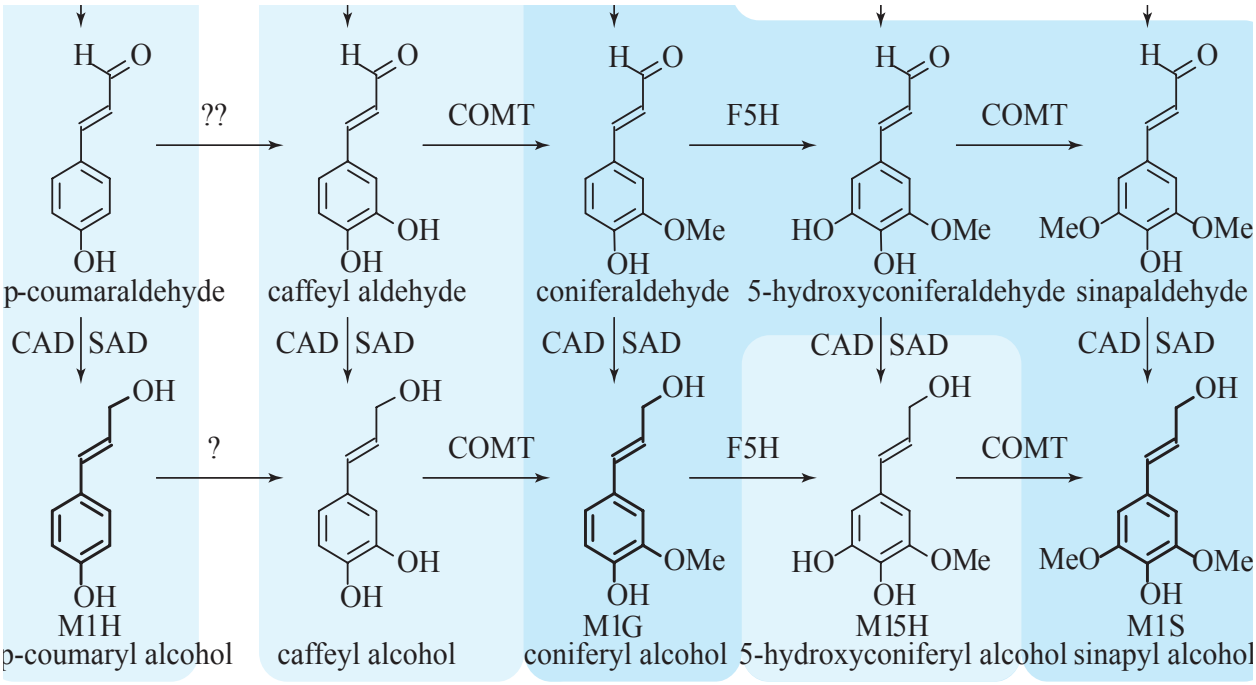


Figure 1.1 The Monolignol Biosynthetic Pathway. The dark blue shaded portion of the pathway represents the route that is predominantly observed in angiosperms for monolignol production. The light blue shaded portions are routes that occur in some plant species but to a minor degree. The white portions have not been observed to play a role in the production of monolignols. Figure was adapted from Boerjan et al. (2003) (3). For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this thesis.

Figure 1.1 (cont'd)



Through numerous studies on mutants and transgenic plants, the major enzymes involved in the biosynthesis of the three principal monolignols have been determined, as reviewed in (2) and (3). However, there is evidence to suggest that lignin biosynthesis is not limited to these three monomer units (5-12). Alternative monolignol precursors have been identified in the lignin of mutants lacking CAD or COMT, both of which are key monolignol biosynthetic genes (2). Down-regulation of the CAD gene in tobacco results in the incorporation of coniferaldehyde into the lignin polymer (5). Similarly, CAD-deficient mutants in pine, maize, and arabidopsis contained normal lignin levels; yet, the lignin composition was greatly altered from that of the wild type. COMT-deficient mutants had 5-hydroxyconiferyl alcohol incorporated into their lignin as a result of COMT deficiency (9,10). These instances demonstrate the plasticity of lignin biosynthesis in that lignin can tolerate variation from the canonical structure, yet still confer the primary function of strength and support to vascular tissue.

Another example of lignin monomer diversity is illustrated by monolignols that undergo modifications prior to incorporation into the lignin polymer. To date, one of the most widely reported modifications to occur is that of monolignol acylation (13,14). This phenomenon has been observed in many plant species, including *p*-coumarates in both C3 and C4 grasses (15-19); *p*-hydroxybenzoates in hardwoods such as aspen, poplar, and willow (20-24); and acetates in abaca, kenaf, palms, and sisal (25-28). Though many accounts of this phenomenon exist in a number of plant species, the functional role of monolignol acylation is unknown.

The identification of acetate groups on lignin remained to be determined until methods that preserved ester linkages were developed. These include nuclear magnetic resonance (NMR) analysis and the "Derivatization Followed by Reductive Cleavage" (DFRC) method, which cleaves ether linkages in the lignin polymer thereby releasing lignin monomers and dimers for analysis (29). DFRC has been used for the analysis of γ-*p*-coumarates in the lignin of grasses because the method does not disturb γ-ester groups on lignin (30). However, a major disadvantage of DFRC is that it utilizes acetate-based reagents, which interfere with the analysis of acetate groups on lignin. Additionally, lignin is commonly pre-acetylated prior to NMR analysis for improved solubility and spectral dispersion (26). Thus, evidence for acetylated lignin was not reported for quite some time.

Acetylated lignin was discovered in *Hibiscus cannabinus*, a non-woody plant commonly known as kenaf, when NMR analysis demonstrated that its phloem (bast) fiber lignin was not only rich in syringyl units, but also extensively acetylated (26). This discovery reinforced the idea that monolignols are pre-acylated prior to incorporation into the lignin polymer. To further analyze acetylation levels in kenaf, an alternative DFRC method was developed, called DFRC′, wherein all acetate-based reagents were substituted with their propionate analogues (30). This

new method has provided conclusive evidence that kenaf bast fibers are extensively acetylated, particularly on syringyl units. In addition, the DFRC′ method has been utilized to analyze acetylated lignin in other plant species such as abaca, sisal and hornbeam (28).

As referenced previously, the observation that kenaf bast fiber lignin contains a high syringyl content was the initial motive for studying this plant system. NMR results for kenaf lignin demonstrated that the molar syringaldehyde:vanillin ratio was 6.0, which supports prior evidence for the high syringyl:guaiacyl ratio observed in hardwoods (26). In addition, it was determined that over 50% of the lignin in kenaf bast fibers was acetylated and from this proportion, about 95% of the acetylation was located on the γ-position of the carbon side chain on syringyl units (Figure 1.2) (26). Though syringyl units were primarily acetylated, acetylation was also observed on guaiacyl units. However, the degree of acetylation on guaiacyl units was marginal when compared to that of syringyl units.

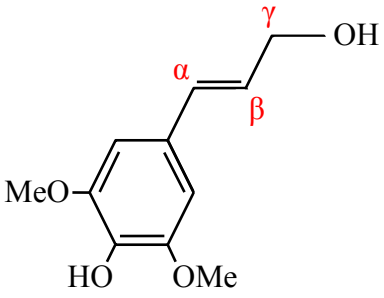


Figure 1.2 Molecular Structure of Sinapyl Alcohol. The carbons of the side chain for sinapyl alcohol are numbered in red.

There are many unexplored questions regarding monolignol acetylation. First, the functional implications of acetylated lignin have not been demonstrated. It has been suggested that lignin acylation may be a process involved in the regulation of lignin structure (31). Plants with high levels of acylation, particularly acetylation, have a greater propensity to forming β-O-4′ linkages rather than β-β′ linkages, which produces a more uncondensed lignin (Figure 1.3).

The advantages of a predominantly β-O-4′ lignin structure are unclear in an evolutionary sense. However, uncondensed lignin is desirable for the commercial purpose of processing biomass to biofuels. Another hypothesis suggests that monolignol acetylation could be a mechanism employed by plants to aid in drought tolerance since acetylation confers greater hydrophobicity to the lignin polymer (28). However, this speculation does not explain why abaca, a plant species whose lignin is highly acetylated, is not drought tolerant. Thus, the functional effects of this particular acylation phenomenon remain poorly understood.

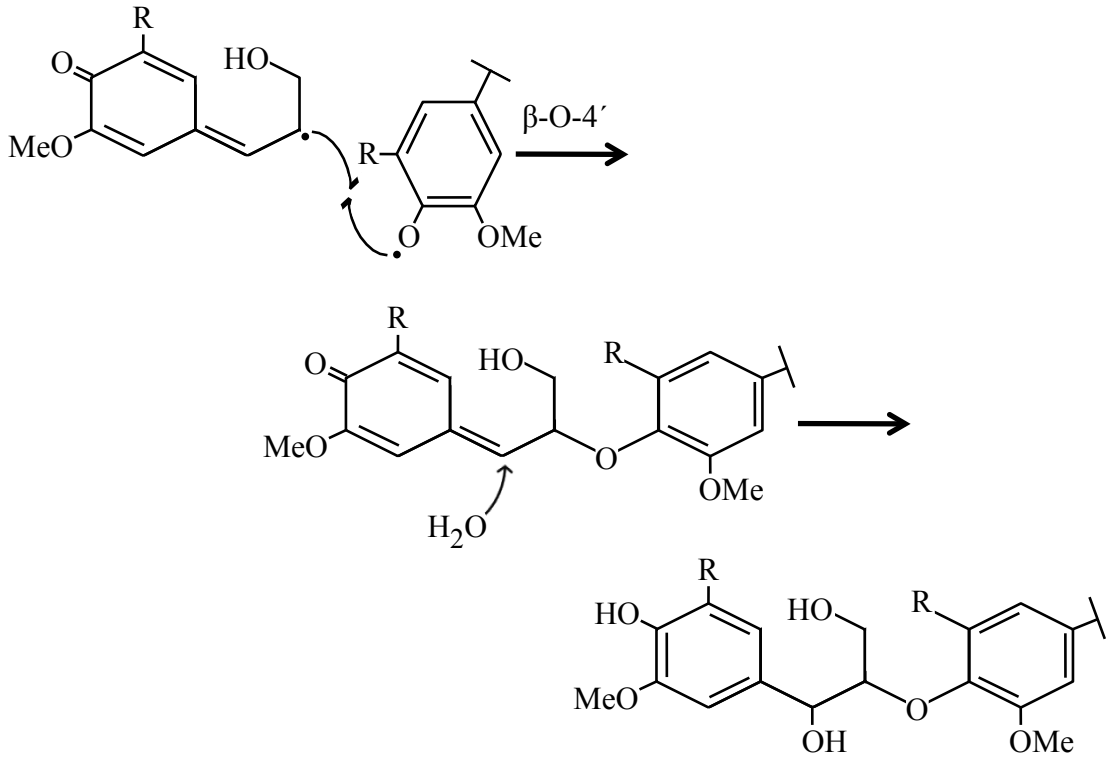
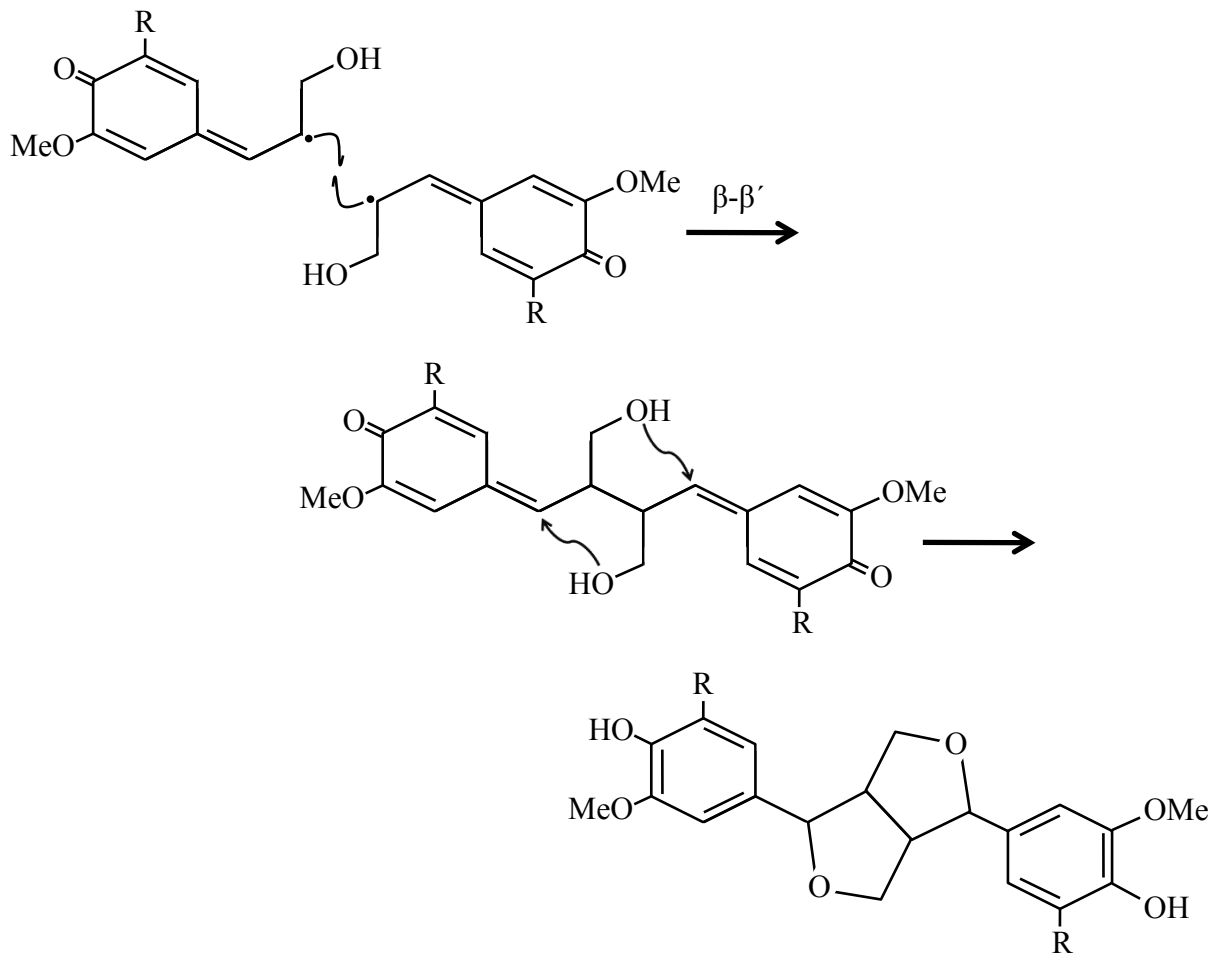


Figure 1.3 General Reaction Scheme for β-O-4′ and β-β′ Dimerization. The β-O-4′ dimerization process produces more linear structures. Monolignols linked via β-β′ dimerization have an additional level of structural complexity, producing a more cross-linked lignin structure. Half-headed arrows represent a single-electron transfer process, whereas full-headed arrows represent a two-electron transfer process. Figure was adapted from Bonawitz and Chapple, (2010) (2).

Figure 1.3 (cont'd)



Secondly, the acetylated monolignols in lignin may indicate that there are alternate biosynthetic pathways to monolignol biosynthesis (28). It is possible that the pre-acylation of monolignol precursors occurs as a step in an alternative pathway that is distinct from the canonical monolignol biosynthetic pathway. It is also possible that acylation occurs on monolignols after they have been synthesized via the usual mechanism. In any case, modifications to the traditional mechanism for monolignol biosynthesis cannot be excluded and should be considered in all interpretations of lignin biosynthesis.

Lastly, though analytical methods have aided in the identification of alternative lignin structure in a number of plants, the enzyme(s) responsible for acylation modifications have not been established. Recently, our group has identified OsPMT, a grass-specific enzyme that catalyzes the synthesis of monolignol *p*-coumarates in rice (19). Identification of the enzymes involved in acylation reactions is critical to understanding the mechanism of lignin acylation and how such modifications can be utilized to improve plant processing for conversion into biofuels. Additionally, manipulation of these enzymes *in planta* would provide a closer look into how monolignol acylation relates to lignin structure and function.

The primary purpose of this study was to identify an enzyme responsible for the acetylation of sinapyl alcohol in kenaf bast fiber lignin. A BAHD acyltransferase, herein called acetyl monolignol transferase (AMT), from kenaf stem was found to catalyze the acetylation of sinapyl alcohol to produce sinapyl acetate. AMT synthesized only trace amounts of coniferyl acetate and did not produce *p*-coumaryl acetate, indicating AMT's specificity for sinapyl alcohol. These biochemical results support data from NMR studies on kenaf bast lignin, which have shown extensive acetylation of sinapyl alcohol and negligible acetylation of coniferyl alcohol (26,28,30,32). To date, there has been no evidence to support the presence of *p*-coumaryl acetates in kenaf lignin (26,28,30,32).

Secondly, the purpose of this study was to produce AMT overexpression transgenic plants to determine the effects of AMT expression on lignin structure and function. The findings presented here will establish a basis for the study of acetylated lignin in kenaf and will also contribute to understanding the mechanism of lignin acylation reactions. Overall, the objective is to understand the mechanism of monolignol acylation and whether changes to acylation levels in plants would prompt facile processing of plant biomass into biofuels.

1.2 MATERIALS AND METHODS

Transcriptional Profiling of Kenaf cDNA Libraries

Plant material was provided by Dr. Ronald Hatfield at the U.S. Dairy Forage Research Center. Saunia Withers at the Michigan State University (MSU) Department of Plant Biology performed a total RNA extraction from kenaf bast fibers using TRIzol Reagent (Invitrogen, Carlsbad, California), followed by cDNA synthesis using the Creator SMART cDNA Library Construction Kit (Clontech, Mountain View, California) according to the manufacturer's instructions. cDNA was submitted to the MSU Genomics Core for Roche 454 sequencing using the 454 GSFLX Titanium Sequencer. Contigs were aligned and compiled in a database, named Kenaf CLC, created by Nick Thrower at the MSU Bioinformatics Core. The acetyl monolignol transferase (AMT) gene candidate was selected based on three criteria: 1) classification as a BAHD acyltransferase family protein, 2) annotation as an acetyltransferase according to The Arabidopsis Information Resource (TAIR) and 3) relative abundance.

Gene Synthesis and Gateway Cloning

AMT from kenaf was submitted to GENEART (Carlsbad, California) to be chemically synthesized with optimized codon usage for protein expression in *Escherichia coli* (*E. coli*). Plasmid DNA for AMT was extracted from a culture grown from the bacterial stab provided by GENEART. AMT was then cloned into the pDONR221 entry vector, followed by cloning into the pDEST17 vector using Invitrogen Gateway system cloning technology. Cloning AMT into the pDEST17 vector produced an expression clone with a N-terminal His_6 tag to be used in protein purification via immobilized metal affinity chromatography (IMAC). The His-tagged clone was then transformed into BL21 pLysS chemically competent cells (Invitrogen).

Heterologous Expression in E. coli

A one-liter bacterial culture containing the N-terminal His-tagged AMT expression clone was grown at 37 °C to an A_{600} between 0.4 and 0.45. Protein expression was induced by the addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG; Roche Applied Technologies, Indianapolis, Indiana). The culture was then grown for approximately 14 h at 18 °C, after which it was centrifuged to obtain a pellet and stored at -80 °C. The culture pellet was resuspended in 20 mL of binding buffer (20 mM Tris-HCl, pH 8, 0.5 M NaCl, and 1 mM 2-mercaptoethanol). The cells were lysed using a French pressure cell press. Soluble and insoluble fractions were separated by centrifugation at 50,000 x g for 30 min at 4 °C. The soluble fraction was syringe filtered prior to storage and the insoluble portion was resuspended in 20 mM Tris-HCl, pH 8. Both fractions were analyzed by SDS-PAGE to confirm that expression had taken place in the induced cultures.

Protein Purification by Fast Protein Liquid Chromatography (FPLC)

Five milliliters of soluble His-tagged AMT protein were loaded onto a 5 mL loop in preparation for FPLC purification. First, the soluble protein was eluted through four stacked 5 mL Hi-Trap desalting columns (GE Healthcare, Uppsala, Sweden) using binding buffer. Protein concentration for each eluted fraction was determined using A_{280}. Fractions with the highest protein concentration were collected to yield 5 mL of protein. This sample was then loaded onto a 1 mL IMAC Hi-Trap HP column (GE Healthcare) and eluted with buffer A (20 mM Tris-HCl, pH 8, 0.5 M NaCl, 1 mM 2-mercaptoethanol, and 25 mM imidazole) and a gradient of buffer B (20 mM Tris-HCl, pH 8, 0.5 M NaCl, 1 mM 2-mercaptoethanol, and 500 mM imidazole). Purified fractions containing protein, as determined by UV absorbance peaks, were analyzed on a SDS-PAGE gel. Protein bands were compared to the expected molecular weight for AMT (ca.

49 kDa). A protein band having the expected molecular weight for AMT was cut from the SDS-PAGE and analyzed by in-gel trypsin digestion and LC/MS/MS at the MSU Proteomics Core. Peptides were identified from MS/MS fragment data by searching against Kenaf CLC and *E.coli* databases using the program Mascot. AMT-containing fractions were collected and buffer exchanged into 100 mM phosphate buffer, pH 6, containing one protease inhibitor tablet (Roche Applied Science) using an Amicon 10K membrane filter (Millipore, Carrigtwohill, Ireland). The enzyme was stored in 20 µL aliquots at -80 °C until use.

Enzyme Activity Assays

Enzyme activity assays were performed in 25 mM phosphate buffer, pH 5.8, 0.5 mM dithiothreitol (DTT), 0.5 mM CoA thioester, 0.5 mM monolignol, and deionized water to a volume of 100 µL. An AMT present reaction and no enzyme control were prepared for each CoA thioester and monolignol combination. AMT reactions were initiated by adding 1 µg of AMT. All reactions were incubated for 45 min and solubilized with an equal volume of methanol. Protein was removed from the AMT reactions by using an Amicon 10K membrane filter. Ultra Performance Liquid Chromatography (UPLC) was used to analyze the flow-through material at 280 nm and 340 nm. The eluting compounds were identified by comparison of their peak retention times to standards for the CoA thioesters, monolignols and monolignol conjugates.

The monolignols (coniferyl and sinapyl alcohol) and monolignol conjugates (coniferyl *p*-coumarate, sinapyl *p*-coumarate, coniferyl caffeate, sinapyl caffeate, coniferyl acetate and sinapyl acetate) were provided by Dr. John Ralph at the University of Wisconsin- Madison (33). CoA thioester synthesis was performed by following a previously described method (34), which

uses a C-terminal His-tagged 4-coumarate-ligase (4CL) from tobacco in the pCRT7/CT TOPO vector (provided by Eran Pichersky, University of Michigan), the acid derivatives, coenzyme A hydrate, and adenosine-5′-triphosphate (ATP). All chemicals for CoA thioester synthesis, as well as acetyl CoA, were purchased from Sigma-Aldrich (St. Louis, Missouri). CoA thioester concentration was quantified by using its maximum absorbance and extinction coefficient.

Enzyme Kinetics

Enzyme kinetics were performed in 25 mM phosphate buffer, pH 5.8, 0.5 mM DTT, 0.005 - 0.05 mM CoA thioester, 0.5 - 2 mM monolignol substrate, and deionized water to a volume of 100 µL. The reactions were initiated with purified AMT in 0.5X BSA (New England Biolabs, Ipswich, Massachusetts). The amount of AMT added to each reaction set was variable and dependent on the rate of product formation for that CoA thioester and monolignol combination. Each 100 µL reaction was divided into 30 µL per time point. The reactions were run for 5, 10, and 20 min and performed in triplicate. All reactions were solubilized with an equal volume of methanol. The CoA thioester substrate used was acetyl CoA and the monolignol alcohol substrates used were sinapyl alcohol and coniferyl alcohol. Peak areas were manually measured from the product peaks corresponding to each enzyme reaction using Empower 2 Software (Waters Corporation, Milford, Massachusetts). The peak areas (A_{280}) were converted to nMoles/sec/mg using extinction coefficients determined from standards. Kinetic parameters (K_m and V_{max}) for each CoA thioester and monolignol substrate combination will be calculated using a custom script from R64, version 2.15.2 (35).

1.3 RESULTS

Identification of the Candidate Gene

The AMT candidate gene was selected based on three criteria. First, I sought a candidate gene classified as a BAHD acyltransferase family protein. This classification was an important consideration because BAHD acyltransferases constitute a super family of proteins that utilize acyl CoA thioesters in acylation reactions (14). It was presumed that the AMT candidate would be a BAHD acyltransferase due to its ability to acylate monolignols using acetyl CoA. An analysis of AMT's peptide sequence demonstrated that it had two conserved domains common to all BAHD acyltransferases: the "HXXXD" and "DFGWG" motifs (Figure 1.4) (14).

MALLRPASLVFTVRRHDPELVVPSKPTPHECKTLSDIDDQDGHRFQIRGLHVYRCNASM
QGKDPVRVIREALAKALVFYYPFAGRIKEGPNRKLMVDCTGEGVLFIEADADVMLEEF
GGSLHPPFPCFKDLLCEPTGSNDLLNSPVLQIQVTRLKCGGFIFAHRFN**HTMSD**AVGLIQ
FMSAMGEIARGAVAPSIPPVWERHLLNARDPPLITCEHHEYDHATATNGTIMPTDNLVH
HSFFFGPTQISALKRLISDNVSCSTFDILTACVWRCRTIAMKLGPDEDVRLICIVNARYKF
NPPLPLGYYGNALGYPAALTTAGELSKKPLEYAVKLVKEAKAKATDEYMKSTADLMV
SRGRPNVNTVRSFLVSDLSRARFREV**DFGWG**KAEFGGPSNGTEIISFYIPSKNKEGKEGI
AVPVCLPASVMESFVKEINSTLADDEATGA

Figure 1.4 Peptide Sequence for Kenaf AMT. The AMT peptide sequence was 444 amino acids in length and had an approximate molecular weight of 49 kDa. The HXXXD and DFGWG motifs common to all BAHD acyltransferases are underlined and in bold.

Second, I sought a candidate gene annotated as an acetyltransferase according to TAIR. The Kenaf CLC database utilizes TAIR to annotate proteins based on their closest homolog in arabidopsis. AMT's closest homolog in arabidopsis was an acetyltransferase gene (Figure 1.5),

```
Majority    ..............L.F.V.R..PEL..P.KPTP.E.K.LSDIDDQ.G.RFQI..
kenaf.pro   ------MALLRPASLVFTVRRHDPELVVPSKPTPHECKTLSDIDDQDGHRFQIRG   49
acetyl.pro MDHQVSLPQSTTTGLSFKVHRQQPELITPAKPTPRELKPLSDIDDQQGLRFQIPV   55

Majority    ...YR.N.S.....PV.VI..ALA.ALV.YYPFAGR..E..NRKL.VDCTGEGVL
kenaf.pro   LHVYRCNASMQGKDPVRVIREALAKALVFYYPFAGRIKEGPNRKLMVDCTGEGVL  104
acetyl.pro IFFYRPNLSS-DLNPVQVIKKALADALVYYYPFAGRLRELSNRKLAVDCTGEGVL  109

Majority    FIEA.ADV.L.E......L.PPFPC...LL....GS.D.LN.P.L..QVTRLKC.
kenaf.pro   FIEADADVMLEEFGGS--LHPPFPCFKDLLCEPTGSNDLLNSPVLQIQVTRLKCG  157
acetyl.pro FIEAEADVSLTELEEADALLPPFPCLDELLFDVEGSSDVLNTPLLLVQVTRLKCR  164

Majority    GFIFA.RFNHTM.D..GL..F.....E.A....APS.PPVW.RHLL.........
kenaf.pro   GFIFAHRFNHTMSDAVGLIQFMSAMGEIARGAVAPSIPPVWERHLLNARD-PPLI  211
acetyl.pro GFIFALRFNHTMTDGAGLSLFLKSLCELAYRLHAPSVPPVWDRHLLTVSASEARV  219

Majority    T..H.EY....A........D..V..SFFF....ISA...L.........F..L.
kenaf.pro   TCEHHEYDHATATNGTIMPTDNLVHHSFFFGPTQISALKRLISDNVSCSTFDILT  266
acetyl.pro THTHREYEDQVAID-AVDTGDPFVSRSFFFSAEEISAIRKLLPPDLHNTSFEALS  273

Majority    ...WRCRTIA....P....RL.CI.N.R.K...PPL..GYYGN....PAA..TA.
kenaf.pro   ACVWRCRTIAMKLGPDEDVRLICIVNARYKFN-PPLPLGYYGNALGYPAALTTAG  320
acetyl.pro SFLWRCRTIALNPDPNTEMRLTCIINSRSKLSNPPLSRGYYGNVFVIPAAIATAR  328

Majority    .L..KPLE.A..L..E.K...T..Y..S...LM..RGRP..........SDL...
kenaf.pro   ELSKKPLEYAVKLVKEAKAKATDEYMKSTADLMVSRGRPNVNTVRSFLVSDLSRA  375
acetyl.pro DLMEKPLEFALRLIQETKSSVTEDYVRSVTALMATRGRPMFVAAGNYIISDLRHF  383

Majority    ....VDFG.WGK...GG............SFY.P.KNK.G..G..V...LP...M
kenaf.pro   RFREVDFG-WGKAEFGGPSNG----TEIISFYIPSKNKEGKEGIAVPVCLPASVM  425
acetyl.pro DLGKVDFGPWGKPVYGGTAKAGIALFPGVSFYVPFKNKKGETGTVVAISLPVRAM  438

Majority    E.FV.E.N..L........
kenaf.pro   ESFVKEINSTLADDEATGA                                     444
acetyl.pro ERFVAELNGVLIKRF                                         453
```

Figure 1.5 Peptide Sequence Alignment of Kenaf AMT and Arabidopsis Acetyltransferase. The AMT candidate only had 49% similarity to the arabidopsis acetyltransferase. Sequences were aligned using MegAlign (DNASTAR). Dots denote non-matching residues. Dashes signify gaps inserted to obtain optimal sequence alignment. The HXXXD and DFGWG motifs are underlined.

which functions in volatile production for plant defense (36). It is important to note that BAHD acyltransferases are well known for serving functionally diverse roles in secondary metabolism, regardless of their conserved domains and overall similarity. Thus, it was not likely that the AMT candidate would have a similar function to the acetyltransferase in arabidopsis. Lastly, due to the prevalence of acetylated syringyl units in kenaf lignin, the candidate gene was selected on the basis of abundance (Table 1.1).

Table 1.1 Abundance Values for Transferase Family Proteins Identified in Kenaf

Representative Definition	**Total Members**
glycosyl transferase family 1 protein	1637
transferase family protein	751
HCT transferase family protein	485
transferase family protein	469
UDP-glucoronosyl/UDP-glucosyl transferase family protein	457
UTP-glucose-1-phosphate uridylyltransferase family protein	309
HCT transferase family protein	248
O-acetyltransferase family protein	232
transferase family protein	***172***
O-acetyltransferase family protein	144
protein arginine N-methyltransferase family protein	90
glycosyltransferase family 14 protein	76
galactosyltransferase family protein	76
HCT transferase family protein	73
glycosyltransferase family 14 protein	65
O-methyltransferase family 2 protein	60
O-acetyltransferase family protein	58
galactosyltransferase family protein	57

Note: The AMT candidate gene is indicated in bold/italics.

Heterologous Expression in E. coli

Using a codon-optimized synthetic gene for kenaf AMT, a clone for a N-terminal His-tagged AMT was produced and utilized for protein expression in *E. coli*. The protein was purified from *E. coli* extracts using IMAC and analyzed by SDS-PAGE. Pre- and post-purification protein samples demonstrated an induced protein in the 49 kDa range, as expected for AMT (Figure 1.6). The un-induced controls did not show induction of a protein of similar size. SDS-PAGE analysis demonstrated that AMT is mostly insoluble. However, protein expression was performed in a 1 L culture to produce enough soluble protein for IMAC purification. Further analysis by LC/MS/MS of tryptic fragments confirmed that the protein band was AMT.

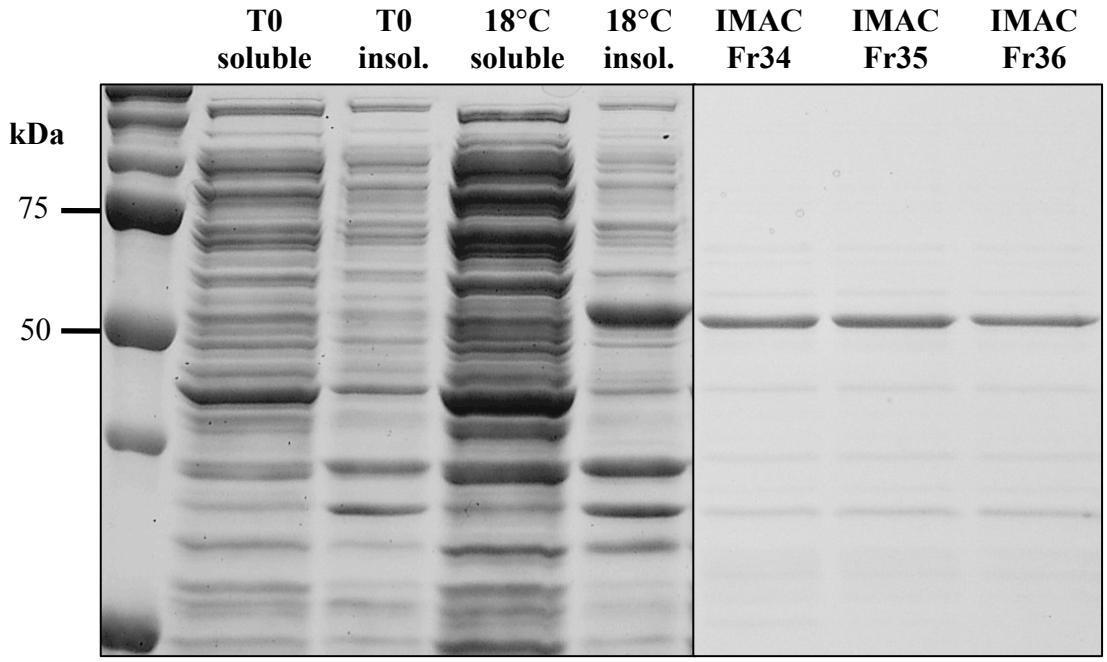


Figure 1.6 Heterologous Expression of AMT in *E. coli*. SDS-PAGE analysis of soluble and insoluble protein fractions for uninduced (T0) and induced (14 h) expression at 18 °C, and IMAC fractions collected for AMT activity assays.

Enzyme Activity Assays

AMT demonstrated a strong preference for sinapyl alcohol as the acyl acceptor. AMT produced sinapyl acetate in a reaction containing 1 μg of AMT, 0.5 mM acetyl CoA, and 0.5 mM sinapyl alcohol (Figure 1.7A). When using coniferyl alcohol, AMT produced only trace amounts of coniferyl acetate (Figure 1.7B). AMT did not synthesize *p*-coumaryl acetate, suggesting that *p*-coumaryl alcohol was not a good substrate for AMT activity. Overall, the substrate specificity observed in the AMT activity assays is similar to the acetylation specificity found in kenaf lignin. When tested with other CoA thioester substrates, AMT demonstrated activity with *p*-coumaroyl CoA and caffeoyl CoA (Table 1.2). However, it is important to note that monolignol *p*-coumarates and caffeates have not been observed in kenaf lignin (26,28,30,32).

Enzyme Kinetics

AMT substrate specificity was analyzed using acetyl CoA as the acyl donor and sinapyl alcohol and coniferyl alcohol as the acyl acceptors. *p*-Coumaryl alcohol was not analyzed due to it being a poor substrate for AMT activity. AMT had a strong affinity for sinapyl alcohol as the acyl acceptor (Table 1.3). The activity observed with coniferyl alcohol and acetyl CoA was too small to obtain the K_m for these substrates. However, an estimated V_{max} was calculated for saturating coniferyl alcohol and varying acetyl CoA. Kinetic parameters were not obtained for varying coniferyl alcohol and saturating acetyl CoA due to little activity with coniferyl alcohol. In these experiments, high concentrations of coniferyl alcohol did not produce enough coniferyl acetate to accurately detect K_m or V_{max}. Thus, AMT substrate specificity for coniferyl alcohol was only examined under saturating monolignol conditions.

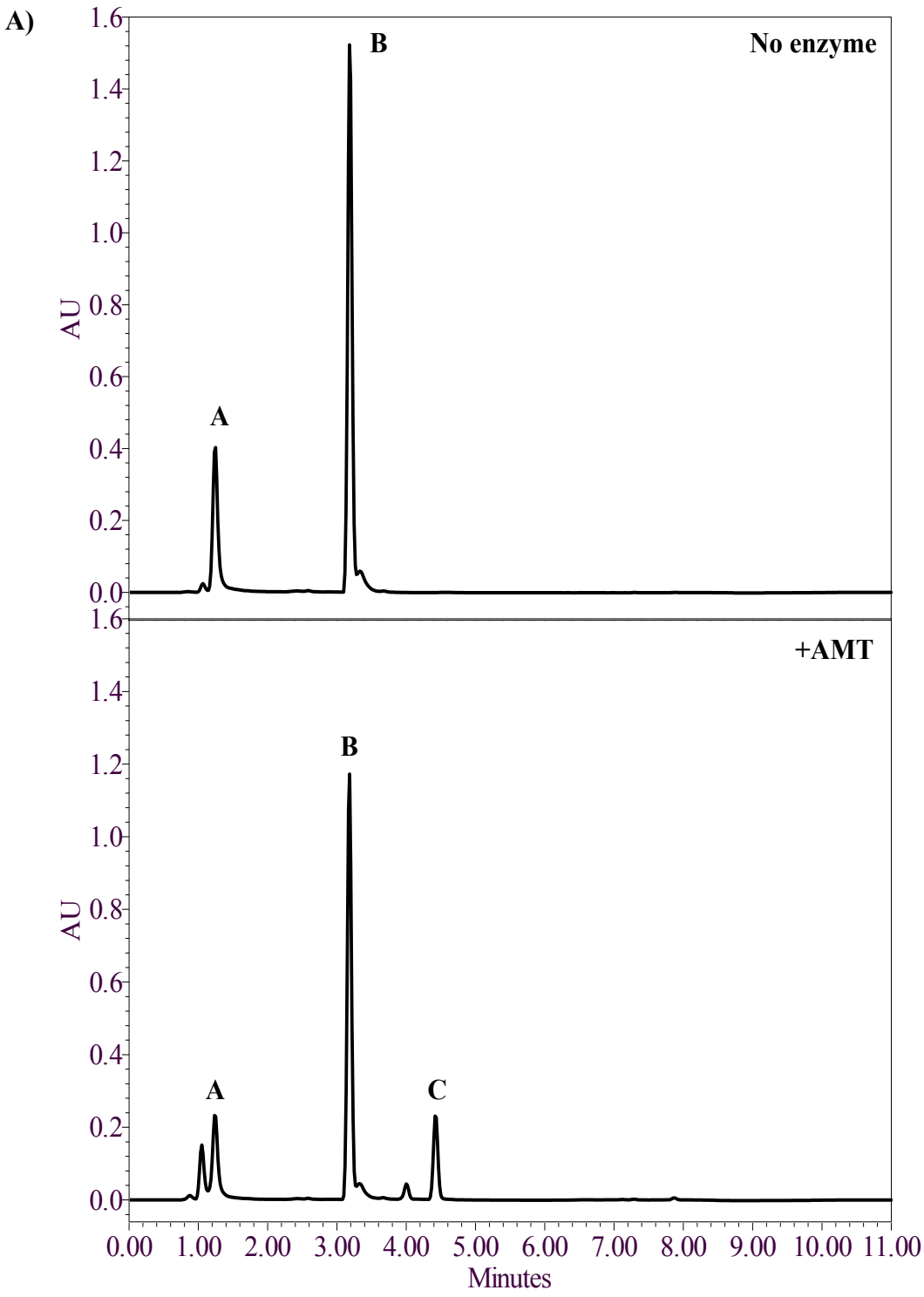


Figure 1.7 HPLC Analyses of AMT Activity Assays. HPLC Chromatograms of AMT activity assays with no enzyme or purified AMT (+AMT). *A)* acetyl CoA **(A)** and sinapyl alcohol **(B)** to make sinapyl acetate **(C)** and *B)* **(A)** and coniferyl alcohol **(D)** to make coniferyl acetate **(E)**.

Figure 1.7 (cont'd)

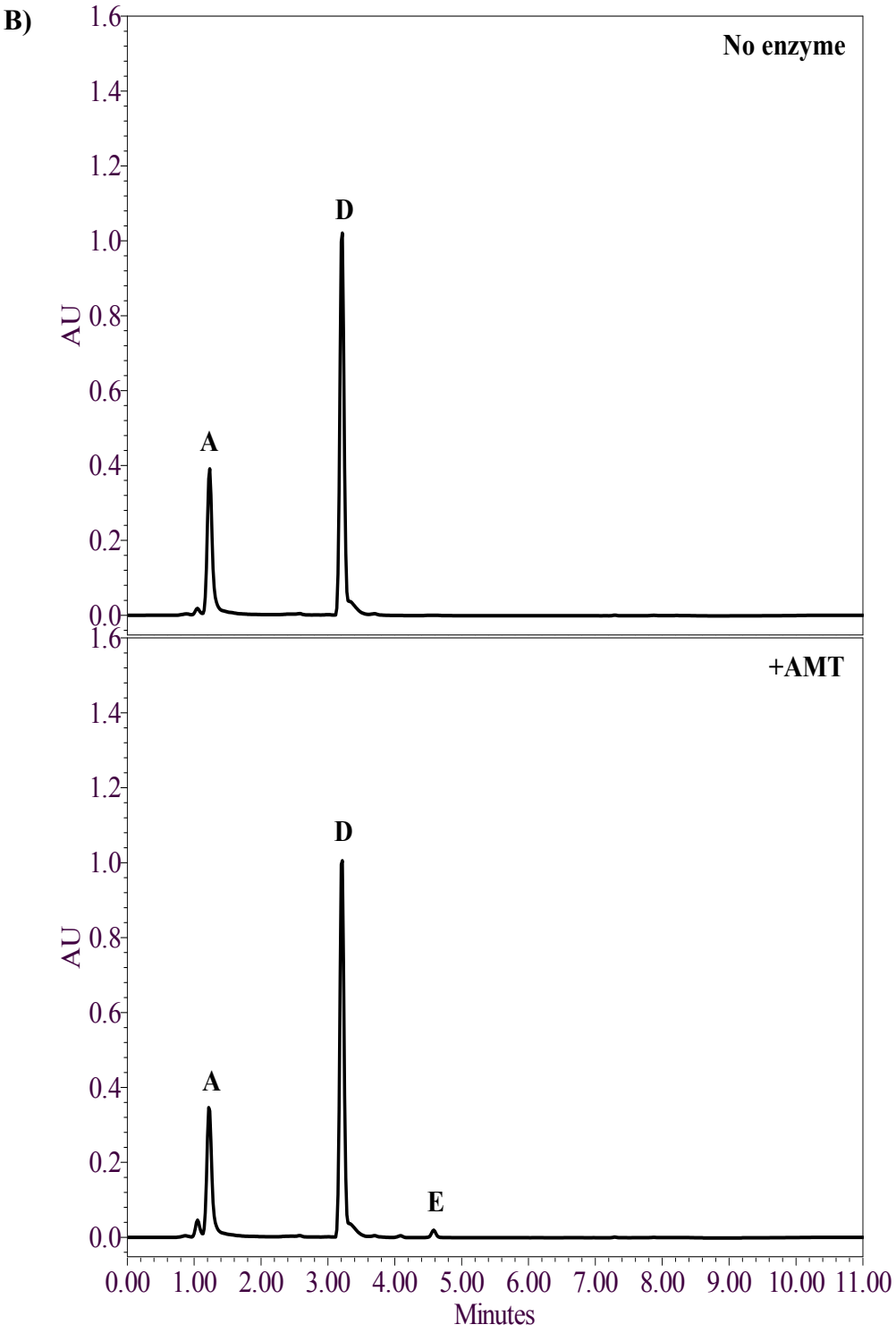


Table 1.2 Enzyme Activity for AMT with Various Substrate Combinations

Monolignol Substrate	CoA Thioester	AMT Activity
Sinapyl alcohol	Acetyl CoA	Yes
Sinapyl alcohol	*p*-Coumaroyl CoA	Yes
Sinapyl alcohol	Caffeoyl CoA	Yes
Sinapyl alcohol	Feruloyl CoA	Yes*
Coniferyl alcohol	Acetyl CoA	Yes*
Coniferyl alcohol	*p*-Coumaroyl CoA	Yes
Coniferyl alcohol	Caffeoyl CoA	Yes
Coniferyl alcohol	Feruloyl CoA	No
p-Coumaryl alcohol	Acetyl CoA	No
p-Coumaryl alcohol	*p*-Coumaroyl CoA	Yes*
p-Coumaryl alcohol	Caffeoyl CoA	No
p-Coumaryl alcohol	Feruloyl CoA	No

* Indicates that a negligible amount of product was formed, which suggests that the substrate combination was a poor match for AMT activity.

Table 1.3 Kinetic Parameters for AMT

Varying Substrate	Saturating Substrate	K_m (μM)	V_{max} (nkat mg^{-1})	K_{cat} (s^{-1})	K_{cat}/K_m (mM^{-1} s^{-1})
Sinapyl alcohol	Acetyl CoA	197.2 ± 36.2	36.2 ± 2.3	1.77	8.98
Acetyl CoA	Sinapyl alcohol	43.5 ± 7.5	30.0 ± 1.3	1.47	33.8
Acetyl CoA	Coniferyl alcohol	NA*	<8.5	NA	NA

* NA indicates that the kinetic parameter was not obtained due to very little AMT activity.

1.4 DISCUSSION

AMT is a BAHD acyltransferase in kenaf that is responsible for catalyzing the acetylation of sinapyl alcohol to produce sinapyl acetate. It has been demonstrated that AMT has a strong preference for sinapyl alcohol as the acyl acceptor and only marginal activity when using coniferyl alcohol. *p*-Coumaryl alcohol was a poor substrate for AMT activity. Interestingly, the substrate specificity observed in the AMT activity assays mimics the acetylation observed in kenaf bast fiber lignin. Additionally, these observations support NMR analyses of kenaf lignin, wherein only a trace amount of acetate groups were found on guaiacyl units when compared to the amount of acetate groups on syringyl units (30).

AMT also demonstrated activity with the acyl donors *p*-coumaroyl CoA and caffeoyl CoA. However, though this activity was observed *in vitro*, it is unlikely that these acylation reactions take place *in planta* because monolignol *p*-coumarates and monolignol caffeates have not been reported in kenaf lignin (26,28,30,32). It may be that *p*-coumarates have not been detected in kenaf lignin due to the limited amount of *p*-coumaroyl CoA present in the cell. *p*-Coumaroyl CoA is the central metabolite in the phenylpropanoid pathway leading to the biosynthesis of not only monolignols but also a wide array of secondary metabolites, such as flavonoids, coumarins, and tannins (37). Thus, the presence of *p*-coumaroyl CoA may be transient in the cell and not abundant enough to produce monolignol *p*-coumarates. The same argument can be made for caffeoyl CoA, which is quickly consumed by COMT and CCoAOMT to drive the synthesis of coniferyl alcohol and sinapyl alcohol.

Del Rio et al. (2007) performed an analysis of naturally acetylated lignin on a large set of plant species, including abaca, kenaf, and sisal (28). Lignin from both angiosperms and gymnosperms was examined using the DFRC´ method and relative abundance values for total

monolignol units, syringyl-to-guaiacyl (S/G) ratio, and acetylated monolignol units were reported. It was demonstrated that native acetylation of monolignol units is a phenomenon that is common among angiosperms. With some exceptions (i.e. bamboo and eucalyptus), monolignol acetylation was predominantly detected on syringyl units, particularly at the γ-carbon of the side chain. Kenaf demonstrated approximately 1.4 times as many total syringyl units than total guaiacyl units. Of this proportion, guaiacyl units were significantly less acetylated than syringyl units at 8.9% vs. 59%, respectively (28). These values demonstrate that acetylation of sinapyl alcohol is strongly preferred in kenaf, even though the amount of total syringyl units and total guaiacyl units detected was relatively similar. These observations strongly suggest that enzyme preference for sinapyl alcohol may be the driving force behind the acetylation of syringyl units to synthesize sinapyl acetates rather than coniferyl acetates.

Coniferyl acetate has been implicated in floral volatile biosynthesis in petunia because it is used as substrate by isoeugenol synthase 1 (PhIGS1) to make isoeugenol, a major component of floral scent in petunia. Dexter et al. (2007) (38) reported on a BAHD acyltransferase called coniferyl alcohol acyltransferase (PhCFAT) that utilizes coniferyl alcohol and acetyl CoA to produce coniferyl acetate in petunia petals. PhCFAT kinetics demonstrated a high affinity for coniferyl alcohol and moderate affinity for sinapyl alcohol- 100% vs. 70%, respectively. To determine whether AMT and PhCFAT were closely related, a protein BLAST was performed to align the AMT and PhCFAT peptide sequences against one another (Figure 1.8). While both enzymes had the HXXXD and DFGWG motifs conserved among all BAHD acyltransferases, AMT and PhCFAT had a low peptide sequence identity (27%). Along with evidence of AMT's low enzyme rate with coniferyl alcohol, this suggests that AMT and PhCFAT have different functions in the plant cell.

```
Majority     ..........F.V.....E.V....P..HE........D.............Y..
Kenaf.pro    MALLRPASLVFTVRRHDPELVVPSKPTPHECKTLSDIDDQDGHRFQIRGLHVYR-   54
Petunia.pro  -----MGNTDFHVTVKKKEVVAAVLPMHHEHWLPMSNLDLLLPPLDFGVFFCYKR   50

Majority     .......KD....I..ALA..LV..Y..AG.............C...GV.F..A.
Kenaf.pro    CNASMQGKDPVRVIREALAKALVFYYPFAGRIKEGPNRKLMVDCTGEGVLFIEAD  109
Petunia.pro  SKINNDTKDDDETIKKALAETLVSFYALAGEVVFNSLGEPELLCNNRGVDFFHAY  105

Majority     AD..L.......HP.......L..............VL..QVT.LKCGG......
Kenaf.pro    ADVMLEEFGGSLHPPFPCFKDLLCEPTGSNDLLNSPVLQIQVTRLKCGGFIFAHR  164
Petunia.pro  ADIELNNLD-LYHPDVSVHEKLIPIKKHG-------VLSVQVTGLKCGGIVVGCT  152

Majority     F.H...DA.....F..A...IAR.........P...R.LLN.R.PP.........
Kenaf.pro    FNHTMSDAVGLIQFMSAMGEIARG-AVAPSIPPVWERHLLNARDPPLITCEHHEY  218
Petunia.pro  FDHRVADAYSANMFLVAWAAIARKDNNINTVIPSFRRSLLNPRRPPQFDDSFIDS  207

Majority     ...........P.D.L...........I..L......N....S......A..W..
Kenaf.pro    DHATATNGTIMPTDNLVHHSFFFGPTQISALKRLISDN-VSCSTFDILTACVWRC  272
Petunia.pro  TYVFLSSPPKQPNDVLTSRVYYINSQEINLLQSQATRNGSKRSKLECFSAFLWKT  262

Majority     ..............L...V....R...........Y.GN.L..P......G.L..
Kenaf.pro    RTIAMKLGPDEDVRLICIVN--ARYKFNPPLPLG-YYGNALGYPAALTTAGELSK  324
Petunia.pro  -IAEGGIDDSKRCKLGIVVDGRQRLRHDSSTTMKNYFGNVLSVPYTEASVGQLKQ  316

Majority     .PL.....LV.........E.........V...RP.............N......
Kenaf.pro    KPLEYAVKLVKEAKAKATDEYMKSTADLMVSRGRP----------NVNTVRSFL  368
Petunia.pro  TPLGKVADLVHTCLDNVANEHHFPSLIDWVELHRPRQAIVKVYCKDECNDEAAIV  371

Majority     VS...R.....V.FGWG...FG....P..G.........PS.NK.G..........
Kenaf.pro    VSDLSRARFREVDFGWGKAEFGG---PSNGTEIISFYIPSKNKEG-----KEGIA  415
Petunia.pro  VSSGLRFPLSQVNFGWGCPDFGSYIFPWGGQTGYVMPMPSPNKNGDWIVYMHLQK  426

Majority     ....L..................D..AT..
Kenaf.pro    VPVCLPASVMESFVKEINSTLADDEATGA                          444
Petunia.pro  KHLDLVETRAPHIFHPLTACYLDLTATY                          454
```

Figure 1.8 Peptide Sequence Alignment of Kenaf AMT and Petunia PhCFAT. Sequences were aligned using MegAlign (DNASTAR). Dots denote non-matching residues. Dashes signify gaps inserted to obtain optimal sequence alignment. The HXXXD and DFGWG motifs are underlined.

To conclude, plants have evolved mechanisms for altering their lignin structure by the utilization of monolignol substitutes or acylation modifications to monolignols. In some cases, plants have substituted traditional monolignol precursors with upstream intermediates in the monolignol biosynthetic pathway by down regulating biosynthetic enzymes. Additionally, in many plants, monolignols can be pre-acylated prior to incorporation into the lignin polymer. In both scenarios, the resulting lignin polymer is structurally different from the classical lignin structure but may confer similar functional and morphological qualities to the plant.

The functional significance of lignin acylation is one of the most intriguing unanswered questions in plant cell wall biology. Elucidating the mechanisms and enzyme(s) involved in monolignol acylation will provide an excellent foundation for uncovering their role in lignin structure and function. Kenaf represents a particularly fascinating plant system due to its high syringyl content and extensive acetylation of syringyl units. The kenaf AMT enzyme is involved in the acetylation of sinapyl alcohol. Future studies on AMT will detail the functional consequences of this gene *in planta.* Ultimately, our objective is to understand the mechanism of monolignol acylation and whether changes to acylation levels in plants would prompt facile processing of plant biomass into biofuels.

1.5 FUTURE STUDIES

AMT has been cloned into three plant expression vectors: pk2GW7, pk7YWG, and pk7WGY (39). These vectors contain the 35S Cauliflower Mosaic Virus (CaMV) promoter and are distinguished from one another in that pk7YWG has a N-terminal yellow fluorescence protein (YFP) fusion tag, pk7WGY has a C-terminal YFP fusion tag, and pk2GW7 does not have a YFP fusion tag. The three vectors were selected because it was not known whether the YFP tag would interfere with AMT expression *in planta*. However, AMT was successfully expressed in a N-terminal His-tagged expression vector, which suggests that expression would likely not be interrupted with a N-terminal tag. The YFP-tagged vectors will be useful in analyzing transgenic plant tissue for localization of AMT expression.

The AMT overexpression vectors were transformed into arabidopsis. Currently, I am in the process of screening arabidopsis seedlings for AMT expression. Plants from this study will be analyzed for: 1) lignin acetylation, 2) lignin composition, 3) cell wall digestibility, and 4) physical and physiological features. The DFRC´ method will be utilized to detect acetylation levels in lignin from AMT transgenic and wild type plants. Higher acetylation levels are expected in the transgenic plants, due to AMT's ability to catalyze acetylation reactions. Particularly, acetylation is expected on syringyl units. Through lignin composition analysis, the plants will be analyzed for syringyl content. Studies have shown that lignin in *Arabidopsis thaliana* has a low S/G ratio (40,41), while kenaf lignin has a high S/G ratio (26,28,32). It would be interesting to determine whether AMT expression in arabidopsis will yield a greater S/G ratio when compared to wild type. A greater S/G ratio in AMT transgenic plants would support the notion that AMT's preference for sinapyl alcohol might be the driving force behind the

formation of lignin with high syringyl content. This may have large implications for biofuels since lignin with high syringyl content may produce uncondensed lignin that is easier to process.

The AMT transgenic plants will be analyzed for cell wall digestibility at the Great Lakes Bioenergy Research Center (GLBRC) Cell Wall Analytical Facility to determine whether AMT expression has the potential to modify lignin structure and generate plants with increased cell wall digestibility. It may be that altering the level of acetylation on lignin may create a lignin structure that provides greater access to cellulose; thus, allowing for increased sugar yields. In addition, observing the physical and physiological characteristics of these plants could provide clues into the functional role of AMT *in planta*. Particularly, I would like to know whether AMT expression has an effect on stem strength or drought tolerance. If a reduction in stem strength is observed, this may suggest that lignin acetylation generates lignin that is easier to break down for access to cellulosic sugars. However, if stem strength is reduced to a significant degree, it may have negative effects on plant structure. These studies will provide crucial information for the implications of altering lignin acylation in plants. Furthermore, since lignin acetylation has been associated with drought tolerance (28), it would be interesting to determine whether the AMT transgenic plants are more drought tolerant than wild type. Overall, these future studies will shed light on AMT's functional role in kenaf and possibly other plant species.

CHAPTER TWO

Characterization of the Sweet Cherry Pollination System

2.1 INTRODUCTION

Self-incompatibility (SI) is a genetically controlled system for mate selection found in angiosperms that prevents inbreeding by the rejection of self-fertilization. SI also promotes out-crossing between genetically unrelated plants of the same species, which generates and maintains genetic diversity within the plant kingdom. This genetic diversity has allowed angiosperms to thrive for millions of years and led to their dominance of most terrestrial ecosystems (42). In gametophytic self-incompatibility (GSI), the SI reaction is determined by the haploid genotype of the pollen and its interaction with the diploid tissue of the pistil. Pollen is rejected if there is a match between the S-allele of the haploid pollen and one of the two *S*-alleles expressed in the pistil. GSI is controlled by a single multi-allelic *S*-locus that includes at least two genes: a pistil-*S* gene and a pollen-*S* gene, which control pistil and pollen specificity, respectively. The close proximity of the pistil-*S* and pollen-*S* genes on the *S*-locus allows for these genes to co-segregate as a single gene. Variants of the *S*-locus are referred to as a "haplotype" (43).

GSI is split into two systems, each of which is classified by its pistil-*S* determinant molecule. In "Papaveraceae-type SI", the pistil-*S* encodes a set of *S*-proteins, which are thought to be involved as signaling molecules that aid in the recognition of pollen-*S* allele specificity (44). The other system is generally referred to as S-RNase-based SI because the pistil-*S* encodes a ribonuclease (*S*-RNase), which exerts its cytotoxicity on pollen with an identical *S*-allele. S-RNase-based SI has also been coined "Solanaceae-type SI" because it was first discovered in *P. inflata* and then extensively characterized in several members of the Solanaceae family, including potato, tomato, and tobacco (45). In spite of this, members of the Rosaceae and Plantaginaceae families also exhibit S-RNase-based SI, which demonstrates how "Solanaceae-type SI" is an inaccurate label for S-RNase-based SI.

Genetic and molecular evidence for GSI in the Solanaceae and *Prunus*, a genus in the Rosaceae, indicates that there are major differences between the SI mechanisms of these groups. First, pollen grains containing two copies of the pollen-*S* allele give rise to different SI phenotypes in the Solanaceae and *Prunus* due to a phenomenon called competitive interaction, wherein breakdown of GSI occurs as a result of polyploidization. This phenomenon does not confer self-compatibility in *Prunus*, whereas in the Solanaceae, self-compatibility has been observed in species that have undergone genome duplication (46-48). Competitive interaction is dictated by the *S*-genotype of the pollen, such that the presence of two distinct *S*-alleles (heteroallelic) in the same pollen grain allows for full compatibility with a style of any *S*-allele type. However, a pollen grain containing two of the same *S*-alleles (homoallelic) is rejected by a style of the same *S*-genotype (Figure 2.1).

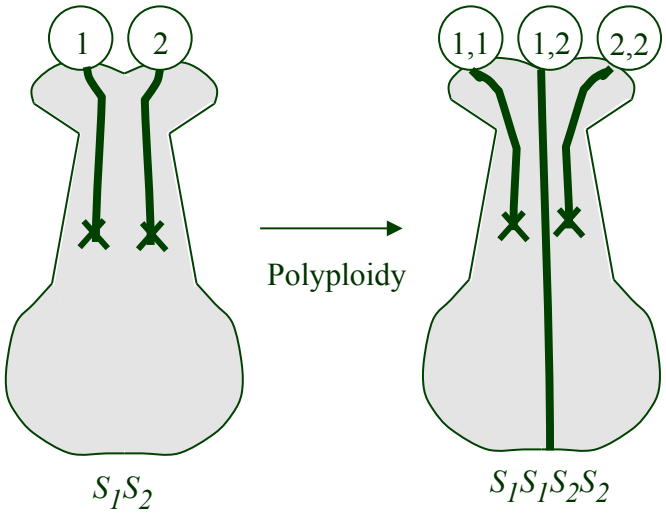


Figure 2.1 Genetic Representation for Competitive Interaction in the Solanaceae. A breakdown in pollen-part SI function confers self-compatibility in heteroallelic pollen. Conversely, homoallelic pollen retains its self-incompatibility and as a result, is rejected by the style. *Prunus* does not become self-compatible as a result of competitive interaction. This figure was adapted from Hauck et al. (2006) (49).

Secondly, studies on competitive interaction have demonstrated that pollen-part mutations (PPM) in *Prunus* and the Solanaceae lead to two distinct outcomes in these groups. PPM in *Prunus* confer self-compatibility, whereas PPM in the Solanaceae are predicted to be pollen-lethal (43,47-50). Studies in *Prunus* found that the breakdown of SI in tetraploid sour cherry was due to the accumulation of nonfunctional *S*-haplotypes in the pistil and pollen. These observations were used to develop the "one-allele-match" model, wherein any single functional *S*-allele match between pollen and pistil would result in an incompatible reaction (Figure 2.2) (49). In contrast, studies in *Nicotiana alata* have not been able to produce "true" pollen-*S* mutants, which suggests that the pollen-*S* is necessary for pollen viability and survival (48).

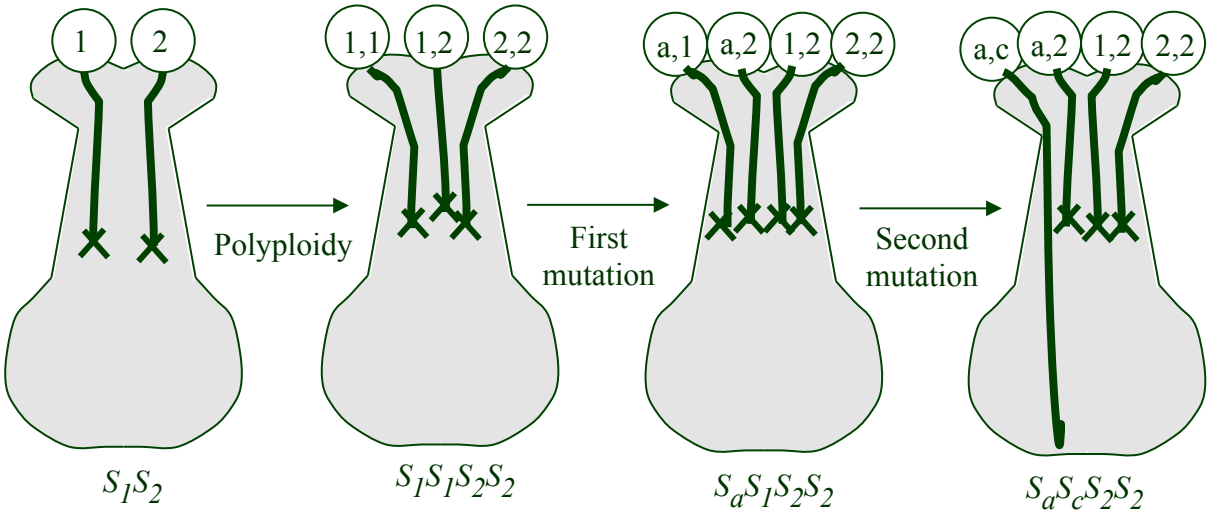


Figure 2.2 Genetic Representation for the "One-Allele-Match" Model in Tetraploid Sour Cherry. Two nonfunctional *S*-haplotypes must be present in order for a breakdown in SI to occur. Any single *S*-allele match between pollen and pistil confers self-incompatibility and subsequently, rejection of pollen by the style. Lowercase letters in the pollen and style indicate nonfunctional *S*-alleles. This figure was adapted from Hauck et al. (2006) (49).

Lastly, there is evidence to suggest that non-*S*-specific factors are involved in the SI mechanisms of Solanaceous species. HT-B, an asparagine-rich protein expressed in the style (51), and the 120 kDa (120K) glycoprotein, an arabinogalactan protein in the style that is taken up by pollen tubes (52,53), have been implicated in contributing to the SI response. RNA interference (RNAi)-mediated suppression of HT-B and the 120K protein in *Nicotiana* and *Solanum chacoense* both produced similar results in that *S*-specific pollen rejection was disrupted (51,54-56). Interestingly, these factors have not been identified in *Prunus.*

This study reports on an analysis of pollination in sweet cherry (*Prunus avium*). Sweet cherry was chosen as our model GSI system due to its regional availability and extensive characterization of its *S*-RNases and *SFB*s (57-60). First, a laboratory pollination method was established in order to determine an accurate timing of pollination in sweet cherry. Second, using the laboratory pollination method, a deep proteome profile of un-pollinated (control) and pollinated cherry style tissue was developed across different time points. It is known that the pollen-*S* encodes a F-box protein (*SFB*) (43,50,61,62), which is a component of the SCF complex that targets proteins for degradation via ubiquitination. Thus, a comprehensive analysis of the sweet cherry proteome may help identify targets of ubiquitination by *SFB*. In addition, this data set can be utilized in the identification of candidate genes involved in GSI. These studies have not yet been described in sweet cherry and as such, the data presented here will provide researchers with a framework for effective laboratory pollinations of sweet cherry and deep proteome profiling data for the identification of *S*-specific and non-*S*-specific factors involved in the SI response of *Prunus*.

2.2 MATERIALS AND METHODS

Plant Material

Four sweet cherry varieties were used for this study: 'Bing' (S_3S_4), 'Brooks' (S_1S_9), 'Emperor Francis' (EF; S_3S_4), and 'NY54' (NY; S_2S_6). All the plant material used in these experiments is planted at the Clarksville Horticultural Experiment Station, Clarksville, Mich., and the Southwest Michigan Research and Extension Center, Benton Harbor, Mich.

Pollen Collection

Cut branches were brought into the lab and forced indoors at room temperature. Open flowers were removed to avoid cross-pollination from field pollinators. Pollen was collected from newly opened flowers and allowed to dry for two days. After two days, the pollen was transferred into a glass vial, sealed in a Drierite-filled bag, and stored at 4 °C until use.

Pollen Germination Assays

Two drops of germination media (10% sucrose, 100 ppm boric acid, 300 ppm calcium nitrate, 200 ppm magnesium sulfate, 100 ppm potassium nitrate, pH 7.3, autoclaved 25 min) were placed onto a microscope slide using a Pasteur pipette. Two to three individual anthers were added to the germination media and crushed with a glass rod to release pollen. Microscope slides were placed on top of a Petri dish and into a plastic container filled with damp paper towels to control humidity. Samples were incubated for 2 h in the sealed, humid container. Cover slips were placed on the microscope slides and germination was observed using a compound microscope. Percent germination was recorded for a subset of free pollen observed for each pollen variety.

Pollination Assays

Pollinations were performed as described in Yamane et al. (2001) (63) with some modifications. Flowers to be used for the pollination studies were emasculated and hand-pollinated when receptive (24 h after emasculation). Pistils were pollinated in the lab with self- and non-self pollen. The control pistils were not pollinated. Pistil collection took place after the indicated time point (3, 6, and 12 h in 2011; 12 and 14 h in 2012). A sample of ten pistils per experiment was collected and transferred to a fixing solution (1 chloroform: 3 (95%) ethanol: 1 glacial acetic acid) for 24 h. The fixing solution was removed and the pistils were rinsed with 10N NaOH then submerged completely in 10N NaOH for 24 h. After removal of NaOH, the pistils were rinsed 5 times with distilled water before being exposed to 0.1% aniline blue solution (in 33 mM K_3PO_4) for 24 h. Pollen tubes were visualized using confocal laser scanning microscopy.

Pollen Tube Analysis via Confocal Laser Scanning Microscopy

An Olympus FluoView FV1000 confocal laser scanning microscope at the MSU Center for Advanced Microscopy was used to analyze stained pollen tubes. Fluorescence images were acquired using the "DAPI" DyeList selection. Differential Interference Contrast (DIC) was used in combination with fluorescence imaging to acquire overlay images of the pollen tubes. Pollen tubes were mounted onto microscope slides in 30% glycerol at the time of analysis. All pollen tubes were analyzed within 1-week of initial staining.

Protein Extraction of Cherry Style Tissue

Pistils were ground to a fine powder in liquid nitrogen, enough to fill up to the 500 µL mark of a 1.5 mL microcentrifuge tube. Approximately 450 µL of 4X SDS buffer (50 mM Tris-HCl, pH

6.8, 20% SDS, 40% glycerol, 0.1% bromophenol blue, and 2% beta-mercaptoethanol) was added to cover the ground plant tissue and immediately vortexed until the plant material was fully dissolved in SDS buffer. The samples were then sonicated for 15 min and stored at 4 °C until use. The cherry protein samples were analyzed for degradation by SDS-PAGE, wherein a standard of volumes was loaded on a SDS-PAGE gel to quantify protein abundance in each sample. The protein extraction experiments were performed in duplicate to obtain two biological replicates for proteome analyses.

Experimental LC/MS/MS

SDS-PAGE gels containing cherry style protein from control, non-self and self pollination experiments were subjected to in-gel tryptic digest according to Shevchenko et al. (1996) (64) with modifications, followed by LC/MS/MS at the MSU Proteomics Core. Data processing of the LC/MS/MS peptide fragments was performed using BioWorks Browser, which generated a Mascot Generic Format (MGF) text file utilized for peptide identification against the cherry database. The results were loaded into Scaffold, v3.6.0 (Proteome Software), a tool that compiles large amounts of proteomic data for visualization and analysis. Scaffold was utilized to filter the proteomic data for the identification of post-translational modifications, such as phosphorylation and ubiquitination, in cherry style protein.

2.3 RESULTS

Pollen Germination Assays

The pollen germination experiments were performed over two harvest seasons. In 2011, the pollen germination range was 20% – 35% for pollen collected from Brooks, EF, and NY54. In 2012, the pollen germination range was 11% – 37% for pollen collected from these same cherry varieties. Pollen with a germination percentage of over 20% was used for the pollination experiments. The cherry variety used for the pollen parent was dependent on the availability of the maternal parent. In 2011, the pollination experiments were EF (S_3S_4) × EF (S_3S_4) and EF (S_3S_4) × Brooks (S_1S_9). In 2012, the pollination experiments were EF (S_3S_4) × EF (S_3S_4), EF (S_3S_4) × Brooks (S_1S_9), Bing (S_3S_4) × EF (S_3S_4), and Bing (S_3S_4) × Brooks (S_1S_9).

Table 2.1 Pollen Germination Percentage Values for Pollen Parents of Pollination Experiments

Pollen Parent	*S*-Genotype	Germination Percent
EF_1*	S_3S_4	25.0
	S_3S_4	36.5
EF_2	S_3S_4	18.8
EF_3	S_3S_4	23.1
EF_4	S_3S_4	18.0
Brooks_1*	S_1S_9	24.5
	S_1S_9	24.3
Brooks_2	S_1S_9	10.5
Brooks_3	S_1S_9	23.4

Note: Pollen parents were numbered to denote separate pollen grain collections. Shaded boxes represent a pollen grain collection that was combined with the preceding pollen sample. The asterisk denotes pollen parent samples that were used in the 2012 pollination experiments.

Pollination and Pollen Tube Growth Analysis

In previous years, pollination experiments were not successful due to a complete lack of pollen grains adhered to the stigma surface. The likely cause of this was that the cherry styles were pollinated before full maturity; thus, affecting the production of exudates at the stigma surface for pollen grain adhesion. In 2011 and 2012, cherry styles were prepared 24 h in advance and examined for exudates production prior to beginning the pollination experiments. A comparative analysis of stained control (unpollinated) and pollinated cherry styles confirmed that the pollination experiments were successful (Figure 2.3). Stained cherry styles from the control experiments revealed an absence of pollen grains on the stigma surface (Figure 2.3A – B). Pollen grains and pollen tubes were identified on stained cherry styles from both incompatible and compatible pollination experiments, indicating that pollen grain germination and pollen tube growth had taken place. Figure 2.3C – D illustrates germinated pollen grains at the stigma surface and pollen tubes growing from the pollen grains in cherry styles from an incompatible pollination. Similar results were observed in cherry styles from compatible pollinations. The pollen grains were circular and spread across the top of the stigma (Figures 2.4A and 2.5). Figure 2.4C – D shows the presence of pollen grains on the styles of two different pollination experiments. These figures also demonstrate the absence and presence of pollen grain germination, indicating that germination did not occur as a direct result of pollen grain adhesion to the stigma. As such, it was critical to analyze a subset of cherry styles from all control and pollination experiments to ensure that pollen grains had germinated. The pollen tubes appeared thin and irregularly shaped in comparison to the stylar transmitting tissue, which was noticeably thicker and less irregular in shape as demonstrated in Figure 2.6.

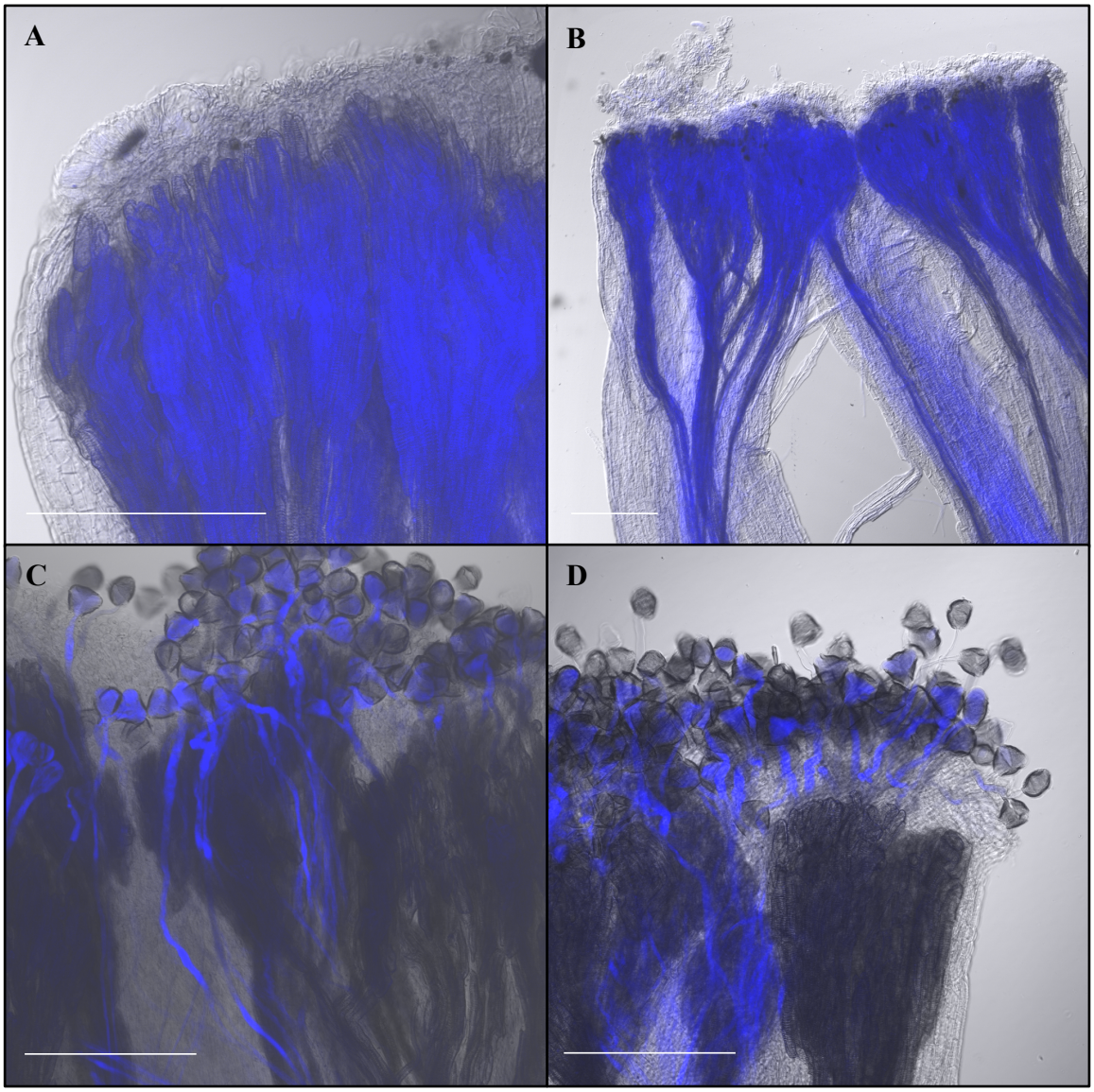


Figure 2.3 Control and Pollinated Cherry Styles Illustration of the difference between control (unpollinated) and pollinated cherry styles. *A)* Bing (S_3S_4) 24 h control, *B)* EF (S_3S_4) 24 h control, *C)* 12 h Bing (S_3S_4) × EF (S_3S_4), and *D)* 24 h EF (S_3S_4) × EF (S_3S_4). For this and all other confocal microscope figures, overlay images were produced from fluorescence and Differential Interference Contrast (DIC) images. All scale bars are 200 µm.

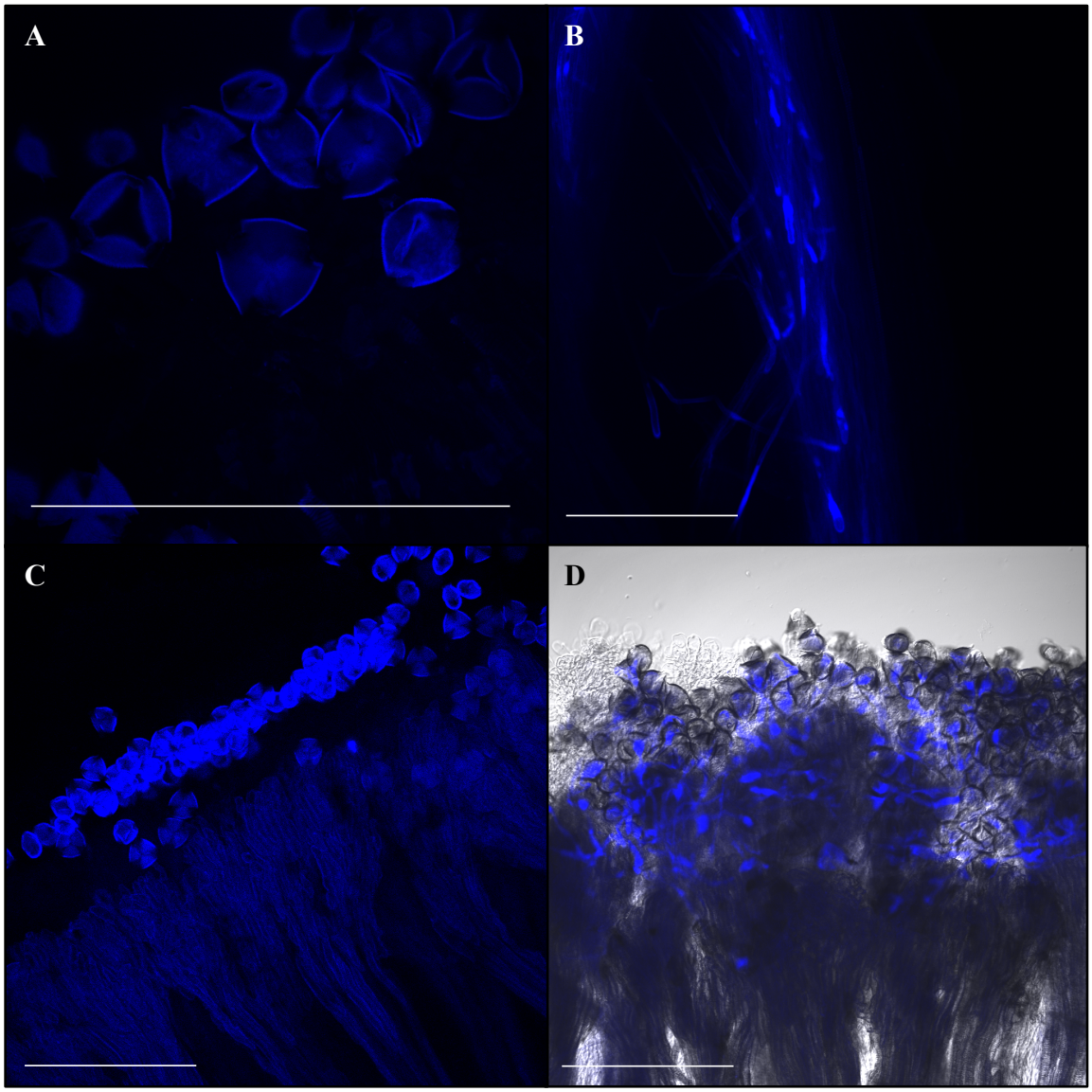


Figure 2.4 Pollen Grain Adhesion on Cherry Style and Pollen Germination Images show adhesion of pollen grains on the stigma surface, pollen germination, and pollen tube growth through the stigma surface. *A)* Pollen grains on the surface of a style from 12 h EF (S_3S_4) × Brooks (S_1S_9), *B)* Pollen tube growth was arrested mid style within 24 h EF (S_3S_4) × EF (S_3S_4), *C)* Non-germinated pollen grains on a style in 12 h EF (S_3S_4) × NY (S_2S_6), and *D)* Germinated pollen grains on a style in 12 h EF (S_3S_4) × EF (S_3S_4).

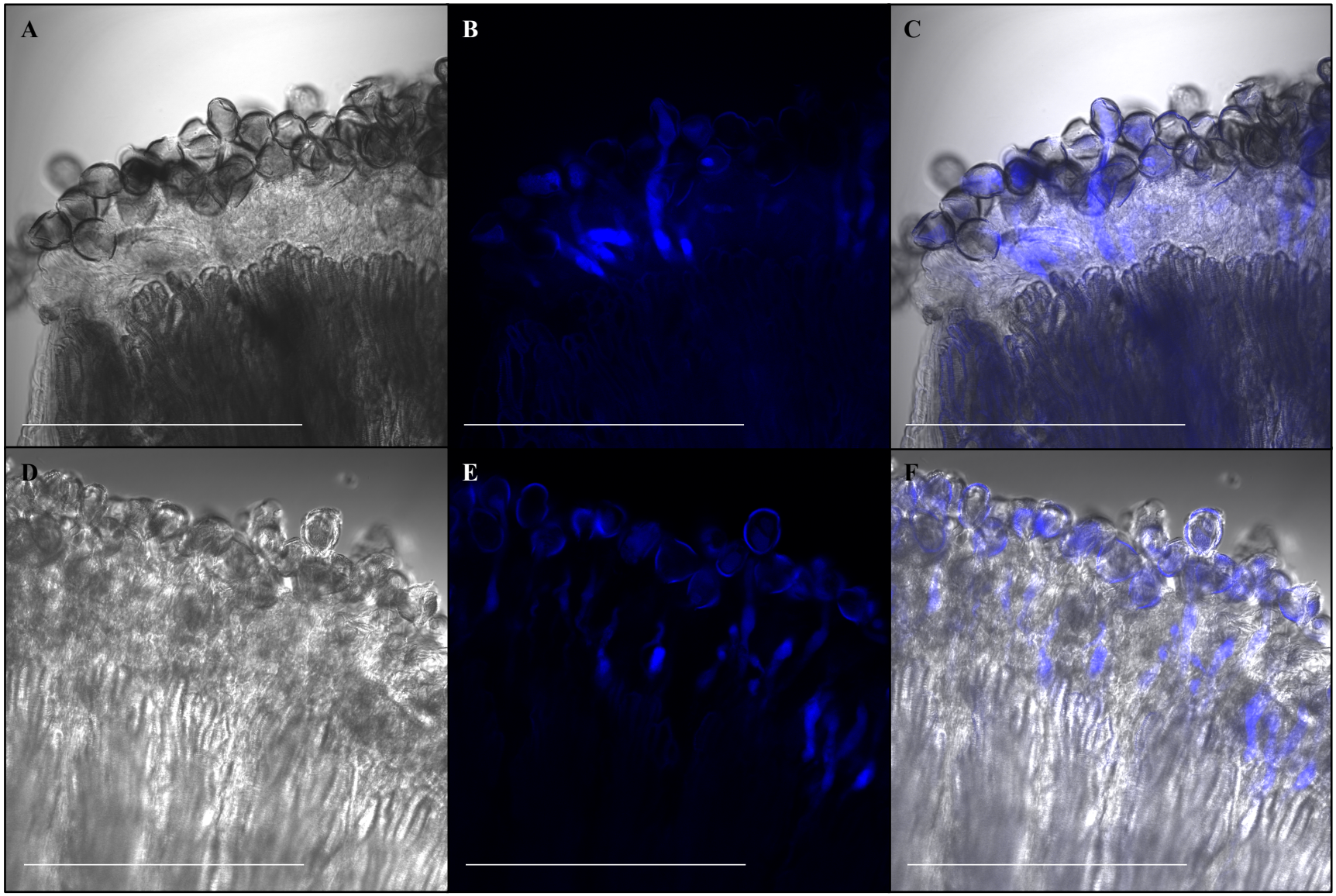


Figure 2.5 Pollen Germination on Cherry Style Surface *A-C)* 24 h EF (S_3S_4) × NY (S_2S_6) and *D-F)* 12 h Bing (S_3S_4) × EF (S_3S_4).

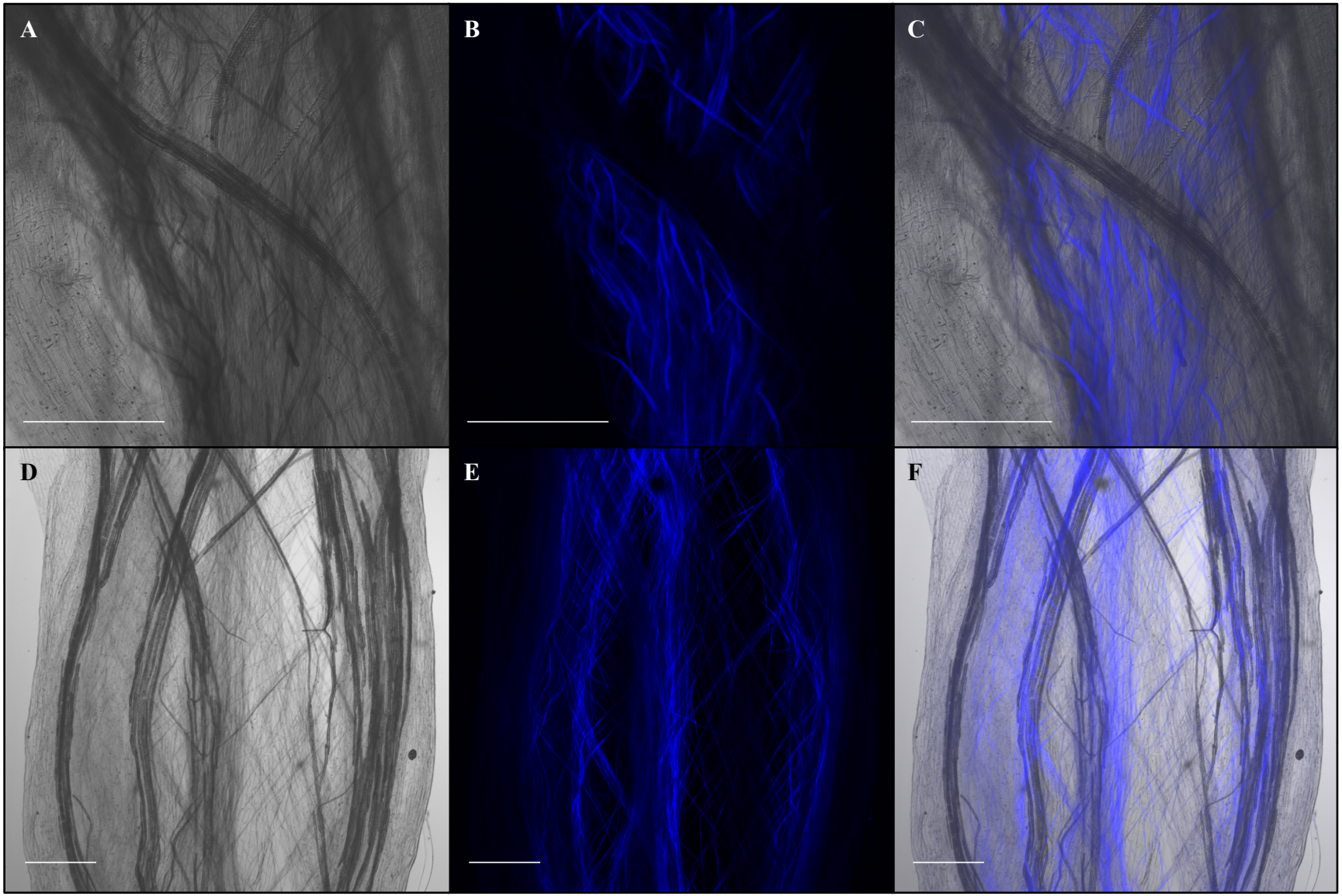


Figure 2.6 Pollen Tube Growth within Cherry Style *A-C)* 24 h EF (S_3S_4) × EF (S_3S_4) and *D-F)* 12 h Bing (S_3S_4) × Brooks (S_3S_4).

Post-Pollination Time Course Analysis

It was crucial to determine an appropriate post-pollination harvest time that would allow for analysis of fully-grown pollen tubes. Initially, pollinated cherry styles were collected at 3 and 6 h post-pollination. However, these time ranges were not sufficient in length to allow for pollen grain germination and subsequent pollen tube growth down the style. These cherry styles had marginal or completely lacked pollen tube growth. In the rare case of pollen tube growth, the pollen grains had begun to germinate but pollen tubes had not yet penetrated the stigma surface. Cherry styles harvested at 12 and 24 h post-pollination revealed that these time ranges were appropriate for detection of pollen tube growth (Figure 2.3C – D). Pollen tube growth in cherry styles from the 12 h compatible and incompatible pollinations was measured at the lower 1/3 to 1/2 portion of the style. At 24 h, full pollen tube growth was observed in cherry styles from the compatible pollination experiments only (Figure 2.7). Cherry styles from the 24 h incompatible pollination experiments exhibited pollen tube growth; however, these pollen tubes did not grow beyond half the length of the style (Figures 2.4B and 2.8C – D). This observation is commonly referred to as the "abortion phenomenon", wherein pollen tube growth is arrested approximately halfway down the style in incompatible pollinations (65).

Proteomics Analysis of Unpollinated and Pollinated Cherry Styles

A representation of the proteomics data can be found in Appendix B. This subset of the proteomics data, along with all other sets for these experiments, illustrates the number of spectra detected within each pollination experiment (control, compatible, and incompatible) for each identified protein. Identified proteins were annotated according to the cherry EST database, which utilizes TAIR to annotate ESTs according to their closest homolog in arabidopsis.

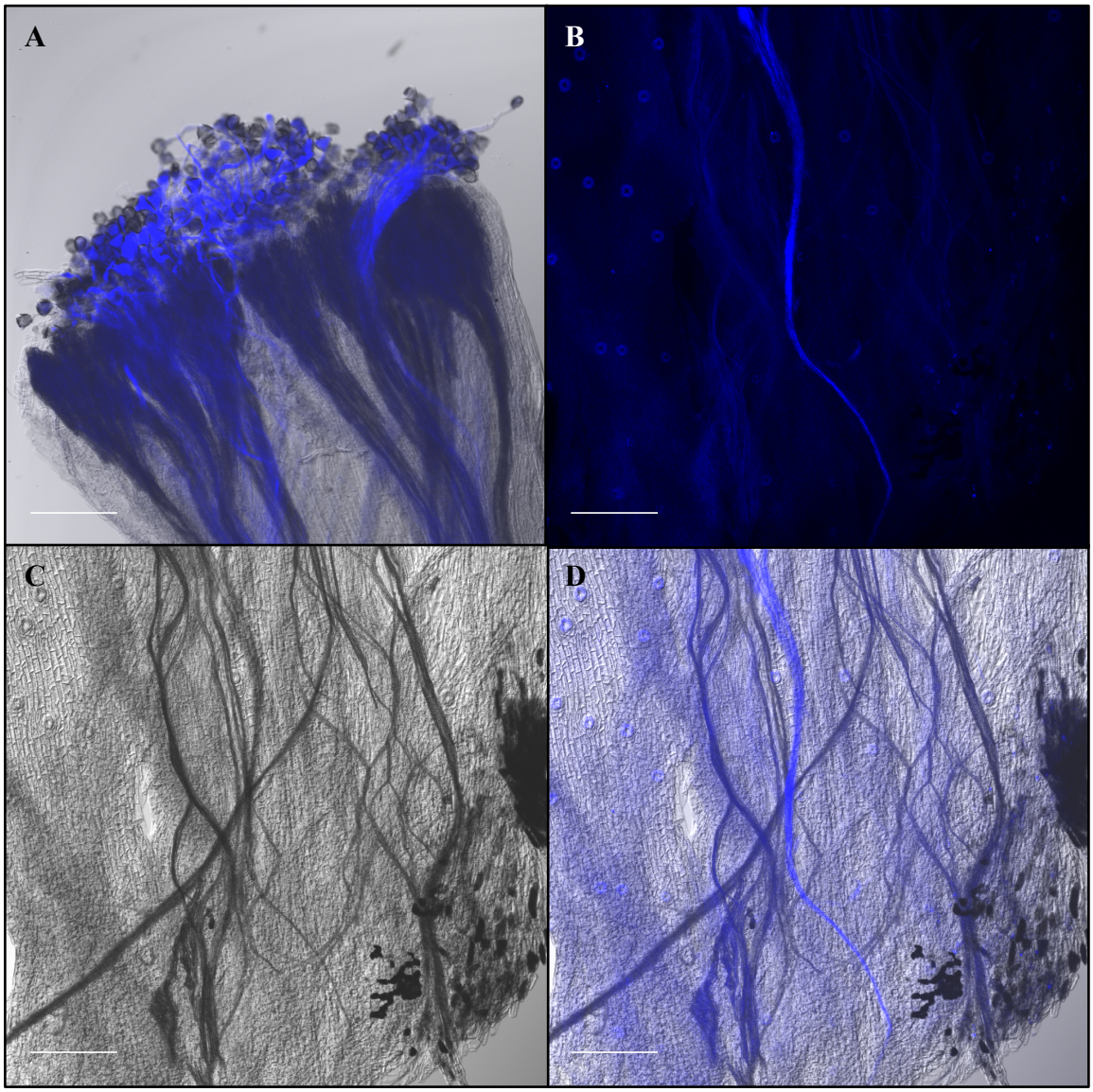


Figure 2.7 Full Growth of Pollen Tube Down Cherry Style *A)* Overlay image of pollen grains on stigma surface and pollen tube growth in 24 h EF (S_3S_4) × Brooks (S_1S_9). Pollen tubes grew down the entire length of the cherry style in 24 h EF (S_3S_4) × Brooks (S_1S_9) as shown by *B)* fluorescence, *C)* DIC, and *D)* fluorescence-DIC overlay images.

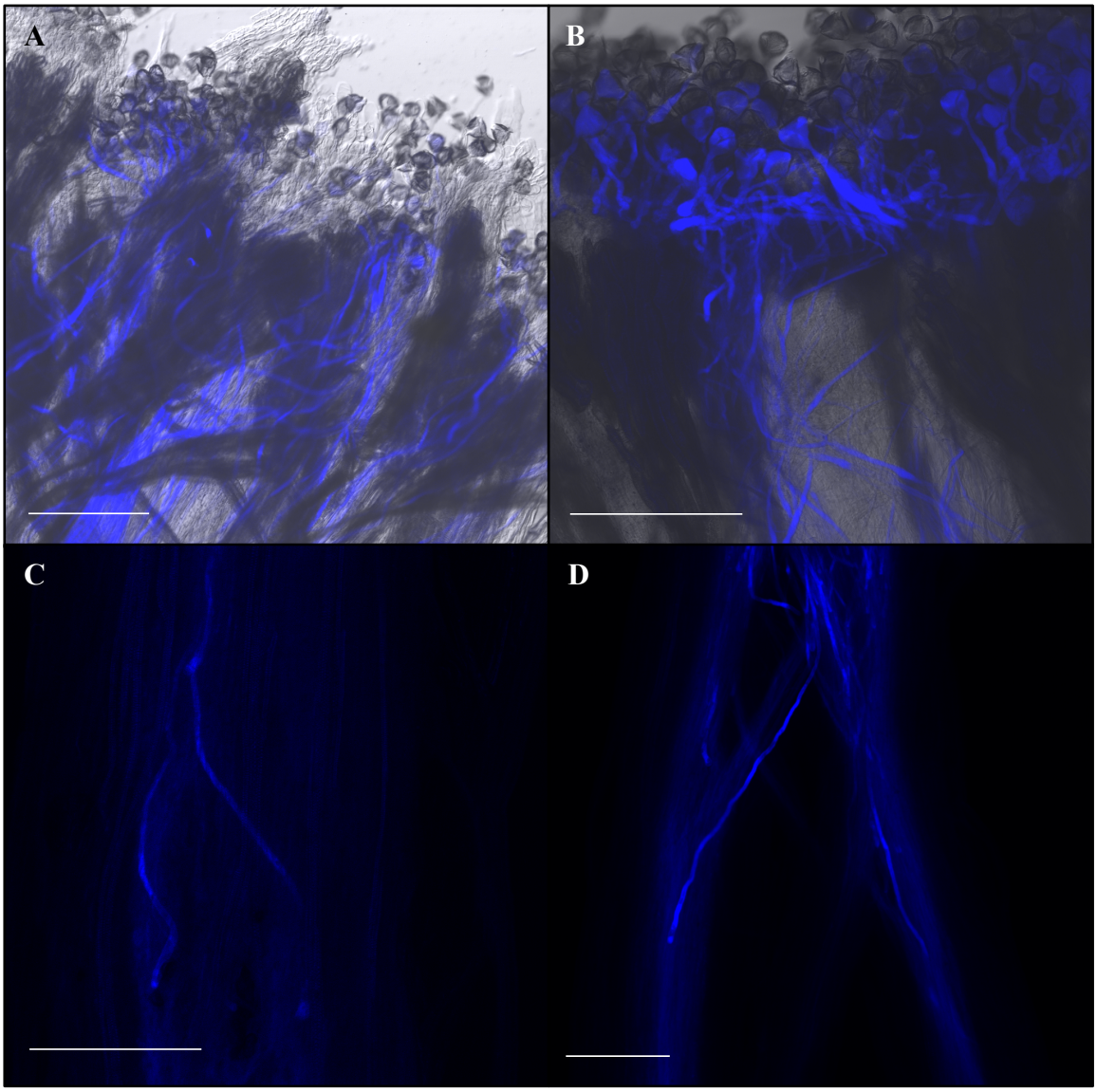


Figure 2.8 Arrest of Pollen Tube Growth at Mid Stylar Region Overlay images of pollen grains on stigma surface and pollen tube growth in incompatible pollinations. *A)* 24 h Bing (S_3S_4) × EF (S_3S_4) and *B)* 24 h EF (S_3S_4) × EF (S_3S_4). Pollen tube growth was arrested at mid style in *C)* 24 h Bing (S_3S_4) × EF (S_3S_4) and *B)* 24 h EF (S_3S_4) × EF (S_3S_4).

2.4 DISCUSSION

Interestingly, there were notable differences in pollen tube growth for the pollination experiments performed in 2012. At 12 h, pollen tube growth for EF (S_3S_4) × EF (S_3S_4) was at 1/3 down the style, whereas in Bing (S_3S_4) × EF (S_3S_4), pollen tube growth was at 1/3 to 1/2 down the style. In pollen tubes for EF (S_3S_4) × Brooks (S_1S_9), growth was observed at 1/2 down the style, while growth was observed at 3/4 down the style in Bing (S_3S_4) × Brooks (S_1S_9) pollen tubes. The same Brooks and EF pollen samples were used in all the pollination experiments performed. Thus, the differences observed in pollen tube growth for each time point experiment cannot be largely attributed to inconsistencies across pollen samples. Also, the Bing pollination experiments were performed days after the EF experiments, which would suggest that pollen viability might have decreased during this time. However, longer pollen tube growth was observed in the Bing experiments, indicating that pollen viability had not decreased significantly. It may be that the Bing styles were more receptive than the EF styles, allowing pollen grains to germinate and begin the growth process sooner in the Bing styles. This could explain why pollen tubes grew longer in the Bing experiments, even though similar time points were used and pollen viability may have been reduced. Nevertheless, it would be interesting to determine whether the maternal parent may be more receptive to pollen from particular cherry varieties.

In depth proteome analysis gave interesting results from the control and experimental pollinations that lend support to the *S*-RNase sequestration model. The sequestration model suggests that *S*-RNases enter the pollen tube by endocytosis and undergo retrograde transport to vacuoles. Studies on *S*-RNase uptake have demonstrated that *S*-RNases enter pollen tubes in a genotype independent manner (66) and are sequestered within a vacuolar compartment in pollen

tubes (65). It has also been shown that the *S*-RNase-containing compartment disrupts late in incompatible pollinations (65). The observation that incompatible pollen tubes exhibit compartment disruption suggests that *S*-RNases are able to exert their cytotoxic activity upon release into the pollen cytosol. A similar mechanism has been found in the cytotoxin ricin from the castor oil plant *Ricinus communis*, which enters the cell by endocytosis and is sorted to a vacuole (65,67). It may be that the mechanism exhibited by ricin is a good model for *S*-RNase compartmentalization in GSI. In previous years, the mono-ubiquitination of S_4-RNase was detected in a proteome analysis of control styles. This finding was fascinating because mono-ubiquitinated proteins in yeast are targeted to the vacuole instead of the proteasome (68). It may be that *S*-RNase mono-ubiquitination may target *S*-RNases to the vacuole, wherein *S*-RNases are sequestered until release into the pollen cytosol.

Other proteins annotated as "aerolysin pore-forming" and "porin" in the proteomics data may be involved in the release of sequestered *S*-RNases in pollen tubes. The finding of an aerolysin protein was interesting in that these pore-forming toxins are secreted as soluble proteins and eventually become a transmembrane channel in the target cell (69). Spectra for the aerolysin protein were 1.5X greater in incompatible styles than control and compatible styles. It could be that an aerolysin protein is activated in incompatible pollinations to form pores for the release of *S*-RNases into the pollen tube cytosol. A bioinformatic analysis of the "porin" protein showed that it was specific to *Prunus* and the proteomics data demonstrated that it was twice as abundant in incompatible styles than control styles. This raises the question of whether the porin protein is specific to *S*-RNase release in *Prunus*. Overall, the candidate proteins presented here, and others in the proteomics data, warrant further study for possible associations with *S*-specific and non-*S*-specific mechanisms in *Prunus*.

To conclude, it is clear that distinct SI mechanisms may exist between *Prunus* and the Solanaceae. In Solanaceous species, competitive interaction confers self-compatibility and PPMs are predicted to be pollen-lethal. Conversely, in *Prunus*, genome duplication does not confer self-compatibility, whereas PPMs confer self-compatibility according to the "one-allele-match" model. In order to gain insight on GSI in *Prunus*, this study provides an in-depth characterization of the sweet cherry pollination system and a deep proteome analysis of cherry style tissue under different experimental conditions. Previous work on this project has also generated an expressed sequence tag (EST) database of cherry styles. Future studies in sweet cherry can utilize the data presented here to establish an accurate timing for laboratory pollinations and explore the sweet cherry proteome and transcriptome for the identification of *S*-specific and non-*S*-specific factors involved in the SI response of *Prunus*.

2.5 FUTURE STUDIES

The next step for characterization of the sweet cherry pollination system will involve cell biological techniques to elucidate on *S*-RNase uptake and sequestration in sweet cherry. First, it will be important to establish whether *S*-RNase uptake into pollen tubes occurs in a *S*-genotype independent manner. Luu et al. (2000) utilized immunocytochemical labeling of pollen tubes from *Solanum chacoense* and *S*-RNase antibodies to determine that *S*-RNases were present in the cytosol of pollen tubes from compatible and incompatible crosses, demonstrating that *S*-RNase uptake occurs in a *S*-genotype independent manner (66). The mechanism of *S*-RNase uptake has not been described in sweet cherry and as such, it would be interesting to confirm whether *S*-RNase uptake in sweet cherry is similar to what has been described in the Solanaceae.

Second, it is critical to find out what occurs to *S*-RNases once they have entered the pollen tube. Goldraij et al. (2006) demonstrated that *S*-RNases were compartmentalized in a *S*-genotype dependent manner within a vacuolar compartment in *Nicotiana alata* pollen tubes (65). Using immunolabeling techniques and *S*-RNase antibodies, *S*-RNase compartmentalization was shown to occur in compatible pollen tubes, while compartment breakdown was observed in incompatible pollen tubes. These findings suggest that the *S*-RNase sequestration model may explain the mechanism of GSI in the Solanaceae. However, the same argument cannot be made for the mechanism of GSI in sweet cherry because *S*-RNase sequestration studies have not been performed in this system. As such, it will be important to determine whether *S*-RNases remain sequestered in the vacuole of pollen tubes in a *S*-genotype dependent manner.

Studies on *S*-RNase uptake and sequestration in sweet cherry can be performed by immunolabeling techniques similar to those described above. Candidate genes can be identified using the cherry EST database for the production of antibodies. These should include *S*-RNases

of different *S*-genotypes and membrane proteins found in the plasma membrane and tonoplast, such as "plasma membrane intrinsic protein" (PIP) and "tonoplast intrinsic protein" (TIP). The nucleotide sequences for these genes can be codon-optimized and synthesized by GENEART for Gateway cloning and subsequent heterologous expression in *E. coli*. Purified soluble protein obtained from *E. coli* expression can be sent to GenScript (Piscataway, New Jersey) for custom antibody production. Additionally, it would be beneficial to obtain an anti-callose antibody from Biosupplies Australia for the identification of callose in the walls of pollen tubes.

The antibodies for *S*-RNases, callose, plasma membrane and tonoplast proteins, along with confocal microscopy to detect immunofluorescence, can be utilized to analyze cherry style tissue for determination of whether *S*-RNase uptake occurs in a *S*-genotype independent manner and whether *S*-RNases are sequestered in a vacuolar compartment within compatible pollen tubes. Cherry styles from different time points can be analyzed by immunofluorescence to obtain a time course for *S*-RNase uptake and compartmentalization. Further, *S*-RNase compartment breakdown might be observed in incompatible pollen tubes, if *S*-RNase compartmentalization is similar to the mechanism described in the Solanaceae.

It may be that *S*-RNase uptake and sequestration in *Prunus* is similar to the Solanaceae. However, there is evidence to suggest that different biochemical factors may be involved in the recognition and rejection processes. HT-B, an asparagine-rich protein expressed in the style, has been implicated in the SI response in *Nicotiana* and *S. chacoense* (51,54-56). Yet, HT-B has not been identified in *Prunus.* Thus, it is important to determine the mechanism of *S*-RNase uptake and sequestration in *Prunus* in order to carry out further analyses on the biochemical factors involved in GSI. These future studies will have significant impacts on plant reproductive biology in that they will help elucidate the mechanism of GSI in *Prunus*.

APPENDICES

APPENDIX A

Identification of an O-Methyltransferase Involved in Monolignol Biosynthesis in Rice

A.1 INTRODUCTION

The significance of lignin acylation has not been established primarily due to not knowing the enzyme(s) responsible for acylation of monolignols. This study addresses the question of the role of lignin acylation by identifying the enzymes involved in acylation reactions. Our laboratory has identified a *p*-coumaroyl monolignol transferase in *Oryza sativa* (OsPMT) as an enzyme implicated in the *p*-coumaroylation of monolignols, a phenomenon that is unique to grasses (17-19,70). OsPMT was most effective at producing sinapyl *p*-coumarate and *p*-coumaryl *p*-coumarate. It has been proposed that OsPMT adds *p*-coumarates onto monolignols after the reduction of hydroxycinnamyl intermediates by F5H and CAD. Interestingly though, OsPMT co-expresses with C3H rather than F5H and CAD. Additionally, while OsPMT has a higher affinity (lower K_m) for sinapyl alcohol, OsPMT has a faster reaction rate (higher V_{max}) with *p*-coumaryl alcohol. This raises the question of whether the transesterification of *p*-coumaroyl CoA by HCT is bypassed and *p*-coumaryl *p*-coumarate is used as substrate for the synthesis of monolignols in grasses. It may be that caffeyl *p*-coumarate or *p*-coumaryl caffeate, the products of 3′-hydroxylation of *p*-coumaryl *p*-coumarate are the targets of 3′-O-methylation activity by an O-methyltransferase (OMT) to produce coniferyl *p*-coumarate and *p*-coumaryl ferulate, respectively. Monolignols would then be produced by removal of the *p*-coumarate ester by OsPMT (Figure A.1).

This study reports on a grass-specific OMT found in *O. sativa*, herein called OsOMT, which is co-expressed with several monolignol biosynthetic genes and catalyzes the synthesis of coniferyl *p*-coumarate and *p*-coumaryl ferulate. The results show that OsOMT may be involved in the 3′-O-methylation of monolignol conjugates in grasses. Furthermore, these observations support the hypothesis that grasses may have an alternative pathway to lignin biosynthesis.

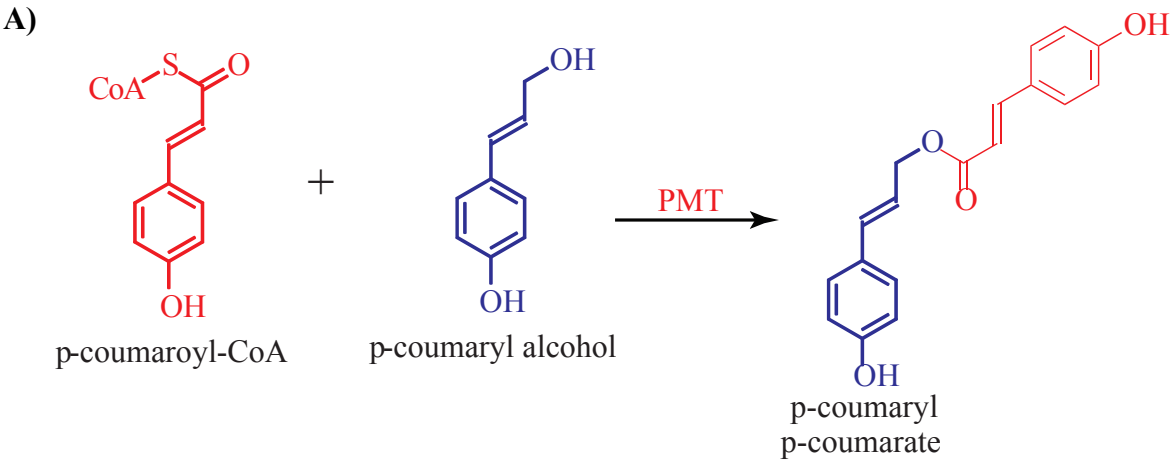
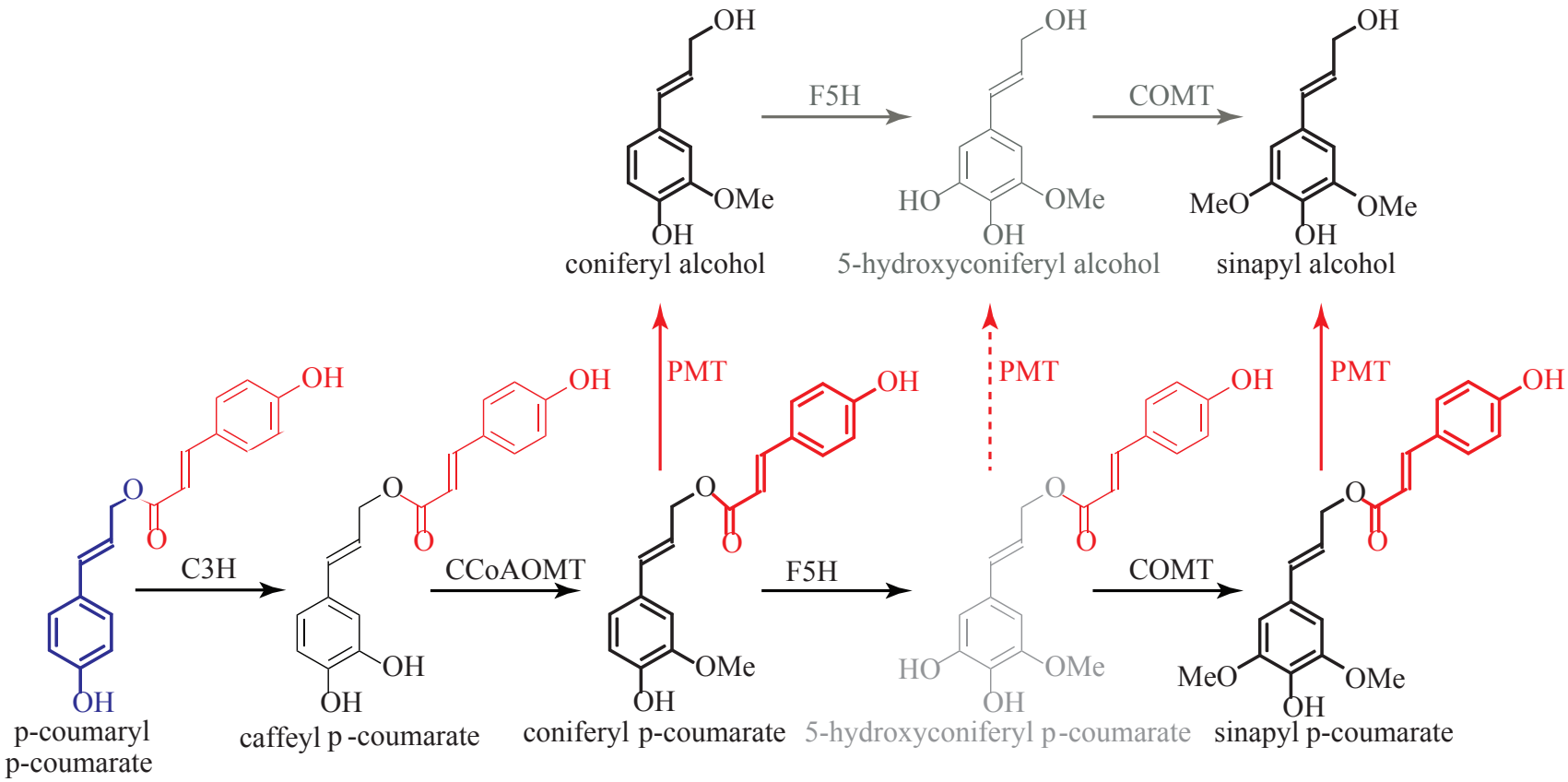


Figure A.1 Hypothetical Monolignol Biosynthetic Pathway in Grasses. *A*, OsPMT effectively utilizes *p*-coumaroyl CoA and *p*-coumaroyl alcohol to produce *p*-coumaryl *p*-coumarate. *B*, *p*-Coumaryl *p*-coumarate is hypothesized to be the target of 3′-hydroxylation by C3H to produce caffeyl *p*-coumarate or *p*-coumaryl caffeate. The reaction scheme presented here implicates the synthesis of caffeyl *p*-coumarate. After C3H activity, the monolignol conjugates would then undergo 3′-O-methylation by an OMT to produce coniferyl *p*-coumarate or *p*-coumaryl ferulate. Monolignols would then be produced by removal of the *p*-coumarate ester by OsPMT. The original CoA thioester compounds and monolignol alcohols are highlighted in red and blue, respectively. The dashed arrow represents a reaction that has not yet been tested using OsPMT. Compounds that are not commonly observed in grass lignin are highlighted in grey. This figure was adapted from Dr. John Ralph.

Figure A.1 (cont'd)

B)



A.2 MATERIALS AND METHODS

Rice Co-Expression Analysis

The Rice Oligonucleotide Array Database (www.ricearray.org) was utilized to identify OMT gene candidates that co-expressed with monolignol biosynthetic genes, such as 4CL. Other genes of interest were annotated as putative O-methyltransferases. Protein BLAST searches were performed to identify OMT gene candidate homologs in arabidopsis.

Gene Synthesis and Gateway Cloning

OsOMT was submitted to GENEART for *E. coli* codon optimization and chemical synthesis. Gateway cloning for OsOMT was performed as previously described (pg. 11) with one modification. The His-tagged expression clone was transformed into BL21 chemically competent cells (Invitrogen, Carlsbad, California). ATOMT1 was cloned into pDEST17 from arabidopsis stem cDNA and transformed into BL21 pLysS chemically competent cells.

Heterologous Expression in E. coli

OsOMT protein expression was performed as described on pg. 12 with some modifications. One 100 mL culture was used to grow the His-tagged OsOMT clone. After IPTG induction, the culture was grown for 4 h at 18 °C. ATOMT1 protein was expressed as described above; however, ATOMT1 was grown for 6 h at 37 °C after IPTG induction.

Protein Purification by FPLC

Five milliliters of soluble His-tagged OsOMT and ATOMT1 protein were used for FPLC purification. FPLC purification was performed as described on pg. 12 with a minor change to the

elution buffer for both protein purifications. The imidazole concentration of buffer A was decreased to 20 mM. Protein bands resulting from each purified fraction were compared to the expected molecular weight for OsOMT and ATOMT1 (ca. 40 kDa). Protein bands at about 40 kDa were extracted from the SDS-PAGE and analyzed by in-gel trypsin digestion and LC/MS/MS. OsOMT peptides were identified against rice and *E.coli* databases using Mascot. ATOMT1 peptides were identified using arabidopsis and *E.coli* databases. Fractions with the proteins of interest were buffer exchanged into 100 mM phosphate buffer, pH 6 (with protease inhibitors) using an Amicon 10K membrane filter. The enzymes were stored at -80 °C until use.

Enzyme Activity Assays

OsOMT and ATOMT1 activity assays were performed in 50 mM phosphate buffer, pH 6.8, 0.5 mM S-adenosyl methionine (SAM), 0.5 mM $MgCl_2$, 0.5 mM caffeyl substrate, and deionized water to a volume of 100 μL. Coupled enzyme assays using OsPMT and either OsOMT or ATOMT1 were performed in 50 mM phosphate buffer, pH 6, 1 mM DTT, 0.5 mM SAM, 0.5 mM CoA thioester, 0.5 mM monolignol, and deionized water to a volume of 100 μL. Enzyme activity assays were performed as previously described with a few modifications. All reactions were incubated for 30 min, stopped with 2 μL of HCl, and then solubilized with a 1:1 addition of 100% methanol. OsPMT was used to synthesize caffeyl *p*-coumarate using caffeyl alcohol and *p*-coumaroyl CoA, and *p*-coumaryl caffeate was synthesized using *p*-coumaryl alcohol and caffeoyl CoA. Caffeyl *p*-coumarate was also provided by Dr. John Ralph (33). CoA thioester synthesis was performed as previously described on pg. 13.

A.3 RESULTS

Identification of the Candidate Gene

Co-expression analysis was performed using the Rice Oligonucleotide Array Database to find an OMT candidate gene that was co-expressed with several monolignol biosynthetic genes. 4CL-2 from rice (LOC_Os02g08100) was utilized as a query for co-expression analysis because it encodes an isozyme that has been implicated in lignin formation (71). Thus, 4CL-2 was expected to be co-expressed with other monolignol biosynthetic genes. The co-expression analysis demonstrated a positive correlation ($R^2 = 0.88$) between 4CL-2 and an OMT candidate gene (LOC_Os08g06100). This candidate gene was annotated as a putative OMT in several grass species within the Poaceae family, such as maize, miscanthus, and sorghum. A protein BLAST of the OMT candidate against arabidopsis generated the most closely related gene with 58% similarity as ATOMT1 (AT5G54160), which was annotated as a caffeate OMT (Figure A.2).

Heterologous Expression in E. coli

A His-tagged OsOMT clone was produced using a codon-optimized synthetic gene, whereas a His-tagged ATOMT1 clone was synthesized from arabidopsis stem cDNA. These clones were utilized for heterologous protein expression in *E. coli*. OsOMT and ATOMT1 were purified from *E. coli* extracts using IMAC and analyzed by SDS-PAGE. Induced protein samples before and after FPLC purification demonstrated protein bands at a molecular weight of approximately 40 kDa, as expected for both OsOMT and ATOMT1 (Figure A.3). The un-induced controls did not show induction of the proteins of interest. SDS-PAGE analysis of the two proteins demonstrated that they are very soluble, even under a short induction time. LC/MS/MS analysis confirmed that the protein bands were in fact OsOMT and ATOMT1.

```
Majority    MGSTA.........D.EA...A.QLAS.S.LPM.LK.A.EL.LLE.........
ATOMT1.pro MGSTAETQLTPVQVTDDEAALFAMQLASASVLPMALKSALELDLLEIMAKNGSP-   54
OsOMT.pro  MGSTAADMAA--AA-DEEACMYALQLASSSILPMTLKNAIELGLLETLQSAAVAG   52

Majority    .........P.E.A.KLP.K.NP.A..M.DR.LRLL.SY.V..C...........
ATOMT1.pro -------MSPTEIASKLPTK-NPEAPVMLDRILRLLTSYSVLTCSNRKLSGDGVE  101
OsOMT.pro  GGGKAALLTPAEVADKLPSKANPAAADMVDRMLRLLASYNVVRCEMEEGADGKLS  107

Majority    R.Y...PVCK.LT.NEDGVS.AAL.LMNQDKVLMESWY.LKDA.LDGGIPFNKAY
ATOMT1.pro RIYGLGPVCKYLTKNEDGVSIAALCLMNQDKVLMESWYHLKDAILDGGIPFNKAY  156
OsOMT.pro  RRYAAAPVCKWLTPNEDGVSMAALALMNQDKVLMESWYYLKDAVLDGGIPFNKAY  162

Majority    GM.AFEYHGTD.RFN.VFN.GM.NHS.I..KK.L..Y.GF.....VDVGGG.GA
ATOMT1.pro GMSAFEYHGTDPRFNKVFNNGMSNHSTITMKKILETYKGFEGLTSLVDVGGGIGA  211
OsOMT.pro  GMTAFEYHGTDARFNRVFNEGMKNHSVIITKKLLDLYTGFDAASTVVDVGGGVGA  217

Majority    T....VS..P...GIN.DLPHVI..AP..PG.EHVGGDMF.SVP.G.DAI.MKWI
ATOMT1.pro TLKMIVSKYPNLKGINFDLPHVIEDAPSHPGIEHVGGDMFVSVPKG-DAIFMKWI  265
OsOMT.pro  TVAAVVSRHPHIRGINYDLPHVISEAPPFPGVEHVGGDMFASVPRGGDAILMKWI  272

Majority    .HDWSDEHC...LKNCY..LPE.GKV...EC.LPE..D.......V.HVD.IMLA
ATOMT1.pro CHDWSDEHCVKFLKNCYESLPEDGKVILAECILPETPDSSLSTKQVVHVDCIMLA  320
OsOMT.pro  LHDWSDEHCARLLKNCYDALPEHGKVVVVECVLPESSDATAREQGVFHVDMIMLA  327

Majority    HNPGGKER.E.EF..LA.A.GF.G.K..........IE..K..
ATOMT1.pro HNPGGKERTEKEFEALAKASGFKGIKVVCDAFGVNLIELLKKL              363
OsOMT.pro  HNPGGKERYEREFRELARAAGFTGFKATYIYANAWAIEFTK               368
```

Figure A.2 Peptide Sequence Alignment of Arabidopsis ATOMT1 and Rice OsOMT. Ibrahim et al. (1998) performed a comparative analysis of amino acid sequences from 26 plant OMTs to determine that plant OMTs have five conserved regions (I-V) in their peptide sequences: region I "LVDVGGGXG", region II "GINFDLPHV", region III "EHVGGDMF", region IV "NGKVI", and region V "GGKERT" (72). Regions I and V are conserved among metal binding OMTs due to their involvement in SAM and metal cation binding, respectively. The conserved motifs are underlined in red. Sequences were aligned using MegAlign (DNASTAR). Dots denote residues that do not match. Dashes signify gaps inserted to obtain optimal sequence alignment.

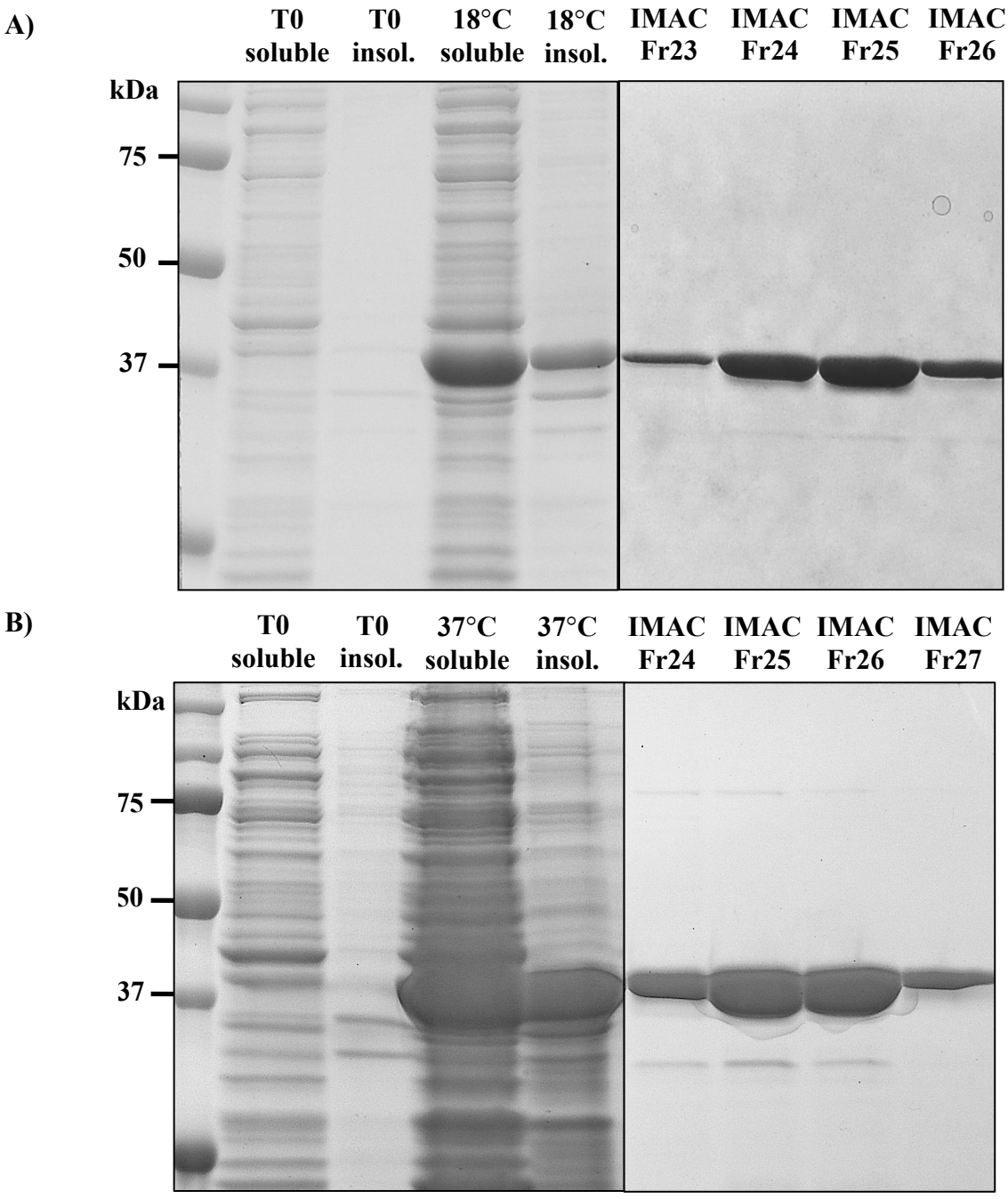


Figure A.3 Heterologous Expression of OsOMT and ATOMT1 in *E. coli*. SDS-PAGE analysis of soluble and insoluble protein fractions for *A*, OsOMT, un-induced (T0) and induced (4 h) expression at 18 °C, and corresponding IMAC fractions, and *B*, ATOMT1, un-induced (T0) and induced (6 h) expression at 37 °C, and corresponding IMAC fractions.

Enzyme Activity Assays

In a previous study (73), it was determined that the OMT candidate gene catalyzed the 3´-O-methylation of caffeic acid. However, it had not been shown whether OsOMT had activity with other caffeyl substrates found in the monolignol biosynthetic pathway, such as caffeyl alcohol, caffeoyl CoA, and caffeyl aldehyde. HPLC analyses demonstrated that OsOMT catalyzes the 3´-O-methylation of caffeyl alcohol and caffeyl aldehyde to produce coniferyl alcohol and coniferyl aldehyde, respectively (Table A.1). OsOMT did not have activity with caffeoyl CoA, even in a reaction containing twice the enzyme amount and incubation time. This observation strongly suggests that not only is caffeoyl CoA a poor substrate for OsOMT activity but also, OsOMT cannot be categorized as a CCoAOMT. ATOMT1 did not show similar activity as with OsOMT, indicating that OsOMT and ATOMT1 have different functions in the plant cell.

Table A.1 Enzyme Activity for OsOMT and ATOMT1 with Various Caffeyl Substrates

Substrate	**OsOMT Activity**	**ATOMT1 Activity**
Caffeic acid	Yes	No
Caffeoyl CoA	No	No
Caffeyl aldehyde	Yes	No
Caffeyl alcohol	Yes	No
Caffeyl *p*-coumarate*	Yes	No
Caffeyl *p*-coumarate**	Yes	No
p-Coumaryl caffeate*	Yes	No

* Indicates that the monolignol conjugate was synthesized in a coupled enzyme assay containing OsPMT and the appropriate CoA thioester and monolignol alcohol substrate.

** Indicates that the monolignol conjugate was chemically synthesized as a standard compound and utilized in a single enzyme assay containing OsOMT and the proper reaction ingredients.

OsPMT was used in a coupled enzyme assay to catalyze the synthesis of caffeyl *p*-coumarate and *p*-coumaryl caffeate. These reactions were coupled with the addition of OsOMT and SAM to determine whether OsOMT could utilize the monolignol conjugates as substrate for 3′-O-methylation activity. It was determined that OsOMT does use caffeyl *p*-coumarate and *p*-coumaryl caffeate as substrates to synthesize coniferyl *p*-coumarate and *p*-coumaryl ferulate, respectively. Nevertheless, it was important to determine whether activity with caffeyl *p*-coumarate to produce coniferyl *p*-coumarate was the result of OsOMT activity on the monolignol conjugate and not caffeyl alcohol. Due to its substrate specificity with caffeyl alcohol, OsOMT could have produced coniferyl alcohol from caffeyl alcohol, which would then be utilized by OsPMT to synthesize caffeyl *p*-coumarate using the *p*-coumaroyl CoA present in the reaction mixture. A chemically synthesized caffeyl *p*-coumarate substrate was utilized in a similar reaction to determine that OsOMT does in fact have activity with caffeyl *p*-coumarate to produce coniferyl *p*-coumarate.

A.4 SIGNIFICANCE

Lignin has been classically described as a biopolymer of three distinct monomer units (H, G, and S). However, as previously illustrated, many studies have identified that plants have the ability to incorporate alternative monolignol precursors into their lignin. These alternative monolignols include monolignols that are acylated prior to incorporation into the lignin polymer. In particular, it has been proposed that grasses have an alternative pathway for monolignol biosynthesis due to the prevalence of *p*-coumarate esters in their lignin. The purpose of this study was to identify an enzyme that could potentially be involved in a novel acylation reaction in rice. OsOMT is a grass-specific enzyme that catalyzes the synthesis of coniferyl *p*-coumarate and *p*-coumaryl ferulate. Co-expression analysis of 4CL-2 demonstrated a strong correlation with OsOMT, which demonstrates a high probability that OsOMT is involved in lignin biosynthesis. In addition, OsOMT is similar to putative OMT genes in other grass species within the Poaceae family. All together, these findings suggest that an alternative pathway for the synthesis of monolignols in grasses may exist. Furthermore, with the identification of OsOMT and future studies on its biochemistry, it will enhance our understanding of acylation reactions in lignin and the functional consequences of this phenomenon.

APPENDIX B

Proteome Analysis of Unpollinated and Pollinated Cherry Styles

The proteomics data displayed below (Table B.1) represents a single replicate of cherry style protein extractions and experimental LC/MS/MS for the 24 h pollination experiments. The individual cherry protein samples were labeled as follows: Bing (S_3S_4) × Brooks (S_1S_9) "BB24", control (unpollinated) styles "BC24", and Bing (S_3S_4) × EF (S_3S_4) "BE24". Since this data subset contained 1507 identified proteins, only the first 100 proteins have been shown. Data in the last three columns represent spectral counts, which are generally used as a measure of protein abundance in pre-digested protein samples.

Table B.1 Proteomics Data for Cherry Pollination Experiments at 24 h

#	Identified Proteins	BB24	BC24	BE24
1	Chain E, Leech-Derived Tryptase InhibitorTRYPSIN COMPLEX	204	327	239
2	| Symbols: ATMETS, ATMS1, ATCIMS | ATCIMS (COBALAMIN-INDEPENDENT METHIONINE SYNTHASE) | chr5:5935773-5939197 FORWARD	225	206	210
3	| Symbols: LOS2 | LOS2 (Low expression of osmotically responsive genes 1); phosphopyruvate hydratase | chr2:15328160-15330865 REVERSE	238	143	251
4	| Symbols: GAPC-2 | GAPC-2; glyceraldehyde-3-phosphate dehydrogenase | chr1:4608462-4610491 REVERSE	177	125	223
5	| Symbols: PR3, PR-3, CHI-B, B-CHI, ATHCHIB | ATHCHIB (BASIC CHITINASE); chitinase | chr3:3962508-3963952 REVERSE	146	119	182
6	| Symbols: | ATP synthase beta chain 2, mitochondrial | chr5:2825740-2828353 FORWARD	149	117	174
7	| Symbols: UBQ6 | UBQ6 (ubiquitin 6); protein binding | chr2:19351771-19352244 FORWARD	139	120	139

Table B.1 (cont'd)

#	Identified Proteins	BB24	BC24	BE24
8	| Symbols: BIP2, BIP | BIP (LUMINAL BINDING PROTEIN); ATP binding | chr5:16824925-16827708 REVERSE	102	127	96
9	| Symbols: ANNAT2 | ANNAT2 (ANNEXIN ARABIDOPSIS 2); calcium ion binding / calcium-dependent phospholipid binding | chr5:25991141-25992780 FORWARD	118	89	118
10	| Symbols: ACT11 | ACT11 (ACTIN-11); structural constituent of cytoskeleton | chr3:3858122-3859615 FORWARD	105	79	122
11	| Symbols: | fructose-bisphosphate aldolase, putative | chr5:963388-964981 REVERSE	86	105	113
12	| Symbols: MEE6, CS1, APX1 | APX1 (ASCORBATE PEROXIDASE 1) | chr1:2438002-2439432 FORWARD	84	85	100
13	| Symbols: BIP1 | BIP1; ATP binding | chr5:10540669-10543278 REVERSE	80	85	96
14	| Symbols: | UTP--glucose-1-phosphate uridylyltransferase, putative / UDP-glucose pyrophosphorylase, putative / UGPase, putative | chr5:5696957-5700847 REVERSE	92	55	100
15	| Symbols: AAC1 | AAC1 (ADP/ATP CARRIER 1); binding | chr3:2605712-2607036 REVERSE	95	44	97
16	| Symbols: | elongation factor 1-alpha / EF-1-alpha | chr5:24306452-24307901 FORWARD	90	55	90
17	| Symbols: AT1G56075.1, LOS1 | LOS1 (Low expression of osmotically responsive genes 1); translation elongation factor/ translation factor, nucleic acid binding | chr1:20971910-20974742 REVERSE	65	86	71
18	keratin 1 [Homo sapiens]	85	42	85

Table B.1 (cont'd)

#	Identified Proteins	BB24	BC24	BE24
19	polyphenyl oxidase 2 precursor	82	46	82
20	aerolysin pore forming?	62	58	83
21	| Symbols: | heat shock protein 81-4 (HSP81-4) | chr5:22694828-22697293 REVERSE	59	81	60
22	| Symbols: ROC3 | ROC3 (rotamase CyP 3); peptidyl-prolyl cis-trans isomerase | chr2:7207944-7208465 FORWARD	47	81	62
23	| Symbols: | glycosyl hydrolase family 17 protein | chr4:9200310-9201457 REVERSE	53	50	66
24	dehydrin cor29	62	49	46
25	| Symbols: SHD | SHD (SHEPHERD); ATP binding | chr4:12551912-12555861 REVERSE	36	61	51
26	| Symbols: MLP423 | MLP423 (MLP-LIKE PROTEIN 423) | chr1:8500642-8501447 REVERSE	26	64	56
27	| Symbols: RCI1, GRF3 | GRF3 (GENERAL REGULATORY FACTOR 3); protein phosphorylated amino acid binding | chr5:15427507-15428515 FORWARD	56	27	62
28	| Symbols: HSP70 | HSP70 (heat shock protein 70); ATP binding | chr3:3991494-3993696 REVERSE	50	40	54
29	| Symbols: ATPDIL2-1, UNE5, MEE30 | ATPDIL2-1/MEE30/UNE5 (PDI-LIKE 2-1); thiol-disulfide exchange intermediate | chr2:19488573-19490753 FORWARD	47	40	57
30	| Symbols: BG3 | BG3 (BETA-1,3-GLUCANASE 3); hydrolase, hydrolyzing O-glycosyl compounds | chr3:21192895-21194024 REVERSE	48	37	58
31	| Symbols: ATP1 | ATPase subunit 1 | chrM:302166-303689 REVERSE	60	35	48
32	| Symbols: | pollen Ole e 1 allergen and extensin family protein | chr2:14641377-14642676 REVERSE	52	43	47

Table B.1 (cont'd)

#	Identified Proteins	BB24	BC24	BE24
33	| Symbols: | malate dehydrogenase, cytosolic, putative | chr1:1189417-1191266 REVERSE	50	35	56
34	| Symbols: TUB1 | TUB1 (tubulin beta-1 chain); structural molecule | chr1:28455039-28457263 REVERSE	38	51	51
35	| Symbols: TPX1 | TPX1 (THIOREDOXIN-DEPENDENT PEROXIDASE 1); antioxidant | chr1:24563187-24564416 REVERSE	51	43	46
36	| Symbols: CRT1 | CRT1 (CALRETICULIN 1); calcium ion binding | chr1:21093687-21096295 REVERSE	55	36	48
37	| Symbols: ATC4H, C4H, CYP73A5 | ATC4H/C4H/CYP73A5 (CINNAMATE 4-HYDROXYLASE, CINNAMATE-4-HYDROXYLASE); trans-cinnamate 4-monooxygenase | chr2:13000938-13002760 REVERSE	34	46	58
38	| Symbols: | glycosyl hydrolase family 3 protein | chr1:29354690-29357762 REVERSE	46	50	40
39	| Symbols: | cytosol aminopeptidase family protein | chr4:15046595-15049310 REVERSE	45	38	53
40	unnamed protein product [Homo sapiens]	61	22	47
41	| Symbols: ATPDIL1-1 | ATPDIL1-1 (PDI-LIKE 1-1); protein disulfide isomerase | chr1:7645756-7648684 FORWARD	44	54	32
42	| Symbols: GR-RBP7, GRP7, CCR2, ATGRP7 | ATGRP7 (COLD, CIRCADIAN RHYTHM, AND RNA BINDING 2); RNA binding | chr2:9272557-9273396 REVERSE	28	44	46
43	| Symbols: | isoflavone reductase, putative | chr4:18266018-18267598 REVERSE	36	38	42

Table B.1 (cont'd)

#	Identified Proteins	BB24	BC24	BE24
44	| Symbols: SAM-2, MAT2 | MAT2/SAM-2 (S-adenosylmethionine synthetase 2) | chr4:796298-797479 REVERSE	31	38	44
45	| Symbols: ACT2/7, ACT7 | ACT7 (actin 7) | chr5:3052810-3054221 FORWARD	30	42	40
46	| Symbols: | cysteine protease inhibitor, putative / cystatin, putative | chr5:19303822-19304190 REVERSE	23	32	62
47	| Symbols: SHD | SHD (SHEPHERD); ATP binding | chr4:12551912-12555861 REVERSE	58	17	35
48	| Symbols: | glycosyl hydrolase family 3 protein | chr1:29354690-29357762 REVERSE	36	25	46
49	| Symbols: TPI, ATCTIMC | ATCTIMC (CYTOSOLIC TRIOSE PHOSPHATE ISOMERASE); triose-phosphate isomerase | chr3:20564771-20567055 FORWARD	38	23	43
50	| Symbols: | succinyl-CoA ligase (GDP-forming) beta-chain, mitochondrial, putative / succinyl-CoA synthetase, beta chain, putative / SCS-beta, putative | chr2:8812655-8814939 FORWARD	30	33	37
51	| Symbols: CSD1 | CSD1 (COPPER/ZINC SUPEROXIDE DISMUTASE 1) | chr1:2827703-2829056 FORWARD	30	29	41
52	| Symbols: | Ras-related GTP-binding family protein | chr5:24124676-24126275 REVERSE	28	29	42
53	| Symbols: | lipid-associated family protein | chr4:18432944-18433575 FORWARD	19	33	47
54	| Symbols: ATARFA1E | ATARFA1E (ADP-RIBOSYLATION FACTOR A1E); GTP binding / phospholipase activator/ protein binding | chr3:23063262-23064520 FORWARD	44	15	40

Table B.1 (cont'd)

#	Identified Proteins	BB24	BC24	BE24
55	| Symbols: ATGUS3 | ATGUS3 (ARABIDOPSIS THALIANA GLUCURONIDASE 3); beta-glucuronidase | chr5:13253142-13255948 REVERSE	47	16	35
56	No hits found	56	2	40
57	| Symbols: | extracellular dermal glycoprotein, putative / EDGP, putative | chr1:787143-788444 FORWARD	38	23	34
58	| Symbols: | ATP synthase delta chain, mitochondrial, putative / H(+)-transporting two-sector ATPase, delta (OSCP) subunit, putative | chr5:4310561-4311944 REVERSE	33	21	41
59	| Symbols: SUS1, ASUS1, ATSUS1 | SUS1 (SUCROSE SYNTHASE 1); UDP-glycosyltransferase/ sucrose synthase | chr5:7050601-7054034 REVERSE	38	24	31
60	| Symbols: ATPQ | ATPQ (ATP SYNTHASE D CHAIN, MITOCHONDRIAL); hydrogen ion transporting ATP synthase, rotational mechanism / hydrogen ion transporting ATPase, rotational mechanism | chr3:19407667-19409097 FORWARD	22	27	44
61	epidermal cytokeratin 2 [Homo sapiens]	39	30	23
62	| Symbols: | 40S ribosomal protein S7 (RPS7A) | chr1:18063522-18064603 REVERSE	28	28	36
63	| Symbols: ATAVP3, AVP-3, AVP1 | AVP1 (vacuolar-type H+-pumping pyrophosphatase 1); ATPase | chr1:5399110-5402180 FORWARD	36	30	25
64	| Symbols: | phosphoglucomutase, cytoplasmic, putative / glucose phosphomutase, putative | chr1:8219935-8224175 FORWARD	29	25	36

Table B.1 (cont'd)

#	Identified Proteins	BB24	BC24	BE24
65	| Symbols: ACLB-2 | ACLB-2 (ATP-citrate lyase B-2) | chr5:20072274-20075421 FORWARD	35	31	24
66	| Symbols: | isocitrate dehydrogenase, putative / NADP+ isocitrate dehydrogenase, putative | chr1:24542751-24545524 FORWARD	38	21	31
67	| Symbols: | malate dehydrogenase (NAD), mitochondrial | chr1:19858634-19860470 REVERSE	20	38	31
68	| Symbols: MTLPD1 | dihydrolipoamide dehydrogenase 1, mitochondrial / lipoamide dehydrogenase 1 (MTLPD1) | chr1:17721101-17722810 REVERSE	31	21	37
69	| Symbols: AHA5 | AHA5 (ARABIDOPSIS H(+)-ATPASE 5); ATPase | chr2:10422602-10426810 FORWARD	28	21	38
70	| Symbols: GST29, ATGSTU18 | ATGSTU18 (GLUTATHIONE S-TRANSFERASE 29); glutathione transferase | chr1:3395740-3396808 REVERSE	31	23	33
71	putative allergen Pru p 1.02	18	34	34
72	| Symbols: PCK1, PEPCK | PCK1/PEPCK (PHOSPHOENOLPYRUVATE CARBOXYKINASE 1); ATP binding / phosphoenolpyruvate carboxykinase (ATP) | chr4:17802968-17806326 REVERSE	27	31	27
73	polyphenyl oxidase 2 precursor	19	42	23
74	| Symbols: PLDALPHA2 | PLDALPHA2 (PHOSPHLIPASE D ALPHA 2); phospholipase D | chr1:19587609-19590220 REVERSE	36	26	21
75	| Symbols: | band 7 family protein | chr1:26297595-26298813 REVERSE	31	10	42
76	keratin 9 [Homo sapiens]	38	17	27

Table B.1 (cont'd)

#	Identified Proteins	BB24	BC24	BE24
77	| Symbols: LHCB5 | LHCB5 (LIGHT HARVESTING COMPLEX OF PHOTOSYSTEM II 5); chlorophyll binding | chr4:6408196-6409492 FORWARD	25	21	36
78	| Symbols: EMB1395, MEE58, SAHH, SAHH1, HOG1 | HOG1 (HOMOLOGY-DEPENDENT GENE SILENCING 1); adenosylhomocysteinase | chr4:8054926-8056671 FORWARD	30	25	26
79	| Symbols: | leucine-rich repeat family protein | chr3:7280936-7282033 FORWARD	22	21	38
80	| Symbols: PHGPX, LSC803, ATGPX6 | ATGPX6 (GLUTATHIONE PEROXIDASE 6); glutathione peroxidase | chr4:7010015-7011324 REVERSE	25	27	29
81	| Symbols: | Identical to Uncharacterized protein At5g48480 [Arabidopsis Thaliana] (GB:Q9LV66); similar to unknown [Populus trichocarpa] (GB:ABK95611.1); contains domain SSF54593 (SSF54593) | chr5:19662040-19662884 FORWARD	18	37	25
82	| Symbols: TUA2 | TUA2 (tubulin alpha-2 chain) | chr1:18521405-18523397 FORWARD	16	31	32
83	| Symbols: | porin, putative | chr3:85761-87619 FORWARD	25	18	36
84	| Symbols: ATAVP3, AVP-3, AVP1 | AVP1 (vacuolar-type H+-pumping pyrophosphatase 1); ATPase | chr1:5399110-5402180 FORWARD	26	30	22
85	| Symbols: | uridylate kinase / uridine monophosphate kinase / UMP kinase (PYR6) | chr5:9276662-9278094 FORWARD	22	22	34

Table B.1 (cont'd)

#	Identified Proteins	BB24	BC24	BE24
86	| Symbols: PIP2D, PIP2;5 | PIP2;5/PIP2D (plasma membrane intrinsic protein 2;5); water channel | chr3:20313095-20314716 FORWARD	33	17	27
87	| Symbols: SAHH2 | SAHH2 (S-ADENOSYL-L-HOMOCYSTEINE (SAH) HYDROLASE 2); adenosylhomocysteinase | chr3:8588020-8589678 REVERSE	33	15	29
88	No hits found	25	27	24
89	| Symbols: | elongation factor 1-alpha / EF-1-alpha | chr5:24306452-24307901 FORWARD	23	24	27
90	| Symbols: | sorbitol dehydrogenase, putative / L-iditol 2-dehydrogenase, putative | chr5:21129046-21130510 FORWARD	19	17	38
91	| Symbols: NDPK1 | NDPK1 (nucleoside diphosphate kinase 1); ATP binding / nucleoside diphosphate kinase | chr4:5923421-5924363 FORWARD	13	24	37
92	| Symbols: ATGSR1 | ATGSR1 (Arabidopsis thaliana glutamine synthase clone R1); glutamate-ammonia ligase | chr5:14950804-14952886 REVERSE	18	29	25
93	aerolysin pore forming?	18	17	36
94	| Symbols: ATTRX H1, ATTRX1 | ATTRX1 (Arabidopsis thaliana thioredoxin H-type 1); thiol-disulfide exchange intermediate | chr3:18962104-18962936 REVERSE	19	25	27
95	| Symbols: PGK1 | PGK1 (PHOSPHOGLYCERATE KINASE 1); phosphoglycerate kinase | chr3:4061134-4063147 REVERSE	30	19	22

Table B.1 (cont'd)

#	Identified Proteins	BB24	BC24	BE24
96	| Symbols: PLD, PLDALPHA1 | PLDALPHA1 (PHOSPHOLIPASE D ALPHA 1); phospholipase D | chr3:5330842-5333481 FORWARD	24	21	25
97	| Symbols: GF14 PHI, GRF4 | GRF4 (GENERAL REGULATORY FACTOR 4); protein phosphorylated amino acid binding | chr1:12867242-12868492 FORWARD	25	15	29
98	| Symbols: | aspartyl protease family protein | chr3:20151269-20153577 REVERSE	23	24	22
99	No hits found	17	22	29
100	| Symbols: | trypsin and protease inhibitor family protein / Kunitz family protein | chr1:6149336-6149926 FORWARD	17	21	30

REFERENCES

REFERENCES

1. Pauly, M. and Keegstra, K. (2008) Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J*, **54**, 559-568.

2. Bonawitz, N.D. and Chapple, C. (2010) The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu Rev Genet*, **44**, 337-363.

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

4. Miao, Y.C. and Liu, C.J. (2010) ATP-binding cassette-like transporters are involved in the transport of lignin precursors across plasma and vacuolar membranes. *Proc Natl Acad Sci U S A*, **107**, 22728-22733.

5. Halpin, C., Knight, M.E., Foxon, G.A., Campbell, M.M., Boudet, A.M., Boon, J.J., Chabbert, B., Tollier, M.T. and Schuch, W. (1994) Manipulation of lignin quality by downregulation of cinnamyl alcohol dehydrogenase. *Plant J*, **6**, 339-350.

6. Atanassova, R., Favet, N., Martz, F., Chabbert, B., Tollier, M.-T., Monties, B., Fritig, B. and Legrand, M. (1995) Altered lignin composition in transgenic tobacco expressing O-methyltransferase sequences in sense and antisense orientation. *The Plant Journal*, **8**, 465-477.

7. Ralph, J., MacKay, J.J., Hatfield, R.D., O'Malley, D.M., Whetten, R.W. and Sederoff, R.R. (1997) Abnormal lignin in a loblolly pine mutant. *Science*, **277**, 235-239.

8. Halpin, C., Holt, K., Chojecki, J., Oliver, D., Chabbert, B., Monties, B., Edwards, K., Barakate, A. and Foxon, G.A. (1998) Brown-midrib maize (bm1) – a mutation affecting the cinnamyl alcohol dehydrogenase gene. *The Plant Journal*, **14**, 545-553.

9. Ralph, J., Lapierre, C., Lu, F., Marita, J.M., Pilate, G., Van Doorsselaere, J., Boerjan, W. and Jouanin, L. (2001) NMR evidence for benzodioxane structures resulting from incorporation of 5-hydroxyconiferyl alcohol into Lignins of O-methyltransferase-deficient poplars. *J Agric Food Chem*, **49**, 86-91.

10. Ralph, J., Lapierre, C., Marita, J.M., Kim, H., Lu, F., Hatfield, R.D., Ralph, S., Chapple, C., Franke, R., Hemm, M.R. *et al.* (2001) Elucidation of new structures in lignins of CAD- and COMT-deficient plants by NMR. *Phytochemistry*, **57**, 993-1003.

11. Sibout, R., Eudes, A., Mouille, G., Pollet, B., Lapierre, C., Jouanin, L. and Seguin, A. (2005) CINNAMYL ALCOHOL DEHYDROGENASE-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of Arabidopsis. *Plant Cell*, **17**, 2059-2076.

12. Van Doorsselaere, J., Baucher, M., Chognot, E., Chabbert, B., Tollier, M.-T., Petit-Conil, M., Leplé, J.-C., Pilate, G., Cornu, D., Monties, B. *et al.* (1995) A novel lignin in poplar trees with a reduced caffeic acid/5-hydroxyferulic acid O-methyltransferase activity. *The Plant Journal*, **8**, 855-864.

13. Smith, D.C.C. (1955) Ester groups in lignin. *Nature*, **176**, 267-268.

14. D'Auria, J.C. (2006) Acyltransferases in plants: a good time to be BAHD. *Curr Opin Plant Biol*, **9**, 331-340.

15. Shimada, M., Fukuzuka, T. and Higuchi, T. (1971) Ester linkages of p-coumaric acid in bamboo and grass lignins. *Tappi*, **54**, 72-78.

16. Hartley, R.D. (1972) P-coumaric and ferulic acid components of cell walls of ryegrass and their relationships with lignin and digestibility. *J Sci Food Agric*, **23**, 1347-1354.

17. Ralph, J., Hatfield, R.D., Quideau, S., Helm, R.F., Grabber, J.H. and Jung, H.J.G. (1994) Pathway of p-coumaric acid incorporation into maize lignin as revealed by NMR. *J Amer Chem Soc*, **116**, 9448-9456.

18. Lu, F. and Ralph, J. (1999) Detection and determination of p-coumaroylated units in lignins. *J Agric Food Chem*, **47**, 1988-1992.

19. Withers, S., Lu, F., Kim, H., Zhu, Y., Ralph, J. and Wilkerson, C.G. (2012) Identification of grass-specific enzyme that acylates monolignols with p-coumarate. *J Biol Chem*, **287**, 8347-8355.

20. Smith, D.C.C. (1955) P-hydroxybenzoate groups in the lignin of aspen (Populus-tremula) *J Chem Soc*, 2347-2351.

21. Nakano, J., Ishizu, A. and Migita, N. (1961) Studies on lignin. XXXII. Ester groups of lignin. *Tappi*, **44**, 30-32.

22. Landucci, L.L., Deka, G.C. and Roy, D.N. (1992) A C-13 NMR study of milled wood lignins from hybrid Salix clones. *Holzforschung*, **46**, 505-511.

23. Sun, R.C., Fang, J.M. and Tomkinson, J. (1999) Fractional isolation and structural characterization of lignins from oil palm trunk and empty fruit bunch fibers. *J Wood Chem Tech*, **19**, 335-356.

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

25. Sarkanen, K.V., Chang, H. and Allan, G.G. (1967) Species variation in lignins. 3. Hardwood lignins *Tappi*, **50**, 587-590.

26. Ralph, J. (1996) An unusual lignin from kenaf. *J Nat Prod*, **59**, 341-342.

27. Lu, F. and Ralph, J. (2002) Preliminary evidence for sinapyl acetate as a lignin monomer in kenaf. *Chem Commun (Camb)*, 90-91.

28. Del Rio, J.C., Marques, G., Rencoret, J., Martinez, A.T. and Gutierrez, A. (2007) Occurrence of naturally acetylated lignin units. *J Agric Food Chem*, **55**, 5461-5468.

29. Lu, F.C. and Ralph, J. (1997) DFRC method for lignin analysys. 1. New method for beta aryl ether cleavage: Lignin model studies. *J Agric Food Chem*, **45**, 4655-4660.

30. Ralph, J. and Lu, F.C. (1998) The DFRC method for lignin analysis. 6. A simple modification for identifying natural acetates on lignins. *J Agric Food Chem*, **46**, 4616-4619.

31. Martinez, A.T., Rencoret, J., Marques, G., Gutierrez, A., Ibarra, D., Jimenez-Barbero, J. and del Rio, J.C. (2008) Monolignol acylation and lignin structure in some nonwoody plants: a 2D NMR study. *Phytochemistry*, **69**, 2831-2843.

32. Lu, F. and Ralph, J. (1998) The DFRC method for lignin analysis. 2. Monomers from isolated lignins. *J Agric Food Chem*, **46**, 547-552.

33. Lu, F. and Ralph, J. (1998) Facile synthesis of 4-hydroxycinnamyl p-coumarates. *J Agric Food Chem*, **46**, 2911-2913.

34. Beuerle, T. and Pichersky, E. (2002) Enzymatic synthesis and purification of aromatic coenzyme A esters. *Anal Biochem*, **302**, 305-312.

35. Team, R.D.C. (2009) R: A Language and Environment for Statistical Computing.

36. D'Auria, J.C., Pichersky, E., Schaub, A., Hansel, A. and Gershenzon, J. (2007) Characterization of a BAHD acyltransferase responsible for producing the green leaf volatile (Z)-3-hexen-1-yl acetate in Arabidopsis thaliana. *Plant J*, **49**, 194-207.

37. Vogt, T. (2010) Phenylpropanoid biosynthesis. *Mol Plant*, **3**, 2-20.

38. Dexter, R., Qualley, A., Kish, C.M., Ma, C.J., Koeduka, T., Nagegowda, D.A., Dudareva, N., Pichersky, E. and Clark, D. (2007) Characterization of a petunia acetyltransferase involved in the biosynthesis of the floral volatile isoeugenol. *Plant J*, **49**, 265-275.

39. Karimi, M., Inzé, D. and Depicker, A. (2002) GATEWAY™ vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci*, **7**, 193-195.

40. Schilmiller, A.L., Stout, J., Weng, J.K., Humphreys, J., Ruegger, M.O. and Chapple, C. (2009) Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in Arabidopsis. *Plant J*, **60**, 771-782.

41. Alejandro, S., Lee, Y., Tohge, T., Sudre, D., Osorio, S., Park, J., Bovet, L., Geldner, N., Fernie, A.R. and Martinoia, E. (2012) AtABCG29 is a monolignol transporter involved in lignin biosynthesis. *Curr Biol*, **22**, 1207-1212.

42. Kear, P.J. and McClure, B. (2012) How did flowering plants learn to avoid blind date mistakes? Self-incompatibility in plants and comparisons with nonself rejection in the immune response. *Adv Exp Med Biol*, **738**, 108-123.

43. Ushijima, K., Yamane, H., Watari, A., Kakehi, E., Ikeda, K., Hauck, N.R., Iezzoni, A.F. and Tao, R. (2004) The S haplotype-specific F-box protein gene, SFB, is defective in self-compatible haplotypes of Prunus avium and P. mume. *Plant J*, **39**, 573-586.

44. Bosch, M. and Franklin-Tong, V.E. (2008) Self-incompatibility in Papaver: signalling to trigger PCD in incompatible pollen. *J Exp Bot*, **59**, 481-490.

45. McClure, B.A. and Franklin-Tong, V. (2006) Gametophytic self-incompatibility: understanding the cellular mechanisms involved in "self" pollen tube inhibition. *Planta*, **224**, 233-245.

46. Brewbaker, J.L. and Natarajan, A.T. (1960) Centric fragments and pollen-part mutation of incompatibility alleles in Petunia. *Genetics*, **45**, 699-704.

47. Golz, J.F., Su, V., Clarke, A.E. and Newbigin, E. (1999) A molecular description of mutations affecting the pollen component of the Nicotiana alata S locus. *Genetics*, **152**, 1123-1135.

48. Golz, J.F., Oh, H.Y., Su, V., Kusaba, M. and Newbigin, E. (2001) Genetic analysis of Nicotiana pollen-part mutants is consistent with the presence of an S-ribonuclease inhibitor at the S locus. *Proc Natl Acad Sci U S A*, **98**, 15372-15376.

49. Hauck, N.R., Yamane, H., Tao, R. and Iezzoni, A.F. (2006) Accumulation of nonfunctional S-haplotypes results in the breakdown of gametophytic self-incompatibility in tetraploid Prunus. *Genetics*, **172**, 1191-1198.

50. Sonneveld, T., Tobutt, K.R., Vaughan, S.P. and Robbins, T.P. (2005) Loss of pollen-S function in two self-compatible selections of Prunus avium is associated with deletion/mutation of an S haplotype-specific F-box gene. *Plant Cell*, **17**, 37-51.

51. McClure, B., Mou, B., Canevascini, S. and Bernatzky, R. (1999) A small asparagine-rich protein required for S-allele-specific pollen rejection in Nicotiana. *Proc Natl Acad Sci U S A*, **96**, 13548-13553.

52. Lind, J.L., Bacic, A., Clarke, A.E. and Anderson, M.A. (1994) A style-specific hydroxyproline-rich glycoprotein with properties of both extensins and arabinogalactan proteins. *Plant J*, **6**, 491-502.

53. Lind, J.L., Bonig, I., Clarke, A.E. and Anderson, M.A. (1996) A style-specific 120-kDa glycoprotein enters pollen tubes of Nicotiana alata in vivo. *Sex Plant Reprod*, **9**, 75-86.

54. O'Brien, M., Kapfer, C., Major, G., Laurin, M., Bertrand, C., Kondo, K., Kowyama, Y. and Matton, D.P. (2002) Molecular analysis of the stylar-expressed Solanum chacoense small asparagine-rich protein family related to the HT modifier of gametophytic self-incompatibility in Nicotiana. *Plant J*, **32**, 985-996.

55. Cruz-Garcia, F., Hancock, C.N., Kim, D. and McClure, B. (2005) Stylar glycoproteins bind to S-RNase in vitro. *Plant J*, **42**, 295-304.

56. Hancock, C.N., Kent, L. and McClure, B.A. (2005) The stylar 120 kDa glycoprotein is required for S-specific pollen rejection in Nicotiana. *Plant J*, **43**, 716-723.

57. Tao, R., Yamane, H., Sugiura, A., Murayama, H., Sassa, H. and Mori, H. (1999) Molecular typing of S-alleles through identification, characterization and cDNA cloning for S-RNases in sweet cherry. *J Amer Soc Hort Sci*, **124**, 224-233.

58. Sonneveld, T., Robbins, T.P., Bošković, R. and Tobutt, K.R. (2001) Cloning of six cherry self-incompatibility alleles and development of allele-specific PCR detection. *Theor Appl Genet*, **102**, 1046-1055.

59. Sonneveld, T., Tobutt, K.R. and Robbins, T.P. (2003) Allele-specific PCR detection of sweet cherry self-incompatibility (S) alleles S1 to S16 using consensus and allele-specific primers. *Theor Appl Genet*, **107**, 1059-1070.

60. Wunsch, A. and Hormaza, J.I. (2004) Cloning and characterization of genomic DNA sequences of four self-incompatibility alleles in sweet cherry (Prunus avium L.). *Theor Appl Genet*, **108**, 299-305.

61. Ushijima, K., Sassa, H., Dandekar, A.M., Gradziel, T.M., Tao, R. and Hirano, H. (2003) Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism. *Plant Cell*, **15**, 771-781.

62. Sijacic, P., Wang, X., Skirpan, A.L., Wang, Y., Dowd, P.E., McCubbin, A.G., Huang, S. and Kao, T.H. (2004) Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature*, **429**, 302-305.

63. Yamane, H., Tao, R., Sugiura, A., Hauck, N.R. and Iezzoni, A.F. (2001) Identification and characterization of S-RNases in tetraploid sour cherry (Prunus cerasus). *J Amer Soc Hort Sci*, **126**, 661-667.

64. Shevchenko, A., Wilm, M., Vorm, O. and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem*, **68**, 850-858.

65. Goldraij, A., Kondo, K., Lee, C.B., Hancock, C.N., Sivaguru, M., Vazquez-Santana, S., Kim, S., Phillips, T.E., Cruz-Garcia, F. and McClure, B. (2006) Compartmentalization of S-RNase and HT-B degradation in self-incompatible Nicotiana. *Nature*, **439**, 805-810.

66. Luu, D.T., Qin, X., Morse, D. and Cappadocia, M. (2000) S-RNase uptake by compatible pollen tubes in gametophytic self-incompatibility. *Nature*, **407**, 649-651.

67. McClure, B. (2006) New views of S-RNase-based self-incompatibility. *Curr Opin Plant Biol*, **9**, 639-646.

68. Torres, M.P., Lee, M.J., Ding, F., Purbeck, C., Kuhlman, B., Dokholyan, N.V. and Dohlman, H.G. (2009) G Protein mono-ubiquitination by the Rsp5 ubiquitin ligase. *J Biol Chem*, **284**, 8940-8950.

69. Iacovache, I., van der Goot, F.G. and Pernot, L. (2008) Pore formation: an ancient yet complex form of attack. *Biochim Biophys Acta*, **1778**, 1611-1623.

70. Grabber, J.H., Quideau, S. and Ralph, J. (1996) P-coumaroylated syringyl units in maize lignin: Implications for β-ether cleavage by thioacidolysis. *Phytochemistry*, **43**, 1189-1194.

71. Ehlting, J., Büttner, D., Wang, Q., Douglas, C.J., Somssich, I.E. and Kombrink, E. (1999) Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms. *Plant J*, **19**, 9-20.

72. Ibrahim, R., Bruneau, A. and Bantignies, B. (1998) Plant O-methyltransferases: molecular analysis, common signature and classification. *Plant Molecular Biology*, **36**, 1-10.

73. Lin, F.Q., Yamano, G., Hasegawa, M., Anzai, H., Kawasaki, S. and Kodama, O. (2006) Cloning and functional analysis of caffeic acid 3-O-methyltransferase from rice (Oryza sativa). *J Pest Sci*, **31**, 47-53.