Analysis of the *Arabidopsis fatty acyl-CoA synthetase5* gene and co-expressed genes reveals an ancient biochemical pathway required for pollen development and sporopollenin biosynthesis

by

Sung Soo Kim

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
(Botany)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

June 2011

ABSTRACT

The structure of the sporopollenin polymer that is the major constituent of exine, the outer pollen wall, remains poorly understood. In flowering plants, sporopollenin precursors are known to be produced in the tapetum and must be deposited and polymerized on the developing microspore during exine formation. Recent characterization of *Arabidopsis thaliana* genes and corresponding enzymes involved in exine formation has demonstrated the role of fatty acid derivatives as precursors of sporopollenin building units. In collaboration with other research groups, we discovered that the *Arabidopsis ACOS5* gene is transiently and exclusively expressed in tapetum cells and that an *Arabidopsis acos5* mutant is completely male sterile with pollen grains that are apparently devoid of sporopollenin. In this thesis, *Arabidopsis* genes encoding potential enzymes that could work with ACOS5 in a common biosynthetic pathway were identified by *in silico* co-expression analyses and functionally characterized. Previous studies and my reverse genetic analyses of selected co-expressed genes including *POLYKETIDE SYNTHASE A* and *POLYKETIDE SYNTHASE B* (*PKSA/B*), and *TETRAKETIDE α-PYRONE REDUCTASE1* and *TETRAKETIDE α-PYRONE REDUCTASE2* (*TKPR1/2*) revealed that mutants in these genes are also compromised in male fertility and sporopollenin deposition. *In vivo* biochemical assays by heterologous expression of PKSA in the yeast *Saccharomyces cerevisiae* showed that the enzyme catalyzes condensation of endogenous fatty acyl-CoAs with malonyl-CoA to generate α-pyrone triketides. Moreover, *in vitro* assays performed by collaborators revealed that the sequential actions of ACOS5, PKSA/B, and TKPR1/2 enzymes on fatty acid substrates generate polyhydroxylated long-chain *α*-pyrones, suggesting that these novel coumpounds are building units of sporopollenin. Phylogenetic analyses showed that these genes are highly conserved in land plants including the moss *Physcomitrella patens*. This work has illuminated the outlines of a conserved novel biosynthetic pathway involved in generating monomer constituents of the sporopollenin biopolymer component of the pollen wall.

PREFACE

Chapter 2 is part of a recently published research paper:

de Azevedo Souza C, **Kim SS**, Koch S, Kienow L, Schneider K, McKim SM, Haughn GW, Kombrink E, Douglas CJ. (2009) A novel fatty Acyl-CoA Synthetase is required for pollen development and sporopollenin biosynthesis in *Arabidopsis*. Plant Cell 21:507-525.

All experiments and data reported in the Results section of this chapter were designed, performed and analyzed by me. All data including figures and tables are extracted from the above paper and reprinted with permission of the American Society of Plant Biologists.

Chapter 3 is part of a recently published research paper:

Kim SS*, Grienenberger E*, Lallemand B, Colpitts CC, Kim SY, de Azevedo Souza C, Geoffroy P, Heintz D, Krahn D, Kaiser M, Kombrink E, Heitz T, Suh DY, Legrand M and Douglas CJ (2010) *LAP6/POLYKETIDE SYNTHASE A* and *LAP5/POLYKETIDE SYNTHASE B* Encode Hydroxyalkyl α-Pyrone Synthases Required for Pollen Development and Sporopollenin Biosynthesis in *Arabidopsis thaliana*. Plant Cell 22: 4045-4066 (* These authors contributed equally to this work)

All experiments and data reported in the Results section of this chapter were designed, performed and analyzed by me. All data including figures and tables are extracted from the above paper and reprinted with permission of the American Society of Plant Biologists.

Chapter 4 is part of a recently published research paper:

Grienenberger E*, **Kim SS***, Lallemand B, Geoffroy P, Heintz D, de Azevedo Souza C, Heitz T, Douglas CJ and Legrand M. (2010) Analysis of TETRAKETIDE α-PYRONE REDUCTASE Function in *Arabidopsis thaliana* Reveals a Previously Unknown, but Conserved, Biochemical Pathway in Sporopollenin Monomer Biosynthesis. Plant Cell 22: 4067-4083 (* These authors contributed equally to this work)

All experiments and data reported in the Results section of this chapter were designed, performed and analyzed by me and the results in sections 4.4.1, 4.4.2 and 4.4.5 are extracted from the above paper and reprinted with permission of the American Society of Plant Biologists.

TABLE OF CONTENTS

LIST OF TABLES

LISTS OF FIGURES

LIST OF ABBREVIATIONS

AHCT	**Anthocyanin O-hydroxycinnamoyltransferase**
AMP	**Adenosine monophosphate**
ATP	**Adenosine triphosphate**
BCIP	**5-Bromo-4-chloro-3-indolyl phosphate**
BEAT	**Benzylalcohol O-acetyltransferase**
bp	**Base pair**
cDNA	**Complementary deoxyribonucleic acid**
CER	**Eceriferum**
CoA	**Coenzyme A**
CYP	**Cytochrome P450**
DAT	**Deacetylvindoline 4-O-acetyltransferase**
DNA	**Deoxyribonucleic acid**
DTT	**Dithiothreitol**
EDTA	**Ethylenediaminetetraacetic acid**
EMBL	**European Molecular Biology Laboratory**
FID	**Flame ionization detector**
GC	**Gas chromatography**
HCBT	**Anthranilate N-hydroxycinnamoyl/benzoyltransferase**
HCT	**Hydroxycinnamoyltransferase**
HCT/HQT	**Hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyltransferase**
His	**Histidine**
IPTG	**Isopropyl β-D-1-thiogalactopyranoside**
JA	**Jasmonic Acid**
Km	**Michaelis-Menten constant**
LAP	**Less adhesive pollen**
LB	**Luria-Bertani**
Leu	**Leucine**
MADS	**MCM1, AGAMOUS, DEFICIENS, SRF**
MS	**Mass spectrometry**
NADH	**Nicotinamide adenine dinucleotide**
NADPH	**Nicotinamide adenine dinucleotide phosphate**
NBT	**Nitro blue tetrazolium chloride**
NTA	**Nitrilotriacetic acid**
PAGE	**Polyacrylamide gel electrophoresis**
PCR	**Polymerase chain reaction**
PHD	**Plant Homeo Domain**
PMSF	**Phenylmethylsulfonyl fluoride**

RNA	**Ribonucleic acid**
RT	**Reverse transcription**
SDS	**Sodium dodecyl sulfate**
SSC	**Saline sodium citrate**
TEM	**Transmission electron microscopy**
TF	**Transcription factor**
TMS	**Trimethylsilyl group**
Trp	**Tryptophan**
Ura	**Uracil**
UV	**Ultraviolet**

ACKNOWLEDGEMENTS

This thesis could not been completed without the guidance and the support of many people. First and foremost, I wish to express my appreciation to my supervisor, Dr. Carl J Douglas, who offered invaluable assistance, support, training, guidance, and hours of academically stimulating conversation and meeting. I would also like to thank my committee members, Dr. George Haughn, Dr. Reinhard Jetter, Dr. Jin-Gui Chen and Dr. Lacey Samuels for their advice on this research project.

I would like to thank all my lab mates, including the past members: Dr. Michael Friedmann, Dr. Björn Hamberger, Dr. Clarice Souza, Dr. June Kim, Dr. Eryang Li, Dr. Apruva Bhargava, Dr. Etienne Grienenberger, Yuanyuan Liu, Teagan Quilichini for sharing their knowledge and experience as well as providing extremely valuable assistance. I also would like to thank Dr. Zhonghua Wang, Dr. Christopher Buschhaus and Yan Cao in Dr. Jetter's lab for their help in chemical analyses and GC maintenance, together with Derrick Horne and Brad Ross in the UBC Bioimaging facility for TEM support. Special thanks to Dr. Santokh Singh and Jarnail Mehroke for thoughtful guidance to improve intellectual communication and teaching skills with various levels of undergraduate students in class.

I am grateful to my collaborators, members of Dr. Michel Legrand's research group in IBMP, France and of Dr. Dae-Yeon Suh's in University of Regina to help make this study successful. I am especially appreciated Che C. Colpitts and Dr. Etienne Grienenberger and for insightful and helpful discussion of biochemistry.

I wish to express my love and gratitude to my beloved family, especially thanks to my wife, KyungHee for her positive thoughts, encouragement and support through the duration of my studies. And finally, I am very happy to complete this thesis, seeing a big and innocent smile of my son, Se-hun.

Chapter 1. General introduction and literature review

1.1 Anther development and microsporogenesis

Microsporogenesis and pollen development are complex processes that take place during flower development, starting with theinitiation of stamen primordia and leading to the formation of two microgametes per mature pollen grain. These developmental events involve precisely controlled cellular processes, such as cell division, cell differentiation, and cell death (Sanders et al., 1999; Scott et al., 2004; Ma, 2005). In *Arabidopsis thaliana*, anther development has been divided into stages based on anatomical, morphological, cellular, and molecular events (Table 1.1; Sanders et al., 1999; Scott et al., 2004; Ma, 2005). At the beginning of anther cell differentiation (Stage 5 in Figure 1.1), four clearly defined locules are established and visible pollen mother cells appear. Subsequent to this, the pollen mother cells undergo meiosis and tetrads are formed, connected by a callose wall (Stage 7 in Figure 1.1). Once the callose wall degenerates to release free microspores (Stage 8 in Figure 1.1), they become vacuolated, and the pollen wall becomes visible. Microspores continue to enlarge and develop, and the tapetum layer, a maternal cell layer that surrounds the inner side of the anther locules, starts to degenerate. At the end of anther development (Stage 10 in Figure 1.1), the tapetum cell layer completely degenerates, and mature pollen grains fill the locules and are released by anther dehiscence (Sanders et al., 1999; Ma, 2005).

Table 1.1 Summary of major events during *Arabidopsis* anther and pollen development.

Anther stage	Major event/ landmark	Pollen wall development	TEM images[1] for Col-0 pollen wall	TEM images[1] for Col-0 pollen
Stage 5	**Anther morphogenesis is complete.** Epidermis, endothecium, middle layer, tapetum and microsporocyte are visible			

Table 1.1 Summary of major events during *Arabidopsis* anther and pollen development. (cont.)

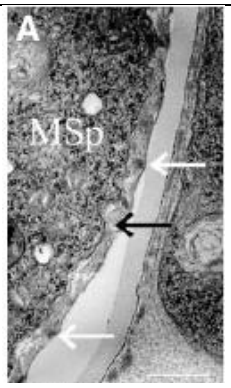
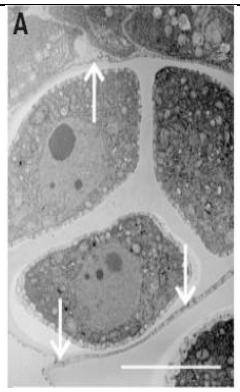
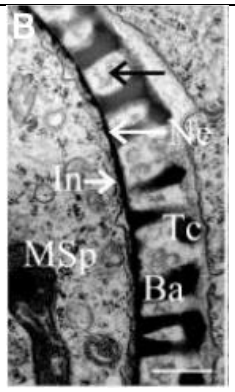
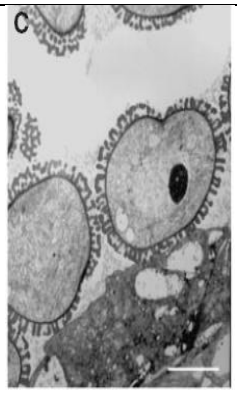
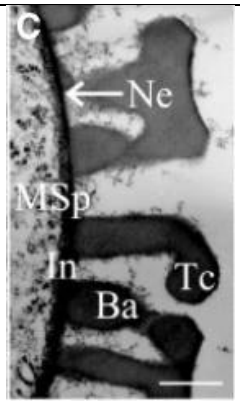
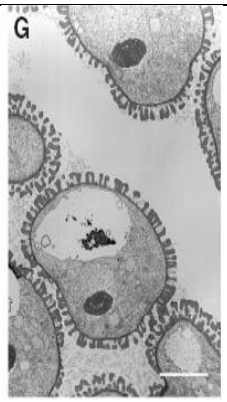
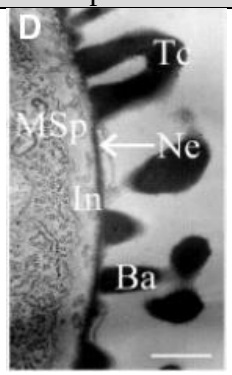
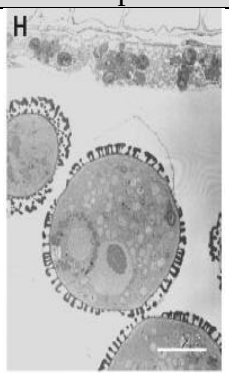
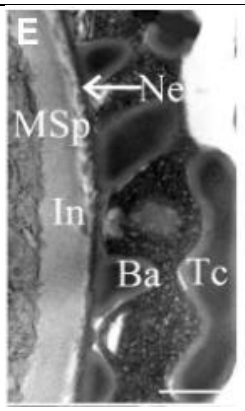
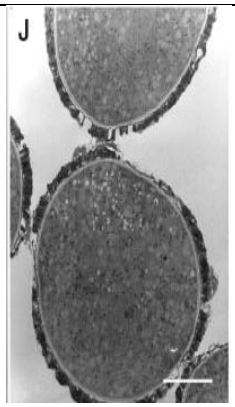
Anther stage	Major event/ landmark	Pollen wall development	TEM images[1] for Col-0 pollen wall	TEM images[1] for Col-0 pollen
Stage 6	**MMC enters Meiosis I.** Callose is deposited. Meiotic cell dissociated from each other and from tapetum.			
Stage 7	**Meiosis is completed.** Tetrad stage (haploid microspore)	**Primexine formation** (Black arrow) **Probacular formation** (White)		
Stage 8	**Microspore is released.** Callose wall is degenerated.	**Exine formation is initiated** Intine, tectum , nexine, and bacular are visible		
Stage 9 - 10	**Tapetum degeneration initiated at stage 10.**			

Table 1.1 Summary of major events during *Arabidopsis* anther and pollen development. (cont.)

Anther stage	Major event/ landmark	Pollen wall development	TEM images[1] for Col-0 pollen wall	TEM images[1] for Col-0 pollen
Stage 11	**Pollen mitotic division occurs (vegetative and generative cell stage).** Tapetum degeneration	**Completion of exine formation** Intine is thicker than previously		
Stage 12	**Tricelluar haploid pollen grain develops.**	**Deposition of extracellular pollen coat**		
Stage 13	**Anther dehiscence occurs.**			
Stage 14	**Senescence of stamen occurs.**			

[1] All TEM images are adapted and reprinted from (Ariizumi et al., 2008), with permission of Oxford University Press (Copyright © 2008 Oxford University Press).
The major events in each anther developmental stages are described based on Sanders et al. (1999). Bar = 500nm for pollen wall and 5µm for pollen. Ba, bacula; In, intine; MSp, microspores; Ne, nexine; Tc, tectum.

A critical event during pollen maturation is the formation of the pollen surface structure. The pollen wall represents one of the most complex plant cell walls, with contributions from both the sporophyte and gametophyte generations (Figure 1.2). The major pollen wall components are a microspore-derived cellulosic primexine synthesized by the developing haploid microspores themselves at the tetrad stage (Blackmore et al., 2007), a thick exine deposited on the outer surface of the primexine largely after the release of free microspores, and a pectocellulosic intine produced by developing

microspores and male gametophytes (Figure 1.2D). In contrast with the primexine and intine, the exine is maternally derived, and exine constituents are produced in the sporophytic tapetum cell layer (Piffanelli et al., 1998). These exine constituents are secreted into the locules and incorporated into the exine by polymerization on the surface

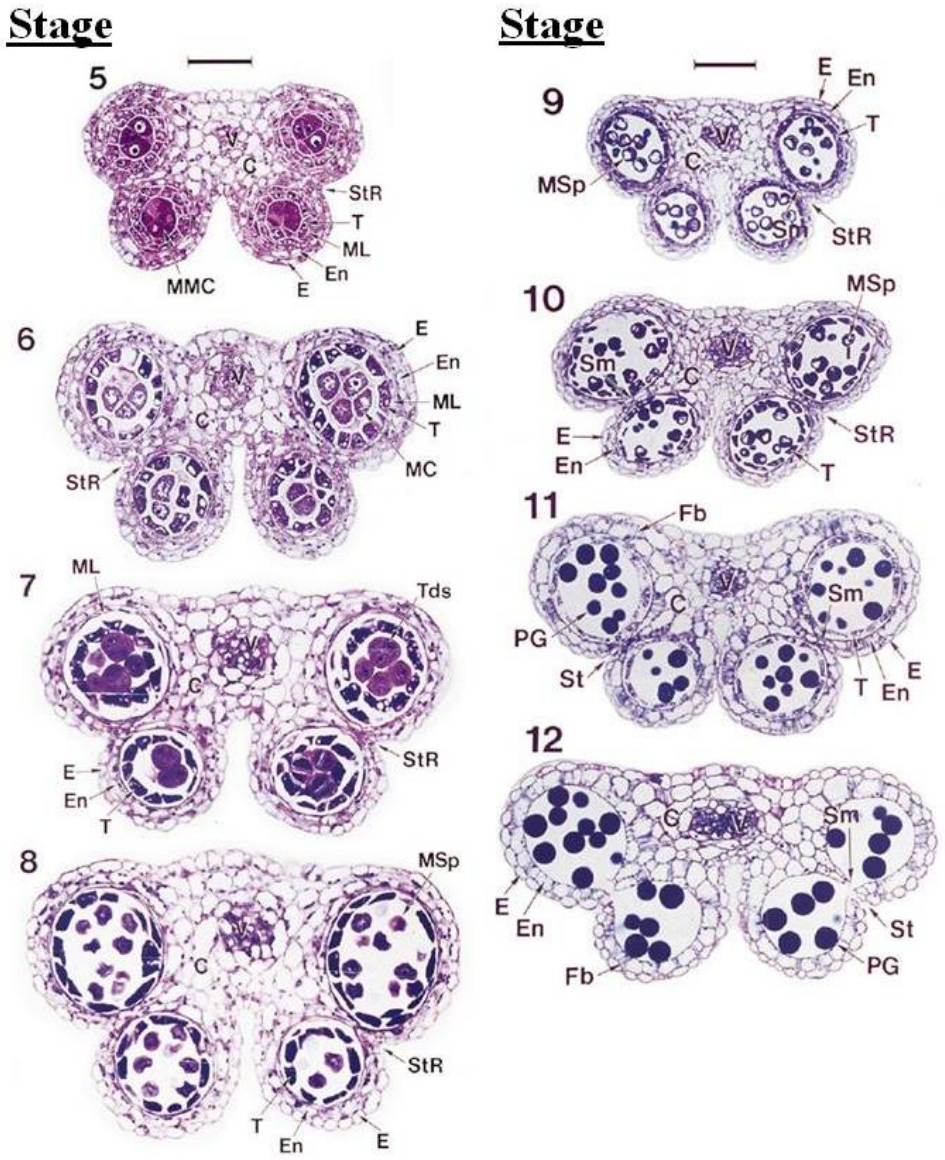


Figure 1.1 Anther cell differentiation.

The flower sections were stained in toluidine blue and anthers were photographed by bright-field microscopy. *C*, connective; *E*, epidermis; *En*, endothecium; *Fb*, fibrous bands; *MC*, meiotic cell; *ML*, middle layer; *MMC*, microspore mother cells; *MSp*, microspores; *PG*, pollen grains; *Sm*, septum; *St*, stomium; *StR*, stomium region; *T*, tapetum; *Tds*,

tetrads; *V*, vascular region. Bar= 50µm (Sanders et al., 1999). Images are adapted and reprinted with permission of Springer (Copyright © 1999, Springer)

of the primexine (Piffanelli et al., 1998; Scott et al., 2004; Ma, 2005), where it often assumes a characteristic reticulate pattern, consisting of baculae and tecta (Figure 1.2B and 1.2C). The final component of the pollen wall is the lipid-rich pollen coat, or tryphine, which is deposited onto the exine surface. Pollen coat components accumulate in tapetum cells and are released into locules during the course of tapetum cell degeneration (Ma, 2005).

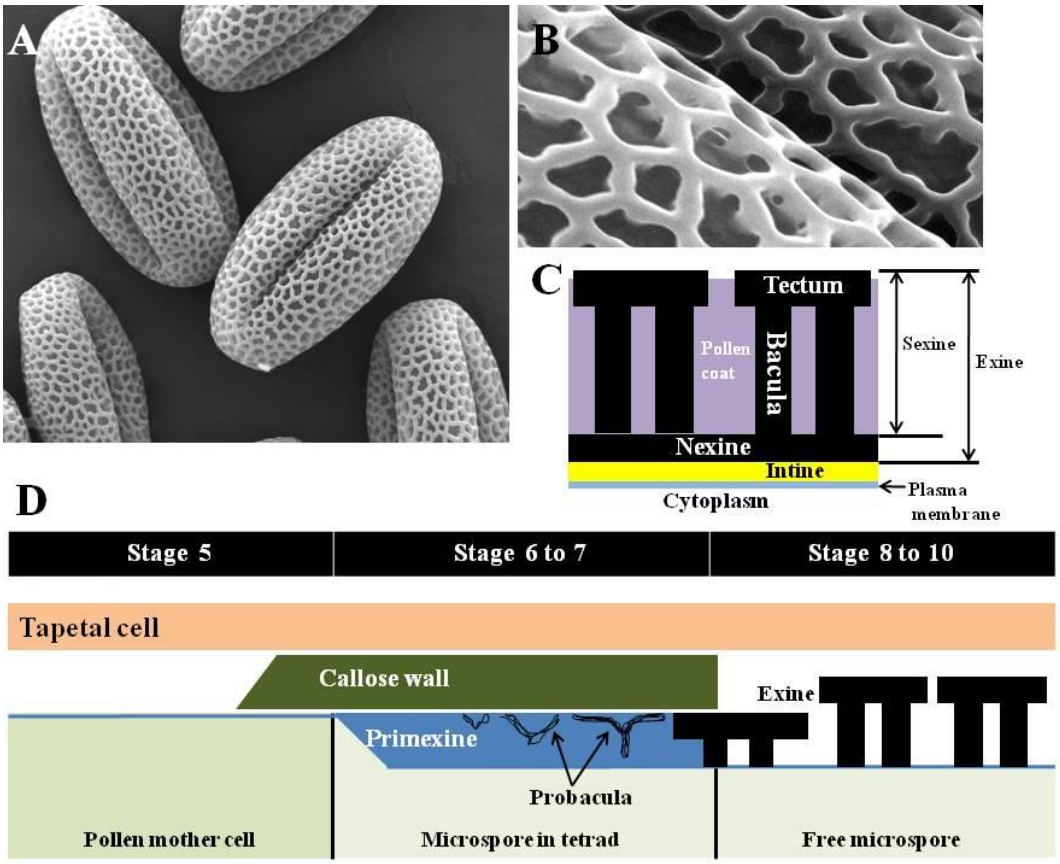


Figure 1.2 *Arabidopsis* (Col-0) pollen wall structure (A and B), schematic diagram of the main features of a mature pollen grain wall (C) and of exine formation (D).

(A) and (B) Scanning electron micrograph, kindly provided by Dr. Etienne Grienenberger. (C) and (D) Diagrams are modified and adapted based on Suzuki et al., (2008).

The pollen wall consists essentially of two layers: the intine and the exine. The intine is mostly synthesized by the haploid microspore itself. However, the tapetum is responsible for the production and secretion of the exine, generally known as a mixture of protein, lipids and aromatic molecules (Sanders et al., 1999; Ma, 2005). After synthesis and deposition of the pollen wall, the tapetum cells are degraded via programmed cell death, and pollen grains continue to develop and mature. Although the exact composition of the exine and other components of the pollen wall are not completely understood, it is known that functional tapetum cells are essential for the development of viable pollen grains (Zhang and Chiang, 1997; Vizcay-Barrena and Wilson, 2006). The major component of the exine is termed sporopollenin, a complex biopolymer which is composed of long-chain fatty acids and poorly characterized phenolic molecules coupled by ester and ether bonds (Scott et al., 2004; Blokker et al., 2006).

1.2 Related biopolymers: suberin, cutin and sporopollenin

Land plants have evolved aliphatic biopolymers that protect their cell surfaces against dehydration, pathogens, and chemical and physical damage. Moreover, to adapt rapid barriers to new physiological and environmental conditions, they have evolved elaborate regulatory networks (Pichersky and Gang, 2000; Bowman et al., 2007; Franke and Schreiber, 2007). Land plants have three types of lipid-based polyesters: cutin, suberin and sporopollenin. They are composed of both fatty acid- and glycerol-based extracellular polymers with phenolic components that represent a barrier of resistance to both water and solutes (Hose et al., 2001; Beisson et al., 2007). Cutin and suberin are better characterized than sporopollenin. The cuticle, which consists of cutin and waxes, is deposited at the outermost area of the epidermis cell wall as a thin membrane-like layer. On the other hand, the location of suberin deposition in plants is highly variable. For example, the outer bark of trees, the outer tissues of stems, the outer and inner sealing tissues of primary roots, and seed coat are known to contain the polymer suberin (Beisson et al., 2007; Soler et al., 2007). Moreover, suberin is located in the Casparian strip, a waterproof band of wall material in the radial (anticlinal) cell walls of the endodermis which has various chemical components described as composed of lignin or suberin or

both (Kolattukudy, 2001).

Cutin is a biopolymer containing mainly ω-hydroxylated and epoxy hydroxylated fatty acids 16 to 18 carbons in length ester-linked into a polymer (Table 1.2; Kolattukudy, 2001). On the other hand, suberin has distinctive domains composed of aliphatic and aromatic constituents (Table 1.2; Kolattukudy, 2001; Soler et al., 2007; Pollard et al., 2008). The aliphatic domain of suberin is a polyester polymer, comprised of mainly ω-hydroxy fatty acids and α,ω-dicarboxylic acids as very long aliphatic constituents with over 20 carbon chain-length (Kolattukudy, 2001; Pollard et al., 2008). Hydroxycinnamic acids in aromatic domains are proposed to covalently link the aliphatic suberin polyester to the cell wall, possibly some monolignols (Kolattukudy, 2001). A recent macromolecular structure model for suberin proposed that the monomeric blocks contain long-chain α,ω-dicarboxyolic acids esterified to glycerol at both ends as the core of the suberin macromolecule (Franke and Schreiber, 2007). These polyol glycerols provide a cross-linked two- and three-dimensional network by the formation of ester linkages to additional α,ω-dicarboxylic acids and ω-hydroxy fatty acids, leading to a rigid and insoluble polymer (Franke and Schreiber, 2007).

Table 1.2 Summary of common cutin, suberin and sporopollenin monomers, identified by functional group and possible biosynthetic reactions for each biopolymer.

Common monomer type	Cutin	Suberin	Sporopollenin
LCFA (C16 ~ C18)	Yes	Yes	Yes (C10~)
VLCFA	No	Yes (~C30)	Yes
Fatty alcohols $CH_3(CH_2)_mOH$	Yes	Yes	N/A
Hydroxy fatty acid	Yes	Yes	Yes
α,ω-dicarboxylic fatty acid	Yes	Yes	N/A

Table 1.2 Summary of common cutin, suberin and sporopellenin monomers, identified by functional group and possible biosynthetic reactions for each biopolymer. (cont.)

Common monomer type	Cutin	Suberin	Sporopollenin
Phenolics	N/A	Ferulic acid / p-hydroxy cinammate	Oxygenated phenylpropanoid (p-coumaric, caffeic acid)
Glycerol	Yes	Yes	N/A
Funtional group	**Cutin**	**Suberin**	**Sporopollenin**
Ketones	Yes	N/A	Yes
Epoxide	Yes	N/A	N/A
Ester	Yes	Yes	Yes
Ether	N/A	N/A	Yes
Possible enzyme reaction	**Cutin**	**Suberin**	**Sporopollenin**
Reduction (carboxylic acid to aldehyde)	N/A	Yes	N/A
Reduction (aldehyde to alcohol)	N/A	Yes	N/A
Esterification (carboxylic acid and alcohol)	Yes	Yes	Yes
Etherification	N/A	N/A	Yes
Hydroxylation by P450	Yes	Yes	Yes
Oxidation by dehydrogenase (alcohol to aldehyde)	N/A	Yes	N/A
Oxidation) by dehydrogenase (aldehyde to carboxylic acid)	N/A	Yes	N/A
Epoxidation	Yes	Yes	N/A

N/A, No information available
VLCFA, Very long chain (longer than 22C) fatty acids

The main constituent of the pollen exine layer is sporopollenin, an extremely robust and durable biopolymer found in the spores of bryophytes and ferns and in pollen exine of seed plants (Bohne et al., 2003). The chemical composition of sporopollenin remains poorly characterized because it is extremely resistant to chemical and biological

degradation procedures (Bubert et al., 2002; Vizcay-Barrena and Wilson, 2006). However, available data are consistent with a sporopollenin polymer consisting of phenolic and fatty acid–derived constituents that are covalently coupled by ether and ester linkages (Table 1.2; Ahlers et al., 2000; Rozema et al., 2001; Bubert et al., 2002; Ahlers et al., 2003; Scott et al., 2004). Recently, a partially male sterile mutant phenotype was described for *Arabidopsis CYP703A2*, a member of cytochrome P450 gene family, and a possible substrate of the CYP703A2 enzyme was reported (Morant et al., 2007). The mutant plants showed impaired pollen wall development with defective exine. Additionally, CYP703A2 hydroxylated medium-chain saturated fatty acids to the corresponding mono-hydroxy fatty acids, with a preferential hydroxylation of lauric acid (C12:0) *in vitro* and *in vivo*. A model for the role of CYP703A2 was proposed to explain its involvement of sporopollenin formation (Figure 1.3). According to the model, monomeric hydroxy fatty acid units generated by CYP703A2 and phenylpropanoid derived molecules like p-coumaric acid and caffeic acid are connected by ether and ester linkages, and they give rise to the backbone structure of the sporopollenin polymer. However, to form a three-dimensional network or cross-linked polymer, additional ether or ester linkages may be needed. This could mean that currently unknown enzymes co-expressed with CYP703A2 contribute to further strengthening of the sporopollenin polymer.

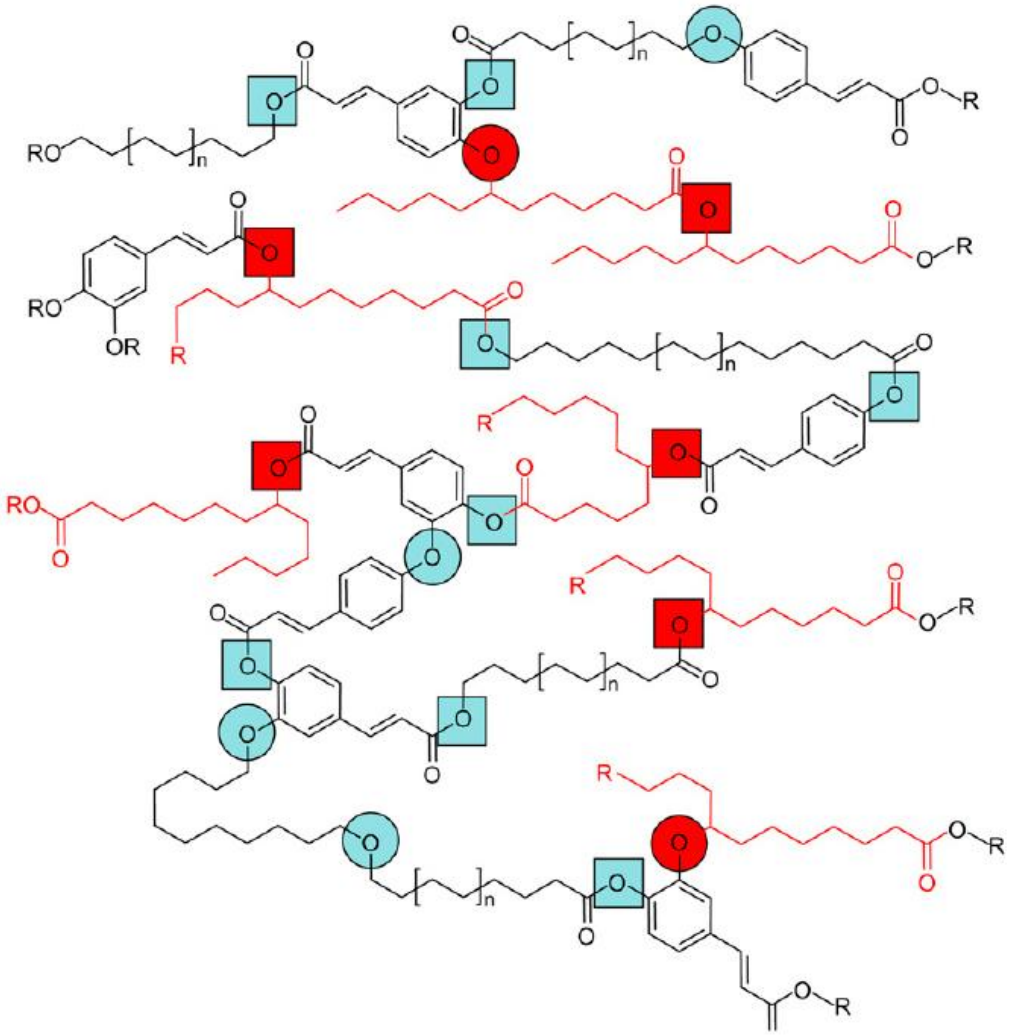


Figure 1.3 Model of the role of CYP703A2 in sporopollenin formation.

Monomeric units derived from CYP703A2-catalyzed hydroxylation of lauric acids are shown in red. The participation of oxygen atoms within these units in ether and ester linkages in the formation of the sporopollenin biopolymer is illustrated by circles and squares, respectively. The p-coumaric and caffeic acid units illustrate the presence of phenylpropanoids in the sporopollenin polymer (Morant et al., 2007). Image was reprinted with permission of American Society of Plant Biologists (Copyright © American Society of Plant Biologists).

1.3 Mutants showing male sterility in *Arabidopsis*

Numerous male sterile or partially sterile mutants that have been isolated and characterized in *Arabidopsis* define some of the key processes in anther and male gametophyte development. Table 1.3 shows some examples of male sterile *Arabidopsis*

mutants that have been characterized.

Male sterile mutants displaying apparently normal microspore and tapetum development, but with defects in pollen wall formation and pollen maturation, have started to shed light on the cell biology and biochemistry of pollen wall, exine, and sporopollenin development and biosynthesis. For example, the *Arabidopsis MALE STERILITY* (*MS2*) gene, defined by the *ms2* male-sterile mutation, is required for exine formation and encodes a predicted fatty acyl-CoA reductase (Aarts et al., 1997). Expression of *MS* in *Escherichia coli* results in the production of alcohols from endogenous *E. coli* fatty acids (Doan et al., 2009), supporting a function in fatty acid metabolism. Loss of function of the *FACELESS POLLEN1*/*WAX2*/*YRE/CER3* gene results in defects in both exine and epicuticular wax deposition, and the gene encodes a putative enzyme of unknown function presumably involved in both wax biosynthesis and pollen wall formation (Ariizumi et al., 2003; Chen et al., 2003; Rowland et al., 2007). Recently, *Arabidopsis* genes encoding the cytochrome P450 enzymes *CYP704B1*, as well as *CYP703A2* described above, have been shown to be involved in the biosynthesis of sporopollenin, since mutants show severe to moderate defects in exine deposition and pollen grain development (Morant et al., 2007; Dobritsa et al., 2009). *In vitro* assays indicate that the enzymes catalyze hydroxylation of medium- to long-chain fatty acid substrates. While CYP703A2 is an in-chain hydroxylase with substrate preference for lauric acid (C12:0), CYP704B1 catalyzes the ω-hydroxylation of long-chain fatty acids (Morant et al., 2007; Dobritsa et al., 2009).

Table 1.3 Summary of male sterile *Arabidopsis* mutants.

Mutant name	Stages of gene expression	Defective main event	Encoding / related metabolism	Phenotype	Reference
spl/nzz (*SPOROCYTLESS/NOZZLE*)	Stage 3, 4 and 5	Sporogenesis	TF (MADS box) Possession of AGAMOUS binding site	Fail to differentiation to male and female sporocyte	(Yang et al., 1999)
ems1/exs (*EXCESS MALE SPOROCYTES /EXTRA SPOROGENOUS CELL*)	Stage 3, 4 and 5	Tapetum differentiation	Putative leucine-rich repeat receptor protein kinase (LRR-RPK)	A tapetum is absent in anther and extra microsporocytes are present	(Zhao et al., 2002)
tpd1 (*TAPETUM DETERMINANT 1*)			Unknown protein of 176 amino acids		(Yang et al., 2003a)
mmd1/duet (*MALE MEIOCYTE DEATH1*)	Stage 6 and 7	Defective meiosis	PHD-finger Nuclear Protein	Aberrant tetrad	(Yang et al., 2003b)
ams (*ABORTED MICROSPORES*)	Stage 7 and 8	Tapetum and microspore development	MYC TF (basic helix-loop-helix (bHLH))	The complete absence of pollen grains due to both the microspores and tapetum disintegration	(Sorensen et al., 2003)
dex1 (*DEFECTIVE IN EXINE PATTERN FORMATION*)	Low and relatively equal amounts throughout the plant	Defective primexine	Membrane protein containing several potential calcium-binding domains	Normal primexine development is disrupted, which affects the conformation of the plasma membrane and sporopollenin deposition.	(Paxson-Sowders et al., 2001)

Table 1.3 Summary of male sterile *Arabidopsis* mutants. (cont.)

Mutant name	Stages of gene expression	Defective main event	Encoding / related metabolism	Phenotype	Reference
myb103 (*R2R3 MYB GENE FAMILY*)	N.A.	Defective exine formation	Action of upstream of MS2 Tapetum Development, callose dissolution and exine formation	Premature degeneration of the tapetum	(Zhang et al., 2007)
ms1 (*MALE STERILITY 1*)	Stage 7 and 8	Defective exine formation	Tapetum specific TF, PHD finger motif	Abnormal pollen wall with aberrant deposition of the exine due to defective development of the tapetum	(Wilson et al., 2001)
myb32 *myb4*	N.A.	Pollen development	Tapetum specific Phenylpropanoid and flavonoid	Aberrant pollen grain due to absence of cytoplasm	(Preston et al., 2004)
ms2 (*MALE STERILITY 2*)	Stage 8, 9, and 10	Defective exine formation	Tapetum specific Fatty acid reductase involved in the synthesis of sporopollenin	No exine in pollen wall	(Aarts et al., 1997)
CYP703A2	Stage 7, 8 and 9 (closed buds)	Defective exine formation	Tapetum specific The conversion of medicum-chain saturated fatty acids to corresponding monohydroxylated fatty acid	Defective exine in pollen wall	(Morant et al., 2007)

Table 1.3 Summary of male sterile *Arabidopsis* mutants. (cont.)

Mutant name	Stages of gene expression	Defective main event	Encoding / related metabolism	Phenotype	Reference
nef1 (*NO EXINE FORMATION1*)	N.A.	Lipid accumulation in the plastid	Tapetum specific Plastid integral membrane protein	Growth defect Aberrant lipid accumulation of the plastid in tapetum	(Ariizumi et al., 2004)
cer1	N.A.	The aberrant pollen coat structure	Conversion of stem wax C30 aldehydes to C29 alkanes	The rehydration deficiency by depleting pollen coat lipids leading to a pollen-stigma signaling failure	(Aarts et al., 1995)
cer6-2	N.A.		Lipid elongation in production >28 Carbons long	Unevenly distributed pollen coat	(Fiebig et al., 2000)
flp1 (*FACELESS POLLEN-1*)	N.A.		Allelic to *CER3*, unknown function	The smooth appearance of the pollen surface is due to excess tryphine. Reduced cuticular wax in stems and siliques	(Ariizumi et al., 2003; Rowland et al., 2007)
scp (*SIDECAR POLLEN*)	N.A.	Defective mitosis in pollen grain	Gametophytic mutation Asymmetric cell division of pollen development	No cell wall separating the two vegetative-like nuclei	(Chen and McCormick, 1996)

Table 1.3 Summary of male sterile *Arabidopsis* mutants. (cont.)

Mutant name	Stages of gene expression	Defective main event	Encoding / related metabolism	Phenotype	Reference
myb26/ms35 (MALE STERILE 35)	N.A.	No anther dehiscence	Endothecial wall thickenings	Disruption of the development of lignified, cellulosic secondary thickenings in the anther endothecium, preventing anther dehiscence.	(Yang et al., 2007b)
myb26	N.A.		Endothecial wall thickenings (allelic to ms35)	No dehiscence process due to the failure that endothecial cell undergo the lignification	(Steiner-Lange et al., 2003)
fad3fad7fad8	N.A.		Deficiency in synthesis of jasmonic acid (JA)	No dehiscence process due to the failure of controlling water transport in the anther	(McConn and Browse, 1996)
dde1 (*DELAYED DEHISCENCE 1*)	Stage 10 and 11		12-oxophytodienoate (OPDA) reductase, in the JA biosynthesis	No dehiscence process due to the failure of controlling water transport in the anther	(Sanders et al., 2000)
dad1 (DEFECTIVE IN ANTHER DEHISCENCE1)	N.A.		Chloroplastic phospholipase A1 in JA biosynthesis	Anthers do not dehisce at flower opening	(Ishiguro et al., 2001)

N.A., No information available

1.4 Phenylpropanoid metabolism-like enzymes

The genes encoding enzymes in phenylpropanoid, flavonoid, and monolignol metabolism have been identified and characterized in many species. Following the generation of complete genome sequence data from *Arabidopsis*, the full set of phenylpropanoid and lignin biosynthetic genes was identified (Costa et al., 2003; Raes et al., 2003; Ehlting et al., 2005). Moreover, the likely sets of true *Arabidopsis* phenylpropanoid gene family members were annotated by sequence similarity searches and gene expression profiling (Raes et al., 2003; Costa et al., 2005; Ehlting et al., 2005). Whereas these phenylpropanoid-like genes are closely related to true phenylpropanoid genes, they likely encode enzymes of unknown specific biochemical function in diverse metabolic pathways in natural product biosynthesis.

Plant adaptation relies on significant metabolic changes that are reflected by the evolution of large gene families. The elaborate decoration of a basic carbon skeleton structure, by using enzymes that catalyze oxidation, reduction, hydroxylation, decarboxylation, glycosylation, and acylation reactions generates the large diversity in plant secondary metabolites. These reactions are well known in phenylpropanoid metabolism, but phenylpropanoid-like enzymes may catalyze similar reactions using different starter carbon skeletons. A well-studied example would be the large family of acyl transferases, BAHD, which is an acronym composed of the first letter from biochemically characterized enzymes of this family (BEAT, AHCT, HCBT, and DAT) (D'Auria, 2006). The HCT enzyme involved in monolignol biosynthesis is part of a large family of “BAHD” acyl tranferases. The BAHD members share sequence identity, conserved motifs and catalytic mechanisms (D'Auria, 2006). The distinct classes of BAHD enzymes are identified by phylogenetic analysis. One subgroup consists of enzymes including HCT/HQT enzymes which are responsible for the formation of hydroxycinnamoyl quinate/shikimate esters in *Nicotiana tabacum* and *Arabidopsis* (D'Auria, 2006). The absence of this enzyme activity results in severe phenotype changes, such as dwarfed plants and deficiency in syringyl lignin units (Hoffmann et al., 2004). Coumaroyl-shikimate esters formed by the action of HCT are known to be intermediates in P450-mediated hydroxylation of coumaric acid to yield caffeic acid (Schoch et al.,

2001). BAHD enzymes in other clades are involved in modification of other secondary metabolites (D'Auria, 2006). Recently, a function of SHT (Spermidine Hydroxycinnamoyl Transferase, At2g19070), one member of the BAHD family was characterized (Grienenberger et al., 2009). *SHT* is specifically expressed in tapetum cells at early flower development stages. SHT catalyzes acylation using hydroxycinnamoyl-CoAs as donor and spermidine as acceptor substrates, generating mono- to trihydroxycinnamoyl spermidine, which is involved in the formation of the pollen coat (Grienenberger et al., 2009).

A key enzyme in phenylpropanoid metabolism is 4-coumarate:CoA ligase (4CL; (Hahlbrock and Scheel, 1989), which generates hydroxycinnamoyl-CoA esters that are central intermediates in the biosynthesis of lignin monomers, flavonoids and other secondary metabolites (Hahlbrock and Scheel, 1989). Previous work in the Douglas lab identified a large set of plant-specific 4CL-like enzymes of mostly unknown function (Souza et al., 2008). The *Arabidopsis* 4CL-like enzyme most closely related to true 4CLs is *ACYL-COA SYNTHETASE5* (*ACOS5*; At1g62940). *ACOS5* is a single copy gene that has homologs in all plants investigated, and such homologs are expressed specifically in anthers (Souza Cde et al., 2008; de Azevedo Souza et al., 2009). The function of *ACOS5* was investigated using a reverse genetic approach which showed that an *acos5* loss of function mutant is male sterile, with complete loss of pollen grain formation. The *acos5* mutation co-segregated with the male sterile phenotype. Normal development of microspores is arrested in the *acos5* mutant at anther development stage 9, and they have apparently defective exine (de Azevedo Souza et al., 2009). A transmission electron microscopy (TEM) of developing *acos5* pollen grains suggests a complete deficiency of sporopollenin and exine. Moreover, the results of *in situ* hybridization experiments demonstrated that *ACOS5* has a transient and tapetum preferred expression pattern and is most highly expressed in the stages immediately preceding the appearance of the visible mutant phenotype.

In this thesis, I report and discuss the further characterization of the *acos5* phenotype and the characterization of a total of five phenylpropanoid-like genes and the corresponding enzymes, which are co-expressed with *ACOS5* and are required for pollen wall formation. Additionally a potential pathway of sporopollenin biosynthesis catalyzed

by these enzymes will be explored in more detail in Chapters 5 and 6.

1.5 Research objectives

1. Determine if enzymes encoded by genes co-expressed with *ACOS5* are required for male fertility.
2. Characterize *ACOS5* co-expressed genes and enzymes that are required for male fertility, and determine possible roles in sporopollenin biosynthesis.
3. Test whether *ACOS5* co-expressed enzymes work in the same biochemical pathway.
4. Investigate the evolutionary conservation of the set of *ACOS5* co-expressed genes in land plants using bioinformatics tools.

Chapter 2. A novel fatty acyl-CoA synthetase (ACOS5) is required for pollen development and sporopollenin biosynthesis in *Arabidopsis thaliana*

2.1 Summary

Acyl-CoA Synthetase (*ACOS*) genes are related to 4-coumarate:CoA ligase (4CL) but have distinct functions. The *Arabidopsis thaliana* ACOS5 protein is in clade A of *Arabidopsis* ACOS proteins, the clade most closely related to true 4CL proteins. This clade contains putative non-peroxisomal ACOS enzymes conserved in several angiosperm lineages and in the moss *Physcomitrella patens*. Although its function is unknown, *ACOS5* is preferentially expressed in the flowers of all angiosperms examined. Together with genetic, phenotypic, bioinformatic experiments performed by a previous student, Dr. Clarice de Azevedo Souza, my data show that an *acos5* mutant produces no pollen in mature anthers, no seeds by self-fertilization, and is severely compromised in pollen wall formation, apparently lacking sporopollenin or exine. Moreover, data from *in vitro* enzymes assays provided by a collaborator, Dr. Erich Kombrink, indicate that recombinant ACOS5 enzyme has a broad *in vitro* preference for medium-chain fatty acids. Based on this work, I propose that *ACOS5* encodes an enzyme that participates in a conserved and ancient biochemical pathway required for sporopollenin monomer biosynthesis that may also include the *Arabidopsis* CYP703A2 and MS2 enzymes.

2.2 Introduction

2.2.1 *4-coumarate:CoA ligase-like* (*4CL-like*) and *acyl-CoA synthetase* (*ACOS*) genes

The enzyme 4-coumarate:CoA ligase (4CL) plays important roles in phenylpropanoid metabolism by generating CoA esters of p-coumaric acid and its derivatives. These activated CoA esters are precursors utilized for the biosynthesis of various plant secondary metabolites such as lignin, flavonoids, suberin and signal molecules in plant-microbe interactions (Hahlbrock and Scheel, 1989). Many adenylate-forming enzymes related to 4CL were identified by sequence homology searches in plant and other genomes. In spite of the remarkable diversity of their substrates, adenylate-forming enzymes use the same two-step catalytic reaction to activate carboxylate substrates. In the first step, adenylation of the carbonyl group of substrates takes place by condensation with ATP to release pyrophosphate (Schneider et al., 2005). The resulting carboxylate adenylate (acyl-AMP) intermediate is very reactive. Thus, it is necessary that this intermediate react with either a thiol, amide or alcohol in the next step to generate the corresponding product by releasing AMP. Adenylate-forming enzymes contain highly conserved putative AMP-binding domains and the adenylate-forming enzyme superfamily is divided into diverse clades containing enzymes such as fatty acyl-CoA synthetases, acetyl-CoA synthetases, 4-coumarate:CoA ligases, chlorobenzoate:CoA ligase, non-ribosomal polypeptide synthetases, and firefly luciferases (Stuible et al., 2000). In *Arabidopsis*, 44 genes have been identified that encode proteins containing the AMP-binding domain (Shockey et al., 2002; Shockey et al., 2003). Among these proteins, several fatty acyl-CoA synthetases, four 4-coumarate:CoA ligases (4CLs), and one acetyl-CoA synthetase have been identified (Ehlting et al., 1999; Fulda et al., 2002; Schnurr et al., 2002; Hamberger and Hahlbrock, 2004). The four isoforms of 4-coumarate:CoA ligase (4CL) presumably constitute the complete enzyme family (Raes et al., 2003), but represent only a portion of the diversity of enzymes related to 4CL in plant genomes.

Using an *in silico* similarity search based on the amino-acid sequences of known *Arabidopsis* genes encoding 4-coumarate:CoA ligase (4CL), the Douglas lab and other labs (Raes et al., 2003; Costa et al., 2005; Ehlting et al., 2005; Souza Cde et al., 2008) identified nine putative genes as members of an *Arabidopsis 4CL-like* gene family which

encode a plant-specific clade of enzymes closely related to true 4CLs. Previous studies have shown that some enzymes encoded by this family of 4CL-like genes activate cinnamic, benzoic, or fatty acid derivatives *in vitro* including precursors of jasmonic acid (JA) (Costa et al., 2005; Schneider et al., 2005). Additionally, recent data indicate that many 4CL-like proteins accept a relatively broad and overlapping range of various medium- and long-chain fatty acid as substrates, instead of the hydroxycinnamate substrates accepted by true 4CL enzymes to form the corresponding CoA esters (Kienow et al., 2008). Based on their activities towards acyl substrates, we have designated those *4CL*-like genes with unknown *in vivo* substrates as *Acyl-CoA Synthetase* (*ACOS*; formerly referred to as the *ACS* genes; Souza Cde et al., 2008).

Phylogenetic analysis of amino-acid sequences revealed five ACOS subclades, each containing at least one ACOS member from each species, suggesting conserved biochemical functions for ACOS enzymes (Figures 2.1). In four of five subclades, almost all proteins contain the PTS1 (peroxisomal target sequence) in their C-termini. On the other hand, subclade A, which includes the *Arabidopsis* gene *ACOS5* (At1g62940), is most closely related to *bona fide* 4CLs and contains single copy genes from *Arabidopsis*, poplar, and rice, which all encode enzymes without predicted PTS1 (Souza Cde et al., 2008). Expression analysis revealed that *ACOS5* is preferentially expressed in flowers, whereas expression of the poplar ortholog is specific to male flowers (Souza Cde et al., 2008). Thus, a function for ACOS5 in anther and/or pollen development has been proposed (Souza Cde et al., 2008).

In this chapter, I further analyzed *acos5-1*, a loss-of-function allele of the *Arabidopsis ACOS5* gene, which was first characterized by a previous student, Dr. Clarice de Azevedo Souza, using light and transmission electron microscopy. In addition, the ability of *ACOS5* to complement the male sterile phenotype in the *acos5-1* background was tested.

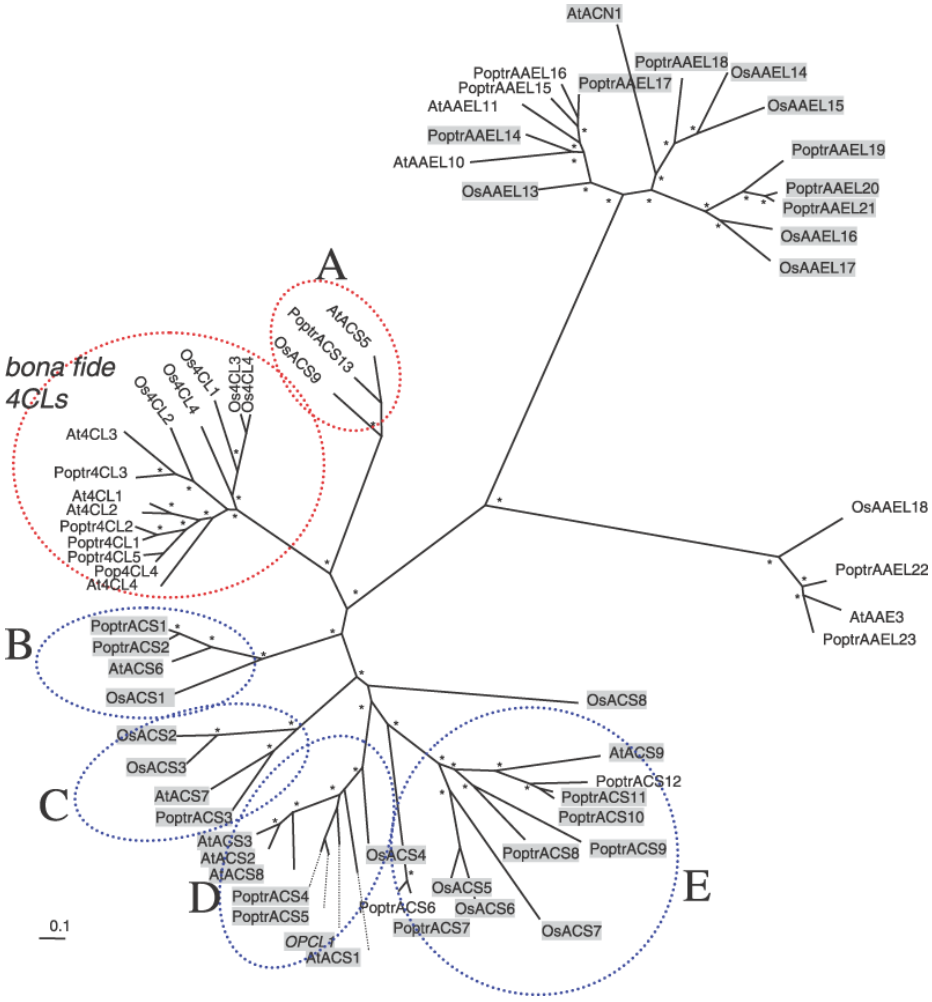


Figure 2.1 Phylogenetic relationships of plant-specific acyl-CoA synthetases (ACSs, ACOSs) from three fully sequenced angiosperm genomes.

Translated nucleotide sequences corresponding to ACS (ACOS5) genes from *Arabidopsis*, poplar and rice were aligned and an unrooted phylogenetic tree generated. Nodes with bootstrap values above 70% are shown by stars. The 4-coumarate:CoA ligase (4CL) and ACS (ACOS) clades A–E discussed in the text are circled and contain at least one representative of each plant species. Protein names in shaded boxes contain the PTS1 peroxisomal target signal. Bar represents 0.1 amino acid changes (Souza Cde et al., 2008). Image was reprinted with permission of New Phytologist (Copyright © New Phytologist (2008)).

2.3 Materials and methods

2.3.1 Plant material and growth conditions

Arabidopsis thaliana wild-type (Col-0) and mutant plants were grown in soil (Sunshine mix 5; Sungrow Horticulture) in controlled environment chambers at 20℃ under long-day conditions (18 h light).

2.3.2 Complementation of *acos5* mutants

A 4368-bp ACOS5 genomic fragment was amplified using the Plantinum Taq DNA polymerase High Fidelity (Invitrogen) with gene-specific primers (Table 2.1) and cloned into pCR8/GW/TOPO (Invitrogen). After verification by sequencing, the fragment was subcloned into the pGWB1 Gateway binary vector (Nakagawa et al., 2007) and introduced into Agrobacterium tumefaciens. Then, *acos5-1* heterozygous plants were transformed using the floral dip method (Clough and Bent, 1998). T1 seeds were sown in half-strength Murashige and Skoog salts (Sigma-Aldrich), supplemented with 1% sucrose and 0.6% agar medium containing 25 mg/L hygromycin. Individual T1 lines were allowed to self pollinate and progeny genotypes were tested with respect to the *ACOS5* locus. The presence of respective transgenes was tested using PCR with the primers given in Table 2-1.

2.3.3 Nucleic acid methods

Genomic DNA extraction was performed using young leaf tissue ground in a bead beater at 4℃, with the use of the Nucleon PhytoPure Kit (Amersham-Pharmacia), according to the manufacturer's instructions. *Arabidopsis* RNA was isolated from tissues frozen in liquid nitrogen and ground to a fine powder by using RNeasy Plant Mini Kit (Qiagen) following the manufacturer's instructions.

2.3.4 Phenotypic analyses

To obtain cross sections of developing anthers, wild-type and homozygous

double mutant inflorescences were fixed in 20ml fixative mixture (4% paraformaldehyde, 2.5% glutaraldehyde and 0.05M sodium phosphate) overnight and directly dehydrated through an aqueous alcohol series (30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% and 100% three times) for 30 min each, without postfixation. Samples were first transferred to a propylene oxide solution in 50% (v/v) ethanol and then two-time washed in only propylene oxide for 30 min. For infiltration of Spurr's epoxy resin (Canemco), propylene oxide solution was replaced with following resin series for 4 hr each: 10%, 25%, 50%, 75%, and 100% twice. For bright-field microscopy, 1 µm sections were cut with glass knives (Leica) on a microtome, mounted on glass slides, heat fixed to the slides and stained with toluidine blue. Sections were photographed using a light microscope.

For TEM, *Arabidopsis* wild-type and *acos5* mutant inflorescences were postfixed in 1% osmium tetraoxide 0.05M sodium cacodylate (pH 6.9) for 30 min and rinsed twice in distilled water. Thin sections (70 nm) were taken using a Leica Ultracut T and Druuker diamond Histoknife. Sections were placed on 100-mesh copper grids and stained for 15 min with 2% uranyl acetate in 70% (v/v) methanol, rinsed thoroughly with water, and stained for 10 min with lead citrate (Sato's Lead). Sections were visualized using a Hitachi H7600 transmission electron microscope.

2.3.5 Accession number

Sequence data from this chapter can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the following accession numbers: *Arabidopsis ACOS5*, At1g62940.

2.3.6 Sequences of primers

Table 2.1 Primers used in this study.

	Genotyping	
Gene	Primer name	Sequence (5' → 3')
At1g62940	CLL4F	TTTGGTACCGTTTAAAAATGGAGTCAAAAG
	EcoR1 reverse	AAAGAATTCCATTGCGGTATCTCCGCA
	dspn1	CTTATTTCAGTAAGAGTGTGGGGTTTTG

Table 2.1 Primers used in this study. (cont.)

	Complementation construct	
Gene	Primer name	Sequence (5' → 3')
At1g62940	ACLL5F	GATTGGTTAAGTTCATACGTTC
	ACLL5R	GCATGAGAAAGCAGCGTG

	Checking complementation	
Gene	Primer name	Sequence (5' → 3')
At1g62940	F CompACLL5	GAG CAA CCA TCT TCC CTT GTG ATT GC
	R CompACLL5	AGA AGC AGG AGA CAA ATG CGG AGA

2.4 Results

2.4.1 The ability of the *ACOS5* gene to complement the male sterile phenotype

To test the ability of the *ACOS5* gene to complement the male sterile phenotype in the *acos5-1* background, I introduced an about 4.4-kb DNA region from an *Arabidopsis* wild-type (Columbia-0 [Col-0]) plant, containing 1.9 kb of the promoter sequence and the complete transcribed region of ACOS5 (Figure 2.2), into *ACOS5 acos5-1* heterozygote plants by *Agrobacterium tumefaciens*–mediated transformation. Four T1 lines harbouring the *ACOS5* transgene were subjected to PCR-aided genotyping, and one was established as being *ACOS5/ACOS5,* one as being *ACOS5/acos5-1,* and two as being *acos5-1 acos5-1*. All plants were fully fertile, suggesting that the introduced *ACOS5* transgene had complemented the *acos5-1* mutation in the two homozygous lines. We further determined the genotypes and phenotypes of 18 T2 progeny from each T1 line that had inherited the ACOS5 transgene (i.e., that was either homozygous or hemizygous for the *ACOS5* transgene, based on hygromycin resistance specified by the T-DNA insertion). For each T1 line, including the two that were homozygous for the *acos5-1* allele and gave rise only to *acos5-1/acos5-1* T2 progeny, all 18 T2 plants were fully fertile, confirming the ability of *ACOS5* to complement the male sterile mutant phenotype.

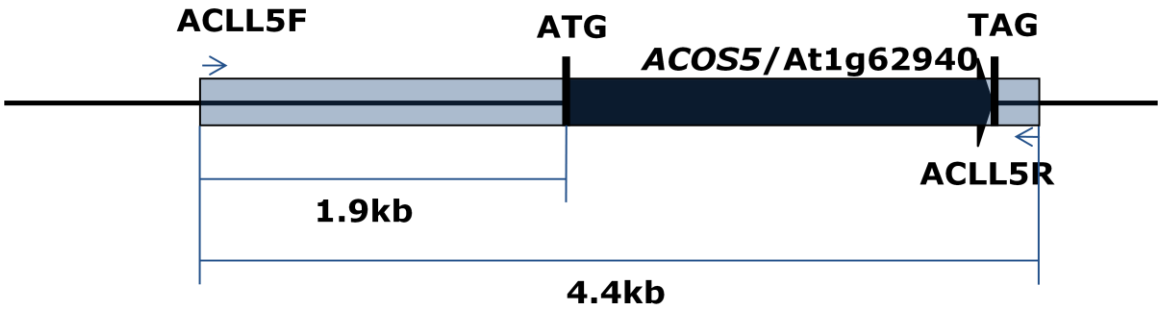


Figure 2.2 The construct used for *acos5* complementation.

Schematic representation of the construct used for the *acos5* complementation test. Shaded boxes show the *ACOS5* transgene, containing 1.9 kb of the promoter sequence, the 3' untranslated region/terminator regions (light gray), and the transcribed region between the start and stop codons (thick black arrow). Primers used to amplify genomic DNA are indicated with arrows above and below the diagram and are shown in Table 2.1 (de Azevedo Souza et al., 2009). Image was reprinted with permission of the American Society of Plant Biologists (Copyright © American Society of Plant Biologists).

2.4.2 Anther and microspore development in the *acos5-1* mutant

To further pinpoint the stage of anther development defective in the *acos5* mutant, I used transmission electron microscopy to gain higher-resolution images of developing anthers in the *acos5* mutant and compared these to the corresponding images from wild-type plants. Figures 2.3A to 2.3D show that, at stage 7, characterized by the presence of tetrads that form after meiosis, wild-type and *acos5* mutant microspores were similar in morphology, with characteristic callose walls. At this stage, tapetal cells in the mutant appeared normal. At stage 8, when free microspores had been released from tetrads following callose wall hydrolysis, massive deposition of a thick sporopollenin-containing exine on the nexine layer was evident in the wild-type anthers (Figures 2.3E and 2.3F). The exine had started to develop into a thick, reticulated wall characteristic of pollen grains, residual primexine was visible in spaces of exine baculae, which are characteristic of this stage (Goldberg et al., 1993; Sanders et al., 1999; Scott et al., 2004; Ma, 2005), and microspores were nonvacuolated. In comparison with the wild type, free *acos5* mutant microspores at stage 8 had a similar nonvacuolated morphology but contained thinner walls that were devoid of the pronounced reticulate exine wall seen in the wild

type (Figures 2.3G and 2.3H). In place of a well-defined exine, these microspores contained an amorphous substance adhering to the nexine that could be residual primexine or unpolymerized sporopollenin precursors. Again, tapetal cells at this stage appeared normal. At stage 9, wild-type pollen grains contained thick, reticulated exine walls and an intine layer. In stage 9 of the *acos5* mutant, many pollen grains were in various stages of lysis and degradation (Figure 2.3K) and had thin cell walls that were devoid of a characteristic exine layer (Figure 2.3L) but often retained amorphous wall material outside the intine and apparent nexine. In other cases, relatively intact pollen grains were observed in *acos5* anthers at stage 9 (Figure 2.3M), but in many cases, pollen walls were completely devoid of wall material outside the intine (Figure2.3N). Again, tapetal cells in the acos5 mutant anthers appeared normal at this stage (Figures 2.3K and 2.3M), suggesting that the defect in *acos5* pollen development is primarily due to the lack of sporopollenin deposition and exine formation, rather than a general defect in tapetal cell development. This analysis pinpoints the defect in microspore development in the *acos5* mutant to the deposition of the sporopollenin-rich exine wall at stages 8 and 9 of anther development, while other aspects of anther development, including tapetal cell development, appear normal.

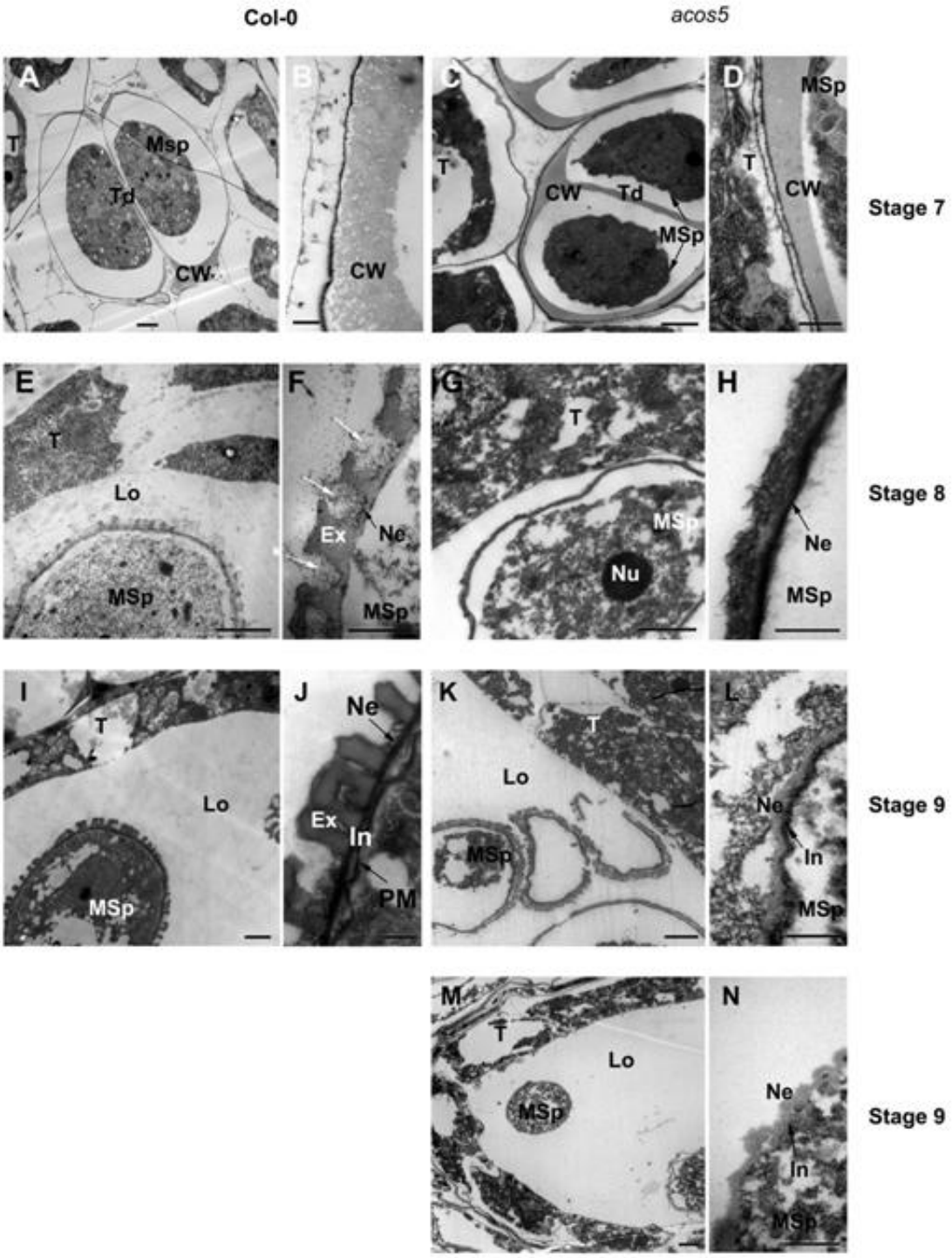


Figure 2.3 Transmission electron micrographs of wild-type (Col-0) and *acos5* mutant anthers, pollen grains, and pollen walls.

(A), (B), (E), (F), (I) and **(J)** Microspore and pollen wall development in Col-0 wild-type

plants. **(C), (D), (G), (H)** and **(K)** to **(N)** Microspore and pollen wall development in *acos5* mutant anthers. **(A)** to **(D)** Stage 7 anthers; **(E)** to **(H)** Stage 8 anthers; **(I)** to **(N)** Stage 9 anthers
White arrows in **(F)** indicate residual primexine in developing baculae of the exine. CW, callose wall; Ex, exine; In, intine; Lo, locule; Msp, microspore; Ne, nexine; Nu, nucleus; PM. Plasma membrane; T, tapetum cell; Td, tetrad. Bars = 2 µm in **(A), (C), (E), (G), (I), (K),** and **(M)** and 500 nm in **(B), (D), (F), (H), (J), (L),** and **(N)** (de Azevedo Souza et al., 2009).

2.5 Discussion

In this chapter, I showed that pollen development in the *acos5* loss-of-function mutant is arrested after release from tetrads and free microspores are devoid of exine (Figure 2.3). Since only one *acos5* allele was available, I used a complementation approach to show that the male sterile and pollen development phenotypes observed are indeed due to loss of function of the *ACOS5* gene. A unique feature of the mutant is the complete absence of pollen grains at anther maturity and consequent complete male sterility. While a number of other *Arabidopsis* male sterile or partial sterile mutants with impaired exine and pollen wall development have been described (Aarts et al., 1997; Morant et al., 2007; Persson et al., 2007), such mutants typically retain variable amounts of abnormal pollen grains at anther maturity and corresponding low levels of fertility and seed development. By contrast, the *acos5* homozygote line never produced even partially fertile siliques, necessitating maintenance of the mutant lines as heterozygote. The exceptionally strong male sterile phenotype and lack of exine formation (Figure 2.3) suggests that ACOS5 plays a crucial role in exine formation and sporopollenin biosynthesis.

Previous analysis of the recombinant ACOS5 protein suggested a lack of activity against hydroxycinnamic acids, substrates typically used by 4CLs (Costa et al., 2005; Kienow et al., 2008). In addition, biochemical assay results from the group of Dr. Kombrink demonstrated ACOS5 *in vitro* activity against oleic acid (C18:1) with kinetic constants comparable to those of other fatty acyl-CoA synthetases (de Azevedo Souza et al., 2009). Hydroxycinnamic acids failed to compete with oleic acid as ACOS5 substrates,

verifying that ACOS5 is not a 4CL. Thus, these data suggest that the *in vivo* substrate of ACOS5 is a fatty acid, consistent with a role for ACOS5 in the biosynthesis of an aliphatic sporopollenin monomeric constituent (de Azevedo Souza et al., 2009). Moreover, competition assay data from Dr. Kombrink showed an *in vitro* preference of ACOS5 for medium-chain fatty acids, including hydroxy fatty acids (de Azevedo Souza et al., 2009). Such hydroxy fatty acids appear to be important sporopollenin constituents, since they provide the second functional group required for the formation of the extensive ether crosslinks in the sporopollenin polymer (Ahlers et al., 2003; Morant et al., 2007). Based on competition assays, the best ACOS5 substrates among those tested were 16-hydroxy hexadecanoic acid (16OH-C16), 8-hydroxy octanoic acid (8OH-C8), and 12-hydroxy octadecanoic acid (12OH-C18) (de Azevedo Souza et al., 2009).

The tapetum contribution to exine synthesis and sporopollenin deposition starts while the microspores are still attached in tetrads and continues through the vacuolated stages until the first pollen mitosis is almost completed (Blackmore et al., 2007). My collaborators showed that the spatio-temporal patterns of ACOS5 gene expression, revealed by *in situ* hybridization and promoter-reporter fusion expression (de Azevedo Souza et al., 2009), are consistent with transient, tapetum-localized functions (occurring maximally at around stage 7 of anther development) at the time of tetrad formation and microspore release, when biosynthesis and secretion of sporopollenin precursors is required. Furthermore, the phenotype associated with loss of ACOS5 function in the *acos5* mutant first appears at stage 8 (Figure 2.3), consistent with defects in deposition of a critical secreted sporopollenin component(s), leading to defective microspores. These microspores, when released from tetrads in stage 8 anthers, fail to develop normal exine and are aborted in development by stage 9 (Figure 2.3). Taken together, these data support the hypothesis that the enzyme encoded by ACOS5 is required for production of sporopollenin constituents in the early steps of exine formation.

The combination of our functional data for *ACOS5* (de Azevedo Souza et al., 2009) and similar data for *CYP703A2* and *MS2* (Aarts et al., 1997; Morant et al., 2007) provides further insight into potential pathway(s) for generation of sporopollenin monomeric units. Based on these data, I propose a working model for the biosynthesis of sporopollenin monomers (Figure 2.4). According to this model, the fatty acyl-CoA ester

product of the ACOS5-catalyzed reaction is a central intermediate used to generate sporopollenin monomers in tapetal cells for export to the locule. Consistent with a central role for ACOS5 in one or more biochemical pathways leading to sporopollenin monomer biosynthesis is the strong sterility phenotype of the *acos5* mutant, and the highly correlated coexpression of *ACOS5* with *Arabidopsis* genes encoding enzymes that could act in the same pathway(s), including CYP703A2 and MS2. As shown in Figure 2.4, one function of ACOS5 could be to regenerate the CoA ester of the proposed hydroxylated fatty acid generated by CYP703A2 (7-hydroxylauryl-CoA). Another possibility is that ACOS5 could function as a plastid-localized acyl-CoA synthetase required to generate a fatty acyl-CoA ester for export into the cytoplasm. However, there is no *in silico* evidence for such localization and also transient ACOS5:YFP fusion protein expression data from collaborator suggest that ACOS5 is localized to cytoplasm rather than plastid (de Azevedo Souza et al., 2009).

In biochemical pathways downstream of ACOS5, reduction and/or modification of the ACOS5-generated fatty acyl-CoA could be required to generate sporopollenin monomers (Figure 2.4). For example, MS2 or other tapetum-expressed reductases could generate fatty aldehyde or alcohol monomeric constituents of sporopollenin from the ACOS5-derived CoA ester, which could then be exported into the locule for incorporation into the sporopollenin polymer. Alternatively, the ACOS5-derived fatty acyl CoA ester could also be used as a starter molecule for incorporation into potentially more complex sporopollenin monomeric constituents, analogous to the incorporation of 4CL-derived p-coumaryl-CoA into flavonoids. In this way, the function of ACOS5 in tapetal cells could be analogous to that of 4CL, which generates hydroxycinnamyl-CoA esters used in distinct branch pathways.

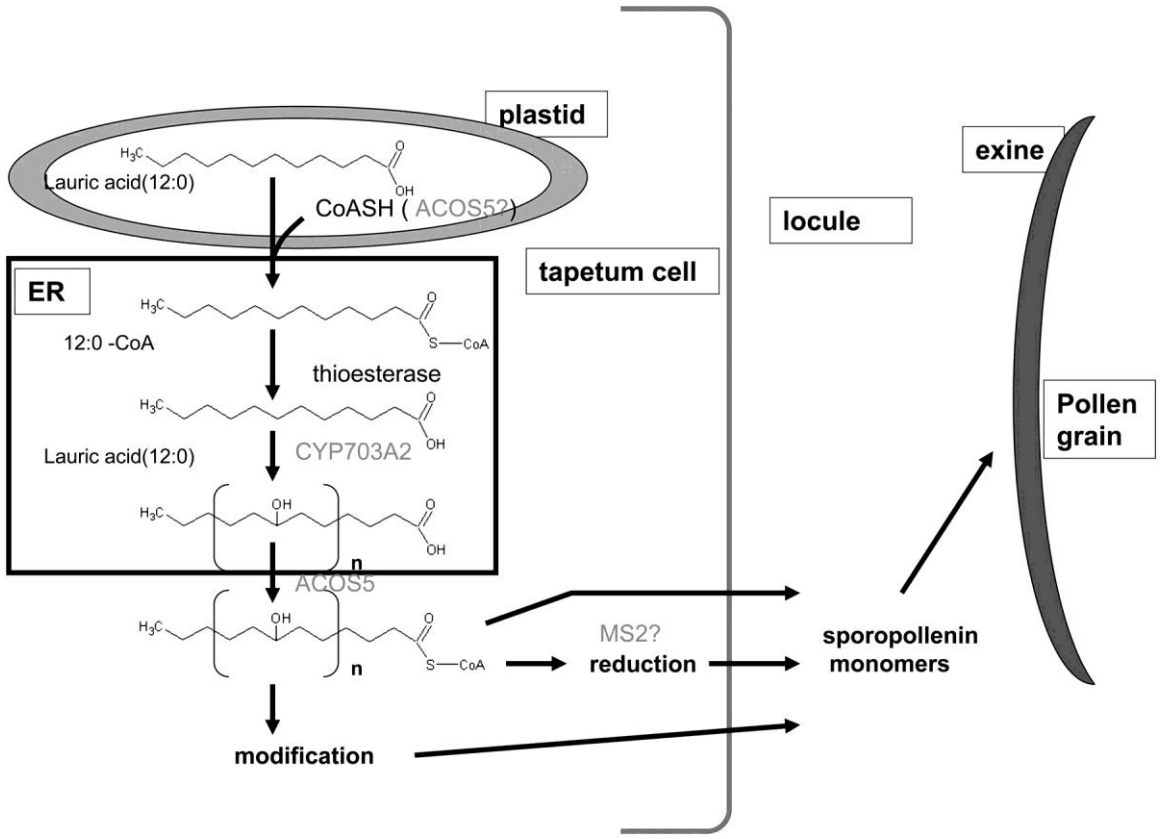


Figure 2.4 Model for the role of ACOS5 in sporopollenin monomer biosynthesis in developing anthers (de Azevedo Souza et al., 2009).

Image was reprinted with permission of the American Society of Plant Biologists (Copyright © American Society of Plant Biologists).

Both *CYP703A2* and *ACOS5* are conserved in land plant lineages, including *Physcomitrella*, but are absent in *Chlamydomonas* (Morant et al., 2007; Souza Cde et al., 2008). Thus, acquisition of an ACOS5- and CYP703A2-dependent sporopollenin biosynthetic pathway appears to be an adaptation that was shared by the common terrestrial ancestor of bryophytes and vascular plants. The ability to generate sporopollenin was likely a key land plant innovation essential for protection of haploid spores from desiccation, UV irradiation, and other stresses of the terrestrial environment, and its evolution likely predated vascular system development and the ability to produce lignin (Bowman et al., 2007). Thus, it is conceivable that the repertoire of 4CL and 4CL-related enzymes now found in land plants (Souza Cde et al., 2008) arose from an ACOS5-

like ancestral enzyme. Further definition of the biochemical pathway involving ACOS5 should not only reveal the nature of sporopollenin monomeric constituents and the sporopollenin polymer but also shed light on the evolution of the diverse polyether and polyester polymers now found in plants.

Chapter 3. *LAP6/POLYKETIDE SYNTHASE A* and *LAP5/POLYKETIDE SYNTHASE B* encode hydroxyalkyl α-pyrone synthases required for pollen development and sporopollenin biosynthesis in *Arabidopsis thaliana*

3.1 Summary

Plant type III polyketide synthases (PKSs) catalyze the condensation of malonyl-CoA units with various CoA ester starter molecules to generate a diverse array of natural products. The fatty acyl-CoA esters synthesized by *Arabidopsis thaliana* ACYL-COA SYNTHETASE5 (ACOS5) are key intermediates in the biosynthesis of sporopollenin, the major constituent of exine in the outer pollen wall. By coexpression analysis, I identified two *Arabidopsis PKS* genes, *POLYKETIDE SYNTHASEA* (*PKSA*) and *PKSB* (also known as *LAP6* and *LAP5*, respectively) that are tightly coexpressed with *ACOS5*. *PKSA* and *PKSB* are specifically and transiently expressed in tapetal cells during microspore development in *Arabidopsis* anthers. Mutants compromised in expression of the *PKS* genes displayed pollen exine layer defects, and a double *pksa pksb* mutant was completely male sterile, with no apparent exine. The biochemical results provided by collaborators indicate that hydroxylated α-pyrone polyketide compounds are the products of sequential action of anther-specific fatty acid hydroxylases, an acyl-CoA synthetase (ACOS5) and two polyketide synthases (PKSA/B).

3.2 Introduction

3.2.1 Polyketides and type III polyketide synthases (PKS)

Polyketide synthases (PKSs) generate a vast array of natural products and are classified as type I, II and III enzymes based on their architectures (Austin and Noel, 2003). Type I PKSs are large multi-domain enzymes consisting of several modules that are comprised of catalytic domains with diverse functions in bacteria and fungi (Austin and Noel, 2003; Watanabe et al., 2007). Type II PKSs, primarily found in bacteria, are separable multi-enzyme complexes that carry out a single set of repeating activities. Each type II PKS contains a minimal set of three subunits containing two β-ketoacyl synthase subunits and an acyl carrier protein subunit to which the growing chain is attached (Austin and Noel, 2003; Watanabe et al., 2007). In addition to being found in bacteria and fungi, type III PKSs are also very common in plants. As homodimers, each subunit of a type III PKS iteratively carries out polyketide synthesis at a single active site (Watanabe et al., 2007). Regardless of their structural differences, all PKSs use a common chemical strategy; they catalyze chain elongation by a decarboxylative condensation reaction, followed by cyclization to generate the final polyketide products (Austin and Noel, 2003).

The type III PKS enzyme superfamily generates a remarkable diversity of polyketide products, due to their ability to accept a broad range of starter-CoA units (Figure 3.1). The best-studied plant type III PKS, chalcone synthase (CHS), as well as CHS-related enzymes such as stilbene synthases, can utilize a wide range of phenylpropanoid-CoAs that are condensed with malonyl-CoAs to generate phloroglucinol-type products with variable functional groups (Austin and Noel, 2003). Moreover a fungal type III polyketide synthase, 2'-oxoalkylresorcylic acid synthase (ORAS) in *Neurospora crassa* as well as a bacterial type III polyketide synthase, alkylresorcinol synthases (ARS) in *Azotobacter vinelandii* are able to generate alkylresorcinols, alkylresorcylic acids and alkylpyrones using a broad range of medium- and long-chain fatty acyl-CoAs as starter units (Funa et al., 2006; Funa et al., 2007). These long-chain alkylresorcinols have been found to be essential for mature cyst formation in *A. vinelandii* by generating an exine structure (Funa et al., 2006).

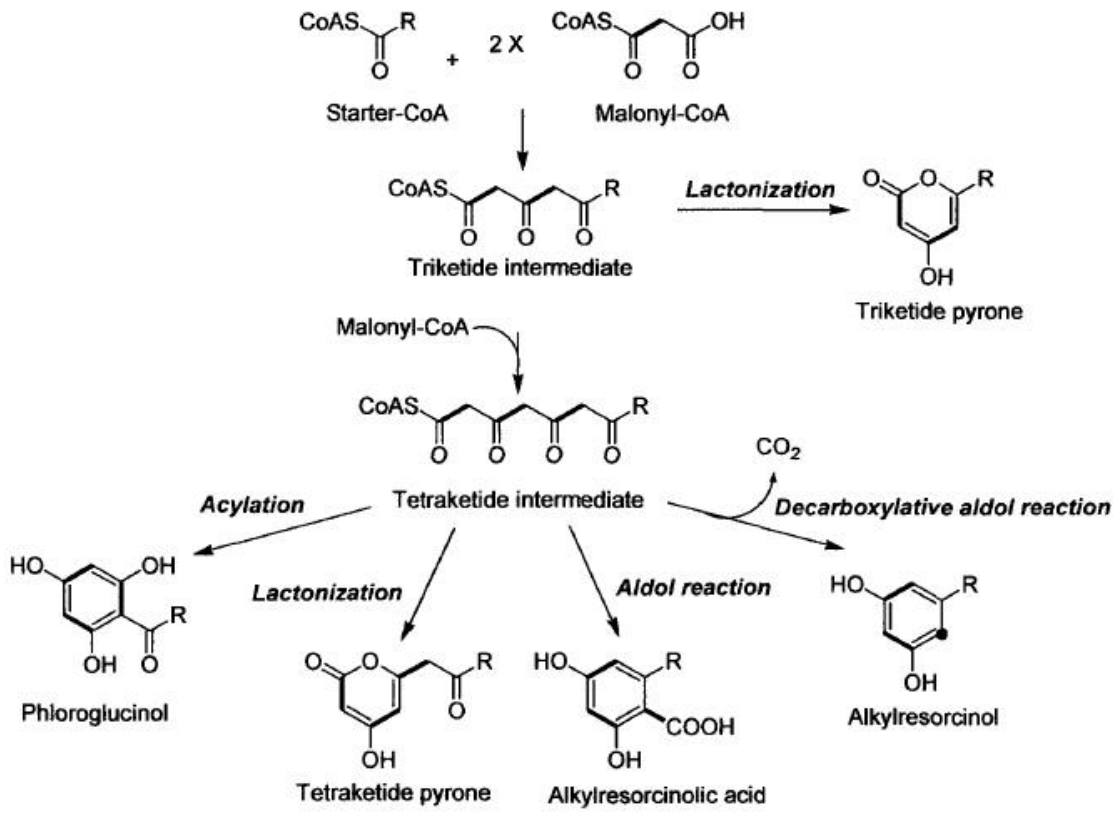


Figure 3.1 Diverse cyclization strategies employed by type III PKSs.

Image was reprinted with permission of Che Caswell Colpitts (Copyright © 2009 C.C. Colpitts).

Type III PKSs are ubiquitous in vascular plants as well as non-vascular plants such as *Physcomitrella patens* (Austin and Noel, 2003; Koduri et al., 2010), and plant-specific type III PKSs synthesize diverse natural products that play import roles in UV photoprotection, anthocyanin flower pigmentation, antimicrobial defense, and pollen fertility. The completed *Arabidopsis thaliana* genome project identified three type III *CHS* gene homologs, *At1g02050, At4g34850* and *At4g00040* as well as the originally described *bona fide CHS* gene, *At5g13930* (Tsai et al., 2006; Wang et al., 2007). Phylogenetic analysis suggested that the two *CHS* homologs At1g02050 and At4g34850 could have undergone functional divergence (Wang et al., 2007), however, until recently little was known about the functions of these CHS-like genes. The first biochemical study of the proteins encoded by At1g02050 (POLYKETIDE SYNTHASE A, PKSA) and At4g34850 (POLYKETIDE SYNTHASE B, PKSB) was performed by Mizuuchi et al.,

(2008) and revealed they can catalyze the unusual formation of alkyl α-pyrones by utliziting fatty acyl-CoAs of up to C20 chain lengths as substrates *in vitro*. Also, a recent genetic study demonstrated mutations in *At1g02050* (*LAP6*) and *At4g34850* (*LAP5*) led to defective exine formation and they are specifically expressed in anthers of stage 9 and 10 buds (Dobritsa et al., 2010).

In this study, I now significantly extend a previous study (Dobritsa et al., 2010) to show that the plant-specific *CHS*-like type III PKSs *LAP6/PKSA* and *LAP5/PKSB* are coexpressed with *ACOS5*, are specifically and transiently expressed in tapetum cells, are required for exine formation and development, and function at specific stages of microspore development. Genetic analyses also showed that *LAP6/PKSA* and *LAP5/PKSB* have overlapping but partially distinct functions in exine development. Furthermore, my collaborators show that *LAP6/PKSA* and *LAP5/PKSB* encode enzymes that catalyze the condensation of hydroxy fatty acyl-CoA esters synthesized by ACOS5 with malonyl-CoAs to yield triketide and tetraketide α-pyrones and that PKSA has a strong *in vitro* preference for medium-chain hydroxy fatty acyl-CoAs that, based on *in vitro* data, may be preferentially synthesized by ACOS5. Combined with the finding that PKSA and PKSB preferentially localize to the endoplasmic reticulum (ER), where they may form part of a metabolon with upstream and downstream enzymes, these data illuminate the important role played by these enzymes in an ancient and evolutionarily conserved biochemical pathway or pathways required for the biosynthesis of polyketide sporopollenin precursors.

3.3 Materials and methods

3.3.1 Plant material and growth conditions

Arabidopsis thaliana Columbia (Col-0) seeds were sterilized and after a cold treatment (2 days at 4°C in the dark) and germinated at 20°C under 70 µmol m^{-2} s^{-1} fluorescent lighting. Twelve days later, the plants were transferred to a growth chamber with a light/dark cycle of 16 hr/8 hr. T-DNA insertion mutants were obtained from SALK (Alonso et al., 2003), and Gabi-Kat (Rosso et al., 2003) collections via The *Arabidopsis* Information Resource and the Nottingham *Arabidopsis* Stock Centre. In progeny,

homozygous insertion lines SALK_134643 for *pksa-1*and GK_089C04 for *pksb-3* were identified by PCR using gene-specific and T-DNA specific primers (Table 3.1).

3.3.2 Phylogenetic and bioinformatic analyses

All aligned sequences obtained are given in Appendix A. Protein sequences were aligned using MUSCLE 3.6 using the default parameters (sequencing clustering; UPGMA, objective score; classic sum-of-pairs score) (Edgar, 2004), and the multiple protein sequence alignments were manually optimized. Aligned sequences are available in Appendix B. To reconstruct phylogenetic trees, maximum likelihood analyses with 1000 bootstrap replicates were performed using PhyML v2.4.4 and default settings (Guindon and Gascuel, 2003) with the JTTmodel of amino acid substitution.

3.3.3 RT-PCR

RNA quality was assessed by visual inspection of rRNA on a 1.2% formaldehyde-agarose (FA) gel and quantified spectrophotometrically, and 2.5 µg RNA/ 20 µL reaction was used to generate first strand cDNA using Superscript II Reverse Transcriptase (Invitrogen) following the manufacturer's protocol. For RT-PCR, gene-specific and intron-spanning primers (Table 3.1) were used in PCR reactions to amplify corresponding cDNA sequences under the following PCR conditions: 95℃ for 3 min, followed by 35 cycles of (94℃ for 30 s, 56℃ for 30 s, 72℃ 1 min) followed by 72℃ for 10 min, using Taq polymerase in a 50 µL total reaction. *Actin2* was used as control.

For quantitative RT-PCR analysis of *PKSA* and *PKSB* expression, 10 ng of cDNA was incubated with 10 µL iQ SYBR Green Supermix (Bio-Rad) and 5 pmol of each forward and reverse primer (Table 3.1) in a total volume of 20 µL. After an initial denaturation step at 95℃ for 3 min, 40 cycles at 95℃ for 15 s, 60℃ for 15 s, and 72℃ for 30 s were followed by a fluorescence reading. A melting curve was generated ranging from 95℃ to 60℃. Threshold cycles (CT) were adjusted manually, and the CT values for a housekeeping control *Actin2* amplified in parallel on each plate were subtracted from CT values obtained for each gene of interest, thus generating normalized CT values (ΔCT). The relative starting quantities of each gene were determined by setting as a base

value the gene with the highest CT value within a tissue panel or treatment series, and relative quantities were calculated using the ΔΔCT method as described in (Hietala et al., 2003). ΔΔCT was calculated using immature flower buds as the highest expressing tissue.

3.3.4 Phenotypic analyses

Tissue fixation, embedding, and sectioning of *Arabidopsis* wild type (Col-0) and *pksa-1 pksb-3* double mutant inflorescences were performed as described in section 2.3.4.

3.3.5 *In situ* hybridization and validation of RNA probes

Arabidopsis Col-0 inflorescences of different developmental stages were fixed in 20 mL scintillation vials in FAA (3.7% formaldehyde, 5% acetic acid, 50% ethanol and 41.3% distilled water) for 3 hr to 4.5 hr. The samples were then dehydrated by immersion in the following ethanol series for 30 min each: 50% twice, 60%, 70%, 85%, 95%, and 100% twice. After ethanol dehydration, 75:25, 50:50, 25:75 ethanol/xylene, and two times of 100% xylene were applied to the samples. The vials containing samples were filled half way with xylene and topped up with paraffin (Paraplast Plus, Sigma) for incubation at 55℃ overnight. Molten paraffin was used to replace the xylene/paraffin solution, and after that, paraffin was replaced at least six times at least 6 hr intervals. Embedded samples were sectioned with a microtome to 8 μm. Sections were floated onto precharged slides using distilled water, dried at 42℃ overnight, and affixed to the slides by raising the temperature of the hot plate to 56 ℃ for 4 hr.

For sense and antisense *PKSA* and *PKSB* probe synthesis, 1209 bp and 1200 bp DNA template corresponding to the *PKSA* and *PKSB* coding region, respectively, were PCR amplified from flower cDNA using gene-specific forward and reverse primers (Table 3.1). A T7 polymerase binding site was incorporated into the forward primer for sense probe amplification and in the reverse primer for antisense probe amplification. In vitro transcription was carried out at 37℃ for 2 hr, and then the RNA was precipitated by adding 2.5 μL of 4 M LiCl and 75 μL of 100% ethanol, and kept at -80℃ for 2 hr. RNA was spun down at 4℃ at maximum speed of microcentrifuge and resuspended in 100 μL

of DEPC-treated water. The RNA probe was then hydrolyzed into fragments between 100 and 150 base pairs long by adding 60 μL 200 mM Na_2CO_3 and 40 μL 200 mM $NaHCO_3$ followed by incubation at 60℃ for 30 min. The mixture was neutralized by the addition of 10 μL of 20% acetic acid. The probe was precipitated using 21 μL of 3 M NaOAC, 2 volumes of 100% ethanol and 1 μL of 20 mg/ml oyster glycogen as carrier at -20℃ for 2-3 hr. 100 μL of 50% deionized formamide was used to dissolve the pelleted probe and then the probe was quantified against digoxygenin (DIG) standard according to manufacturer's instructions.

For hybridization of probes into sections on slides, first paraffin was removed by immersing slides in xylene twice for 10 min, and 100% ethanol twice for 2 min each. Sections were hydrated by immersion in 95%, 90%, 80%, 60%, and 30% ethanol, 0.85% NaCl, and then 1X PBS (0.13 M NaCl, 3 mM NaH_2PO_4, 7 mM Na_2HPO_4) for 2 min each. Slides were incubated for exactly 30 min at 37℃ with 1 μg/mL proteinase K in 100 mM Tris-HCl, pH 7.5, and 50 mM EDTA, and washed with 1x PBS again at the end. Slides were then dehydrated in 0.85% NaCl, 30%, 60%, 80%, 90%, 95%, and 100% ethanol for 1 min each and stored at 4℃ in a closed box with a few drops of ethanol soaking the paper until further processing. Hybridization was done overnight at 55℃ with a DIG-labeled RNA probe (10-50 ng) in 200 μL of hybridization buffer (10 mM Tris-HCl, pH7.5, 1 mM NaCl, 50% deionized formamide, 7% dextransulfate, 1x Denhardt's solution (Sigma), 50 mg/mL yeast tRNA (Roche)). Slides were washed in 2x SSC (1x SSC is 0.15 M NaCl, 0.015 M sodium citrate) for 5 min, four times in 0.2X SSC at 55℃ for 30 min each, once at 37℃ and then in 1x PBS for 5 min at RT.

Immunological detection of the hybridization probe was performed as follows: slides were covered for 45 min with 1 mL of 1% blocking reagent (Boehringer Mannheim) in 100 mM maleic acid, pH 7.5, and 150 mM NaCl, then incubated for 45 min in 1 mL of BSA solution (1% BSA, 0.3% Triton X-100, 100 mM Tris-HCl, pH 7.5, and 150 mM NaCl) after washing with 2 mL of fresh BSA solution. The slides were then incubated for 1.5 hr with 1 mL of diluted (1:1250) antibody conjugated in BSA solution, followed by three washes in BSA solution for 20 min each. For color reaction, the slides were washed in TNM-50 (100 mM Tris-HCl pH 9.5, 100 mM NaCl, 50 mM $MgCl_2$) twice for 15 min each. To activate the color reaction, slides were incubated overnight with 20 uL of the

BCIP/NBT (Roche) per 1ml of TMN-50 in substrate buffer in the dark. 10 mM Tris-HCl, pH 8.0, and 5mM EDTA were added to stop the color reaction.

To validate specificity of RNA antisense probes, I performed DNA gel blotting. PCR-amplified cDNA clones of *PKSA* and *PKSB* (Table 3.1) were blotted to positively charged nylon membranes, Hybond-XL (GE Healthcare Life Sciences), using 10x SSC as the transfer buffer. Transfer of cDNA to the nylon membrane was monitored via ethidium bromide. The digoxigenin-labeled probes were prehybridized and hybridized using 1x Denhardt's solution, 50% deionized formamide, 10% dextran sulfate, 0.2 mg/mL of salmon sperm DNA. The blots were washed twice using 2x SSC and 0.1% SDS at 55℃ for 10 min each, three times using 2x SSC at 55℃ for 30 min each, and twice using 0.2x SSC at 55℃ for 10 min each. The hybridized probes were detected by the addition of BCIP/NBT color reagent (Roche).

3.3.6 Cloning for PKSA and PKSB protein expression

Arabidopsis total RNA was extracted from flower buds with the RNeasy Plant Mini kit (Qiagen) following the manufacturer's protocol. First strand cDNA was generated from RNA (2.5 μg) using SuperScript II reverse transcriptase (Invitrogen), and PCR-amplification of *PKSA* and *PKSB* was carried out with the primers containing NcoI and EcoRI sites (Table 3.1). Because PKSA coding sequence contains EcoRI restriction enzyme site, both the 3rd (AAT) and 4th (TCT) codon sequences from start codon (ATG) were replaced with AAC encoding Asn and TCG encoding Ser in PKSA forward primer, FCHSL1NcoI (Table 3.1), respectively, based on codon usage table for *E. coli W3110* (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=316407). The PCR products were digested with restriction enzymes and subcloned into pET-32a(+) or pET-28a(+) expression vector (Novagen) for further heterologous expression in *E. coli*.

3.3.7 Accession number

Sequence data from this article can be found in the Arabidopsis Genome Initiative, GenBank/EMBL, or other databases under the following accession numbers: *Arabidopsis PKSA*, At1g02050; *Arabidopsis PKSB*, At4g34850; *Arabidopsis Actin2*,

At3g18780; *Arabidopsis CHSL2*, At4g00040; *Oryza sativa CHSL1*,LOC_Os10g34360 (http://rice.plantbiology.msu.edu/index.shtml); *O. sativa CHSL2*, LOC_Os07g22850 (http://rice.plantbiology.msu.edu/index.shtml); *Physcomitrella patens CHS10*, protein ID 149790 (http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html); *Pinus radiata CHS1*, AAB80804; *Nicotiana sylvestris CHSLK*, CAA74847; *Silene latifolia Chs*, AB182106; *Populus trichocarpa CHSL4*, protein ID 551991 (http://genome.jgi-psf.org/poplar/poplar.home.html); *P. trichocarpa CHSL5*, protein ID 591704 (http://genome.jgi-psf.org/poplar/poplar.home.html); *P. trichocarpa CHSL6*, protein ID 556583 (http://genome.jgi-psf.org/poplar/poplar.home.html); and *P. trichocarpa CHSL*, protein ID 200918 (http://genome.jgi-psf.org/poplar/poplar.home.html).

3.3.8 Sequences of primers

Table 3.1 Primers used in this study.

Genotyping

Gene	allele	Primer Name	Sequence (5' → 3')
PKSA(At1g02050)	pksa-1	PKSa1F	GCA TCC ACC ATC TTT CTT CC
		PKSa1R	GGG GTT GTT CTC AGC AAT GT
PKSB(At4g34850)	pksb-1	PKSb1F	GAG GAA TTC AAT GGG AAG CAT CGA TGC TGC
		PKSb1R	AAC CCG TTA TGA GAA GAT CCA A
	pksb-3	CHSL3LP	TGT AAC ACC AGG TCC AAA AGC
		CHSL3RP	TGA AGG AGG ATC CAC AGT GAC
T-DNA specific		GK specific	ATA TTG ACC ATC ATA CTC ATT GC
		LBb1.3	ATTTTGCCGATTTCGGAAC

RT-PCR

Gene	allele	Primer Name	Sequence (5' → 3')
PKSA(At1g02050)	pksa-1	RCHSL1 RT	TTA GGA AGA GGT GAG GCT GCG G
		FCHSL1 RT	ATG TCG AAT TCT AGG ATG AAT GGT GTT G
PKSB(At4g34850)	pksb-1	FPKSB RT	GAG GAA TTC AAT GGG AAG CAT CGA TGC TGC
		RPKSB RT	CTC AAG CTT TCA GAC ATC AAG GTT TCG AG
	pksb-3	FCHSL3 RT	ACT CGT CTC TGC AAG ACA
		RCHSL3 RT	TGT AAC ACC AGG TCC AAA AGC
Actin2		Actin2-RT-FW	CCAGAAGGATGCATATGTTGGTGA
		Actin2-RT-RW	GAGGAGCCTCGGTAAGAAGA

Table 3.1 Primers used in this study. (cont.)

Quantitative RT-PCR

Gene	Primer name	Sequence (5' → 3')
PKSA(At1g02050)	FCHSL1-qRT	TAA GCA GCA AAT CCA CAA GGC GTG
	RCHSL1-qRT	CGT TTT GCA CAA GTG TTC
PKSB(At4g34850)	FCHSL3-qRT	TGT TCT GGG CGG TTC AT
	RCHSL3-qRT	CCT CAC TTT CTT GCT CTC CT
Actin2	Actin2-RT-FW	CCAGAAGGATGCATATGTTGGTGA
	Actin2-RT-RW	GAGGAGCCTCGGTAAGAAGA
β-tubulin	tubulin-RT-F	CGT GGA TCA CAG CAA TAC AGA GCC
	tubulin-RT-R	CCT CCT GCA CTT CCA CTT CGT CTT C

In situ hybridization

Gene	Primer name	Sequence (5' → 3')
PKSA(At1g02050)	CHSL1F-Sense	CAT AAT ACG ACT CAC TAT AGG ATG TCG AAT TCT AGG ATG AAT G
	CHSL1R-Sense	GGA AGA GGT GAG GCT GCG
	CHSL1R-Anti	CAT AAT ACG ACT CAC TAT AGG TTA GGA AGA GGT GAG GCT GCG
	CHSL1F-Anti	A TGT CGA ACT CGA GGA TGA ATG
PKSB(At4g34850)	CHSL3F-Sense	CAT AAT ACG ACT CAC TAT AGG ATG GGA AGC ATC GAT GCT G
	CHSL3R-Sense	GAC ATC AAG GTT TCG AGC GAT
	CHSL3R-Anti	CAT AAT ACG ACT CAC TAT AGG TCA GAC ATC AAG GTT TCG AGC GAT
	CHSL3F-Anti	A TGG GAA GCA TCG ATG CTG

Cloning

Gene	Primer name	Sequence (5' → 3')
PKSA(At1g02050)	F CHSL1 NcoI	CCA TGG CTA TGT CGA ACT CGA GGA TGA ATG
	R CHSL1 EcoRI	GAA TTC TTA GGA AGA GGT GAG GCT GCG
PKSB(At4g34850)	F CHSL3 NcoI	CCA TGG CTA TGG GAA GCA TCG ATG CTG
	R CHSL3 EcoRI	GAA TTC TCA GAC ATC AAG GTT TCG AGC GAT

DNA blotting

Gene	Primer name	Sequence (5' → 3')
PKSA(At1g02050)	FPKSABlot	ATGTCGAATTCTAGGATGA
	RPKSABlot	AGGAAGAGGTGAGGCT
PKSB(At4g34850)	FPKSBBlot	ATGGGAAGCATCGATGCTG
	RPKSBBlot	GACATCAAGGTTTCGAGCGATAA

3.4 Results

3.4.1 Analysis of *ACOS5* co-expression genes

Previously we reported that *ACOS5* encodes a fatty acyl-CoA synthetase required for sporopollenin biosynthesis in *Arabidopsis* (de Azevedo Souza et al., 2009). To define other potential enzymes in the sporopollenin biosynthetic pathway, I used data mining tools to identify coexpressed genes in public global gene expression data sets. Using the Correlated Gene Search tool (http://prime.psc.riken.jp), I queried 237 microarray experiments in the Tissue and Development data set, using a cutoff Pearson coexpression coefficient (r^2) of 0.80. This analysis identified 35 coexpressed genes, most of unknown function. Among these coexpressed genes, several have been shown to encode enzymes involved in sporopollenin biosynthesis, such as MS2 (At3g11980), CYP703A2 (At1g01280), and DRL1 (At4g35420) (Aarts et al., 1997; Morant et al., 2007; Tang et al., 2009). Among the coexpressed genes, I focused on those that encode enzymes that could utilize the potential fatty acyl-CoA product(s) generated by the ACOS5 as substrate(s), and therefore might be important in sporopollenin biosynthesis. Two of the most promising potential candidates were genes annotated as encoding plant-specific type III PKSs, *LAP6/PKSA* (At1g02050) and *LAP5/PKSB* (At4g34850), which were previously reported to generate triketide and tetraketide α-pyrone compounds using fatty acyl-CoAs (up to 20 carbon chain length) as starter substrates (Mizuuchi et al., 2008) and were strongly coexpressed with *ACOS5* (r^2 for *LAP6/PKSA* of 0.94; r^2 for *LAP5/PKSB* of 0.99). To facilitate their description and represent their enzymatic function for sporopollenin biosynthesis in the following paragraphs, genes and proteins corresponding to *At1g02050* and *At4g34850* are referred to as *PKSA* and *PKSB*, respectively, consistent with the established nomenclature of Mizuuchi et al., (2008).

3.4.2 *PKSA* and *PKSB-like* genes are conserved in land plant lineage

To investigate a potential conserved function for *PKSA* and *PKSB* genes in pollen wall development, I performed *in silico* searches of the full genome sequences of *Arabidopsis*, poplar (*Populus spp*), rice (*Oryza sativa*) and *Physcomitrella patens* using

PKSA and *PKSB* as queries to retrieve potential *PKSA*- and *PKSB*-related PKS genes in these plants (Table 3.2). I also retrieved the PKS-related *Nicotiana sylvestris CHSLK*, *Silene latifolia Chs*, and *Pinus radiata CHS1*genes, previously shown to have high expression in male flowers or anthers (Atanassov et al., 1998; Walden et al., 1999; Ageez et al., 2005) and generated an un-rooted maximum likelihood tree of aligned CHS and PKS-related protein sequences. This analysis, shown in Figure 3-2, indicated that *PKSA* and *PKSB* are located in two distinct *PKS* subclades that are clearly distinct from the more distantly related clade containing the *bona fide Arabidopsis* and *Physcomitrella CHS* genes and other putative *CHS* genes from poplar and rice. Each subclade including either *PKSA* or *PKSB* contains homologs from the fully sequenced poplar and rice genomes, as well as the rice or *Silene* homologs known to be expressed in male organs. According to this analysis, the *Pinus CHS1* gene is a *PKSA/B* homolog basal to the angiosperm *PKSA* and *PKSB* clades, and the *Physcomitrella PKS* (*CHS10*) and *CHS* genes are basal to the tracheophyte *PKSA/B* and *CHS* clades, respectively (Figure 3.2). These data indicate that *PKSA/B* clade PKS genes arose early in land plant evolution and may have common roles in male organ or spore development in land plant lineages.

Table 3.2 Putative *PKSA* and *PKSB* orthologs and expression in other species.

Species	Gene name	Accession or gene model	Expression
Arabidopsis thaliana	*PKSA*	At1g02050	Tapetum[1]
	PKSB	At4g34850	Tapetum[1]
	CHSL2[2]	At4g00040	Flower and leaf[3]
Oryza sativa	*CHSL1*	Os10g34360	Immature panicle[4]
	CHSL2	Os07g22850	Immature panicle[4]
Physcomitrella patens	*CHS10*[5]	e_gw1.304.37.1, Protein ID149790	Sporophyte[6]

Table 3.2 Putative *PKSA* and *PKSB* orthologs and expression in other species. (cont.)
Table was reprinted with permission of American Society of Plant Biologists (Copyright © American Society of Plant Biologists).

Species	Gene name	Accession or gene model	Expression
Pinus radiata	*CHS1*	AAB80804	Male cone[7]
Nicotiana sylvestris	*NSCHSLK*	CAA74847	Anther[8]
Silene latifolia	*SlChs*	AB182106	Male flower[9]
Populus trichocarpa	*CHSL4*[2]	LG_II:10548880-10550149	N/A
Populus trichocarpa	*CHSL5*[2]	scaffold_40:395399-396653	N/A
Populus trichocarpa	*CHSL6*[2]	LG_IV:15418736-15420441	N/A
Populus trichocarpa	*CHSL7*[2]	LG_IX:2399625-2400948	N/A

[1] This study
[2] Tsai et al., New Phytologist 2006
[3] http://bar.utoronto.ca/
[4] http://mpss.udel.edu/rice/
[5] Jiang et al., Phytochemistry 2006
[6] http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?LID=23755&PAGE=1
[7] Walden AR et al., Plant physiol 1999
[8] Atanassov I et al., Plant Mol Biol 1998
[9] Ageez A et al., Genes Genet Syst 2005
N/A, no information available

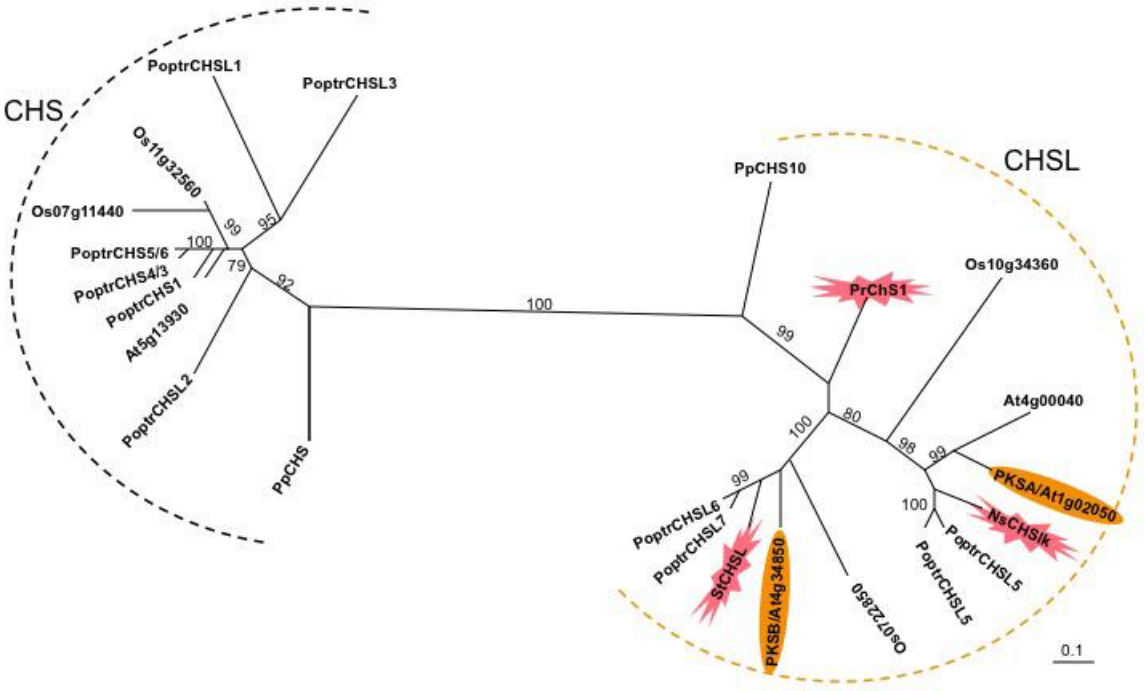


Figure 3.2 Phylogenetic analysis of CHS and CHS-like (CHSL) protein sequences from land plants.

Sequences retrieved from the complete genome sequences of *Arabidopsis* (At), poplar (Poptr), rice (Os), and *Physcomitrella* (Pp), as well as selected sequences from *Silene* (St), pine (Pt) and *Nicotiana sylvestris* (Ns) were included in the maximum-likelihood (ML) tree built using 1000 bootstrap replicates in PhyML 2.4.4. Bootstrap values are indicated on branches (out of 100). The *Arabidopsis* PKSA and PKSB proteins are highlighted (ovals). The proteins encoded by genes known to be expressed in tapetum cells during anther development are indicated by a flash. Protein sequences used in this analysis are given in Appendix A., and the alignment is available in Appendix B. Bar = 0.1 amino acid substitutions (Kim et al., 2010).

3.4.3 Tapetum-specific expression of PKSA and PKSB proteins

To test possible functions of the *Arabidopsis* PKSA and PKSB proteins in male organ development, first I used quantitative reverse transcription PCR to analyze the expression of *PKSA* and *PKSB* in different *Arabidopsis* organs. Expression profiles are shown in Figure 3.3. Both genes were preferentially expressed in flowers, and *PKSB* transcripts were detected exclusively in this organ.

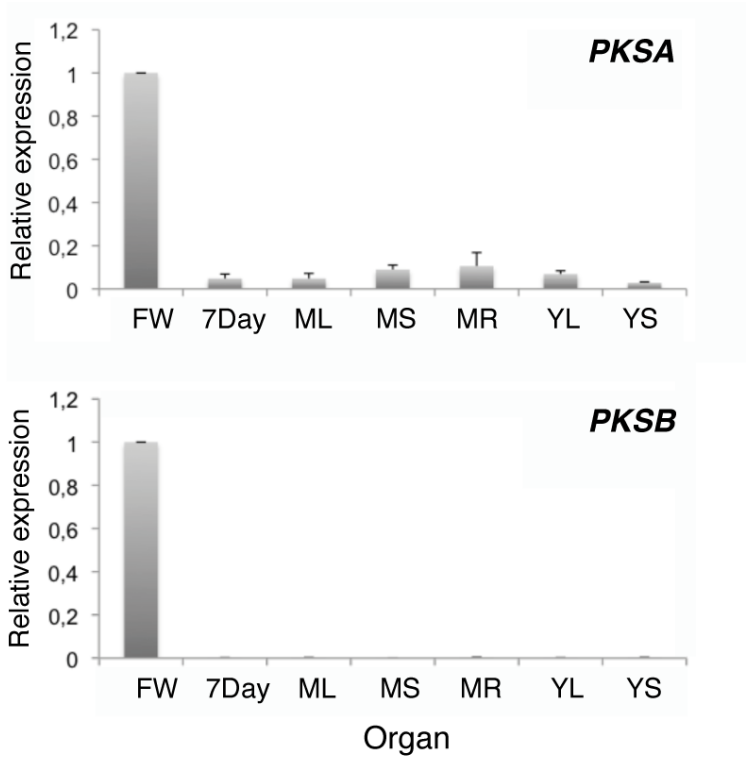


Figure 3.3 Developmental expression profiles of *PKSA* and *PKSB*.

Quantitative reverse transcription-PCR analysis of relative *PKSA* and *PKSB* expression levels in various *Arabidopsis* organs. Expression was calculated using the ΔΔCT method and is represented relative to the organ with the highest level of expression (flowers), set at 1.0. *Actin2* was used as a reference gene. Bars represent standard deviations from the means of triplicate determinations. 7Day, 7-day old seedlings; FW, flower; ML, mature leaf; MR, mature root; MS, mature stem; YL, young leaf; YS, young stem (Kim et al., 2010).

To investigate the spatio-temporal patterns of *PKSA* and *PKSB* expression in anthers, *in situ* hybridization experiment was performed. Because *PKSA* and *PKSB* share 62.9% nucleotide identity and could potentially cross-hybridize, the specificity of the *PKSA*- and *PKSB*-derived riboprobes used for *in situ* hybridization analysis was demonstrated by DNA gel blotting, showing that both probes hybridized specifically with the target templates, with no detectable cross-hybridization (Figure 3.4). To elucidate the specific expression patterns of *PKSA* and *PKSB*, these RNA probes were hybridized to sections of developing wild-type flowers (Figure 3.5) focusing on anther stages 6-11 as

defined by Sanders et al., (1999), during which free microspores are generated and sporopollenin-containing exine is deposited. Both *PKSA* and *PKSB* were strongly and transiently expressed in the tapetum cell layer of developing anthers. Whereas the hybridization signal was largely specific to tapetal cells, low levels of signal appeared to be present in stage 7 tetrads, so that a function in microspores cannot be excluded. Interestingly the temporal patterns of expression in the developing tapetum were slightly different. *PKSA* expression was first detected at stage 6, and by stage 7, strong hybridization was detected in the tapetum. By stage 8, *PKSA* hybridization to tapetum cells had weakened, and at stage 11 it had disappeared. By contrast, *PKSB* expression was initiated at stage 7 and maximal hybridization was seen in the tapetum at stage 8.

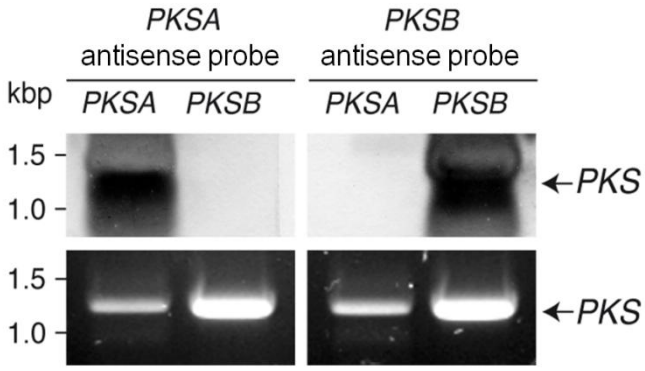


Figure 3.4 DNA gel blot showing specificity of PKSA or PKSB RNA probes used for *in situ* hybridization.

PKSA and *PKSB* digoxigenin labeled antisense probe were hybridized against a blot of PCR amplified full-length PKSA and PKSB coding sequences (above). The Gel Red stained gel showing prior to DNA gel blot transfer is shown below (Kim et al., 2010).

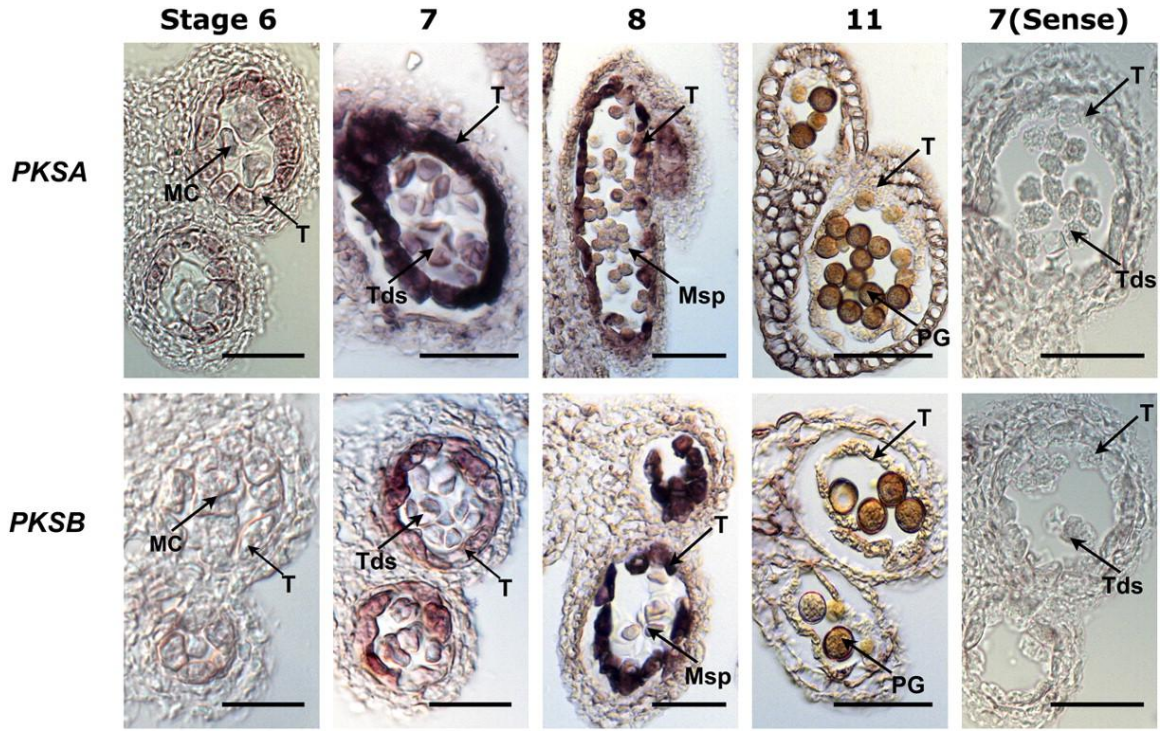


Figure 3.5 Transient tapetum-specific expression of *PKSA* and *PKSB*.

PKSA and *PKSB* mRNAs were localized by *in situ* hybridization to sections taken from developing anthers of wild-type (Col-0) flowers using gene-specific antisense probes for *PKSA* and *PKSB* and control sense probes. Stages of anther development are according to Sanders et al. (1999). Dark precipitates indicate hybridization of the probe. MC, meiotic cell; T, tapetum; Tds, tetrads; Msp, microspores; PG, pollen grain. Scale bars = 70 μm (Kim et al., 2010).

3.4.4 Identification and phenotypic analysis of *PKSA* and *PKSB* loss-of-function alleles

To test the roles of PKSA and PKSB in pollen development and male fertility, both T-DNA insertion lines *pksa-1* for *PKSA* and *pksb-3* for *PKSB* were obtained from public collections (Alonso et al., 2003). The locations of each T-DNA insertion in the *PKSA* and *PKSB* genes were verified by sequencing analysis (Figure 3.6A). *PKSA* and *PKSB* expression in the insertion lines was assayed by RT-PCR, using template cDNAs derived from both wild type and mutant flowers (Figure 3.6B), and no *PKSA* or *PKSB* expression was detected. This analysis suggested that each of the two alleles is a loss-of-

function allele of *PKSA* or *PKSB*. Homozygous lines for either *pksa-1* or *pksb-3* are fertile and there were no obvious morphological differences between each homozygous and wild-type.

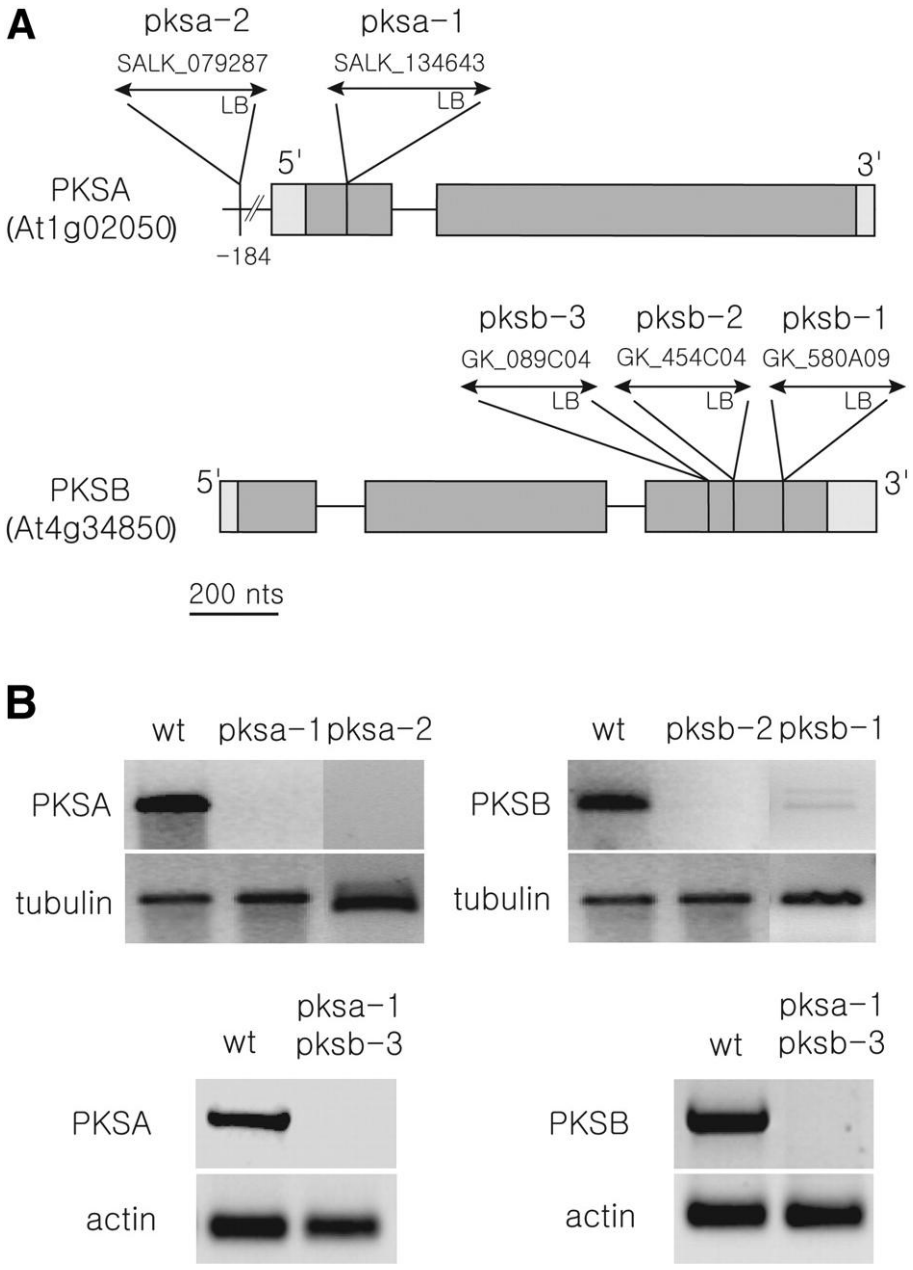


Figure 3.6 Molecular characterization of *pksa*, *pksb* and *pksa pksb* insertion alleles and effects on gene expression.
(A) Position of T-DNA insertions in the different mutant lines is shown. The cartoons for

gene model were drawn base on TAIR (The Arabidopsis Information Researce) database search (http://www.arabidopsis.org). Grey boxes denote exons, with lighter gray indicating 5' and 3' untranslated regions. Thin horizontal lines denote introns. The location of T-DNA and absence of mRNA in *pksa-1* and *pksb-3* were verified by me and those in *pksa-2*, *pksb-1* and *pksb-2* were verified by our collaborator, Dr. Michel Legrand's lab.
(B) RT-PCR analysis of gene expression in flower buds of wild-type (wt, Col-0) and single and double mutants. No amplicon was detectable in mutant samples. *TUBULIN* and *ACTIN* are shown as positive controls (Kim et al., 2010).

3.4.5 PKSA and PKSB have partially redundant function in male fertility

Since the *Arabidopsis PKSA* and *PKSB* genes encode PKSs with similar *in vitro* activities (Mizuuchi et al., 2008; Dobritsa et al., 2010), biochemical redundancy between PKSA and PKSB was a strong possibility. To test this, I generated a homozygous double mutant, *pksa-1 pksb-3*, identified within F2 populations derived from crossing the corresponding homozygous *pksa* and *pksb* lines. Initial phenotypic examination of *pksa-1 pksb-3* mutant plants (Figure 3.7) revealed anthers devoid of visible pollen, male sterility, and siliques devoid of seeds. Whereas no visible pollen was ever observed in plants homozygous for the *pksa-1 pksb-3* alleles, *pksa-1 pksb-3* flowers were female-fertile when pollinated with wild-type pollen. There were no other obvious morphological differences between the *pksa-1 pksb-3* mutant and wild-type plants grown to maturity (Figure 3.7).

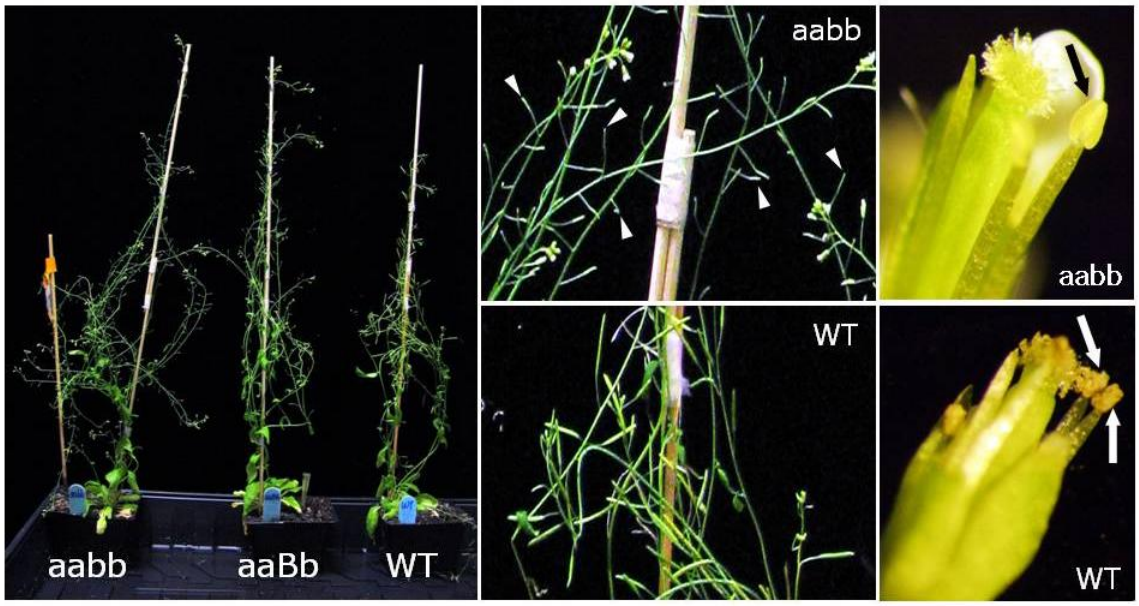


Figure 3.7 Phenotypic characterization of plants segregating for *pksa* and *pksb* alleles.

Plants homozygous for *pksa*-1 and *pksb*-3 alleles were crossed, and an F1 individual self-pollinated to generate a population segregating for *pksa-1* and *pksb-3* alleles. Plants were subjected to PCR-aided genotyping and grown to maturity. *PKSA* alleles are symbolized by A (WT) and a (*pksa-1*), and *PKSB* alleles are symbolized by B (WT) and b (*pksb-3*). Phenotypes of mature aabb, aaBb, and AABB (WT) plants are shown. Double mutant plants (aabb) had no pollen in anthers (black arrow) and undeveloped siliques (arrowheads), resulting in a complete absence of seeds, whereas Col-0 anthers had abundant pollen at this stage (white arrows). However, there are no other obvious morphological differences except that they flowered for a longer time (Kim et al., 2010). Image was reprinted with permission of the American Society of Plant Biologists (Copyright © American Society of Plant Biologists).

Within the *pksa-1 pksb-3* F2 population, I identified plants homozygous for *pksa-1* and heterozygous for *pksb-3*. Interestingly, these plants displayed a partially sterile phenotype, with smaller siliques containing fewer seeds or empty siliques. Therefore, I quantified seed set in siliques taken at random from plants with different *PKSA* and *PKSB* allelic combinations (Table 3.3; n = 15 siliques for each genotype). *pksa-1* and *pksb-3* single mutants and plants heterozygous for *pksa-1* and homozygous for *pksb-3* show slightly reduced numbers of seeds relative to wild-type control plants, whereas plants homozygous for *pksa-1* and heterozygous for *pksb-3* showed clear reduction in fertility, with most siliques having 10 or fewer seeds.

Table 3.3 Quantification of seed set in plants with different *PKSA* and *PKSB* allelic combinations.

	Number of siliques containing the following numbers of seeds								
Genotype[a]	>71[b]	70-61	60-51	50-41	40-31	30-21	20-11	10-1	0
AABB	6.0	9.0	0	0	0	0	0	0	0
AAbb	4.3	8.3	2.3	0	0	0	0	0	0
Aabb	3.7	10.3	1.0	0	0	0	0	0	0
AaBb	5.3	9.7	0	0	0	0	0	0	0
aaBB	1.0	10.0	4.0	0	0	0	0	0	0
aaBb	0	0	0	0	0.7	0.7	2.7	10.0	1.0
aabb	0	0	0	0	0	0	0	0	15.0

[a] At1g02050 (*PKSA*) alleles are symbolized by A (WT) and a (*pksa-1*), and At4g34850 (*PKSB*) alleles are symbolized by B (WT) and b (*pksb-3*).

[b] Values represent the average number of siliques on one branch from the indicated genotype that contains seeds numbering within the given range. Genotype AABB (wild type) contained at least 61 seeds in each silique. On the other hand, genotype aabb (double mutant *pksa-1 pksb-3*) had no seeds in any siliques. Genotype aaBb showed significantly reduced fertility with most siliques having 10 or fewer seeds (Kim et al., 2010).

3.4.6 Anther and microspore development in the *pksa-1 pksb-3* double mutant

To determine the point at which pollen development was impaired in the completely male sterile *pksa-1 pksb-3* double mutant, I examined developing anthers in the double mutant background relative to wild-type by light microscopy (Figure 3.8), using the stages of anther development defined by Sanders et al. (1999). Microspore and anther development in *pksa-1 pksb-3* plants appeared normal through stage 8, when individual microspores could be seen, indicating that the callose wall had degenerated, and releasing microspores from tetrads in a normal manner. However, by stage 9 to 11, aberrant microspore development in *pksa-1 pksb-3* anthers relative to wild type was observed. Free microspores appeared to have thin walls and aberrant structures, and locules had accumulated debris of defective pollen grains (Figure 3.8). By stage 12,

mature pollen grains were observed in locules of wild-type plants, while most *pksa-1 pksb-3* anthers were devoid of pollen. In a smaller number of *pksa-1 pksb-3* mutant anthers in stages 9-12 (lower *pksa-1 pksb-3* panels in Figure 3.8) some pollen grains appeared to be still in the tetrad stage or were fused with each other, and the tapetum layers were highly enlarged and vacuolated, and failed to undergo programmed cell death.

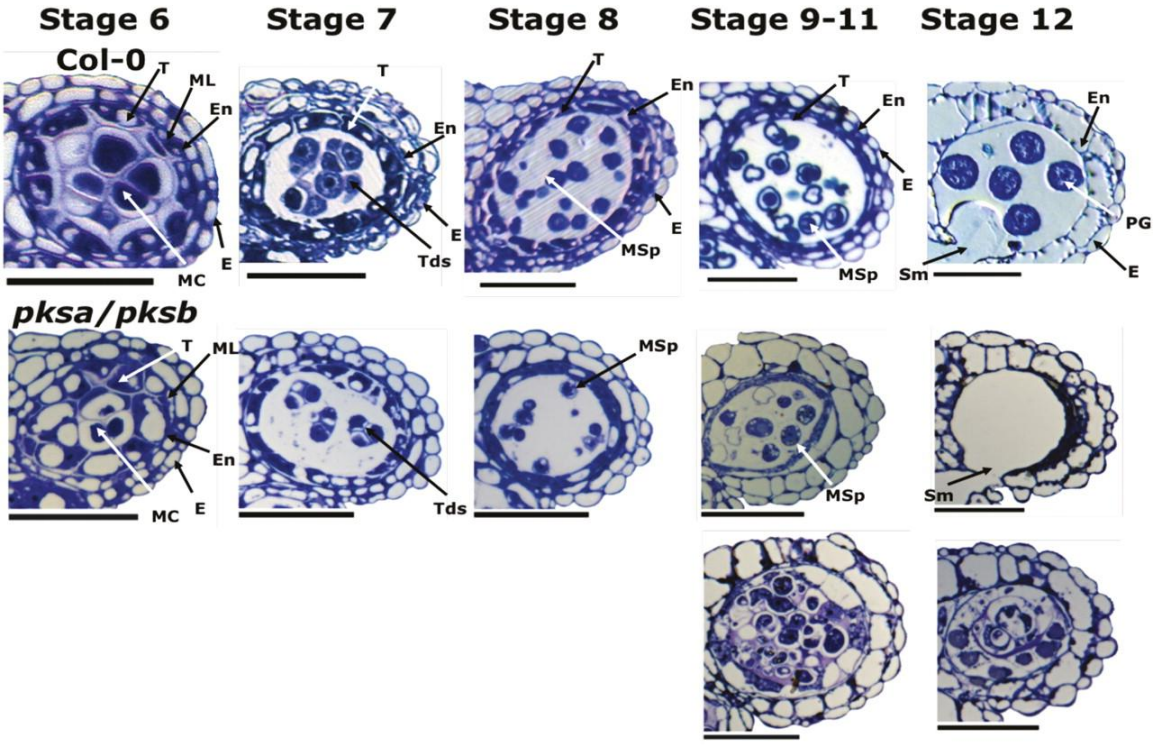


Figure 3.8 Phenotypic characterization of anther and microspore development in wild-type (Col-0) and *pksa-1 pksb-3* flowers.

Anther cross sections (1 µm) were taken from developing flowers of wild type plants and *pksa pksb* mutant plants and stained with toluidine blue. Numbers indicate anther developmental stages according to Sanders et al. (1999). The two panels at the bottom illustrate anther phenotypes at stages 9-12 occasionally found in *pksa pksb* mutants.E, epidermis; En, endothecium; MC, meiotic cell; ML, middle layer; MSp, microspores; PG, pollen grains; Sm, septum; T, tapetum; Td; tetrad. Scale bars = 40 µm (Kim et al., 2010). Image was reprinted with permission of the American Society of Plant Biologists (Copyright © American Society of Plant Biologists).

I used TEM to examine *pksa-1 pksb-3* microspore development at greater resolution. In agreement with light microscopy observations, at stages 5 to 7, microspore and tapetum development were normal in the mutant, and free microspores were observed at stage 8 (Figure 3.8 and Figure 3.9). However, at stage 9 *pksa-1 pksb-3* microspores had thin cell walls apparently devoid of an exine and by stage 12, locules were mostly devoid of pollen grains, although occasional defective pollen grains were observed (Figure 3.9).

I next used TEM to examine *pksa-1 pksb-3* mutant anthers and microspores at stages 9 and 11 at greater detail, relative to wild type. Figure 3.10 shows that at stage 9, wild-type anthers contained uninucleate microspores with a thick, reticulate exine, intact tapetum and an electron-dense cuticle on the outer anther epidermis, which exhibited hair-like protrusions (Figure 3.10A, D, G, J and M). By contrast, aberrant microspore development was observed at the same stage in *pksa-1 pksb-3* anthers (Figure 3.10B, E and K). While uninucleate microspores, an intact tapetum and a normal anther epidermis with an intact cuticle were present (Figure 3.10E, H and N), microspores were devoid of a recognizable exine (Figure 3.10E). At higher magnification, a defective, thin fibrillar matrix, presumably defective exine, was apparent on mutant microspores (Figure 3.10K). At stage 11 most *pksa-1 pksb-3* mutant anthers contained only a few degenerating pollen grains (Figure 3.8 and Figure 3.9). However, in a subset of *pksa-1 pksb-3* mutant individuals, enlarged pollen grains were present at this stage (Figure 3.10C) that had an abnormal exine structure without a thick and reticulated wall (Figure 3.10F and 3.10L). Under higher magnification, it was also apparent that the locules of these anthers were filled with misshapen structures and an electron-dense network, possibly unpolymerized sporopollenin precursors (Figure 3.10I and 3.10L). In these individuals, the tapetum and anther outer wall epidermis and cuticle appeared to be normal (Figure 3.10I and 3.10O).

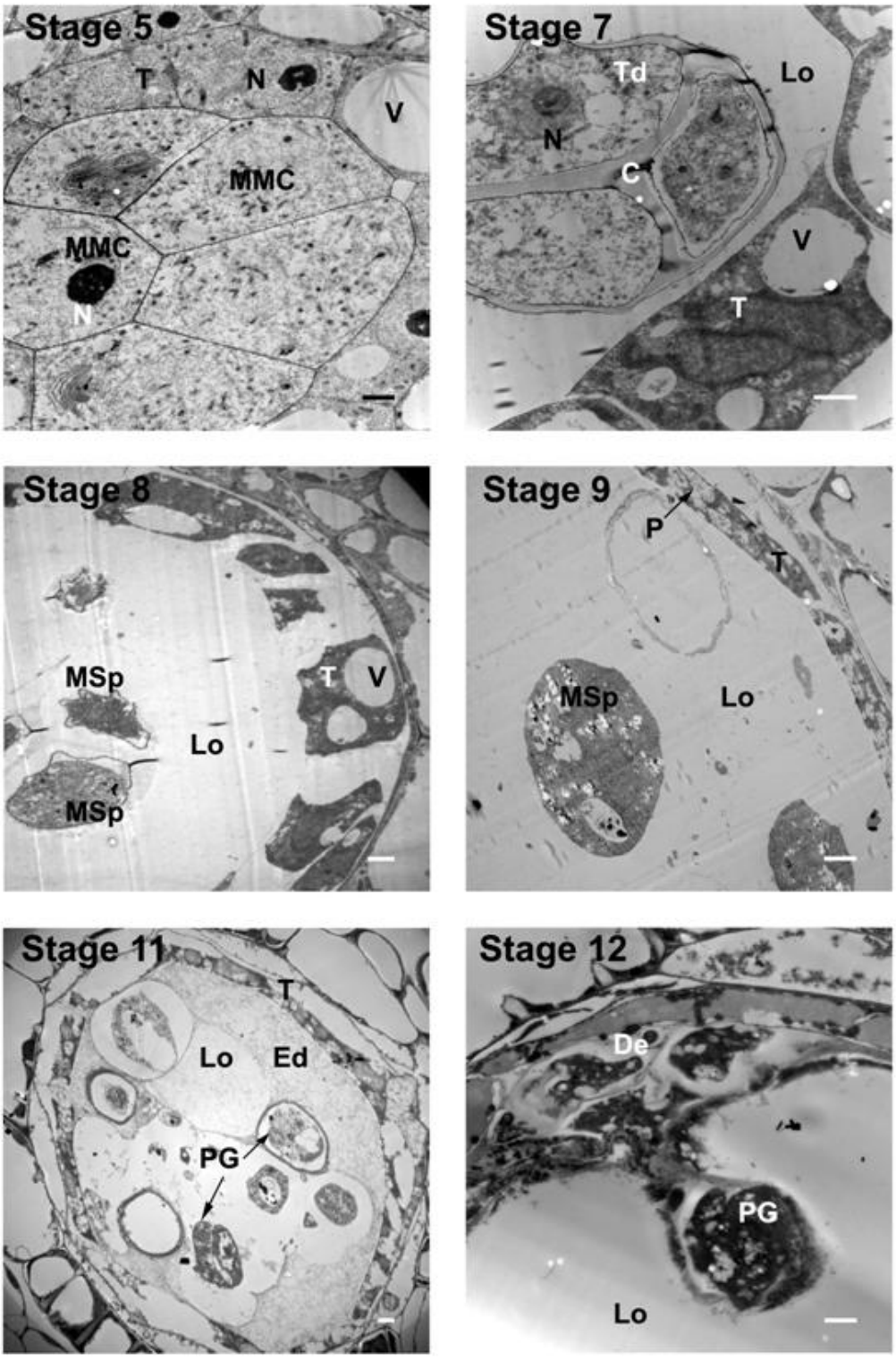


Figure 3.9 Anther development in the double mutant *pksa-1 pksb-3*.

C, callose wall; De, debris of defective pollen grains; Ed, electron-dense material; Lo,

locule; MMC, microspore mother cell; MSp, microspore; N, nucleus; P, plastid filled with plastoglobuli; PG, pollen grain; T, tepetal cell; Td, tetrad; V, vacuole. Scale bar = 2μm (Kim et al., 2010).

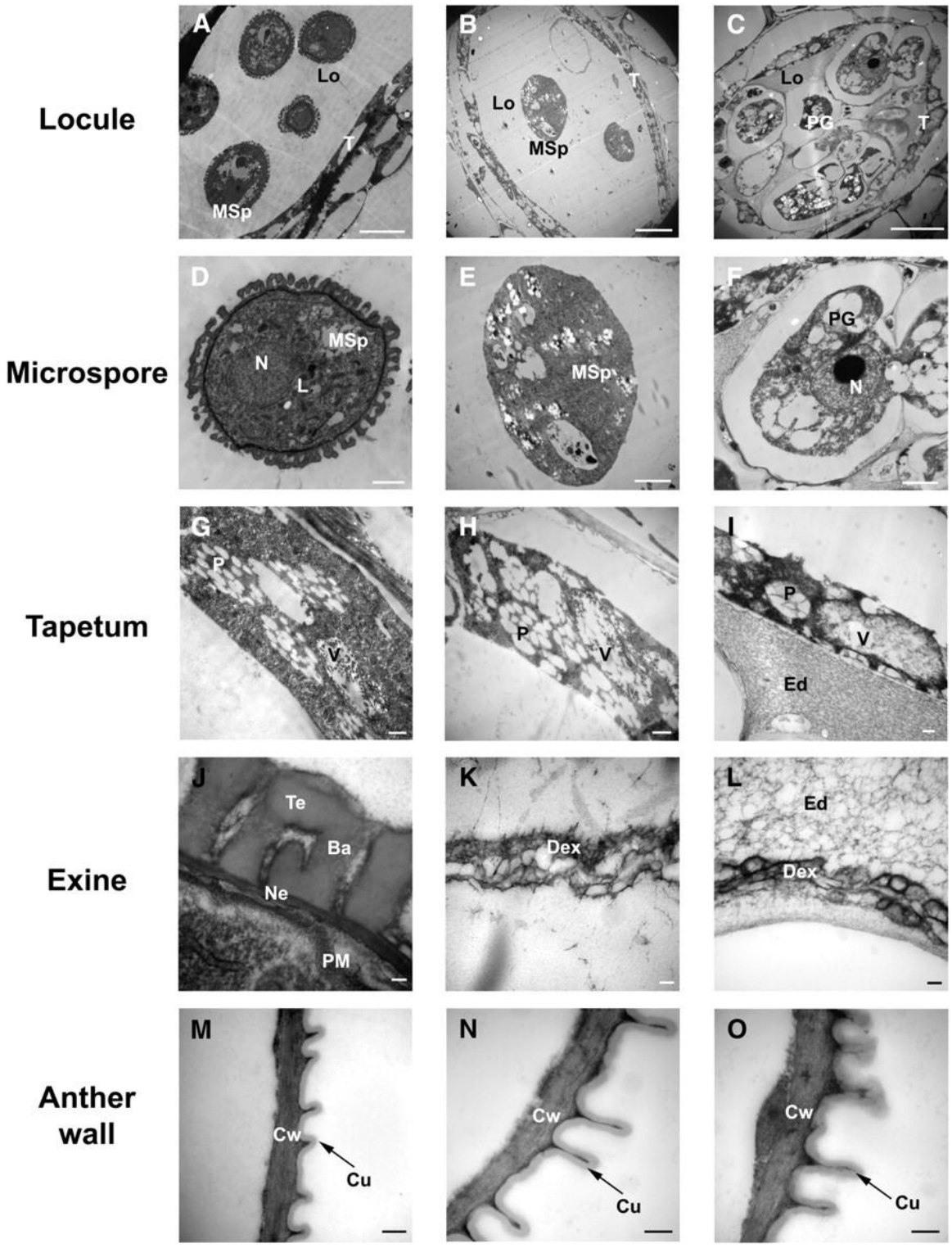


Figure 3.10 Transmission electron micrographs of wild-type (Col-0) and *pksa-1 pksb-3* anthers and pollen.

(A), **(D)**, **(G)**, **(J)** and **(M)** Microspore structure, tapetum structure, exine formation, and

outer wall of anther epidermis at anther stage 9 in Col-0 wild-type plants.
(B), **(E)**, **(H)**, **(K)** and **(N)** Microspore structure, tapetum structure, exine formation, and outer wall of anther epidermis at anther stage 9 in *pksa-1 pksb-3* plants.
(C), **(F)**, **(I)**, **(L)** and **(O)** Pollen grain structure, tapetum structure, exine formation, and outer wall of anther epidermis at anther stage 11 in *pksa-1 pksb-3* plants.

Ba, bacula; Cu, cuticle; Cw, cell wall; Dex, defective exine structure; Ed, electron-dense material; Ex, exine; Lo, locule; MSp, microspore; Ne, nexine; P, plastid filled with plastoglobuli; PG, pollen grain; PM, plasma membrane; T, tapetal cell; Te, tectum; V, vacuole containing electron-dense material. Scale bars = 10 μm in (A) to (C), 2 μm in (D) to (F), 500 nm in (G) to (I) and (M) to (O) and 100 nm in (J) to (L) (Kim et al., 2010).

3.5 Discussion

In my work, I documented defects in exine and pollen formation leading to complete loss of fertility in the *pksa pksb* double mutant. My detailed phenotypic analysis of the double mutant revealed new features of its loss-of-function phenotype, providing insights into functions of the wild-type enzymes. In the majority of anthers observed, defective microspore development was first observed at stage 9 (Figure 3.7), consistent with the timing of transient *PKSA* and *PKSB* expression (Figure 3.4) in the tapetum, and the timing of exine formation (Blackmore et al., 2007). High-resolution TEM images of stage 9 wild type and double mutant anthers (Figure 3.9) showed that mutant microspores completely lacked exine, which was replaced by an amorphous material similar to other mutants defective in sporopollenin biosynthesis (*acos5*, *abcg26*, *drl1/tkpr1*; de Azevedo Souza et al., 2009; Quilichini et al., 2010; Grienenberger et al., 2010). No abnormalities in tapetum cells were observed, and anther epidermal wall cuticle deposition was similar to that in wild type plants. Thus, the *pksa pksb* defect appears to be highly specific to exine formation and sporopollenin biosynthesis, consistent with *PKSA* and *PKSB* tapetum-specific expression patterns. Thus, unlike the fatty acid hydroxylase CYP704B1, which is involved in generating both sporopollenin and cutin precursors in developing rice anthers (Li et al., 2010), PKSA and PKSB appear to be specific in generating sporopollenin constituents.

In some of the *pksa pksb* mutant anthers, additional more complex microspore

and locule phenotypes were observed at stage 9 and later stages. The presence of aberrant microspores showing signs of developmental arrest and cell fusion suggests that lack of PKSA/PKSB-derived sporopollenin constituents affects cell surface properties of developing microspores at the tetrad and subsequent stages. The densely staining material found in the locules of such *pksa pksb* mutant anthers (Figures 3.7 and 3.9) was never observed in wild-type anthers and could represent high levels of unpolymerized fatty acid derived precursors and/or material that normally co-polymerizes with such sporopollenin constituents to form the exine. Abnormal accumulation of such potentially lipophilic material could also result in the abnormal microspore cellular structure and apparent cell fusions observed in these anthers, and the complex phenotype of the double mutant could explain the extensive changes in the anther metabolome observed in *lap5/pksa lap6/pksb* mutants (Dobritsa et al., 2010). Taken together, my results indicate an essential function for PKSA- and PKSB-derived products in sporopollenin biosynthesis and microspore development.

The phylogenetic analysis (Figure 3.2) I performed showed that the plant *PKS* clade containing *PKSA* and *PKSB* is clearly distinct from the clade containing the *bona fide Arabidopsis* and *Physcomitrella CHALCONE SYNTHASE (CHS)* genes and other putative *CHS* genes from poplar and rice. This relationship between the true *CHS* genes and the *PKS CHS-like* (*CHSL*) genes, also observed by others (Mizuuchi et al., 2008; Wu et al., 2008; Dobritsa et al., 2010), and is similar to the relationship between ACOS5 and true 4CL enzymes (Souza Cde et al., 2008; de Azevedo Souza et al., 2009). The *CHSL* clade containing *PKSA* and *PKSB* has two distinct sub-clades of angiosperm *PKS* genes that are related to either *Arabidopsis PKSA* or *PKSB*, with at least one *PKSA* and one *PKSB* homolog found in each of the fully sequenced poplar and rice genomes. In addition, the *Nicotiana sylvestris PKSA* homolog *CHSlk* (Atanassov et al., 1998) and *Silene latifolia PKSB* homolog *CHSL* (Ageez et al., 2005) are known to be expressed in male reproductive organs, and the wheat and triticale homologs of the *Arabidopsis PKS*s also have anther and tapetum-preferred expression patterns (Wu et al., 2008).

The *CHSL* sub-clade containing *Arabidopsis PKSA* and *PKSB* contains both pine and *Physcomitrella* members, both of which occupy positions at or near the base of the clade (Figure 3.2). The *Pinus radiata ChS1* gene is specifically expressed in male cones

and is likely tapetum-expressed (Walden et al., 1999). I surveyed the expression pattern of the *Physcomitrella PKS* (*CHS10*) gene (Jiang et al., 2006; Koduri et al., 2010), an apparent *PKSA* and *PKSB* homolog (Figure 3.2), by assessing transcript abundance in a *Physcomitrella patens* cDNA database (http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?LID=23755&PAGE=1). This analysis showed that *PpCHS10* cDNAs are found exclusively in a library derived from RNA extracted from the sporophyte. Together, these data suggest a conserved function in sporopollenin monomer biosynthesis for PKSA and PKSB and their apparent orthologs in land plants, and indicate that this function arose early in land plant evolution prior to the divergence of bryophytes and tracheophytes. The exine in pollen walls may have evolved a more elaborate structure in seed plants, based on specialized functions of the duplicated *PKSA* and *PKSB* genes found in common within the angiosperm lineage.

Type III polyketide synthases (PKSs) produce secondary metabolites that play a variety of roles in plants. The studies performed in the lab of our collaborator, Dr. Michel Legrand and published in Kim et al. (2010), together with the recent report of Mizuuchi et al. (2008) shows that both PKSA and PKSB produce triketide and tetraketide α-pyrones by condensation with long chain fatty acyl-CoAs (up to 20 carbon chain lengths) and malonyl-CoA. The two PKSs have unusually broad substrate specificities as compared to typical plant type III PKSs (Mizuuchi et al., 2008). In contrast to their results and those of Mizuuchi et al., (2008), Dobritsa et al. (2010) failed to observe activity of recombinant LAP5/PKSA or LAP6/PKSB against fatty acyl substrates greater than C12 in length. The reason for this discrepancy is not clear, but activity against C16 to C18 fatty acyl-CoA substrates was consistently observed in multiple assays using their recombinant enzyme preparations (Kim et al., 2010). Moreover, the results from another collaborator, Dr. Dae-Yeon Suh, also show that, *in vitro*, PKSA preferentially catalyzes condensation of hydroxy fatty acyl-CoAs, which may be sequentially generated by CYP703A2 (Morant et al., 2007) and/or CYP704B1 (Dobritsa et al., 2009) with ACOS5 (de Azevedo Souza et al., 2009) to produce corresponding hydroxy tri- and tetraketide α-pyrones (Kim et al., 2010).

Since ACOS5 also accepts a broad range of fatty acid substrates (de Azevedo Souza et al., 2009), the exact nature of its *in vivo* products that could be used as potential

in vivo PKSA and PKSB substrates remains unclear. However, ACOS5 shows highest activity against medium-chain hydroxylated fatty acids (de Azevedo Souza et al., 2009), and could thus generate medium-chain hydroxy fatty acyl-CoAs, for which PKSA has a strong substrate preference. Thus, an attractive hypothesis is that the hydroxylated acyl chains generated by ACOS5 and a suite of cytochrome P450 hydroxylases (Morant et al., 2007; de Azevedo Souza et al., 2009; Dobritsa et al., 2009; Li et al., 2010) are incorporated into multi-hydroxy tri- and tetraketide α-pyrones products *in vivo*. This model is supported by data from Dr. Michel Legrand's group showing the preferential subcellular localization of PKSA and PKSB to the ER (Kim et al., 2010). Since the P450 hydroxylases are likely ER-localized, they may form metabolons with the ER-associated PKSs to form alkyl α-pyrones.

Both PKSA and PKSB recombinant enzymes catalyze the decarboxylative condensations of fatty acyl-CoAs with malonyl-CoA *in vitro* to generate tri- and tetraketide α-pyrones (Mizuuchi et al., 2008). Based on SEM data provided by our collaborator Dr. Michel Legrand (Kim et al., 2010), **s**ingle *pksa* and *pksb* mutants, display subtle changes in exine patterning and deposition (data not shown) but are fertile. By contrast, the *pksa pksb* double mutant produced very small amounts of pollen and was completely male sterile (Figures 3.7 and 3.9), consistent with the observations of Dobritsa et al. (2010). Whereas these data suggest that they have partially redundant functions in exine formation, I present several lines of evidence suggesting that PKSA and PKSB could fulfill different *in vivo* functions. First, exine patterning defects in the *pksa* and *pksb* mutants were slightly different, with *pksb* mutants showing more pronounced defects including ectopic globular exine protrusions (Kim et al., 2010). Secondly, results from quantification of seed set in siliques of plants with different *PKSA* and *PKSB* allelic combinations (Table 3.1) showed that partial addition of PKSA activity to the *pksb pksb* background in *PKSA pksa pksb pksb* plants partially restored fertility relative to fully sterile *pksa pksa pksb pksb* plants, whereas addition of PKSB activity to *pksa pksa pksb pksb* plants (*pksa pksa PKSB pksb* plants) did not have this effect (Table 3.1). One interpretation of these data is that PKSA plays a more critical role in exine formation than does PKSB.

Recent microarray analyses comparing gene expression in *Arabidopsis* wild-type

anthers relative to that in *sporocyteless*/*nozzle* or *excess microsporocytes1*/*extra-sporogenous cells* mutants suggest that SPL and/or EMS1 positively regulate transcription of *ACOS5*, *PKSA* and *CHSL2* (At4g00040), all of which showing highly decreased expression levels in the *spl/nzz* and *ems1/exs* mutants (Wijeratne et al., 2007). By contrast, expression of *PKSB* was unaffected in the mutant backgrounds (Wijeratne et al., 2007). My results show that the timing of maximal tapetum-localized *PKSA* and *PKSB* mRNA accumulation differs over the course of anther development (Figure 3.4), further supporting distinct regulatory control and potential specialized functions of the two genes.

It is noteworthy that another *Arabidopsis PKS* gene, *CHSL2* (At4g00040) is 79% similar to *PKSA*. However, when expressed in bacteria, the cognate recombinant protein displayed no activity with the various fatty acyl-CoAs and malonyl-CoA used as substrates (Dr. Michel Legrand, personal communication). These results, together with expression data from public databases showing that *CHSL2* is expressed at later stages of flower bud development, indicate that CHSL2 performs an unknown enzymatic function distinct from that of PKSA.

Recently, Dobritsa et al. (2010) suggested that LAP5/PKSB and LAP6/PKSA could play roles both in the synthesis of alkylpyrones and in synthesis of phenolic constituents of sporopollenin in exine (Dobritsa et al., 2010). Metabolic profiling of developing anthers indicated that several flavonoids, including chalcone, naringenin, dihydrokaempferol and isorhamnetin 3-sophoroside, were significantly reduced in single mutants and undetectable in the double mutant. However, such flavonoids are not likely to be direct products of LAP5/PKSB or/and LAP6/PKSA. Whereas LAP5/PKSB and LAP6/PKSA are closely related to CHS, the recombinant enzymes do not exhibit CHS activity (Dobritsa et al., 2010). Furthermore, whereas CHS activity and flavonoid biosynthesis appear to be crucial to pollen development in some plants (Mo et al., 1992; van der Meer et al., 1992; Fischer et al., 1997; Hofig et al., 2006; Schijlen et al., 2007), complete disruption of *CHS* expression in *Arabidopsis*, leading to the absence of foliar anthocyanins and flower flavonoids, has no impact on male or female fertility (Burbulis et al., 1996), suggesting that CHS-generated flavonoids are not required for normal *Arabidopsis* pollen development.

One possible reason for reduced flavonoid levels in *pksa*, *pksb*, and *pksa pksb* mutants is that the defective exine walls of pollen grains in these mutants affect subsequent deposition of pollen coat tryphine that contains phenolics, flavonoids, fatty acid derivatives, and proteins (Piffanelli et al., 1998; Scott et al., 2004; Grienenberger et al., 2009). Thus, reduction or absence of flavonoids could be the indirect consequence of reduced deposition of flavonoid-containing tryphine. Consistent with this, pollen coat deposition is affected in the *pksa* and *pksb* mutants, and is more severely deficient in the *pksb* mutant (data not shown). The more severe tryphine defect in *lap5*/*pksb* observed in TEM images is consistent with the much greater reduction of flavonoid accumulation *lap5*/*pksb* anthers relative to *lap6/pksa* anthers (Dobritsa et al., 2010).

PKSA and *PKSB* are strongly co-expressed in tapetum cells with *ACOS5* (de Azevedo Souza et al., 2009), ACOS5 preferentially generates hydroxy fatty acyl-CoAs but has no activity towards hydroxycinnamic acids (de Azevedo Souza et al., 2009), and *PKSA* preferentially accepts hydroxy fatty acyl-CoAs to generate tri- and tetraketide α-pyrones (Kim et al., 2010). These observations suggest that the most plausible *in vivo* substrates for PKSA and PKSB are hydroxy fatty acyl-CoAs rather than phenolics such as hydroxycinnamoyl-CoAs, and that alkyl pyrones are natural products generated by PKSA and PKSB that are required for sporopollenin biosynthesis. The biosynthesis of the tri- and tetraketide α-pyrones in plants has not been well studied and their presence in *Arabidopsis* has not been reported. Thus, while *in vitro* biochemical assays indicate that the *in vivo* functions of PKSA and PKSB are to catalyze condensation of malonyl-CoA with hydroxy fatty acyl-CoAs generated by ACOS5 to generate alkyl pyrones, this hypothesis requires further testing. For example, reduction of keto functions after elongation of the chain by PKSA/PKSB, would prevent the formation of the α-pyrone ring, and alkyl phloroglucinols could be formed as sporopollenin natural products.

The tri- and tetraketide α-pyrones generated *in vitro* by PKSA and PKSB also contain a ketone group on the α-pyrone rings and an additional ketone on the alkyl carbon chain, in the case of the tetraketide. These carbonyl groups could be the targets for further reduction to generate even more highly hydroxylated polyketides. One candidate enzyme for such activity is the reductase encoded by the *Arabidopsis DIHYDROFLAVNONOL 4-REDUCTASE LIKE1* (*DRL1*)/*TETRAKETIDE α-PYRONE REDUCTASE (TKPR1*)

(At4g35420) gene, which is tightly co-expressed with *ACOS5*, *PKSA*, and *PKSB* and is required for male fertility and exine formation (Tang et al., 2009).

The remarkable conservation of *Arabidopsis PKSA* and *PKSB*, *ACOS5* (de Azevedo Souza et al., 2009) and *CYP703A2* (Morant et al., 2007) genes in land plants suggests that the encoded enzymes are part of an ancient sporopollenin biosynthetic pathway. The *in vitro* biochemical functions of PKSA and PKSB that we explored in detail, together with similar data on ACOS5, and CYP703A2, the coordinated co-expression of these genes in the tapetum during free microspore stages of anther development, and the ER localization of PKSA and PKSB support the hypothesis that this pathway catalyzes sequential modification of fatty acid starter molecules to generate integral sporopollenin components of the pollen exine. Based on these new data, I propose a reaction sequence localized to the ER leading from hydroxylation of medium to long chain fatty acids (catalyzed by CYP703A2), to fatty acyl-CoA formation (catalyzed by ACOS5), and condensation of fatty acyl-CoAs with malonyl-CoA (catalyzed by PSKA and PSKB) to generate triketide and tetraketide α-pyrone sporopollenin precursors. Work reported in Chapter 4 and recently published (Grienenberger et al., 2010) extends this work and shows that reductases encoded by tapetum-expressed *DRL*/*TKPR* genes are also part of this pathway, and accept the tetraketide α-pyrone products generated by PKSA and PKSB *in vitro* to form reduced derivatives that appear to be sporopollenin precursors.

Chapter 4. Analysis of TETRAKETIDE α-PYRONE REDUCTASE (TKPR) function in *Arabidopsis thaliana* reveals a novel and conserved biochemical pathway in sporopollenin monomer biosynthesis

4.1 Summary

In this chapter, I show that two *Arabidopsis* genes encoding oxidoreductases, *TKPR1* and *TKPR2*, are co-expressed with *ACOS5* and are specifically and transiently expressed in tapetal cells during microspore development. The null mutant *tkpr 1-1* displayed severe pollen exine layer defects, was male sterile and was shorter in stature than wild-type (Col-0) plants and had smaller leaves. Phylogenetic studies indicated that the two reductases belong to a large reductase/dehydrogenase gene family and cluster in two distinct clades with putative orthologs from several angiosperm lineages and the moss *Physcomitrella patens*. Recombinant proteins produced in bacteria reduced the carbonyl function of tetraketide α-pyrone compounds synthesized by PKSA/B rather than the CoA esters to generate aldehyde or alcohol. Thus the proteins were therefore named TETRAKETIDE α-PYRONE REDUCTASE1/2 (TKPR1/2) (previously called DRL1 and CCRL6, respectively). TKPR activities, together with those of ACOS5 and PKSA/B, identify a conserved biosynthetic pathway leading to hydroxylated α-pyrone compounds that were previously unknown to be sporopollenin precursors.

4.2 Introduction

4.2.1 DIHYDROFLAVONOL 4-REDUCTASE (DFR) and DIHYDROFLAVONOL 4-REDUCTASE-LIKE1 (DRL1)

Flavonoids are comprised of a relatively diverse family of aromatic molecules generated by the sequential decarboxylative addition of three acetate units from malonyl-CoA to a p-coumaryl-CoA starter molecule derived from the general phenylpropanoid pathway. These flavonoids compounds are normally classified to six major subgroups such as the chalcones, flavones, flavonols, flavandiols (leucoanthocyanidins), anthocyanins, and proanthocyanidins (condensed tannins) (Winkel-Shirley, 2001). Dyhydroflavonol 4-reductase (DFR) is the first committed enzyme of anthocyanin biosynthesis in the flavonoid pathway (Shimada et al., 2004). DFR acts on dihydroflavonols, generating leucoanthocyanidins by reduction of a ketone to a hydroxyl group (Figure 4.1). These leucoanthocyanidins are converted into colored anthocyanidins via an oxidation step catalyzed by anthocyanidin synthase (ANS), a 2-oxoglutarate iron-dependent oxygenase. Subsequent reduction of anthocyanidins by anthocyanindin reductase (ANR) produces condensed tannins or proanthocyanidins (Xie et al., 2003). *DFR* genes have been cloned from a variety of plants and DFR is considered to regulate carbon flux into anthocyanin biosynthesis (Bernhardt et al., 1998; Tanaka et al., 1998; Itoh et al., 2002; Shimada et al., 2004).

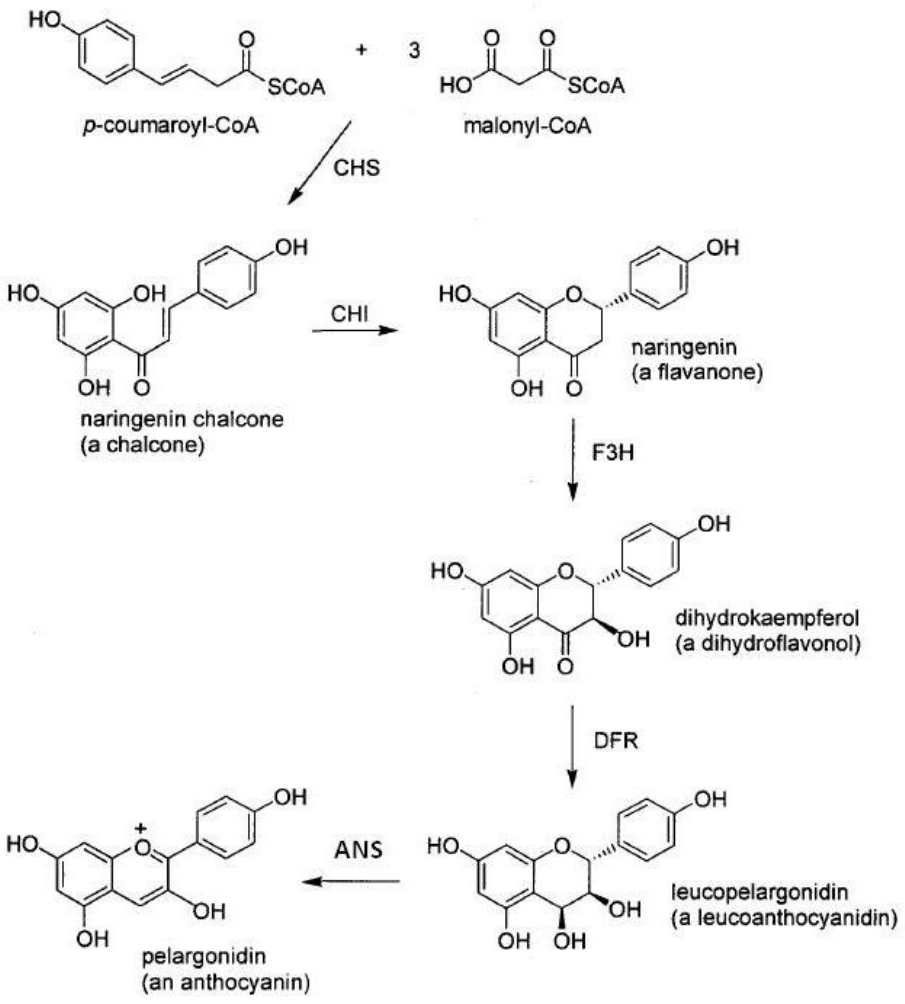


Figure 4.1 Overview of the flavonoid biosynthesis pathway.
Enzyme names are abbreviated as follows: chalcone synthase (CHS), chalcone isomerise (CHI), flavanone 3-hydroxylase (F3H), dihydroflavonol 4-reductase (DFR), and anthocyanidin synthase (ANS). Image was reprinted with permission of Che Caswell Colpitts (Copyright © 2009 C.C. Colpitts).

The completed *Arabidopsis thaliana* genome has revealed that in addition to *DFR*, *Arabidopsis* has a family of four additional related genes (Yuan et al., 2007). True *DFR* (At5g42800) was first identified by its role in synthesis of brown pigments in the seed coat, a phenotype collectively termed *transparent testa* (*tt*) (Shirley et al., 1995). Other *DFR*-like genes in *Arabidopsis* have been verified to have different biochemical and physiological functions. *BANYULS* (*BAN*, At1g61720) encodes anthocyanidin reductase (ANR) mentioned above (Winkel-Shirley, 2001; Xie et al., 2003). In addition,

studies on another *DFR*-like gene, *BEN1* (At2g45400) suggest that BEN1 could be a brassinosteroid reductase that catalyzes the conversion of typhasterol (TY), castasterone (CS) and brassinolide (BL) to biologically inactive 6-OHTY, 6-OHCS and 6-OHBL, respectively (Yuan et al., 2007). Thus, it appears that this DFR-like enzyme can act on an classes of molecules entirely different from DFR, but one that shares structural similarity in carbon-ring structure to flavonoids. In a recent study of the *DFR-like* gene *DRL1* (At4g35420) published while my research was in progress, this gene was shown to be essential for pollen wall development (Tang et al., 2009), suggesting that this DFR-like enzyme also acts on an different class of molecules than DFR. The likely substrate diversity of DFR and DFR-like enzymes also has been highlighted by the phylogenetic analysis and amino acid alignment of DFRs and DFR-like proteins in various plant species. These studies showed that proteins related to DFR contain not only a putative NADPH-binding domain but also a variable putative substrate specificity domain between well-conserved regions (Johnson et al., 2001; Shimada et al., 2004). The data were interpreted to indicate that DFR-like enzymes with minor amino acid difference in the presumed substrate-binding region could have different substrate preferences (Johnson et al., 2001).

In Chapter 3 and in Kim et al. (2010), we demonstrated that two *Arabidopsis* polyketide synthases, POLYKETIDE SYNTHASE A/B (PKSA/B), play critical roles in sporopollenin biosynthesis, acting downstream of ACOS5. *pksa pksb* double mutants are completely male sterile and lack an exine. *In vitro*, both proteins accept fatty acyl-CoA esters synthesized by ACOS5 and condense them with malonyl-CoA to yield triketide and tetraketide α-pyrones as reaction products (Kim et al., 2010). Here, I show by *in situ* hybridization of mRNAs of two *Arabidopsis* oxidoreductases, one encoded by the *DRL1* gene previously described by Tang et al. (2009) and the other annotated as *CINNAMOYL COA REDUCTASELIKE6* (*CCRL6*) (Hamberger, 2007), are coexpressed with *ACOS5*, *PKSA*, and *PKSB* in anther tapetum cells. Our collaborator showed that the recombinant enzymes produced in bacteria accept the tetraketide α-pyrones produced by PKSA and PKSB as substrates to reduce the carbonyl function on the tetraketide alkyl chain to a secondary alcohol function (Grienenberger et al., 2010). Phylogenetic studies showed that the oxidoreductases belong to a gene family conserved from moss to flowering plants.

Since similar gene conservation holds true for *PKSA*, *PKSB*, and *ACOS5*, it appears that the whole biosynthetic pathway leading from medium or long-chain fatty acids to sporopollenin units is highly conserved and may have been a key determinant in the evolution of land plants.

4.3 Materials and methods

4.3.1 Plant material and growth conditions

Arabidopsis thaliana Columbia (Col-0) seeds were sterilized and after a cold treatment (2 days at 4°C in the dark) and germinated at 20°C under 70 µmol m^{-2} s^{-1} fluorescent lighting. Twelve days later, the plants were transferred to a growth chamber with a light/dark cycle of 16 hr/8 hr. T-DNA insertion mutants were obtained from SAIL (Alonso et al., 2003) via The Arabidopsis Information Resource. In progeny, homozygous insertion lines SALK_837_D01 for *tkpr-1* was identified by PCR using gene-specific and T-DNA–specific primers (Table 4.1).

4.3.2 Identification and characterization of *TKPR1* insertion mutant

The T-DNA insertion line *tkpr1-1* (SAIL_837_D01) was identified using the Salk Institute T-DNA Express *Arabidopsis* gene-mapping tool (Alonso et al., 2003). Homozygous plants were identified in PCR reactions on genomic DNA with the primers 5RP and 5LP in combination with LB1 which is left border primer of T-DNA insertion. Primers are listed in Table 4.1. Crosses of wild-type pollen to homozygous *tkpr1* mutant plants were performed to obtain F2 generation plants. The patterns of *TKPR1* T-DNA insertion allele segregation in the F2 generations were tested by chi-square statistical analysis of observed phenotypes and genotypes using Graph-Pad software (http://graphpad.com/quickcalcs/chisquared1.cfm), with expected values based on Mendelian segregation, observed values on the F2 population, and two degrees of freedom (genotypes) or one degree of freedom (phenotypes).

4.3.3 Complementation of *tkpr1* mutant

A 3593bp *DRL1* genomic fragment was amplified using the Plantinum Taq DNA polymerase High Fidelity (Invitrogen) with the gene-specific primers (Table 4.1) and cloned into pCR8/GW/TOPO (Invitrogen). After verification by sequencing, the fragment was subcloned into pGWB1, Gateway binary vector (Nakagawa et al., 2007). The transformants were selected using 25 mg/L gentamycin, 25mg/L rifampicin and 50 mg/L kanamycin. Verified transformant was introduced into *acs5* and *dfrl1* heterozygous plants using the floral dip method (Clough and Bent, 1998). Mature plants were harvested for seeds, and seeds (T1) were sown in ½ MS (Murashige and Skoog) salts (Sigma Aldrich), supplemented with 1% sucrose and 0.6% agar medium containing 25 mg/L hygromycin.

4.3.4 Phylogenetic studies

The *Arabidopsis TKPR1* (At4g35420) gene was used in BLAST searches to identify potential homologs in the genomes of *Arabidopsis* (TAIR, http//www.arabidopsis.org), poplar (Joint Genomics Institute, *Populus trichocarpa* v.1.1; http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) rice (The Institute for Genome Research; http://www.tigr.org/tdb/e2k1/osa1/), *Physcomitrella patens* (Joint Genomics Institute, *Physcomitrella patens* v.1.1 http://genome.jgi-psf.org//Phypa1_1/Phypa1_1.home.html), and *Chlamydomonas reinhardtii* (Joint Genomics Institute *Chlamydomonas reinhardtii* v. 3.0; http://genome.jgi-psf.org/Chlre3/Chlre3.home.html) All sequences obtained are given in Appendix C. Protein sequences were aligned using MUSCLE 3.6 using the default parameters (sequencing clustering; UPGMA, objective score; classic sum-of-pairs score) (Edgar, 2004), and the multiple protein sequence alignments were manually optimized. Aligned sequences are available in Appendix D. To reconstruct phylogenetic trees, maximum likelihood analyses with 1000 bootstrap replicates were performed using PhyML v2.4.4 and default settings (Guindon and Gascuel, 2003) with the JTT model of amino acid substitution.

4.3.5 Microscopy

Tissue fixation, embedding, and sectioning of *Arabidopsis* wild type (Col-0) and *tkpr1-1* mutant inflorescences were performed as described in section 2.3.4.

4.3.6 RT-PCR

RNA quality was assessed by visual inspection of rRNA on a 1.2% formaldehyde-agarose (FA) gel and quantified spectrophotometrically, and 2.5 µg RNA/ 20 µL reaction was used to generate first strand cDNA using Superscript II Reverse Transcriptase (Invitrogen) following the manufacturer's protocol. For RT-PCR, gene-specific and intron-spanning primers (Table 4.1) were used in PCR reactions to amplify corresponding cDNA sequences under the following PCR conditions: 95℃ for 3 min, followed by 35 cycles of (94℃ for 30 s, 56℃ for 30 s, 72℃ 1 min) followed by 72℃ for 10 min, using Taq polymerase in a 50 µL total reaction. *Actin2* was used as control.

For quantitative RT-PCR analysis of *TKPR1, TKPR2* and *At1g25460* expression, 10 ng of cDNA was incubated with 10 µL iQ SYBR Green Supermix (Bio-Rad) and 5 pmol of each forward and reverse primer (Table 4.1) in a total volume of 20 µL. After an initial denaturation step at 95℃ for 3 min, 40 cycles at 95℃ for 15 s, 60℃ for 15 s, and 72℃ for 30 s were followed by a fluorescence reading. A melting curve was generated ranging from 95℃ to 60℃. Threshold cycles (CT) were adjusted manually, and the CT values for a housekeeping control *Actin2* amplified in parallel on each plate were subtracted from CT values obtained for each gene of interest, thus generating normalized CT values (ΔCT). The relative starting quantities of each gene were determined by setting as a base value the gene with the highest CT value within a tissue panel or treatment series, and relative quantities were calculated using the ΔΔCT method as described in (Hietala et al., 2003). ΔΔCT was calculated using immature flower buds as the highest expressing tissue.

4.3.7 *In situ* hybridization and validation of RNA probes

Tissue fixation, embedding, hybridization, and signal detection were performed as described in section 3.3.5. For sense and antisense *TKPR1* and *TKPR2* probe synthesis, 981 bp and 966 bp DNA template corresponding to the *TKPR1* and *TKPR2* coding region, respectively, were PCR amplified from flower cDNA using gene-specific forward and reverse primers (Table 4.1).

To validate specificity of RNA antisense probes, I performed DNA gel blotting. PCR-amplified cDNA clones of *TKPR1* and *TKPR2* (Table 4.1) were blotted to positively charged nylon membranes, Hybond-XL (GE Healthcare Life Sciences), using 10x SSC as the transfer buffer. Transfer of cDNA to the nylon membrane was monitored via ethidium bromide. The probe hybridization, and signal detection were performed as described in section 3.3.5.

4.3.8 Production of TKPR1 recombinant proteins and activity assay *in vitro*

cDNAs of *TKPR1* were amplified and ligated to the pET-28a expression vector containing His-tag (Novagen). After confirmation of the sequences, each plasmid was transformed into BL21 (DE3). The cells harboring the plasmid were cultured in 4 mL of LB medium containing 50 ug/mL of kanamycin at 37°C overnight. The culture was diluted 1:100 in LB medium containing 50 ug/ml of kanamycin and cultured to an A_{600} of 0.5 at 37°C. Then, the each protein was induced by adding IPTG to a final concentration of 1mM. The culture was further incubated at 25°C for 20 hr to allow for overproduction to occur. The *E.coli* cells were harvested by centrifugation at 5000 x g for 10 min. The cell pellets were washed once with PBS buffer (140 mM NaCl, 3 mM KCl, 10 mM Na_2HPO_4, and 2mM KH_2PO_4, pH of 7.4) and centrifuged at 5000 x g for 10min. The cell pellets were stored at -80°C to be purified subsequently.

Cell pellets were suspended in 20 mM potassium phosphate buffer, pH 7.4, containing 300 mM NaCl, 0.1% Triton X-100, 5 mM imidazole, β-mercaptoethanol,10% glycerol, 1 mM PMSF and 1 mg/mL lysozyme and incubated for 30 min at 4°C. Cells were disrupted by sonication and centrifuged at 10,000 g for 30 min. The supernatant was gently mixed up with 50% Ni-NTA agarose (Qiagen) for 1 hr at 4°C. The lysate-Ni-NTA

mixture is loaded into an Econo-Pac column (Bio-Rad). After washing with 20 mM potassium phosphate buffer, pH 7.4, containing 300 mM NaCl and 40 mM imidazole, the recombinant proteins were eluted with 20 mM potassium phosphate buffer, pH 7.4, containing 30 mM NaCl and 300 mM imidazole. To concentrate and reconstitute proteins into enzyme assay buffer (100 mM Na_2HPO_4/KH_2PO_4, pH 6.25 for CCR buffer condition) or 100 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 2.5 mM DTT for ACOS5 buffer condition) for comparison between CCR1 and TKRP1, Ultra-15 filter (Amicon) was used, following manufacturer's instructions. Protein concentration was estimated by SDS-PAGE and determined more accurately using Nanodrop spectrophotometer. The enzyme purity was assessed by SDS-PAGE.

Enzymatic activity against synthetic alkyl phloroglucinol was tested by incubating 100 µM substrate in a reaction mixture containing, 1 mM NADPH and 10 µg enzyme in 500 µl assay buffer (100mM Na_2HPO_4/KH_2PO_4, pH 6.25) for 1 hr at 30°C. To extract potential reaction products, 1 mL chloroform ($CHCl_3$) was directly added to the reaction mixtures vials, vigorously vortexed, before using for GC-MS aided product identification.

4.3.9 Identification of products by GC-MS

For GC-MS analyses, extraction from reaction mixture, derivatization of organic compounds and running condition of samples in GC-MS are described by Wang *et al.*, (2010).

4.3.10 Accession numbers

Sequence data from this article can be found in the Arabidopsis Genome Initiative or GenBank/EMBL databases under the following accession numbers: At4g35420 (*TKPR1*), At1g68540 (*TKPR2*), At1g02050 (*PKSA*), At4g34850 (*PKSB*), At1g62940 (*ACOS5*) and At3g18780 (*ACTIN2*).

4.3.11 Sequences of primers

Table 4.1 Primers used in this study.

Qantitative RT-PCR

Gene	Primer name	Sequence (5' → 3')
AT4g35420	qRT DFRL2F	CAGAGATCCAGGAAATGAGAAGAAAC
	qRT DFRL2R	AAGCACCGGAGAAGCAGTATGGAA
Actin2	Actin2-RT-FW	CCAGAAGGATGCATATGTTGGTGA
	Actin2-RT-RW	GAGGAGCCTCGGTAAGAAGA

Genotyping

Gene	Primer name	Sequence (5' → 3')
AT4g35420	5RP	AAAGAATTCCATTGCGGTATCTCCGCA
	5LP	GAAGAAACTTGCGCACCTATG
	LB1	GCCTTTTCAGAAATGGATAAATAGCCTTGCTTCC

Complementation construct

Gene	Primer name	Sequence (5' → 3')
AT4g35420	DFRL1F	GCAATCCAAAGGGAATCGAAA
	DFRL1R	CGATTCTGTGTTTACGAATGCTG

In situ hybridization

Gene	Primer name	Sequence (5' → 3')
AT4g35420	DFRL RTPCR1F	GATCCAGGAAATGAGAAGAAAC
	T7 DFRL 2R	CATAATACGACTCACTATAGGGTTTCTCAAACCTCTTGGGG
	T7 DFRL 1F	CATAATACGACTCACTATAGGGATCCAGGAAATGAGAAGAAAC
	DFRL RTPCR2R	GTTTCTCAAACCTCTTGGGG
AT1g68540	T7 CCRL6-Sense	CAT AAT ACG ACT CAC TAT AGG ATG TCT GAG TAT TTG GTA ACT GG
	R CCRL6-Sense	TTA GAG CAG ACC CTT CTT CTG AAA AC
	F CCRL6-Anti	ATG TCT GAG TAT TTG GTA ACT GG
	T7 CCRL6-Anti	CAT AAT ACG ACT CAC TAT AGG TTA GAG CAG ACC CTT CTT CTG AAA AC

4.4 Results

4.4.1 Genes involved in pollen cell wall formation are tightly co-regulated

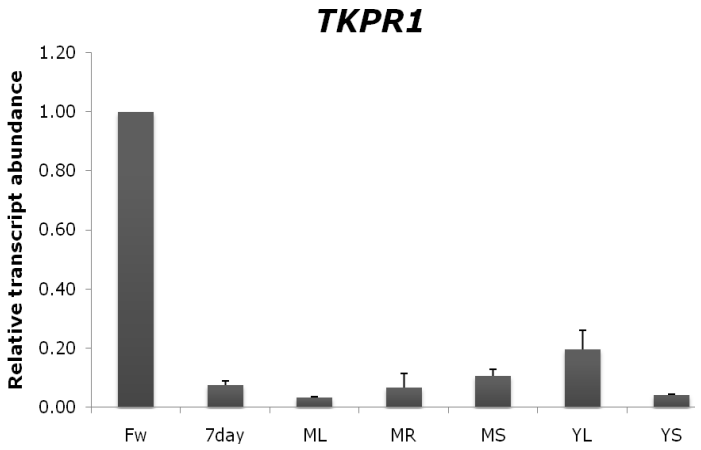
Pollen wall formation requires exquisite and coordinated spatio-temporal regulation of numerous biosynthetic genes by specific transcriptional regulators (Alves-

Ferreira et al., 2007; Yang et al., 2007a). In particular, many genes involved in exine biosynthesis in *Arabidopsis* have been shown to be repressed by the MALE STERILITY1 (MS1) transcription factor in wild type plants, and to be over-expressed in *ms1* mutant (Ito et al., 2007; Yang et al., 2007a). Examination of *Arabidopsis* microarray data showed that, during development of wild type and *ms1* anthers, several uncharacterized genes are co-regulated with genes involved in exine formation. These genes may therefore encode unknown players of sporopollenin biosynthesis (http://bbc.botany.utoronto.ca/ntools/cgi-bin/ntools_expression_angler.cgi;) (Toufighi et al., 2005). Among the genes co-expressed during pollen development, two were annotated as *CHALCONE SYNTHASE-LIKE* and identified in Chapter 3 and Kim et al., (2010) as encoding PKSA and PKSB. PKSA and PKSB catalyze the condensation of fatty acyl-CoA esters produced by ACOS5 with malonyl-CoA to yield tri- and tetraketide α-pyrone compounds as reaction products (Kim et al., 2010). Several other tightly co-regulated genes were annotated as oxido-reductases/dehydrogenases that all have unknown physiological substrates (http://www.arabidopsis.org/). Among them, the proteins encoded by *At4g35420* (*DRL1;* Tang et al., 2009) and At1g68540 (*CCRL6;* Hamberger et al., 2007) contain putative NAD(P)H-binding domains. Both proteins display sequence similarity with two well-characterized plant oxidoreductases: DFR, an enzyme of anthocyanin synthesis (52% and 43% similarity, respectively) (Shirley et al., 1992), and cinnamoyl-CoA reductase (CCR) that is involved in lignin biosynthesis (53% similarity for both At4g35420 and At1g68540 encoded proteins) (Lacombe et al., 1997). These plant enzymes belong to a superfamily whose members are also encountered in microbial and mammalian kingdoms and share a conserved N-terminal sequence that is likely involved in the interactions with NAD(P)H (Baker and Blasco, 1992; Lacombe et al., 1997). *DRL1* has been shown to be required for male fertility (Tang et al., 2009), but neither the *DRL1* expression pattern nor its exact role in pollen wall formation have been described in detail.

4.4.2 At4g35420 and At1g68540 expression profiles during flower development

I explored publicly available microarray databases, such as Genevestigator (https://www.genevestigator.com/gv/index.jsp) (Hruz et al., 2008) and the *Arabidopsis* eFP browser (http://bbc.botany.utoronto.ca/efp/cgi-bin/efpWeb.cgi) (Winter et al., 2007), to determine the tissue expression patterns of At4g35420 (*DRL1*) and At1g68540 (*CCRL6*). The results of this search indicated that both were preferentially expressed in young flower buds, in accordance with their high coregulation scores, similar to the expression patterns reported for *PKSA*, *PKSB*, and *ACOS5* (de Azevedo Souza et al., 2009; Kim et al., 2010). To facilitate their designation in the following paragraphs, genes corresponding to At4g35420 (*DRL1*) and At1g68540 (*CCRL6*) were named *TETRAKETIDE α-PYRONE REDUCTASE1* (*TKPR1*) and *TKPR2*, respectively, in anticipation of the enzymatic activities described later for the two corresponding proteins.

Measurements of relative mRNA abundance by quantitative RT-PCR in RNA preparations from various organs confirmed the flower-specific expression of *TKPR1* and *TKPR2* in contrast with At1g25460, a close homolog (Figure 4.2) that displayed a strikingly different expression pattern.



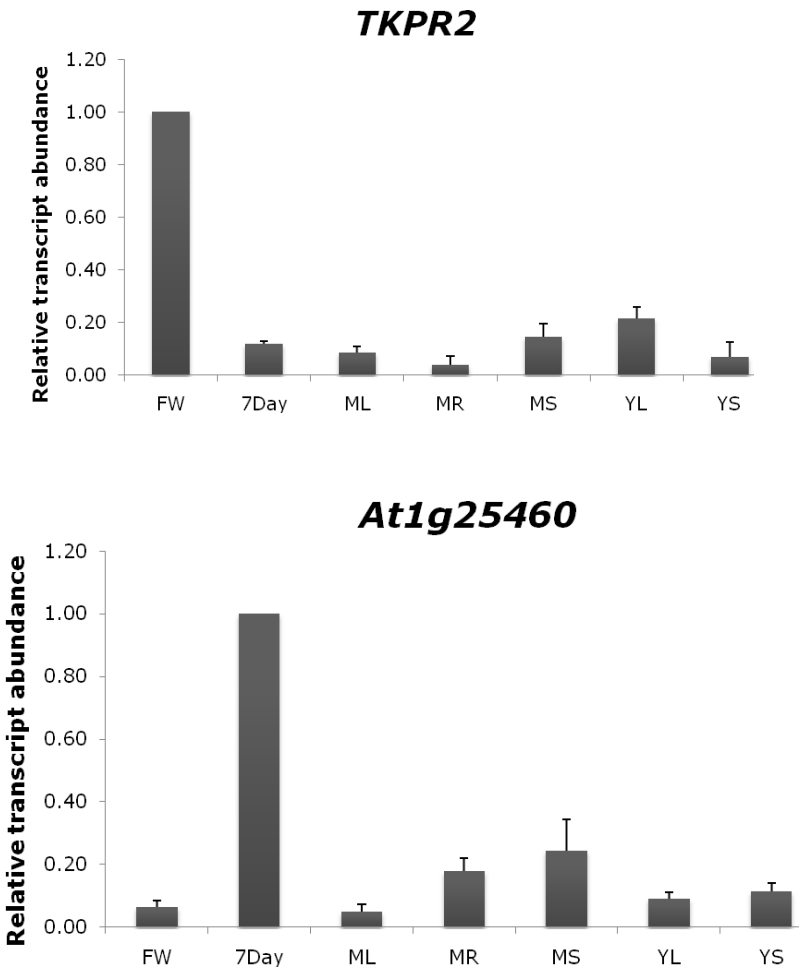


Figure 4.2 Developmental expression profile of *TKPR1, TKPR2 and At1g25460*.

Quantitative RT-PCR analysis of relative *TKPR1, TKPR2 and* At1g25460 expression levels in various *Arabidopsis* organs. Expression was calculated using the ΔΔCT method and is represented relative to the organ with the highest level of expression (TKPR1 and TKPR2; flowers, At1g25460; 7day), set at 1.0. Actin 2 (At3g18780) was used reference as a reference gene. Bars represent standard deviations from the means of triplicate determinations. 7Day, 7-day old seedlings; Fw, flower; ML, mature leaf; MR, mature root; MS, mature stem; YL, young leaf; YS, young stem (Grienenberger et al., 2010).

To determine the precise sites of *TKPR1* and *TKPR2* expression in flower tissues, *in situ* hybridization experiments were performed and showed the specific accumulation of transcripts in the anthers (Figure 4.3). *In situ* hybridization experiments at different

stages of anther development demonstrated the tapetum-specific expression of both *TKPR* genes. The two genes displayed similar but distinct temporal expression patterns. The highest hybridization signal for both was found at stage 7 of anther development (Figure 4.3); however, the *TKPR1* expression pattern was broader over developmental time. The specificity of the *TKPR1*- and *TKPR2*-derived riboprobes used for *in situ* hybridization analysis was demonstrated by DNA gel blotting, showing that both probes hybridized specifically with the target templates, with no detectable cross-hybridization (Figure 4.4).

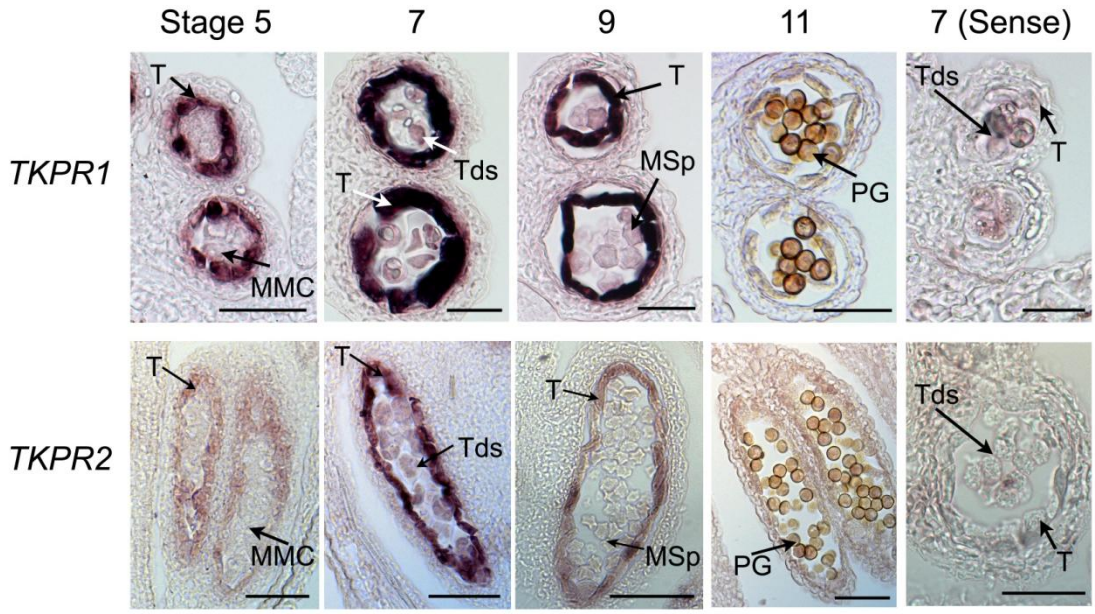


Figure 4.3 Transient tapetum-specific expression of *TKPR1* and *TKPR2*.

TKPR1 and *TKPR2* mRNA were localized by *in situ* hybridization to sections taken from developing anther locules of wild-type (Col-0) flowers. *TKPR1* and *TKPR2* localization were detected by using a gene-specific antisense probe and control sense probe. Stages of anther development are according to Sanders *et al.* (1999). Dark precipitates indicate hybridization of the probe. Stage 5 locules show hybridisation signal in the tapetum. Stage 7 shows highest hybridisation signal in both of TKPR1 and TKPR2 tapetum as well as tetrads. Tapetum signal weakens by stage 9 and disappears by stage 11. MMC, microspore mother cells; Tds, tetrads; T, tapetum; MSp, microspores; PG, pollen grain. Scale bars=70µm (Grienenberger et al., 2010).

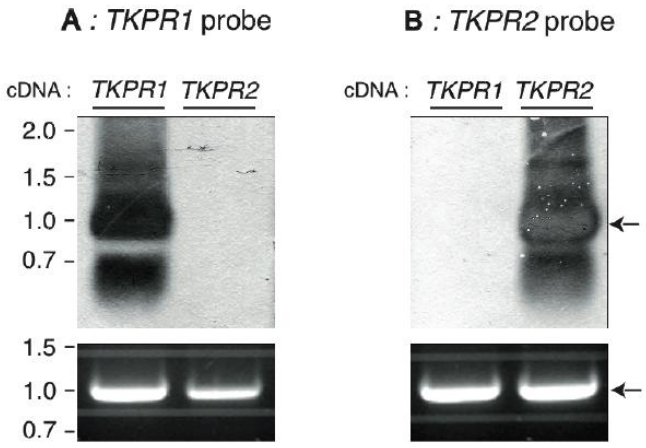


Figure 4.4 Specificity of TKPR nucleotidic probes.

Specificity of probes used for *in situ* hybridization was evaluated by DNA gel blotting. Coding sequences of *TKPR* transcripts were amplified by PCR using gene-specific primers. Amplicons of 0.95 kb predicted size were visualized on Gel Red-stained 0.7% agarose gels (lower panels) before transfer onto nylon membranes. Blots were hybridized separately with digoxigenin-labelled riboprobes corresponding to coding sequences of *TKPR1*(A) or *TKPR2* (B). An indication of size positions in kb is given on the left. Arrows indicate positions of amplicons on gel and blot. Data show the absence of cross-hybridization between the two transcripts (Grienenberger et al., 2010).

4.4.3 Genetic and phenotypic analysis of *tkpr1-1*

T-DNA insertion alleles of *tkpr1-1* in *At4g35420/TKPR1* and *tkpr2-1* in At1g68540/*TKPR2* were obtained from public collections (Alonso et al., 2003) and homozygous populations were generated. Since only *tkpr1-1* homozygous line showed a sterile phenotype with siliques devoid of seeds in initial phenotypic examination, further genetic and phenotypic analysis focused on *tkpr1-1*. There were no other obvious morphological differences between *tkpr2-1 and* wild-type (Col-0). The location of the T-DNA insertion in the fourth exon of *TKPR1* gene was verified (Figure 4.5) and *TKPR1* expression in *tkpr1-1* mutant plants assayed by RT-PCR, using template cDNA derived from both wild type and mutant flowers. This analysis suggested that *tkpr1-1* is a null allele of *TKPR1* (Figure 4.5). I allowed F1 heterozygote plants derived from pollination

with wild-type pollen to self-pollinate and analyzed the resulting F2 population for co-segregation of the male sterile phenotype with *tkpr1-1*. The results showed that the mutant phenotype was inherited in a Mendelian fashion, with one quarter of the F2 progeny displaying complete male sterility (χ^2 = 4.596; p>0.1; n=89), showing that the mutant phenotype is caused by a mutation at a single locus. In the F2 population generated from this cross, the male sterile phenotype co-segregated with *tkpr1-1* (19/89 *tkpr1-1 TKPR1* homozygotes male sterile, 39/89 *tkpr1-1* heterozygotes and 31/89 wild type (Col-0), strongly suggesting that the male sterile phenotype, and complete block in pollen formation in the *tkpr1-1* mutant line is caused by loss of function of the *TKPR1* gene.

Examination of *tkpr1-1* mutant plants (Figure 4.6) revealed anthers devoid of visible pollen, and it was completely male sterile, with no seeds recovered from siliques derived from mutant plants. While no pollen was ever observed in plants homozygous for the *tkpr1-1* allele, *tkpr1-1* flowers were female fertile when pollinated with wild-type pollen. In addition to this male sterile phenotype, mature *tkpr1-1* plants were consistently shorter than wild-type (Col-0) plants and had smaller leaves (Figure 4.6).

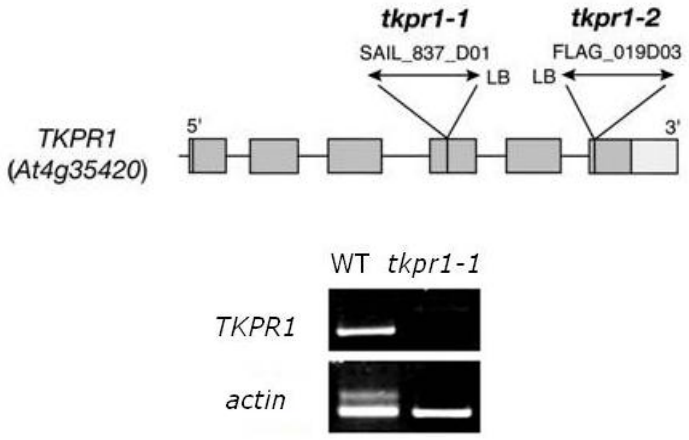


Figure 4.5 Molecular characterization of *tkpr1* insertion alleles and effects on gene expression.

Location of the T-DNA insertion in the fourth exon of *TKPR1* in SAIL_837_D01 is shown above. The cartoon for gene model was drawn base on TAIR (The Arabidopsis Information Researce) database search (http://www.arabidopsis.org). The other allele,

tkpr1-2 was identified and characterized by our collaborator, Dr. Michel Legrand's lab. RT-PCR analysis of *TKRP1* expression in wild-type and SAIL_837_D01 (*tkpr1-1*) flowers is shown below. Expression was evaluated using intron-spanning primers on either side of the T-DNA insertion (Grienenberger et al., 2010).

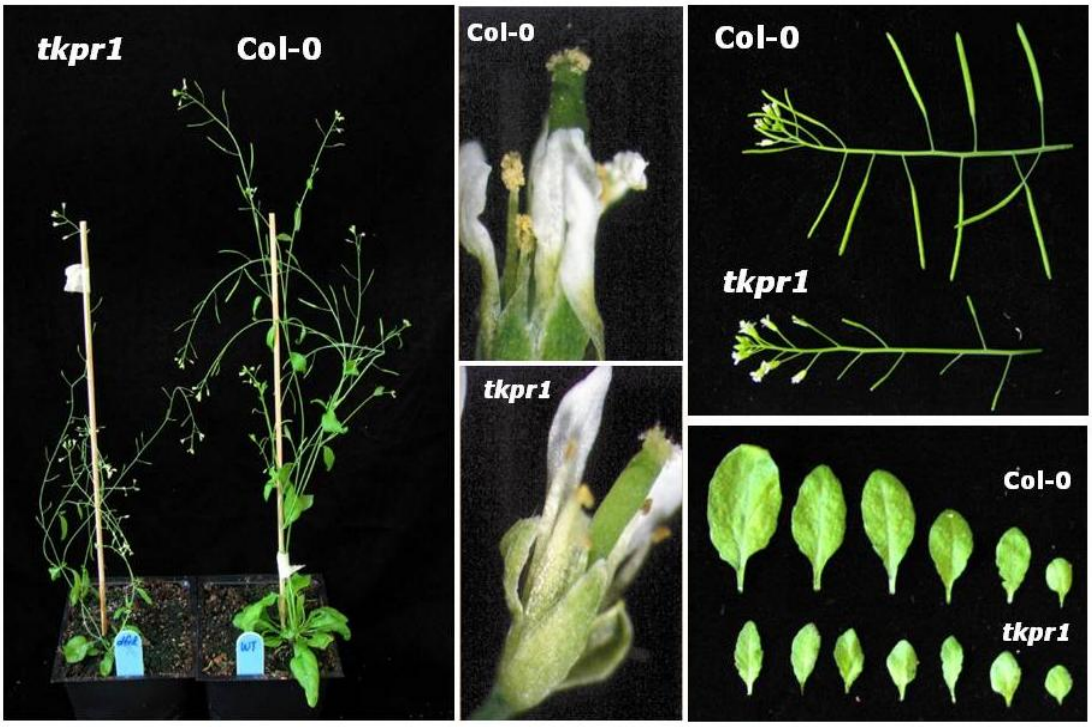


Figure 4.6 Phenotypic characterization of *tkpr1* plant.

Vegetative phenotype of the *tkpr1* mutant is shown. Mutant plants are smaller in size and have smaller rosette leaves than wild-type. Mature wild-type (Col-0) and *tkpr1*flowers are shown. Mutant anthers are devoid of pollen and no pollen grains were observed attached to the stigma. All *tkpr1* siliques are undeveloped and no seeds were recovered from from siliques.

4.4.4 Complementation analysis

To test for the ability of the *TKPR1* gene to complement the male sterile and morphological phenotypes in the *tkpr1-1* background, I PCR-amplified an approximately

3.6-kb region from Col-0 genomic DNA containing 577bp of sequence upstream of the *TKPR1* start codon, and the complete transcribed region (Figure 4.7), introduced the construct into a T-DNA vector, and transformed *TKPR1-1 tkpr1-1* heterozygote plants by *Agrobacterium* mediated transformation. The genotypes of 14 transgenic lines harboring the *TKPR1* transgene were characterized by PCR using primers to differentially detect the presence of the potentially complementing transgene and the T-DNA insertion in the *TKPR1* gene of the 14 T1 plants investigated, three *tkpr1-1*/*tkpr1-1* homozygote T1 lines were fully fertile with normal morphology, showing that the *TKPR1* transgene had complemented the *tkpr1-1* mutation. Taken together, these analysis show that a mutation in the *TKPR1* gene causes the male sterile phenotype observed in the *tkpr1-1* mutant.

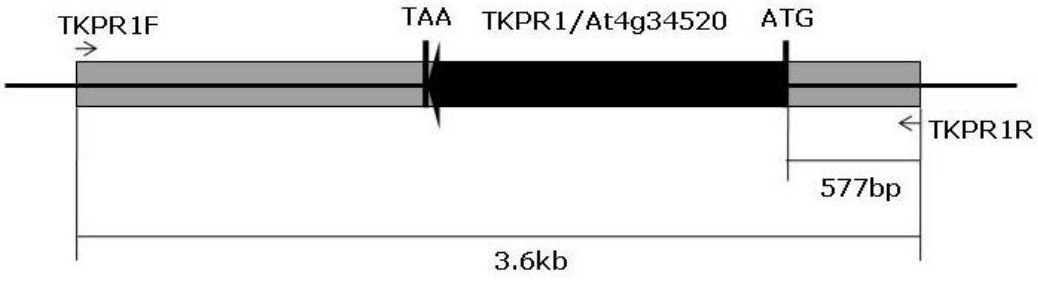


Figure 4.7 Schematic representations of the construct for *tkpr1* complementation test.

Gray shaded box shows *TKPR1* transgene, containing 577bp of promoter sequence and 3'UTR/terminator.

4.4.5 Anther and microspore development in the *tkpr1* mutant

To determine the point at which pollen development was impaired in the *tkpr1* mutant, I examined developing anthers in the *tkpr1-1* mutant background and compared their development to that seen in wild-type anthers (Figure 4.8). Anther, microspore, and pollen development proceeded normally through stage 8. However, development of microspores was arrested in stage 9. *tkpr1-1* anthers and pollen grains were subsequently lost and presumably degraded while anthers devoid of visible pollen grains continued to develop.

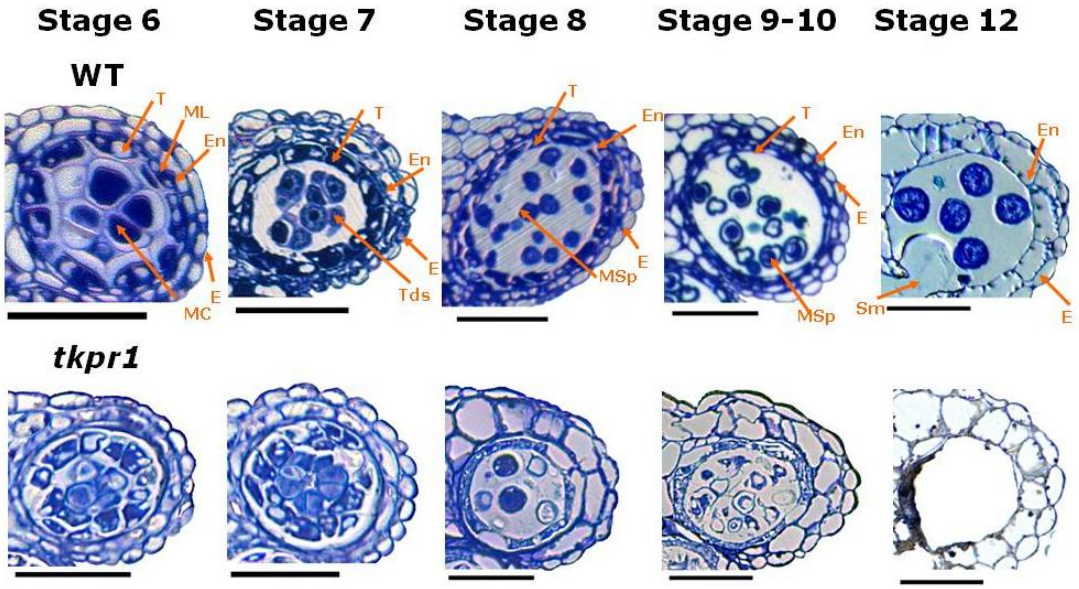


Figure 4.8 Phenotypic characterization of anther and microspore development in wild-type (Col-0) and *tkpr1* flowers.

Anther cross sections (1µm) were taken from developing flowers of the genotypes indicated. Numbers indicate anther developmental stages according to Sanders *et al*. (1999). E, epidermis; En, endothecium; MSp, microspores; Sm, septum; T, tapetum. Scale Bar=40um.

Since a primary defect in the *tkpr1* mutant appeared to be formation of the pollen wall, which was arrested at stage 9 of anther development, *tkpr1-1* anthers and microspores were examined in detail by TEM and compared to wild-type at stage 9 of development (Sanders et al., 1999). Figure 4.9 shows that in wild-type anthers, a thick reticulate exine with distinct baculae and tecta had formed around free, uninucleate microspores. In mutant flowers, however, the tapetum was highly vacuolated (Figure 4.9F) and microspores were profoundly affected with an exine structure that appeared very thin (Figure 4.9D) and completely disorganized without baculae and tecta (Figure 4.9H). In contrast to the cytoplasmicaly dense microspores in wild-type anthers (Figure 4.9C), *tkpr1-1* microspores at this stage were disorganized, largely devoid of cytoplasm, and showed signs of rupture. Finally, *tkpr1-1* locules contained a fibrillar, electron-dense network (Figure 4.9F) that was never observed in wild-type anthers and could represent

unpolymerized sporopollenin precursors. In contrast, cell walls of wild-type and mutant anthers were similar with visible superficial cuticle layer (Figures 4.9I and 4.9J).

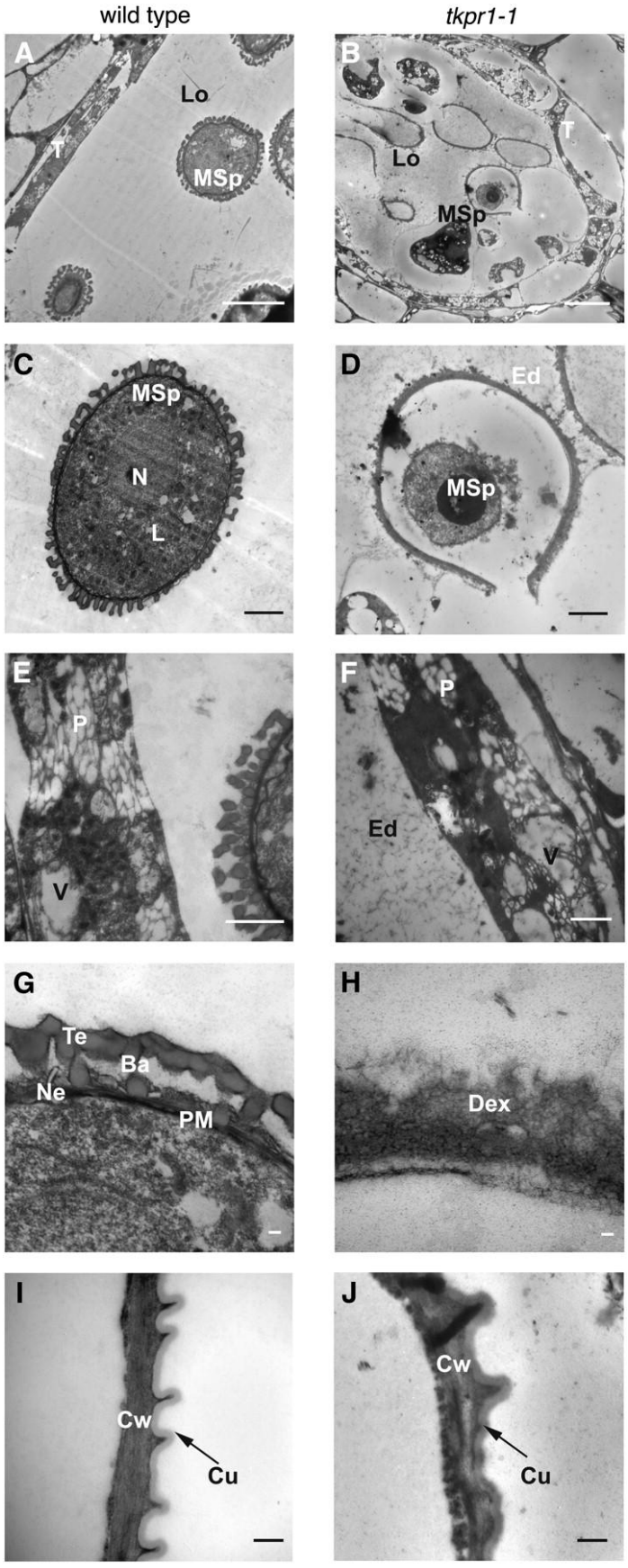

wild type
tkpr1-1
A
Lo
T
MSp
B
T
Lo
MSp
C
MSp
N
L
D
Ed
MSp
E
P
V
F
P
Ed
V
G
Te
Ba
Ne
PM
H
Dex
I
Cw
Cu
J
Cw
Cu

Figure 4.9 Exine formation is impaired in *tkpr1-1* anthers.

Wild type (**[A]**, **[C]**, **[E]**, **[G]**, and **[I]**) or *tkpr1-1* (**[B]**, **[D]**, **[F]**, **[H]**, and **[J]**) plants at stage 9 of anther development (Sanders et al., 1999). Details are shown for anther locule (**[A]** and **[B]**), microspore (**[C]** and **[D]**), tapetum (**[E]** and **[F]**), exine (**[G]**and **[H]**), and anther wall (**[I]** and **[J]**).
Ba, baculae; Cu, cuticle; Cw, cell wall; Dex, defective exine structure; Ed, electron-dense material; L, lipid droplets; Lo, locule; MSp, microspore; N, nucleus; Ne, nexine; P, plastid filled with plastoglobuli; PM, plasma membrane; T, tapetal cell; Te, tectum; V, vacuole containing electron-dense material. Bars = 10 µm in **(A)** and **(B)**, 2 µm in **(C)** to **(F)**, 500 nm in **(I)** and **(J)**, and 100 nm in **(G)** and **(H)** (Grienenberger et al., 2010).

4.4.6 Testing TKPR1 enzymatic function *in vitro*

Sequence analysis of *TKPR1* (At4g35420) suggested that it encodes a protein possessing a consensus NADPH/NADH binding motif (Tang et al., 2009). Therefore, considering the fact that this putative oxido-reductase gene is tightly co-expressed with *ACOS5* and the polyketide synthases *PKSA* and *PKSB* in tapetal cells during pollen development (de Azevedo Souza et al., 2009; Kim et al., 2010), it could act downstream of PKSA and/or PKSB, utilizing a tri- and tetraketide compounds or alkyl phloroglucinol products by condensation of fatty acyl-CoAs with malonyl-CoA (Mizuuchi et al., 2008).

First, the coding region of *TKPR1* was cloned in a vector that introduced a His-tag and was expressed in *E. coli*. TKPR1 recombinant protein was purified by affinity chromatography on Ni-NTA agarose (Figure 4.10).

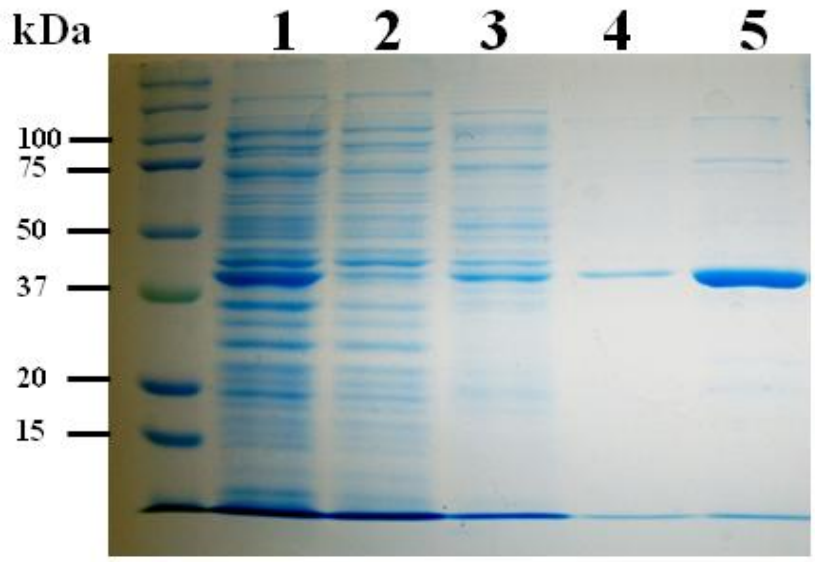


Figure 4.10 Analysis of recombinant protein preparations at different steps of purification.

Bacterial protein extracts were prepared and purified as described in Methods section. Protein preparations were analyzed by electrophoresis on SDS-polyacrylamide gels and Coomassie Blue staining. Purification steps of TKPR1 are illustrated. Lane 1, soluble protein from induced bacteria that was further fractionated; lane 2, Flow through fraction; lane 3, washed fraction; lane 4, eluates, lane 5, concentrated eluates.

One of possible compounds generated from PKSA/B condensation, alkyl phloroglucinol, which contains an alkyl chain bearing a carbonyl function was chemically synthesized by Yan Cao in Dr. Jetter Reinhard's lab and incubated with TKPR1 (Figure 4.11 and Figure 4.12). Assay mixtures were analyzed by GC-MS and the ketoalkyl phloroglucinol compounds gave rise to strong signals (characteristic mass (m/z) at 369 and 509) but no reduction products could be detected (Figure 4.12 and Figure 4.13), thus indicating that the alkyl phloroglucinol compound is not a substrate of TKPR.

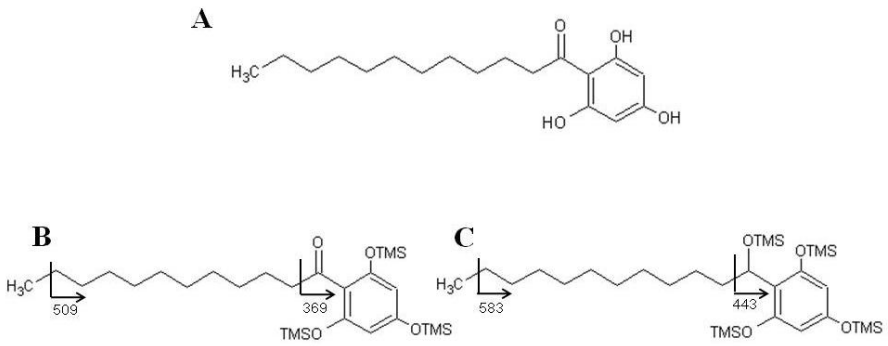


Figure 4.11 Structure of a synthetic alkyl phloroglucinol tested as a TKPR1 substrate and predicted MS fragmentation patterns.

(A) Synthetic alkyl phloroglucinol kindly provided by Yan Cao (UBC). Chemical formula: $C_{18}H_{28}O_4$; Exact Mass: 308.20; Molecular weight: 308.41; m/z 308.20 (100.0%), 309.20 (19.6%), 310.21 (1.9%).
(B) Predicted fragmentation pattern of the alkyl phloroglucinol. After derivatization with TMS, total mass is 524.
(C) Predicted fragmentation pattern of reduced hydroxyl alkyl phloroglucinol. After derivatization with TMS, total mass is 598.

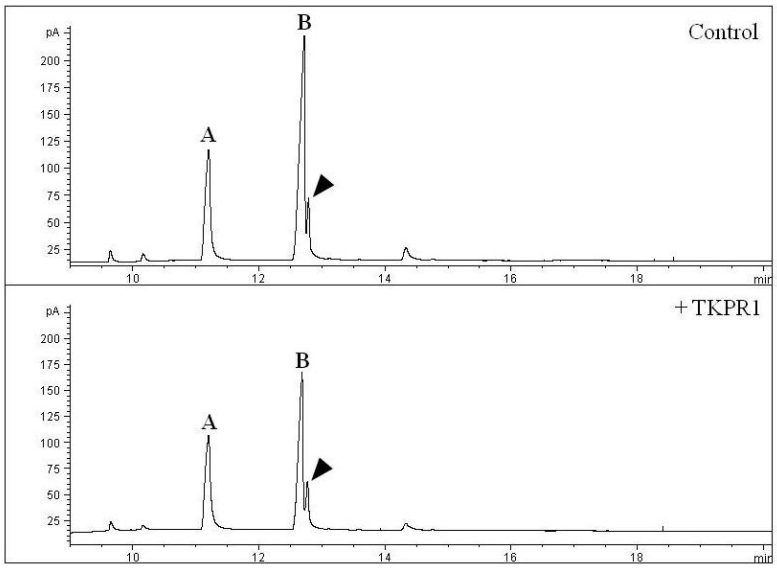


Figure 4.12 GC analysis of reaction products generated by TKPR1 incubation with a synthetic alkyl phloroglucinol.

The synthetic alkyl phloroglucinol shown in Figure 4.11 was incubated in optimized reaction conditions (described in 4.3.9) without recombinant TKPR1 (upper panel) and with TKPR1 (lower panel), and the mixture analyzed by GC-FID. The alkyl phloroglucinol was synthesized together with the major unknown products such as A and B. Arrow heads indicate the synthetic alkyl phloroglucinol.

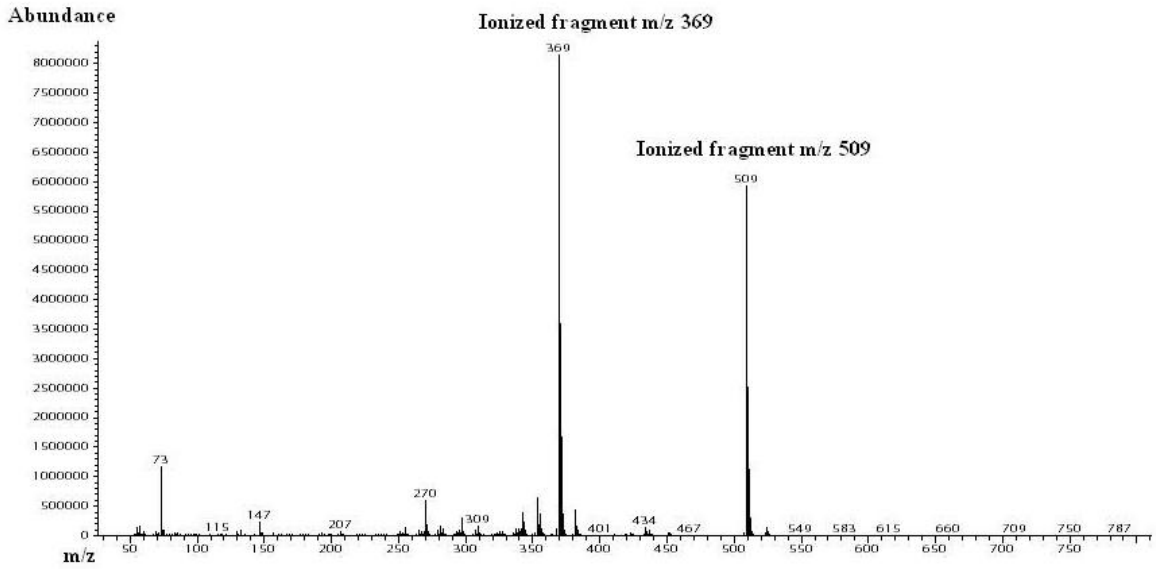


Figure 4.13 Mass spectra of a peak indicated by arrowhead in the GC chromatogram of Figure 4.12.

The m/z values of the GC peak observed in the mass spectrum shown are identical to the predicted values of the unreduced alkyl phloroglucinol substrate added (Figure 4.11 B).

4.4.7 Phylogenetic analysis of *TKPR* genes

TKPR1 and *TKPR2* belong to a gene superfamily that includes members of mammalian, bacterial, and plant origins (Baker et al., 1990; Baker and Blasco, 1992; Lacombe et al., 1997). Plant family members whose functions are unknown have been referred to as *DRL* (Tang et al., 2009) and *CCR-like (CCRL)* (Hamberger, 2007) since the functions of *DFR* in anthocyanin synthesis and *CCR* in lignin biosynthesis had been characterized several years ago (Shirley et al., 1992; Lacombe et al., 1997). Alignment and phylogenetic analysis of *DFR*, *CCR* and related genes from several plant taxa including those from the fully sequenced genomes of *Arabidopsis*, poplar, rice, and the moss *Physcomitrella patens* (Figure 4.14) showed that known and putative *DFR*, *CCR*, and *ANR* genes fall into distinct clades in flowering plants, but do not contain Physcomitrella representatives. Furthermore, *TKPR1* and *TKPR2* define two new clades of plant reductases, formerly annotated as *DRL* and *CCRL* genes, that are distinct from the *DFR*, *ANR*, and *CCR* clades (Figure 4.14). These two new clades both include

homologs from all plant species, ranging from the moss *Physcomitrella patens* to various angiosperms. Strikingly, most of the species examined possess a single homolog in each clade. This indicates that these genes appeared early in evolution and were likely present in a common land plant ancestor (Figure 4.14; see the list of genes in Table 4.2), and may be enzymes ancestral to CCR and DFR. As previously reported, *Arabidopsis ACOS5*, *PKSA*, and *PKSB* genes show similar patterns of phylogenetic conservation (de Azevedo Souza et al., 2009; Kim et al., 2010), and, like *ACOS5*, *PKSA*, and *PKSB* homologs, a *TKPR* homolog is expressed in male organs of rice (*Oryza sativa*). These results suggest conservation of the biosynthetic pathway involving ACOS5, PKS, and TKPR activities that leads to sporopollenin precursors in land plants, and that this pathway was present early in land plants, including the last common ancestor of bryophytes and angiosperms.

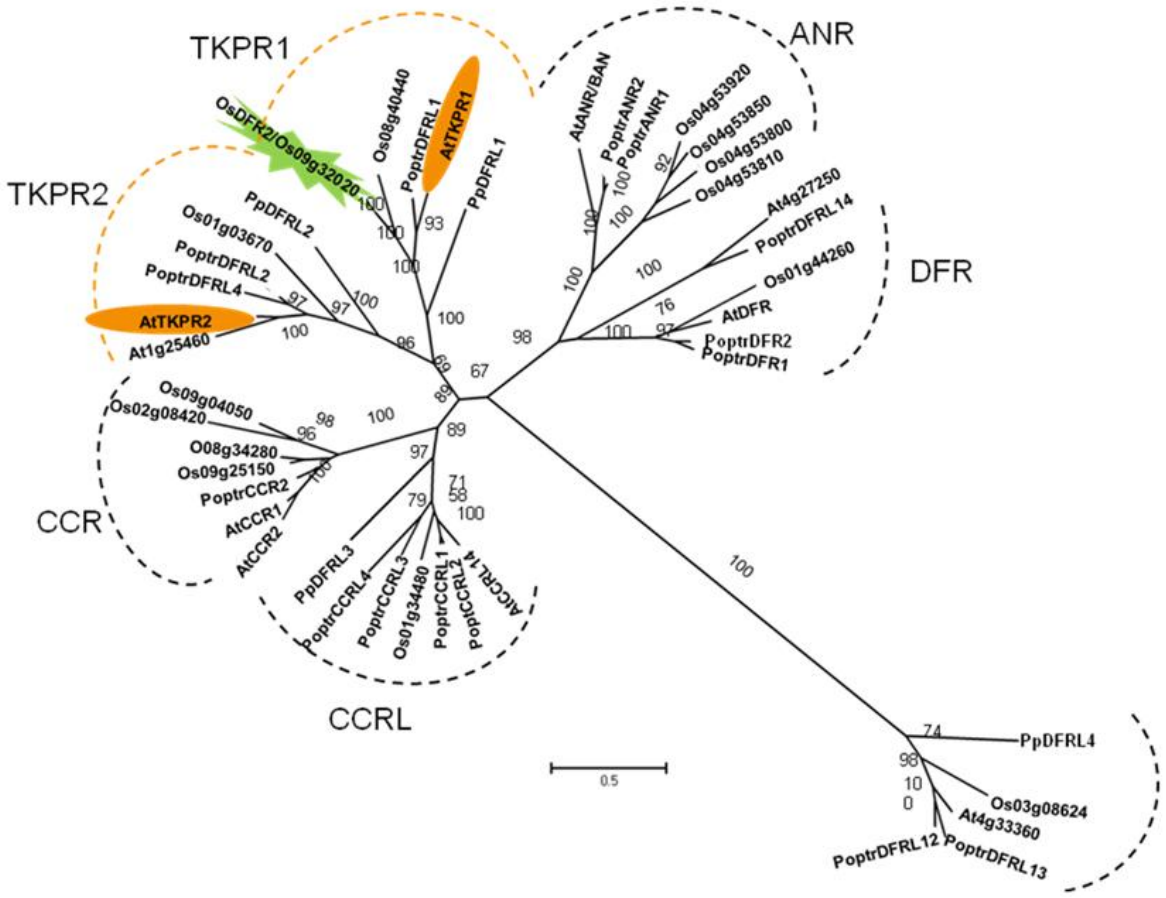


Figure 4.14 Phylogenetic analysis of DFR and DFR related protein sequences in *Arabidopsis*, poplar, rice, and Physcomitrella.

An unrooted maximum-likelihood tree was constructed from aligned sequences using PhyML 2.4.4. Bootstrap values (1000 replicates) above 70% are shown on branches. Clades highlighted in orange contain *Arabidopsis* genes that are co-expressed *in silico* and relative to exine development. ANR, anthocyanin reductase; CCR, cinnamyl CoA reductase; CCRL, CCR-like; DFR, dihydroflavonol reductase; DFRL; DFR-like, TKPR, tetraketide α-pyrone reductase. The protein encoded by a rice gene annotated as OsDFR2 (GenBank Accession AF134807) and expressed in tapetum cells during anther development is indicated by a flash. Protein sequences used in this analysis are given in Appendix C., and the alignment is available in Appendix D. The scale represents 0.5 amino acid changes.

Table 4.2 Putative TKPR orthologs and expression in other species.

Species	Gene name	Accession or gene model	Expression
Arabidopsis thaliana	At*TKPR1*	At4g35420	Tapetum[1]
Populus trichocarpa	Poptr*DFRL1*	fgenesh4_pm.C_LG_VIII000606	N/A
Oryza sativa	Os*DFR2*	AF134807; Os09g32020	Tapetum[2]
			Immature panicle[3]
	No name	Os08g40440	Immature panicle[3]
Physcomitrella patens	Pp*DFRL1*	e_gw1.144.123.1 Ppa.26189 (PHYPADRAFT_86351)	Sporophyte[4]
Arabidopsis thaliana	At*TKPR2*	At1g68540	Tapetum, Tetrad[1]
	No name	At1g25460	Young Seed, silique[5]
Populus trichocarpa	Poptr*DFRL2*	fgenesh4_pg.C_LG_VIII001076	N/A
	Poptr*DFRL4*	estExt_fgenesh4_pg.C_LG_X1136	N/A
Oryza sativa	No name	Os01g03670	Tetrad[6]
			Immature panicle[6]
Physcomitrella patens	Pp*DFRL2*	estExt_Genewise1.C_1140128 Ppa.18009 (PHYPADRAFT_215362)	Protonema[6]

[1] This study
[2] Yau et al., Sexual Plant Reprod 2005
[3] http://mpss.udel.edu/rice/
[4] http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?LID=23755&PAGE=1
[5] http://bar.utoronto.ca/
[6] Huang et al., Plant physiol 2009
N/A, no information available

4.5 Discussion

In this chapter, I used reverse genetics and co-expression analysis to identify candidate oxidoreductase genes required for male fertility, exine formation and sporopollenin deposition during anther development. *TKPR1* and *TKPR2* were identified by their tight co-expression with *ACOS5* and are closely related to the known enzymes DFR and CCR, but they have likely different functions in sporopollenin synthesis. Loss of function mutants in *TKPR1* display apparently identical phenotypes of to *acos5* with respect to pollen development, with arrest of pollen maturation after release from tetrads

and apparent lack of exine formation (de Azevedo Souza et al., 2009; Figure 4.8), whereas *tkpr2* is fertile and produced pollen grains exhibiting only slightly modified exine patterns (data not shown). *In situ* hybridization data in my study reveals that the two genes share a highly similar transient, tapetum-localized expression over the course of anther development. However, *TKPR1* expression is initiated at an earlier stage of anther development (Figure 4.3), suggesting that TKPR1-catalized sporopollenin monomer synthesis is essential at an early step of exine formation, possibly for initiating the reticulated exine pattern.

As shown in phylogenetic analysis (Figure 4.14), TKPR1 lies within a clade of NADPH-dependent reductases similar to proteins encoding CINNAMYL-CoA REDUCTASE (CCR) and DIHYDROFLAVONOL 4-REDUCTASE (DFR), well known enzymes in lignin monomer and flavonoid metabolism. The close structural and phylogenetic relationship between plant *CCR* and *DFR* genes has been previously noted (Lacombe et al., 1997; Yau et al., 2005), and the close relationship of *DFR* to the *Arabidopsis BANYLUS*/*ANTHOCYANIDIN REDUCTASE* (*BAN*/*ANR*) gene, encoding a reductase involved in condensed tannin biosynthesis (Xie et al., 2003) and has also been noted (Devic et al., 1999). First TKPR1 function could be analagous to that of DFR, reducing a hypothetical ketone group generated from a condensing reaction catalyzed by a co-expressed PKSA/B (Kim et al., 2010) to a corresponding hydroxyl group. This reaction would be consistent with presence of oxygenated aliphatic moieties in sporopollenin, and the extensive ether and ester bonds in the polymer that would rely on polyhydroxylated sporopollenin monomers (Ahlers et al., 2000; Ahlers et al., 2003; Morant et al., 2007). Alternatively, based on the known substrates of CCR and DFR enzymes, TKPR1 could act in a manner analogous to CCR, by reducing a fatty-acyl CoA ester generated by ACOS5 to the aldehyde, or possibly on to the primary alcohol, in a manner analogous to CER4, and *Arabidopsis* primary alcohol forming fatty acyl-CoA reductase involved in cuticular wax biosynthesis (Rowland et al., 2006). Finally, TKPR1 could have a true DFR activity and use dihydroflavanols, such as taxifolin, as substrates to generate leucoanthocyanin products.

In this chapter, I explored one possibility by which TKPR1 could participate in downstream of PKSA/B of the sporopollenin biosynthesis pathway. TKPR1 was tested

for activity against a synthetic alkyl phloroglucinol containing an alkyl chain bearing a ketone group, which was reported as one of the products generated by PKSA/B *in vitro* (Mizuuchi et al., 2008). No reduction products could be detected by GC-MS (Figure 4.12). This result indicates that TKPR1 may act on another substrate such as the tri- or/and tetraketide α-pyrones generated by PKSA and PKSB. Additionally to test the alternative possibilities that TKPR catalyzes the reduction of acyl-CoA esters to the aldehyde or alcohol, or reduction of taxifolin (dihydroquercetin) to leucoanthocyanidin, I carried out preliminary assays of recombinant TKPR against a set of cinnamoyl CoAs, fatty acyl CoAs, and taxifolin. I employed a spectrophotometric assay to monitor consumption of NADPH and used recombinant *Arabidopsis* CCR1 as a positive control for reduction of hydroxycinnamoyl-CoAs. While some TKPR reductase activity, judged by NADPH consumption, was observed for both cinnamoyl CoAs and fatty acyl CoAs, it was difficult to draw concrete conclusions about the biological relevance of these results since they could not be compared quantitatively to activity against potential optimal TKPR substrates (eg tetraketide α-pyrones). It would be beneficial to repeat these assays to determine what products are made, test enzyme kinetic parameters against these substrates, and compare TKPR result to results from both the *Arabidopsis* DFR and CCR1 enzymes as positive controls.

Recently biochemical data for two proteins, which were reported by our collaborator, showed that TKPR1 and TKPR2 catalyze the reduction of the ketone group on the alkyl chain of tetraketide α-pyrones, generated *in vitro* by the sequential reactions of ACOS5 and PKSA/B, to secondary alcohols while leaving the lactone ring unaffected (Grienenberger et al., 2010). These data provide evidence that the *in vivo* functions of these enzymes are likely to be tetraketide α-pyrone reductases and that TKPR activity creates a new alcohol function on putative alkyl α-pyrone sporopollenin precursors, whose biosynthetic origin is thus distinct from those of the hydroxyl groups introduced in ω- and in-chain positions of fatty acids by CYP450 enzymes in generating such precursors (Morant et al., 2007; Dobritsa et al., 2009). However, the two reductases do not fulfill exactly the same functions *in vivo*. One evidence of different functions comes from the distinct subcellular localization of TKPR1 and TKPR2 analyzed by transient expression of GFP fusion proteins (Grienenberger et al., 2010). Most TKPR1 is localized

to the ER, similar to PKSA and PKSB (Kim et al., 2010), while TKPR2 is primarily cytosolic and not associated with ER. This ER-localization of TKPR1 suggests that it may form a metabolon composed of CYP703A2, PKS A/B and TKPR1 to yield polyhydroxylated α-pyrones in the ER. Cytosolic TKPR2 would act on different substrates in the cytoplasm, but likely one of similar structure containing a ketone group.

Putative rice, poplar, and *Physcomitrella TKPR1* orthologs are present in the *TKPR1* clade, and the rice *DFR2*/Os09g32020 gene within this clade is transiently expressed in rice anther tapetum cells during microsporogenesis (Yau et al., 2005). While Yau *et al.*, (2005) speculated that the rice *DFR2*/Os09g32020 gene may be required for flavonoid biosynthesis during pollen development, results in this study together with our collaborator's data suggest instead that *TKPR* and its orthologs encode enzymes in a pathway for sporopollenin aliphatic monomer biosynthesis. Data in support of this hypothesis are: 1) the conservation of *TKPR* genes in land plants including *Physcomitrella*, 2) the tight co-expression of *TKPR* with *MS2*, *CYP703A2*, *ACOS5* and *PKSA/B*, all required for sporopollenin biosynthesis, 3) the highly similar loss of function mutant phenotypes and transient tapetum localized expression patterns of the *Arabidopsis ACOS5* and *TKPR1* genes, and 4) the ability of ACOS5, PKSA/B, and TKPR1/2 to catalyze three sequential reactions leading from medium chain fatty acids to reduced alkyl α-pyrones *in vitro*.

Of the *Arabidopsis* genes known or inferred to be required for sporopollenin biosynthesis, *CYP703A2*, *ACOS5*, *PKSA/B* and *TKPR1/2* are conserved in land plant lineages including *Physcomitrella* but are absent in *Chlamydomonas* (Morant et al., 2007; de Azevedo Souza et al., 2009; Kim et al., 2010). This provides strong support for an ancient origin of the biochemical pathway(s) defined by these genes, which probably arose early in land plant evolution and was likely shared by a common ancestor of bryophytes and angiosperms. Thus, acquisition of sporopollenin biosynthesis was likely a key land plant innovation essential for protection of haploid spores from desiccation, UV irradiation, and other stresses of the terrestrial environment and its evolution likely pre-dated vascular system development (Bowman et al., 2007). Further definition of the pathway defined by *ACOS5*, *PKSA/B* and *TKPR1/2* at the biochemical level should not only reveal details regarding the nature of sporopollenin monomeric constituents and

sporopollenin structure, but also shed light on the evolution of the diversity of enzymes that generate the diversity of polyether and polyester polymers now found in plants.

Chapter 5. The introduction of sporopollenin biosynthesis enzymes CYP703A2, ACOS5, PKSA and TKPR1 into the yeast *Saccharomyces cerevisiae* in different combinations

5.1 Summary

Sporopollenin is a poorly characterized mixed aliphatic and aromatic polymer with ester and ether linkages found in the pollen exine layer. Recently, our studies including my work reported in Chapters 2, 3, and 4 have shown that polyhydroxylated α-pyrone polyketide compounds, generated by the sequential action of *Arabidopsis thaliana* ACOS5, PKSA/B and TKPR1/2 are potential and previously unknown sporopollenin precursors. The yeast *Saccharomyces cerevisiae* offers a useful system for expression of heterologous genes, including cytochrome P450 (*CYP*) genes, and using sets of dual expression vectors with different selectable makers it is possible to express various combinations of genes in a single stain. This system allows, for example the reconstruction of plant natural product biosynthetic pathways in yeast (Ro et al., 2004).

In this Chapter, I used the yeast expression system to express different combinations of sporopollenin biosynthetic enzymes to test different models of potential sporopollenin biosynthetic pathways *in vivo*. Because hydroxy fatty acids and acyl-CoAs are the preferred substrates for ACOS5 and PKSA/B *in vitro*, CYP703A2 was first introduced alone to test the ability of CYP703A2 to generate hydroxy fatty acids. Subsequently, a CYP703A2/ACOS5 dual expresser was generated and PKSA and TKPR1 activities added alone or in combination with CYP703A2/ACOS5. Product identification was achieved by GC-FID and GC-MS. My results indicate that heterologous expression of PKSA can catalyze condensation of endogenous fatty acyl-CoAs with malonyl-CoA to generate α-pyrone triketides, while CYP703A2 and ACOS5 could not access endogenous substrates.

5.2 Introduction

Sporopollenin, a heterogeneous biopolymer found in the pollen exine layer and in spores of mosses, is composed of polyhydroxylated unbranched aliphatic and phenolic constituents as the main monomeric units (Scott et al., 2004). These units are covalently coupled by ether linkages in order to provide the characteristic high resistance to chemical degradation (Scott et al., 2004). In a previous study, an exine defective, partially male sterile mutant of *Arabidopsis* CYP703A2, a member of cytochrome P450 super gene family was reported (Morant et al., 2007). CYP703A2 hydroxylated medium-chain saturated fatty acids to the corresponding mono-hydroxyl fatty acids, with a preferential hydroxylation of lauric acid (C12:0) at the C-7 position *in vitro* and *in vivo* (Morant et al., 2007). A model for the role of CYP703A2 was proposed to explain its involvement of sporopollenin formation, suggesting that it is involved in generating hydroxy fatty acid components of mixed fatty acyl-phenolic sporopollenin polymer (Morant et al., 2007).

Recently, we showed that an evolutionariy conserved metabolic pathway involving fatty acyl-CoA condensation and extension followed by reduction, catalyzed by POLYKETIDE SYNTHASE A and B (PKSA/B) and and TETRAKETIDE α-PYRONE REDUCTASE (TKPR) enzymes respectively, leads to polyketide α-pyrone compounds that are essential sporopollenin precursors (Grienenberger et al., 2010; Kim et al., 2010). Based on these results, we proposed a model for sporopollenin precursor biosynthesis that includes different possibilities for the sequential actions of the enzymes involved in α-pyrone sporopollenin unit synthesis. After CoA ester formation by ACOS5, PKSs can catalyze acyl-CoA condensation with malonyl CoA and extension to generate tri- and tetraketide α-pyrones (Figure 5.1, route 1). CYP450 hydroxylases involved in sporopollenin biosynthesis have been shown to be active on free fatty acids (Morant et al., 2007; Dobritsa et al., 2009), thus implicating putative thioesterase and CoA ester regeneration steps upstream and downstream of the hydroxylation step as shown on the left side of the scheme (Figure 5.1, route A). We have shown that hydroxy fatty acyl-CoAs are the most efficient substrates for PKSA and PKSB, consistent with an ER localization (Kim et al., 2010), and can yield various hydroxy polyketides (route 2). These latter compounds might also arise from the hydroxylation of the alkyl α-pyrones by CYP450s (route B), but this possibility remains to be examined experimentally. Finally,

reduction of the carbonyl function of the hydroxy tetraketide α-pyrones by TKPRs gives rise to an additional hydroxyl function in the alkyl chain of the polyketide products, generating more highly hydroxylated polyketides that are proposed constituents of a sporopollenin polymer highly cross-linked by ester and ether bonds. The association of TKPR1 with ER, where hydroxylation by P450s takes place, may be important in this respect.

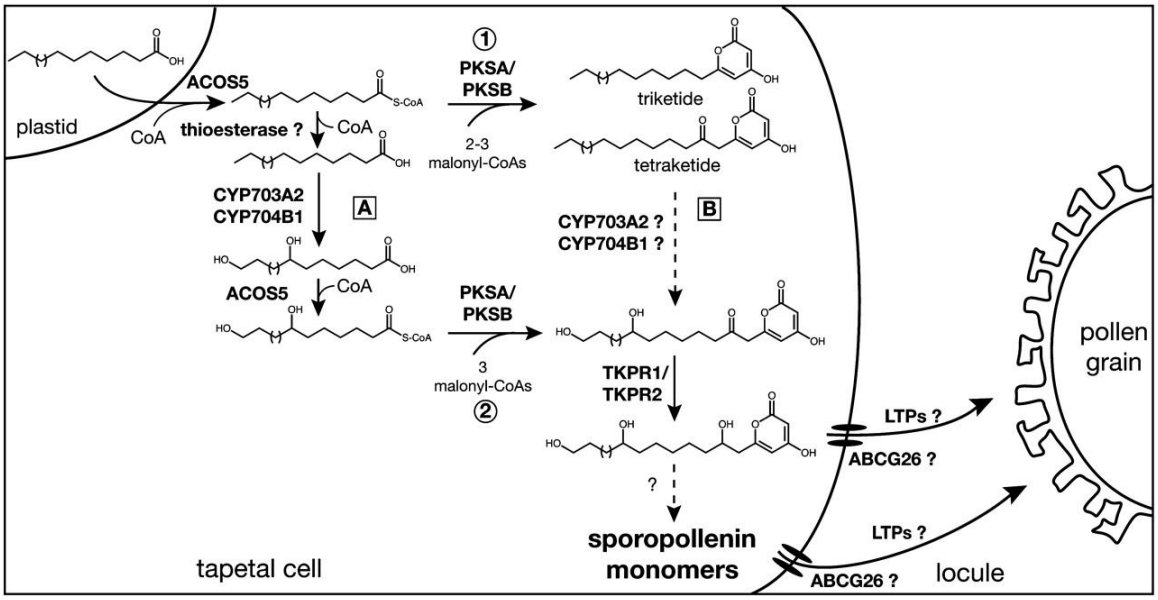


Figure 5.1 A putative scheme of sporopollenin biosynthesis (Grienenberger et al., 2010).

Fatty acid biosynthesis is an essential metabolic process. Unlike plants, which use the plastid as their fatty acid biosynthetic site, the yeast *Saccharomyces cerevisiae* synthesizes fatty acids predominantly in the cytoplasm. Fatty acid biosynthesis in yeast is carried out by a multifunctional type I protein composed of two polypeptide chains (α and β) (Lynen, 1980; Schweizer and Hofmann, 2004), whereas in plant plastids and bacteria, it is catalyzed by a type II fatty acid synthase (FAS), which is dissociable and typically consists of more than 8 separate proteins. The major fatty acids produced in yeast are

palmitoleic acid (C16:1), palmitic acid (C16:0), oleic acid (C18:1) and stearic acid (C18:0), which are required to generate lipids for biological membranes (Dittrich et al., 1998).

In this study, the metabolic changes in yeast *Saccharomyces cerevisiae* containing CYP703A2, ACOS5, PKSA and TKPR1 in different combinations were examined by GC-MS.

5.3 Materials and methods

5.3.1 Construction of CYP703A2, ACOS5, PKSA and TKPR1 in pESC vector

Coding regions of *CYP703A2, ACOS5*, *PKSA* and *TKPR1* were amplified by PCR using Phusion High Fidelity DNA polymerase (New England Biolabs). Each gene-specific primer was designed with an appropriate restriction enzyme site and a yeast consensus sequence for proper initiation of translation (Cigan and Donahue, 1987; Romanos et al., 1992). The sequence for primers is listed in Table 5-1. The PCR fragements were subcloned into pGEM-T vector for sequence verification as well as manipulation. After the digestion with a restriction enzyme, the products were subcloned into pESC vector containing one of four different auxotrophic selectable markers (*HIS3*, *TRP1*, *LEU2*, or *URA3*) for expression in yeast.

5.3.2 Yeast strains, culture and heterologous expression in yeast

Various strains of the yeast *Saccharomyces cerevisiae* were used. Untransformed YPH499 strain (*MAT*a, *ura3-52, lys2-801, ade2-101, trp1-*Δ*63, his3-*Δ*200 leu2-* Δ1) and WAT11 (*MATa*; *ade*2–1; *his*3–11,-15; *leu*2–3,-112; *ura*3–1; *can*^{R}; *cyr*$^{+}$) were maintained in YPAD medium containing 0.08 g/L adenine hemisulfate salt, 10 g/L yeast extract, 10 g/L Bactopeptone, and 20 g/L dextrose. For solid medium, 15 g/L of agar was supplemented. In the WAT11 strain, the yeast reductase was replaced by the ATR1 reductase from *Arabidopsis thaliana*, controlled by the GAL10-CYC1 promoter (Pompon et al., 1996). Transformed strains with pESC were screened and maintained in each amino acid dropout minimal medium. For example dropout synthetic minimal medium

contained 6.7 g/L yeast nitrogen base without amino acids (BD science), 1.3 g/L His-dropout amino acid powder (Sigma), and 20 g/L dextrose (for synthetic dextrose [SD]-His dropout medium) or galatose (for synthetic galactose [SG]-His dropout medium). All the yeast expression vector stocks were independently transformed into yeast strains by the polyethylene glycol-LiAc method (Gietz et al., 1992). Briefly, a fresh yeast culture (OD_{600} = 0.25) was grown in complete YPAD medium for 5 hr. The cells were collected, washed twice with water and resuspended in 1.5 mL of a 0.1 M lithium acetate (LiAc) solution in TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA). For transformation, salmon sperm was added as DNA carrier (100 µg from a10 mg/mL solution in TE after 10 min boiling) to 1 ug of plasmid DNA. 100 µL of competent yeast cells and 250 µL of 50% polyethylene glycol 3500 in LiAc solution in TE buffer were added. The mixture was incubated for 1 hr at 42℃. After centrifugation, the transformed yeast cells were collected and then plated on SD minimum medium. Strains transformed with pESC were grown for 2 to 3 days at 30℃.

For gene induction, transformed yeast strains were subcultured in 10 mL of SD minimum dropout medium overnight. Cell densities were determined at OD_{600} and cell aliquots sufficient to obtain on OD_{600} of 0.4 in 50 mL of SG induction medium calculated. Culture pellets from these aliquiots were resuspended in 1 mL of SG induction medium and inoculated into 50 mL of SG medium. For exogenous fatty acid feeding experiment, 1 mM each fatty acid was added to SG induction medium with 150 µL 100% ethanol. Yeast cells were cultured at 28℃ on a shaker at 200 rpm for 20 hr to 24 hr. The cell pellets were washed once with PBS buffer (140 mM NaCl, 3 mM KCl, 10 mM Na_2HPO_4, and 2 mM KH_2PO_4, pH of 7.4) and centrifuged at 1500 x g for 15 min. The cell pellets were stored at -80°C for further studies.

Time-dependent sampling was performed aseptically during 24 hr by taking out 1 mL aliquots from main culture. Samples were diluted 10x immediately in corresponding culture medium, and then subjected to duplicate absorbance determination in a spectrophotometer at 600 nm. Diluted cell-free medium was used to establish background readings and set zero absorbance levels. Values were averaged and corrected for dilution.

5.3.3 SDS-PAGE and immuno-blotting

Fresh or frozen yeast cell pellets were resuspended in 500 µL breaking buffer (50 mM sodium phosphate, pH 7.4, 1 mM EDTA, 5% glycerol and 1 mM PMSF) and centrifuged at 1500 g for 5 min at 4℃ to pellet cells. After resuspending the cells in a volume of breaking buffer to obtain an OD_{600} of 50-100, an equal volume of 0.4-0.6 mm acid-washed glass beads (Sigma) were added and vortexed four times for 30 sec, followed by 30 sec on ice to lyse the cells. Prepared soluble total protein samples were separated on 10% polyacrylamide gels and either stained with Coomassie blue or transferred to PVDF membrane (GE Healthcare) for immunoblot analysis. The membranes were incubated in blocking buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl, 0.1% Tween-20) supplemented with 10% (w/v) non-fat dried milk powder. Alkaline Phosphatase (AP)-conjugated antibodies used, depending on target proteins, were anti-FLAG (Sigma) and anti-cMyc (Invitrogen), which were diluted at 1:1000 and 1:2000 in blocking solution, respectively. Immuno-detection was performed by adding 200 µL NBT/BCIT (Roche) in 20 mL staining buffer (100 mM Tris-HCl, pH9.5, 100 mM NaCl, 50 mM $MgCl_2$).

5.3.4 Chemical analysis of yeast expression product

Total fatty acids in yeast cell extracts were transmethylated by adding 1 mL methanolic-HCl and incubating at 80℃ for 1hr to 2hr. To obtain organic extracts, 1.5 mL hexane was added followed by vortexing. The organic phases were pooled and evaporated under nitrogen gas. For GC-MS analyses, extraction from reaction mixture, derivatization of organic compounds and running condition of samples in GC-MS are described by Wang *et al.*, (2010), except the following: the oven temperature was programmed for 2 min at 50℃, followed by a 40℃ min^{-1} ramp to 120℃, held at 120℃ for 2 min, increased by 2℃ min^{-1} to 225℃, and held at 320℃ for 10 min.

5.4 Results

5.4.1 CYP703A2 and/or ACOS5 expression in yeast

The cDNAs for *CYP703A2* and *ACOS5* were amplified, cloned into yeast vectors, and expressed to reconstruct the potential entry point of the sporopollenin monomer biosynthesis pathway in yeast cells. To generate *CYP703A2*, *ACOS5*, and *CYP703A2*/*ACOS5* dual-expressing yeast strains, each cDNA was cloned into the pESC-His vector under the control of Gal1 promoter for *ACOS5* and Gal10 promoter for *CYP703A2*, where they were expressed as fusions to the cMyc epitope tag (*ACOS5*) and FLAG epitope tag (*CYP703A2*). Functional expression of the two genes in yeast was verified by immunoblot analysis, using monoclonal anti-FLAG and anti-cMyc antibodies to detect epitope-tagged CYP703A2 and ACOS5, respectively (Figure 5.2). The CYP703A2 recombinant protein was not detected in the cytosolic fraction but rather in cell debris, indicating insoluble status embedded in microsomes, while ACOS5 was detected in both fractions of transformed yeast strains (Figure 5.2). Growth rates of yeast strains cultured in SG induction medium were monitored from initial OD 0.4 for 24 hr. Empty vector control strains increased up to OD_{600} of 5.32±0.09 while the cell growth of *ACOS5* sole-, *CYP703A2* sole- and *CYP703A2/ACOS5* dual-expression strains was repressed, reaching ODs of 2.01±0.05, 2.00±0.08 and 2.00±0.02 at 24 h, respectively.

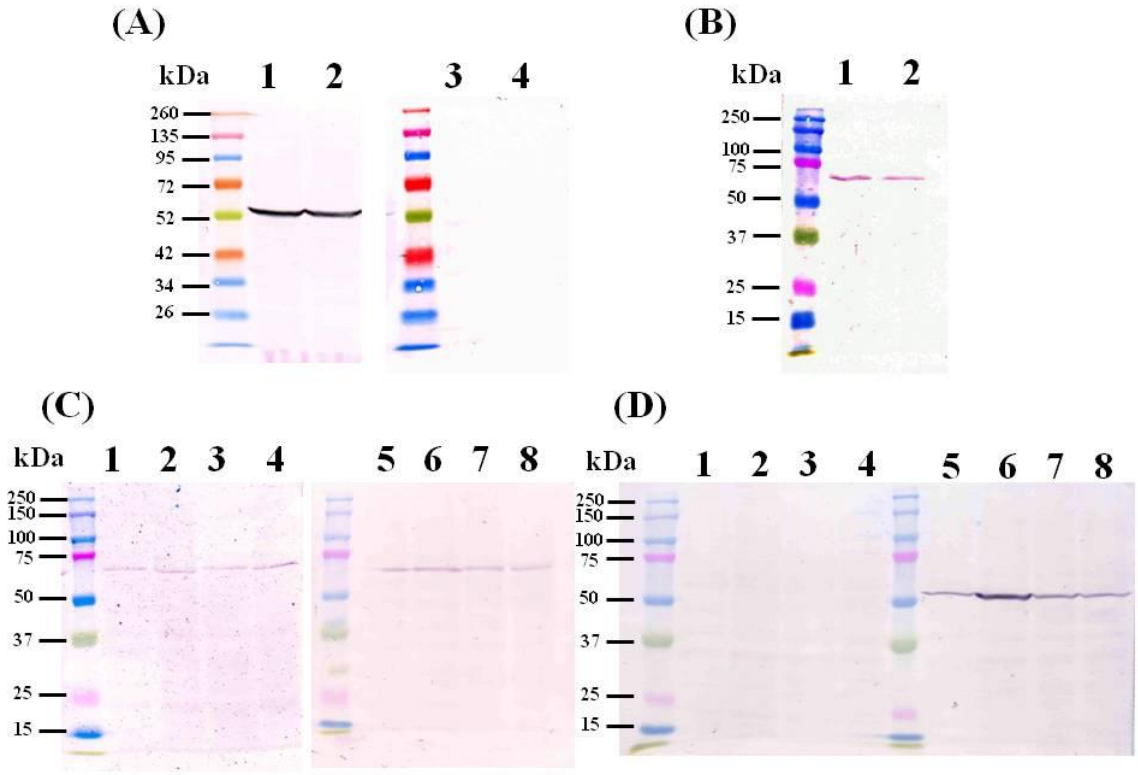
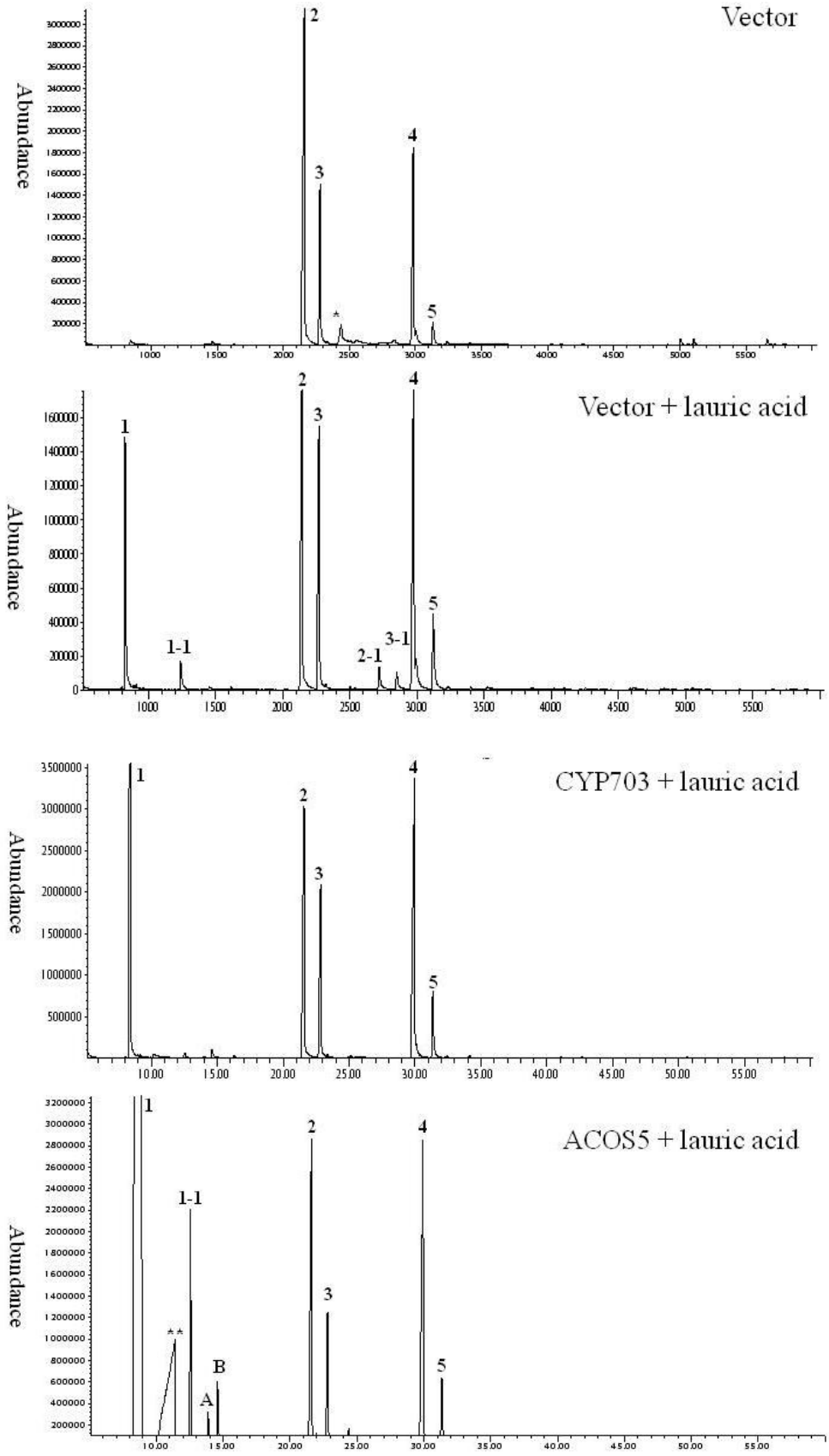


Figure 5.2 Immunoblot detection of CYP703A2 and ACOS5 proteins in transgenic yeast strains.

(A) Immunoblot analysis of proteins from cell debris (lane 1 and 2) and cytosolic fraction (lane 3 and 4) in the *CYP703A2* expresser, reacted with a FLAG monoclonal antibody to detect CYP703A2. Lane 1 and 3, 16 hr culture; lane 2 and 4, 24 hr culture
(B) Immunoblot analysis of proteins from cytosolic fraction in *ACOS5* expresser reacted with a cMyc monoclonal antibody to detect ACOS5. Lane 1, 16 hr culture; lane 2, 24 hr culture
(C) Immunoblot analysis of proteins from cytosolic fraction (lane 1, 2, 3 and 4) and cell debris (lane 5, 6, 7 and 8) in *CYP703A2/ACOS5* dual expresser, reacted with a cMyc monoclonal antibody to detect ACOS5. Lane 1 and 5, 3 hr culture; lane 2 and 6, 9 hr culture; lane 3 and 7, 19 hr culture; lane 4 and 8, 24 hr culture
(D) Immunoblot analysis of proteins from cytosolic fraction (lane 1, 2, 3 and 4) and cell debris (lane 5, 6, 7 and 8) in the *CYP703A2/ACOS5* dual expresser, reacted with a FLAG monoclonal antibody to detect CYP703A2. Lane 1 and 5, 3 hr culture; lane 2 and 6, 9 hr culture; lane 3 and 7, 19 hr culture; lane 4 and 8, 24 hr culture

To assay the fatty acid (FA) metabolic profile in each strain, lipophilic compounds were extracted with hexane and then derivatized. The FA

derivatives were identified by comparing their gas chromatography-mass spectrometry (GC-MS) characteristics with literature data. CYP703A2 preferentially catalyzes hydroxylation of lauric acid (C12) at the C-7 position *in vitro* (Morant et al., 2007). Because in yeast cells medium-chain fatty acids such as C12 FA are expected to be mostly intermediate products maintained as thioester conjugates to the acyl carrier protein (ACP) during fatty acid synthesis, *Arabidopsis* CYP703A2 may not be able to access endogenous medium-chain yeast FAs proposed to be the starting points for the sequential reactions of sporopollenin biosynthesis by Grienenberger *et al.*, (2010). Thus, first I tested the ability of CYP703A2 to use C12 FA (lauric acid) fed exogenously to the growth media. Surprisingly, no new peaks were present in the chromatogram of the *CYP703A2* expression strain compared with the empty vector strain under these conditions (Figure 5.3). The possibility that CYP703A2 might preferentially accept long-chain fatty acids *in vivo* was tested by feeding palmitic acid (C16:0) to the growth media. Whereas C16 FA was accepted as a substrate by CYP703A2 in yeast microsomes with 5 to 10% efficiency to generate hydoxy C16 FA derivatives (Morant et al., 2007), I could not identify hydroxyl C16 FAs. In parallel, the FA composition of *ACOS5* and *CYP703A2/ACOS5* dual-expressing strains was also assayed after feeding C12 FA. In both strains two new peaks; corresponding to myristic acid (C14:0) and tetradecenoate (C14:1) were observed in GC chromatograms (Figure 5.3). These C14 FA derivatives were never observed in the absence of exogenously supplied C12 FA.


Vector
Abundance
2
3
4
5
*
Vector + lauric acid
Abundance
1
1-1
2
3
2-1
3-1
4
5
CYP703 + lauric acid
Abundance
1
2
3
4
5
ACOS5 + lauric acid
Abundance
1
1-1
**
A
B
2
3
4
5

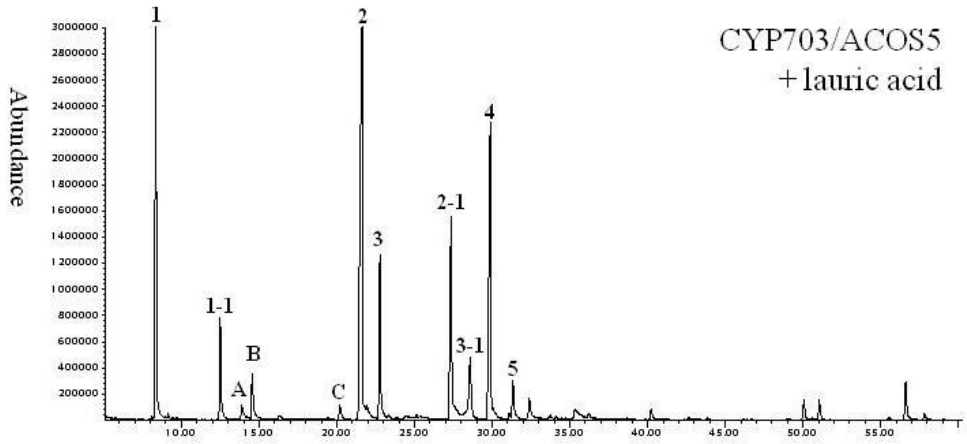


Figure 5.3 GC analysis of fatty acid metabolites produced by vector control and transgenic yeast strains.

GC chromatograms of extracts from different yeast strains are shown. Fatty acid derivatives were converted into methyl esteresters or trimethylsilyl esters prior to GC analysis.
1, dodecanoic acid methyl ester (C12); 1-1, dodecanoic acid trimethylsilyl estertrimethylsilyl ester (C12); 2, 9-hexadecenoic acid methyl ester(C16:1); 2-1, 9-hexadecenoic acid trimethylsilyl estertrimethylsilyl ester (C16:1); 3, hexadecanoic acid methyl estermethyl ester (C16); 3-1, hexadecanoic acid trimethylsilyl ester (C16); 4, 9-octadecenoic acid methyl estermethyl ester (C18:1); 5, octadecanoic acid methyl estermethyl ester (C18); A, 11-tetradecenoic acid methyl ester (C14:1); B, tetradecanoic acid methyl ester (C14); C, tetradecanoic acid trimethylsilyl ester; *, 9-hexadecenoic acid (C16:1); **, dodecanoic acid (C12).

5.4.2 Triple and quadruple expression in yeast

The yeast strain transformed with pESC-HIS::*CYP703A2/ACOS5* was co-transformed together with pESC-Ura::*TKPR1*, Ura::*PKSA* or Ura::*PKSA/TKPR*, to generate three different yeast strains: two triple-expressing strains expressing *CYP703A2/ACOS5/TKPR1* and *CYP703A2/ACOS5/PKSA*, and one quadruple-expressing strains expressing all candidate genes, *CYP703A2/ACOS5/PKSA/TKPR1*. Immunoblot analysis failed to detect PKSA recombinant protein in either triple or quadruple expressers using the anti-cMyc antibody (Figure 5.4). However, FLAG-tagged TKPR1 was detected in the cytosolic fraction as well as in cell debris from both triple- and quadruple-transformed yeast strains (Figure 5.4). The OD_{600} of each strain increased from the initial value of 0.4 up to 5.60±0.09 for empty vector control, 5.71±0.10 for

CYP703A2/ACOS5/TKPR1 expresser, 4.32±0.21 for *CYP703A2/ACOS5/PKSA* expresser and 4.43±0.05 for *CYP703A2/ACOS5/PKSA/TKPR1* expresser, respectively. These data suggest that accumulation of TKPR1 or/and PKSA can rescue the repressed growth rate due to the expression of *CYP703A2* or/and *ACOS5* in yeast cells. Moreover addition of the *PKSA* gene to both *CYP703A2/ACOS5* and *CYP703A2/ACOS5/TKPR1* expressers slightly reduced the growth rates. Therefore, it is likely that PKSA enzyme is present even though I could not detect the fusion tagged version using the anti-cMyc antibody.

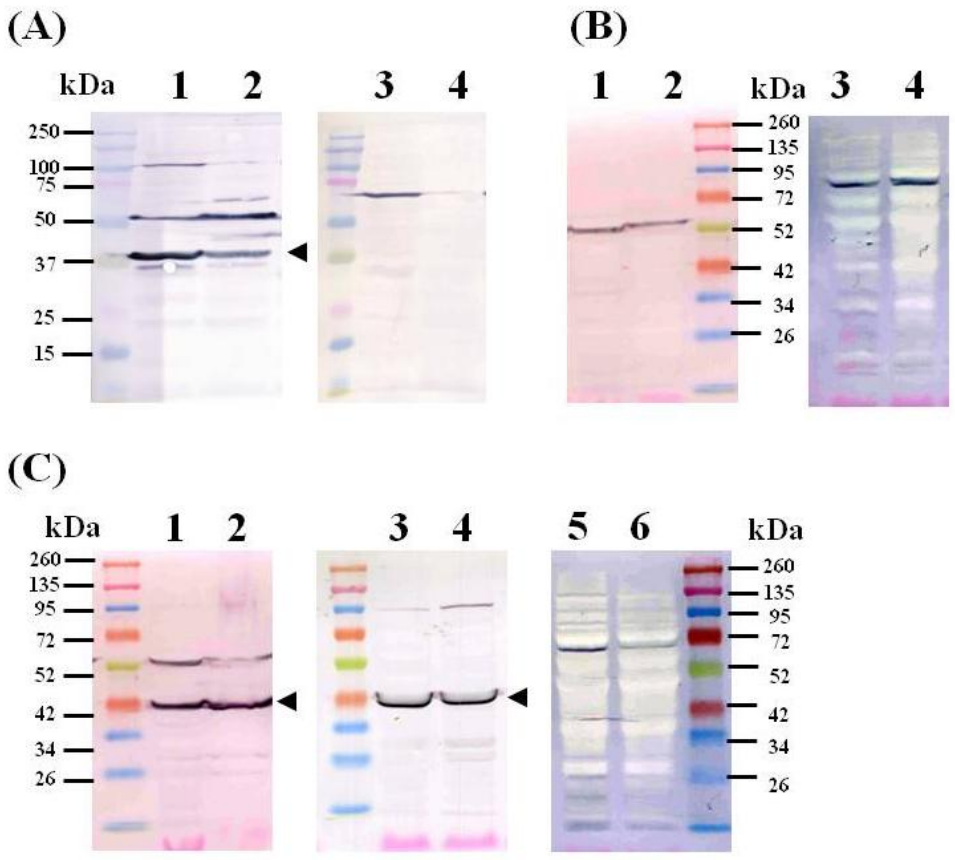


Figure 5.4 Immunoblot detection of CYP703A2, ACOS5 and TKPR1 proteins in transgenic yeast strains.

(A) Immunoblot analysis of proteins from cell debris (lane 1 and 2) and cytosolic fraction (lane 3 and 4) in *CYP703A2/ACOS5/TKPR1* triple expresser, reacted with a FLAG monoclonal antibody to detect CYP703A2 and TKPR1 (lane 1 and 2) as well as cMyc monoclonal antibody to detect ACOS5 (lane 3 and 4).
(B) Immunoblot analysis of proteins from cell debris (lane 1 and 2) and cytosolic fraction

(lane 3 and 4) in *CYP703A2/ACOS5/PKSA* triple expresser, reacted with a FLAG monoclonal antibody to detect CYP703A2 (lane 1 and 2) as well as cMyc monoclonal antibody to detect ACOS5 and PKSA (lane 3 and 4).
(C) Immunoblot analysis of proteins from cell debris (lane 1 and 2) and cytosolic fraction (lane 3, 4, 5 and 6) a in *CYP703A2/ACOS5/PKSA/TKPR1* quadruple expresser, reacted with a FLAG monoclonal antibody to detect CYP703A2 and TKPR1 (lane 1, 2, 3 and 4) as well as cMyc monoclonal antibody to detect to detect ACOS5 and PKSA (lane 5 and 6). Lane 1, 3 and 5, 9 hr culture; lane 2, 4 and 6, 24 hr culture; black arrow head indicates TKPR1.

To explore the potential products generated by the sequential actions of the sporopollenin monomomer biosynthetic enzymes in triple or quadruple expressers *in vivo*, total lipophilic compounds ranging from medium FA derivatives to very long chain FA derivatives were profiled by GC-MS. Whereas triple expression of *CYP703A2/ACOS5/TKPR1* did not result in metabolites changes compared with the empty vector strain, the *CYP703A2/ACOS5/PKSA* and *CYP703A2/ACOS5/PKSA/TKPR1* expressers generated two unique peaks in GC chromatograms (Figure 5.5). Each peak had three fragments of the same characteristic masses (183, 198 and 211 m/z) while the two peaks had two different total masses (392 and 394 m/z). Since these novel peaks were present only in PKSA expressing strains, they are likely to be polyketide products, supporting the presence of active PKSA enzymes in the strains (Figure 5.6). Interestingly the total level of these novel peaks were much lower in the *CYP703A2/ACOS5/PKSA/TKPR1* expresser than in the *CYP703A2/ACOS5/PKSA* expresser (Figure 5.5). These data indicate that the TKPR1 enzyme could act on the product generated by PKSA, leading to new compounds not detected in this metabolic profile. Alternatively, TKPR1 could metabolize endogenous PKSA substrates into unknown products, making them less available for PKSA activity.

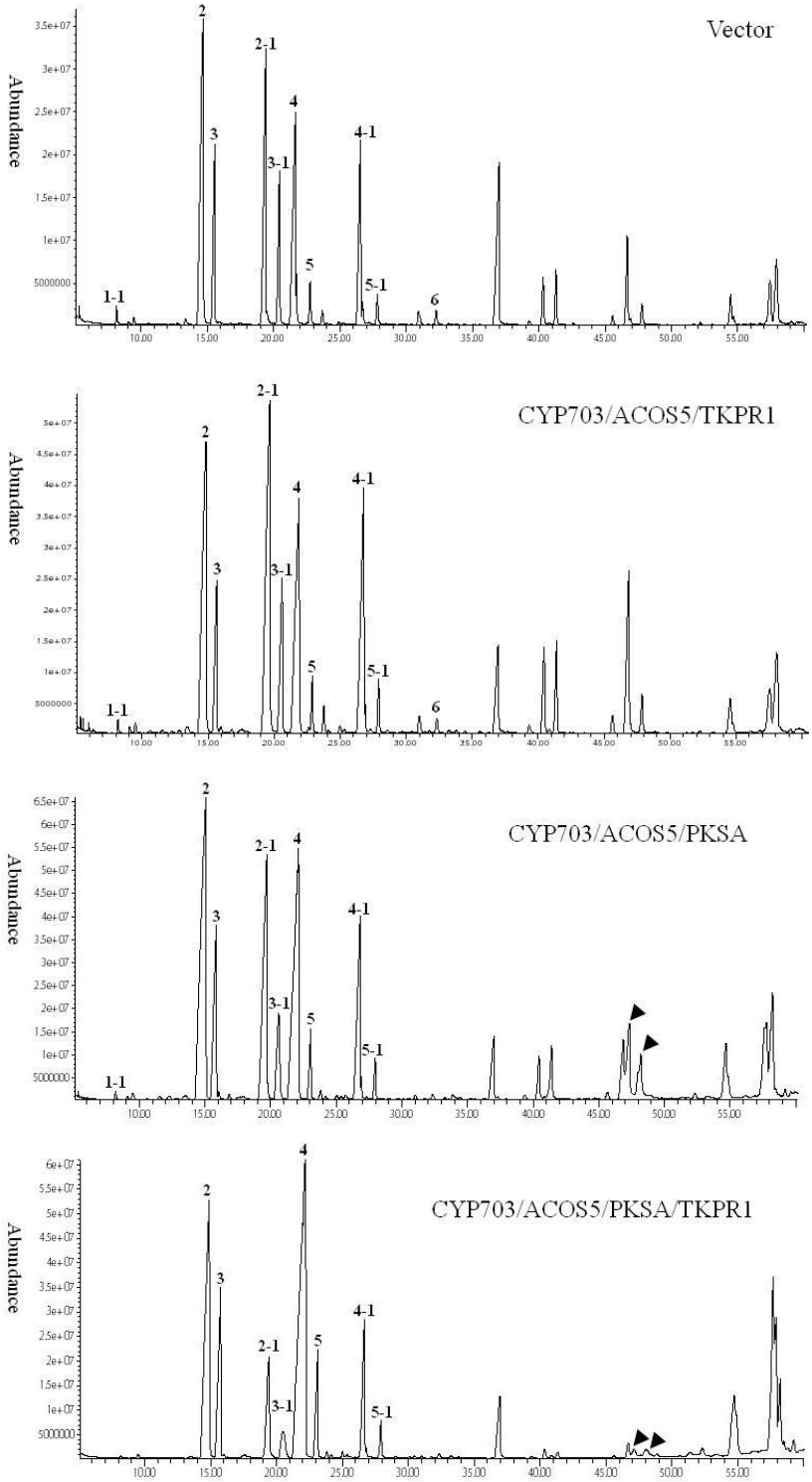


Figure 5.5 GC analysis of fatty acid metabolites produced by vector control and transgenic yeast strains.

GC chromatograms of extracts from different yeast strains are shown. Fatty acid derivatives were converted into methyl esters or trimethylsilyl esters prior to GC analysis.

1-1, dodecanoic acid trimethylsilyl ester (C12); 2, 9-hexadecenoic acid methyl ester (C16:1); 2-1, 9-hexadecenoic acid trimethylsilyl ester (C16:1); 3, hexadecanoic acid methyl ester (C16); 3-1, hexadecanoic acid trimethylsilyl ester (C16); 4, 9-octadecenoic acid methyl ester (C18:1); 4-1, 9-octadecenoic acid trimethylsilyl ester (C18:1); 5, octadecanoic acid methyl ester (C18); 5-1, octadecanoic acid trimethylsilyl ester (C18); 6, 9-hexanedioic acid 2,3-bis ester; arrow heads indicate novel compounds generated by PKSA.

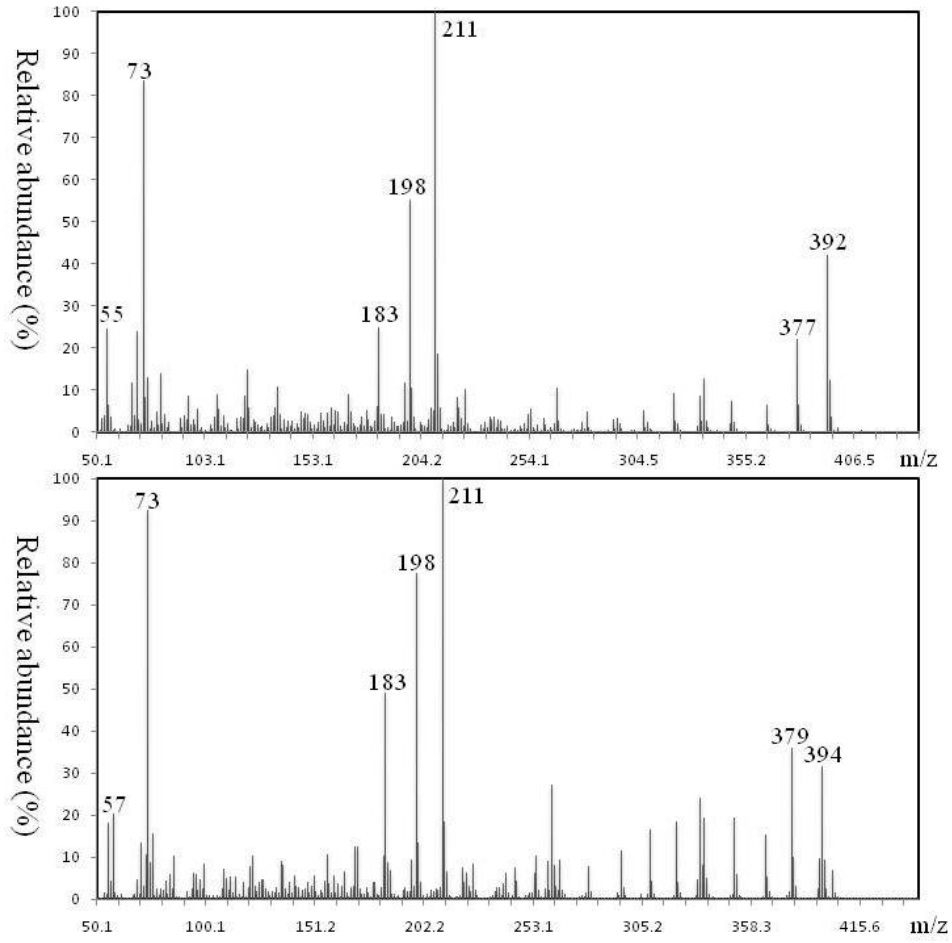


Figure 5.6 Mass spectra of two novel metabolites found in yeast strains expressing *CYP703A2/ACOS5/PKSA* or *CYP703A2/ACOS5/PKSA/TKPR1*.

Two mass spectra have three common characteristic peaks: 183, 198 and 211, together with different possible total molecular weight, 392 and 394, respectively. Upper spectrum was obtained at approximately 47.3 min of retention time and lower spectrum at approximately 48.3 min.

5.4.3 PKSA and/or TKPR1 expression in yeast

In vitro studies have shown that PKSA produces tri- and tetraketide α-pyrones by condensation of fatty acyl CoAs with malnoyl-CoAs (Mizuuchi et al., 2008; Dobritsa et al., 2010; Kim et al., 2010) and that TKPR1 reduces the keto function of tetraketide α-pyrone compounds synthesized by PKSA *in vitro* (Grienenberger et al., 2010). To determine the *in vivo* products generated by PKSA or/and TKPR1 using potential substrates present in wild-type yeast strains, *PKSA*, *TKPR1*, and *PKSA/TKPR1* dual-expressing yeasts were generated. As with previous experiments 5.4.2, immunoblots failed to detect PKSA proteins in either *PKSA* or *PKSA/TKPR1* expressers, whereas a high expression level of *TKPR1* was found in soluble lysates as well as from cell debris from both *TKPR1*- and *PKSA/TKPR*-transformed yeast strains (Figure 5.7). The cell densities of each strain were measured at four time points during 24 hr culture. In most yeast strains, growth rates reached approximately 5 at OD_{600}, except for PKSA/TKPR1 expresser showing 2.24±0.15. GC-MS analysis revealed that a total of six unidentified compounds were exclusively present in *PKSA/TKPR1* dual-expressing yeasts (Figure 5.8). Figure 5.9 illustrates mass spectra containing not only the same characteristic mass (183, 198 and 211 m/z) reported in 5.4.2 but also different total mass (338, 366, 392, 394, 420 and 422 respectively). These new compounds could be potential *in vivo* compounds generated by PKSA or PKSA as well as TKPR even though they could not be identified by comparison to the library of mass spectra or theliterature.

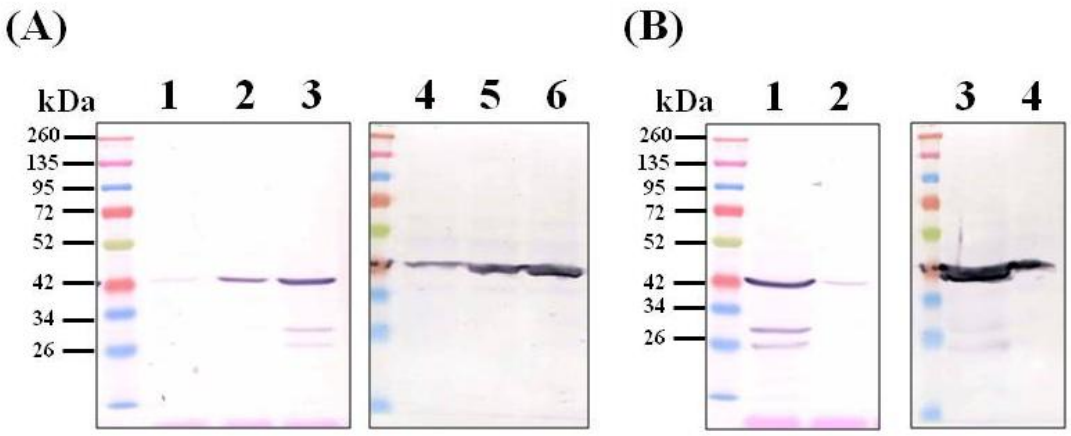


Figure 5.7 Immunoblot detection of TKPR1 proteins in transgenic yeast strains.

(A) Immunoblot analysis of proteins from cytosolic fraction (lane 1, 2 and 3) and cell debris (lane 4, 5 and 6) and in *PKSA/TKPR1* dual expresser, reacted with a FLAG monoclonal antibody to detect TKPR1. Lane 1 and 4, 12 hr culture; lane 2 and 4, 18 hr culture; lane 3 and 6, 24 hr culture
(B) Immunoblot analysis of proteins from cytosolic fraction (lane 1 and 2) and cell debris (lane 3 and 4) and in *TKPR1* expresser, reacted with a FLAG monoclonal antibody to detect TKPR1. Lane 1 and 3, 24 hr culture; lane 2 and 4, 9 hr culture

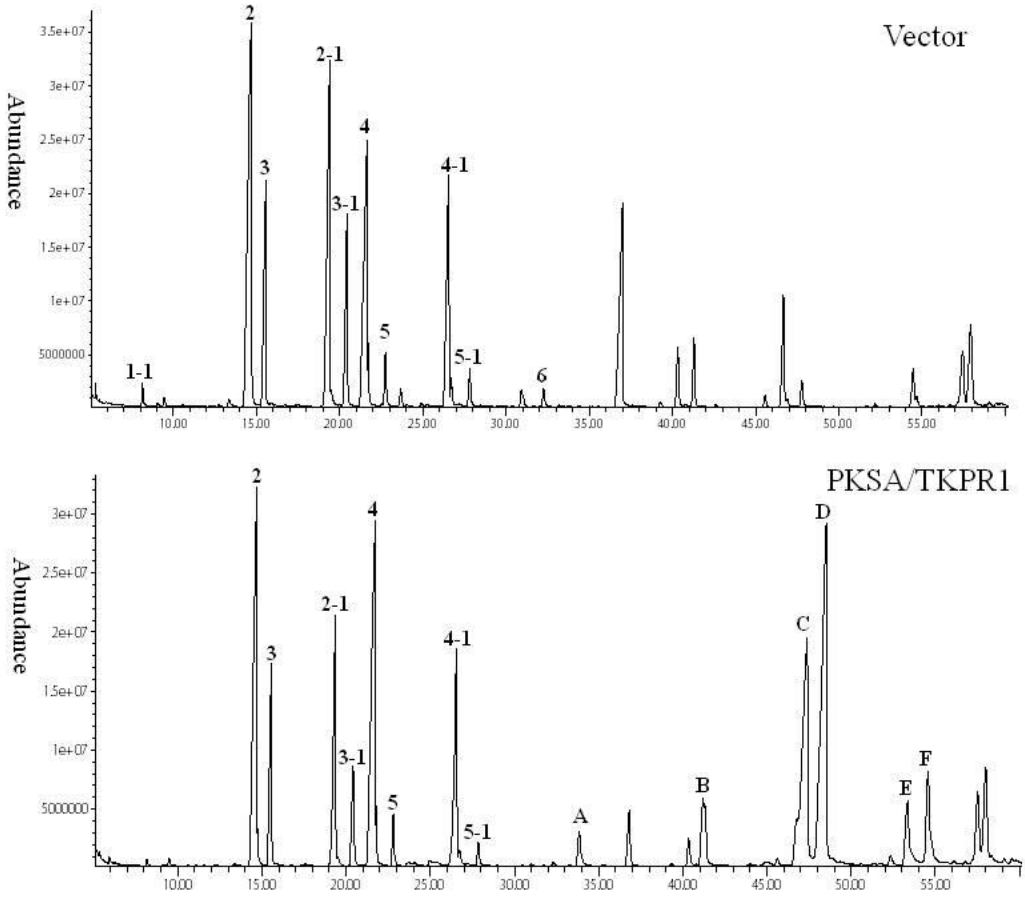
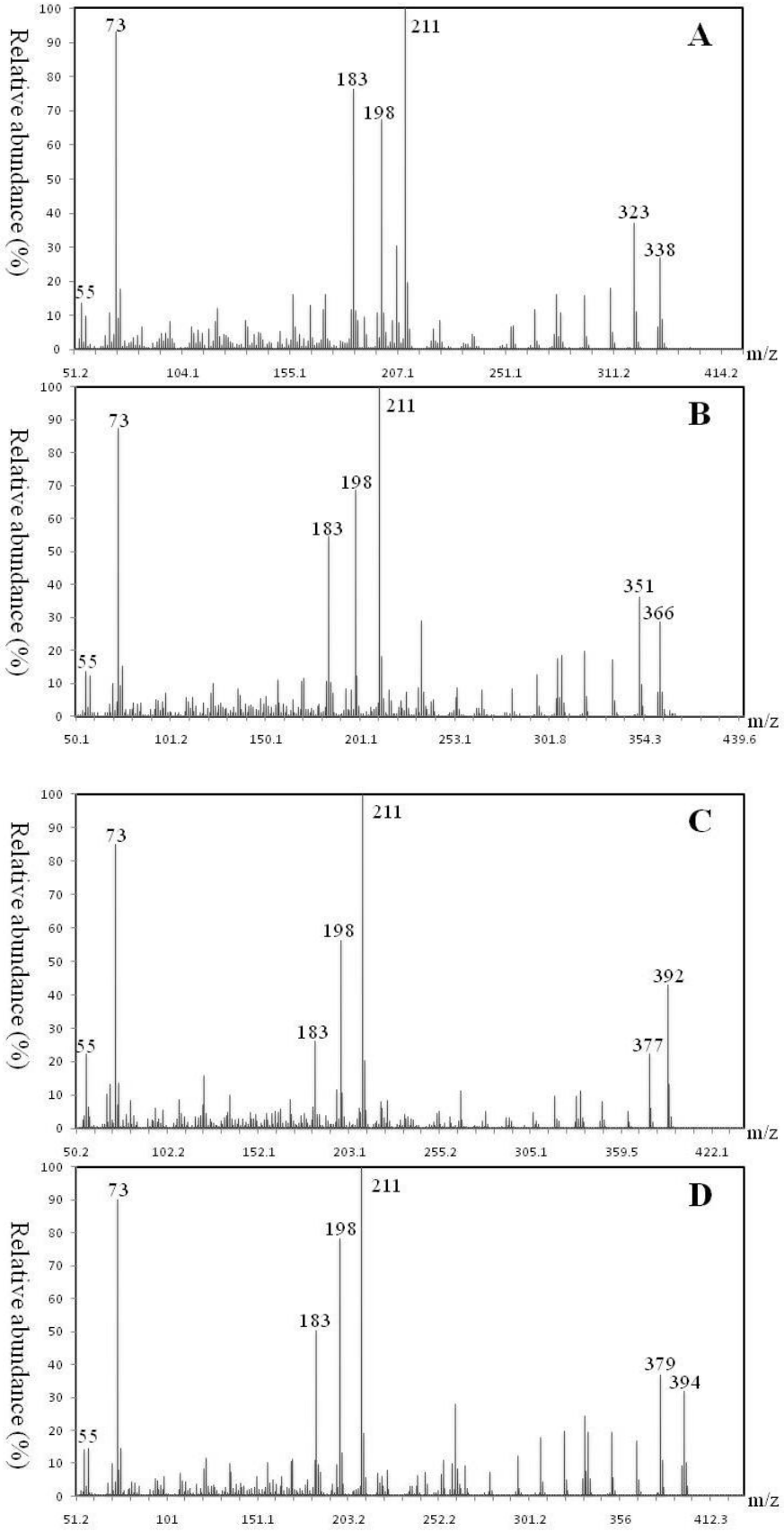


Figure 5.8 GC analysis of fatty acid metabolites produced by vector control and *PKSA/TKPR1* expressing yeast strains.

Fatty acid derivatives were converted into methyl esters or trimethylsilyl esters prior to GC analysis. Total six new compounds (A to F) were detected in PKSA/TKPR1 expresser. 1-1, dodecanoic acid trimethylsilyl ester (C12); 2, 9-hexadecenoic acid methyl ester (C16:1); 2-1, 9-hexadecenoic acid trimethylsilyl ester (C16:1); 3, hexadecanoic acid methyl ester (C16); 3-1, hexadecanoic acid trimethylsilyl ester (C16); 4, 9-octadecenoic acid methyl ester (C18:1); 4-1, 9-octadecenoic acid trimethylsilyl ester (C18:1); 5, octadecanoic acid methyl ester (C18); 5-1, octadecanoic acid trimethylsilyl ester (C18); 6, 9-hexanedioic acid 2,3-bis ester.


A
Relative abundance (%)
73
211
183
198
323
338
55
m/z
51.2
104.1
155.1
207.1
251.1
311.2
414.2
B
Relative abundance (%)
73
211
198
183
351
366
55
m/z
50.1
101.2
150.1
201.1
253.1
301.8
354.3
439.6
C
Relative abundance (%)
73
211
198
392
183
377
55
m/z
50.2
102.2
152.1
203.1
255.2
305.1
359.5
422.1
D
Relative abundance (%)
73
211
198
183
379
394
55
m/z
51.2
101
151.1
203.2
252.2
301.2
356
412.3

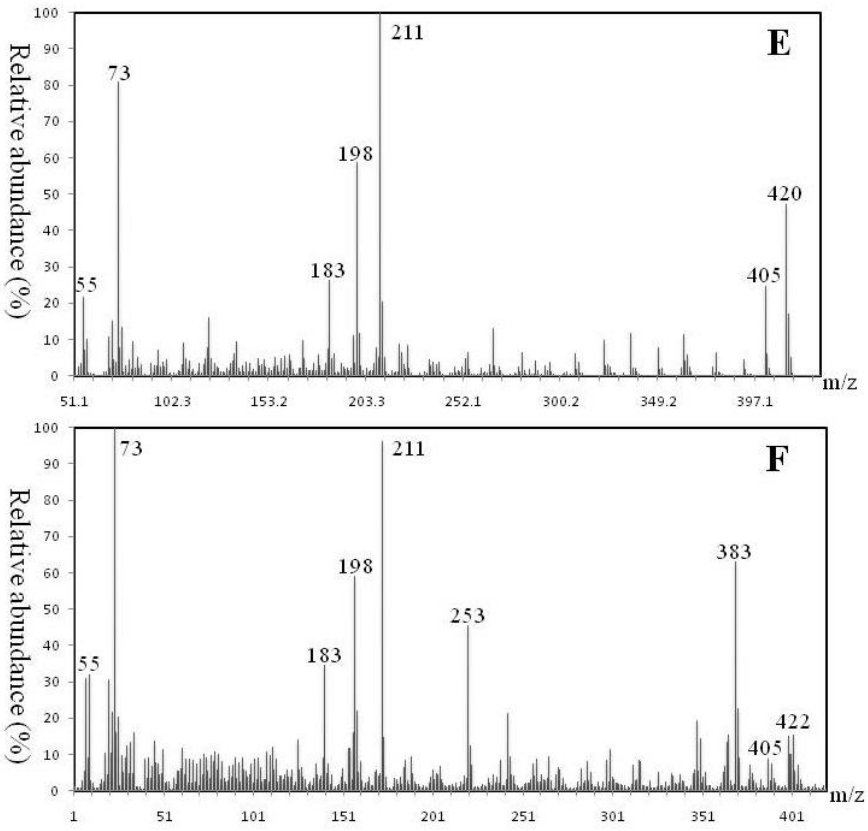
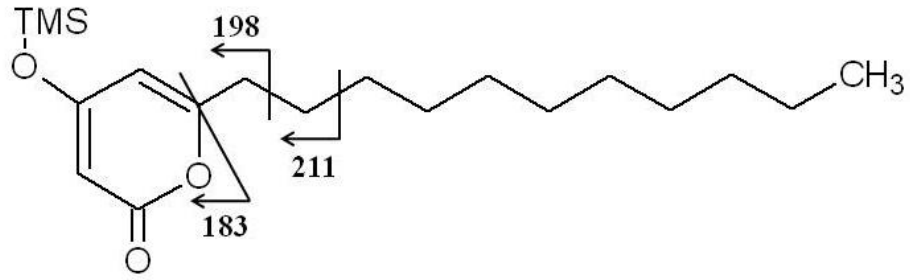


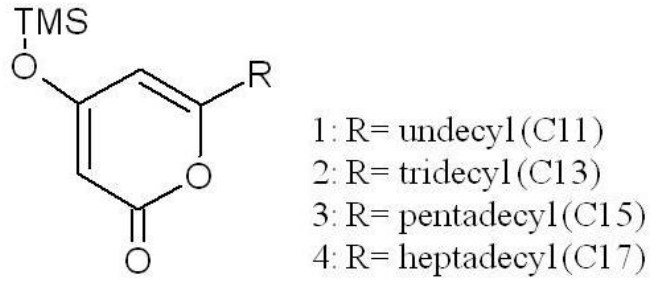
Figure 5.9 Mass spectra of novel peaks from the GC chromatogram obtained from the yeast strain expressing *PKSA/TKPR1*.

Each mass spectrum has three common characteristic peaks at m/z 183, 198 and 211, together with different possible total molecular weights of 338, 366, 392, 394, 420 and 422, respectively. Spectra were obtained at 33.89 min (A), 41.21 min (B), 47.30 min (C), 48.35 min (D), 53.34 min (E) and 54.33 min (F) of retention time.

A



B



C

	R (M.W.)	Pyrone +TMS (M.W.)	Total M.W.	Total M.W present in mass specturm
1 (C11)	155	183	338	338
2 (C13)	183		366	366
3 (C15)	211		394	394 and 392
4 (C17)	239		422	422 and 420

Figure 5.10 Predicted GC-MS fragmentation patterns and estimated molecular weights of α-pyrones containing various alkyl chain lengths.

(A) Based on characteristic m/z values from mass spectra in Figure 5.9, predicted fragments are drawn for triketide compound (A) and possible alkyl chains on the R position (B). Calculated molecular weights of alkyl α-pyrones are 338, 366, 394 and 422 from R-group alkyl chain lengths of C11, C13, C15 and C17, respectively. These values correspond to total molecular weight (m/z) present in Figure 5.9. Both 392 and 420 (m/z) would be unsaturated alkyl chains containing one double bond, of which start molecules are likely 9-hexadecenoic acid (C16:1) and 9-octadecenoic acid methyl ester (C18:1).

5.5 Discussion

Based on phenotypic, genotypic, *in vitro* biochemical data (Grienenberger et al., 2010; Kim et al., 2010), I postulated that CYP703A2, ACOS5, PKSA and TKPR1 work in the same biochemical pathway *in vivo* (Figure 5.1). However, as a reductase, it is

possible that *in vivo* TKPR1 acts directly on the CoA ester product of ACOS5 rather than on the polyketide product of PKSA as shown in Figure 5.1, which predicts that the *in vivo* fatty acid CoA ester of ACOS5 is the *in vivo* substrate of PKSA.

I used a yeast system for expressing up to four heterologous genes to test different models of biosynthetic pathways involving these four enzymes. *In vitro* assays indicate that hydroxy fatty acids are the preferred substrates of ACOS5 and PKSA (de Azevedo Souza et al., 2009; Grienenberger et al., 2010). Thus, expression of CYP703A2 was tested first for its ability to generate hydroxy fatty acids in yeast. In spite of strong growth inhibition by *CYP703A2* expression, no significant change in the fatty acid profile was detected by GC-FID and GC-MS in this yeast strain. Feeding of exogenous lauric acid (C12), which is the most highly preferred CYP703A2 substrate *in vitro* (Morant et al., 2007) did not result in hydroxy lauric acid accumulation in this strain. One possible explantion for these results is a lack of CYP703A2 activity due to misfolded protein or protein aggregates. In eukaryotes, properly folded cytochrome P450 enzymes are often localized to the ER membrane, anchored by a short hydrophobic sequence that exposes their enzymatically active domain into the cytoplasm, and misfolding or mislocalization could lead to loss of enzyme activity.

Alternatively, endogenous or exogenously added C12 may not have been accessible to the enzyme in these strains if it does not exist in the acid form at significant concentrations. For example, fatty acids are not freely available in plants but are generally present as triacylglycerol or CoA ester forms (Morant et al., 2007). Even though a small amount of endogenous lauric acid derivative (dodecanoic acid trimethylsilyl ester) was detected in vector- control yeast cells by GC-MS (compound 1-1, Figure 5.5), it is not known if this derivative is derived from free lauric acid. To further test for CYP703A2 activity in this strain, CYP703A2-containing microsomes could be extracted and incubated with free lauric acid to determine if the hydroxy-lauric acid can be produced *in vitro*.

While no significant change in the fatty acid profile was detected in the CYP703A2 expresser, *ACOS5* expression altered the yeast fatty acid profile (Figure 5.3). In *Saccharomyces cerevisiae*, palmitic acid (C16:0), palmitoleic acid (C16:1), stearic acid (C18:0) and oleic acid (C18:1) are the major fatty acyl constituents (Tuller et al., 1999),

while medium and short chain fatty acids, especially those less than C14, comprise less than 2% of total fatty acids (Tuller et al., 1999). Normally, only traces of endogenous medium chain fatty acids were detected in this study (Figure 5.3 and 5.5). However, it was striking that both the *ACOS5* expresser and the *CYP703A2/ACOS5* dual expresser accumulated tetradecanoic acid (C14) as well as 11-tetradecenoic acid (C14:1) upon feeding exogenous lauric acid (C12) (Figure 5.3). These data suggest that ACOS5 can generate accumulation of C14 fatty acyl-CoA esters from high levels of lauric cytoplasmic acid (C12). This is consistent with *in vitro* assays of ACOS5 activity that show a clear substrate preference for C10 to C14 fatty acids (de Azevedo Souza et al., 2009). This accumulation of C14 fatty acid derivatives was not observed in expresser without feeding lauric acid (data not shown). In the future, it would be beneficial to test whether accumulated C14 fatty acid derivatives are derived from exogenous lauric acid by feeding radio-labeled lauric acid to the growth media.

Novel products were found in *PKSA* expressing strains, and GC-MS analysis indicated chemical structures consistent with triketide α-pyrone accumulation in these strains (Figure 5.10). This is consistent with previous studies suggesting that PKSA catalyzes the condensation of malonyl-CoA units with fatty acyl-CoAs of various chain lengths *in vitro* (Mizuuchi et al., 2008; Dobritsa et al., 2010; Kim et al., 2010). Based on the total masses of the putative α-pyrone polyketides that accumulated, it appears that PKSA accepts C12:0, C14:0, C16:0, C16:1, C18:0 and C18:1 fatty acids *in vivo* to generate triketide α-pyrones. Pentadecyl triketide α-pyrones, generated from condensation of C16:0 or C16:1 with malonyl-CoA were the major products in these yeast strains, suggesting that C16 and C16:1 are preferable substrates for PKSA *in vivo*. These results are consistent with our *in vitro* data (Kim et al., 2010) and the recent report (Mizuuchi et al., 2008), showing the production of triketide and tetraketide α-pyrones by condensation with long-chain fatty acyl-CoAs (up to C20 chain length) and malonyl-CoA. Interestingly while C12 and C14 fatty acid derivatives were not detected by GC, undecyl and tridecyl triketides α-pyrones generated by condensation reaction with C12 and C14, respectively, were present in the PKSA/TKPR1 expresser (Figure 5.8). This suggests that PKSA might have a higher affinity for medium chain fatty acyl-CoAs than acyl-CoA thioesterases catalyzing the hydrolysis of acyl-CoAs to the free acids and CoAs for

further reaction.

The estimated molecular weights of triketides and tetraketides as well as the congruence of predicted and observed fragmentation patterns of triketide α-pyrones suggests that PKSA generated triketide α-pyrones, rather than tetraketide α-pyrones (the TKPR1 substrate *in vitro*; Grienenberger et al., 2010), accumulate in *PKSA*-expressing yeast strains. Moreover, no significant changes in the metabolite profiles were detected in strains expressing *TKPR1*, such as *CYP703A2/ACOS5/TKPR1* and *CYP703A2/ACOS5/PKSA/TKPR1*expresser, even though TKPR1 was clearly present in both soluble and in cell debris fractions of these strains. The likely reason for the lack of TKPR1-specific products in these strains is that TKPR1 substrates were absent. We previously observed that the tri- to tetraketide α-pyrone ratio is inversely correlated with the level of activity of the PKSA enzyme preparation, suggesting that the triketide α-pyrone represents a derailment reaction product due to incomplete catalysis (Kim et al., 2010). Thus, heterologously expressed *Arabidopsis* PKSA apperars to perform only two round of condensation with malonyl-CoA in yeast cells, generating a trikedtide α-pyrones that do not provide substrates for TKPR1. While tetraketide alkyl-α-pyrones were not generated in yeast strains expressing *PKSA*, these results support an *in vivo* function of PKSA in generating alkyl-α-pyrones. In addition, my results suggest that TKPR1 does not reduce the acyl-CoA esters to aldehydes or alcohols *in vivo*, as postulated in Chapter 4.

Chapter 6. Conclusion and future directions

Formation of pollen and spore walls requires the deposition of sporopollenin to protect the gametophyte from desiccation, temperatures, UV light and mechanical damages. The biochemically and physically extremely resistant nature of sporopollenin has been an obstacle to precise determination of the constituents of exine. A few chemical analyses by nuclear magnetic resonance (NMR) have demonstrated that sporopollenin is composed of polyhydroxylated unbranched aliphatic and phenolic constituents as the main monomeric units. These units are covalently coupled by ester and ether linkages to provide the characteristic high resistance to chemical degradation. Our reverse genetic approach to elucidate the pathway of sporopollenin monomer biosynthesis provides evidence that medium to long-chain (i.e. C12 to C18) hydroxy fatty acids as well as tri- and tetraketide α-pyrones are likely important precursors of building blocks of sporopollenin. Moreover, the remarkable conservation of *CYP703A2* (Morant et al., 2007), *ACOS5* (de Azevedo Souza et al., 2009), *PKSA/B* and *TKPR1/2* genes in land plants suggests that these enzymes are part of an ancient sporopollenin biosynthetic pathway.

We showed that fatty acid modification such as hydroxylation (CYP703A2), CoA esterification (ACOS5), condensation with malonyl-CoA (PKSA and PKSB) and reduction of a carbonyl group (TKPR1 and TKPR2) are key steps for sporopollenin biosynthesis in the tapetum. However, confirmation that these sequential biochemical reactions also take place *in vivo* is still needed. We proposed different possibilities for the sequential actions of CYP703A2, ACOS5, PKSA/B and TKPR1/2, based on *in vitro* biochemical assays and subcellular localization (Grienenberger et al., 2010). In Chapter 5, yeast strains expressing various combinations of CYP703A2, ACOS5, PKSA and TKPR1 were generated to test the potential sequential pathway *in vivo*. Kinetic parameters provided by collaborators have indicated that hydroxy fatty acids and hydroxy fatty acyl-CoA are preferred substrates of ACOS5 (de Azevedo Souza et al., 2009) and PKSA (Kim et al., 2010), respectively. Thus, my hypothesis is that fatty acid hydroxylation is the first reaction in the sequence leading to sporopollenin monomer biosynthesis. Unfortunately, possibly due to the failure of CYP703A2 and ACOS5 to gain access to appropriate

endogenous substrates in yeast cells, these two putative sequential reactions could not be reproduced *in vivo*. In the future, it would be interesting to test the sequential pathway in yeast cells expressing PKS and TKPR fed with exogenous hydroxy fatty acyl CoAs such as 12-OH-C18-CoA and 16-OH-C16-CoA. If these compounds penetrate yeast cell plasma membrane and are able to be accessed by PKS and TKPR, I expect that the hydroxy fatty acyl CoAs would be used as substrates by these enzymes to generate hydroxy alkyl α-pyrones. Expression of PKSA in the yeast showed PKSA can catalyze condensation of a broad range of endogenous fatty acyl-CoAs *in vivo* (Chapter 5). In such cells with increased pool sizes of the preferred substrates (i.e. hydroxy fatty acyl CoAs), PKSA may be much less active in catalyzing condensation of endogenous fatty acyl-CoAs due to the competition with the exogenously fed substrates. This could confirm not only the importance of hydroxylation of fatty acids for the pathway but also the *in vivo* preference of PKSA for hydroxy fatty acyl CoAs. In addition, we could test an alternative pathway (route B, Figure 5.1), which postulates hydroxylation reactions occurring on polyketides rather than on fatty acids, by using the CYP703A2/PKSA dual-expresser and CYP703A2/PKSA/TKPR1 triple-expresser. While CYP703A2 has been shown to in-chain hydroxylate lauric acid (C12:0) (Morant et al., 2007), we cannot rule out the possibility that CYP703A2 hydroxylates alkyl α-pyrones in the last step of potential sequential reaction. If hydroxyl alkyl α-pyrones were present in metabolic profiles of CYP703A2/PKSA expressers, it would support the alternative pathway in which these novel alkyl α-pyrones are the true substrates to CYP703A2 *in vivo*.

Recently, *CYP704B1*, which is also tightly coexpressed with *ACOS5*, was implicated in exine formation. Recombinant CYP704B1 catalyzes the ω-hydroxylation of C16 to C18 long-chain fatty acids, especially showing preference of unsaturated C18 fatty acids (Dobritsa et al., 2009). This indicates that CYP704B1 prefers longer carbon chain length fatty acids than CYP703A2 and can hydroxylate them at a different carbon position. Thus, two cytochrome P450s could provide different building blocks for sporopollenin synthesis. However, phenotypic analysis of the double mutant *cyp703a2 cyp704b1* did not show either an additive or new phenotype (Dobritsa et al., 2009). This suggests that *in planta* the hydroxylation steps involving CYP704B1 and CYP703A2 take place in a common pathway and that their catalytic functions may generate a

common sporopollenin monomer. It would be interesting to obtain further information about the relationship of CYP704B1 to CYP703A2 by transferring *CYP704B1* into a *cyp703a2* mutant and *vice versa*. If *cyp703a2* transgenic line harboring the *CYP704B1* transgene is fully fertile with normal exine formation, it would support the interpretation that the two hydroxylation reactions are in the same pathway.

Even if the composition of sporopollenin is not well known yet, it has been consistently reported that sporopollenin is produced largely from fatty acids and phenylpropanoid precursors in various plant species (Piffanelli et al., 1998; Domínguez et al., 1999). In order to test the functional conservation of the set of co-expressed enzymes based on the results of the comparative genomics approach, it would be interesting to test the ability of *ACOS5, PKSA/B* and *TKPR1/2* homologues from other plant species to complement the *Arabidopsis* mutations. As an initial attempt, the poplar *PoptrACOS13* gene, which is preferentially expressed in the male flowers and is grouped in same clade with *ACOS5* (Souza Cde et al., 2008), was chosen for heterologous complementation. The complementation construct was composed of a 2-kb promoter region of *ACOS5*, a 2.5-kb coding sequence of *PoptrACOS13* and a 2.6-kb terminator region of *ACOS5*. This construct could be transformed into an *ACOS5/acos5* heterozygous plant in the future. If the introduced *PoptrACOS13* transgene complements the male sterile phenotype in the *acos5* mutant background, transgenic plants will be fully fertile. This would suggest that the catalytic function of PoptrACOS13 in sporopollenin biosynthesis is conserved in *Populus trichocarpa*. In addition, a construct with the poplar *PKSA/B* or *TKPR1/2* gene could be designed and tested in a similar way.

While both *CYP703A2* and *ACOS5* orthologs could not be identified in the green alga *Chlamydomonas reinhardtii*, they are conserved in land plant lineages, including *Physcomitrella* (Morant et al., 2007; de Azevedo Souza et al., 2009). In mosses, a haploid spore is produced by the meiotic division of spore mother cells (archesporial cells) in the archesporium, the inner most tissue of the capsule in the sporophyte. The moss spore walls consist of three layers: the outermost perine, a separating layer, the exine and the inner intine. The major component of the exine layer of the spores is sporopollenin (Brown and Lemmon, 1984). Thus, acquiring the ability to generate the sporopollenin polymer to protect haploid spores was likely a key land plant innovation essential for

protection of haploid spores from various stresses of the terrestrial environment.

The first draft genome of *Physcomitrella patens* was released by the JGI (Joint Genome Institute, http://genome.jgi-psf.org//Phypa1_1/Phypa1_1.home.html). By searching for genes by sequences similarity as well as by phylogenetic analysis, I confirmed the existence of candidate *ACOS5, PKSA/B* and *TKPR1/2* homologues in *Physcomitrella patens* (de Azevedo Souza et al., 2009; Grienenberger et al., 2010; Kim et al., 2010). Moreover, recently the ppgs cDNA library (http://www.ncbi.nlm.nih.gov/UniGene/library.cgi?ORG=Ppa&LID=23755), which is derived from green sporophytes containing archesporial cells and developing spores (Dr. Dae-Yeon Suh, personal communication) has been made available. Potential moss orthologs of *Arabidopsis* genes involved in sporopollenin biosynthesis, such as *PpACOS6*, *PpCHS10* and *PpDFRL1* genes were found exclusively in the ppgs library and were absent in other libraries from gametophytic tissue (Che C. Colpitts et al., submitted). Thus, to test the functional conservation of the putative moss orthologs, I could use an approach similar to that described for *PoptrACOS13* poplar genes. These *Physcomitrella* genes could be placed under the control of the respective *Arabidopsis* promoters and these candidate genes will be introduced into the *Arabidopsis* mutant lines to test heterologous complementation.

Ability of the moss homologues to complement the *Arabidopsis acos5, pksa pksb* and *tkpr1* mutations would support the idea that these genes encode enzymes that form part of an ancient biochemical pathway common to land plant, required for sporopollenin biosynthesis.

References

Aarts, M.G., Keijzer, C.J., Stiekema, W.J., and Pereira, A. (1995). Molecular characterization of the *CER1* gene of A*Arabidopsis* involved in epicuticular wax biosynthesis and pollen fertility. Plant Cell **7,** 2115-2127.

Aarts, M.G., Hodge, R., Kalantidis, K., Florack, D., Wilson, Z.A., Mulligan, B.J., Stiekema, W.J., Scott, R., and Pereira, A. (1997). The *Arabidopsis* MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. Plant J **12,** 615-623.

Ageez, A., Kazama, Y., Sugiyama, R., and Kawano, S. (2005). Male-fertility genes expressed in male flower buds of *Silene latifolia* include homologs of anther-specific genes. Genes Genet Syst **80,** 403-413.

Ahlers, F., Lambert, J., and Wiermann, R. (2003). Acetylation and silylation of piperidine solubilized sporopollenin from pollen of *Typha angustifolia* L. Z Naturforsch C **58,** 807-811.

Ahlers, F., Bubert, H., Steuernagel, S., and Wiermann, R. (2000). The nature of oxygen in sporopollenin from the pollen of *Typha angustifolia* L. Z Naturforsch C **55,** 129-136.

Alonso, J.M., Stepanova, A.N., Leisse, T.J., Kim, C.J., Chen, H., Shinn, P., Stevenson, D.K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C.C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D.E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W.L., Berry, C.C., and Ecker, J.R. (2003). Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301,** 653-657.

Alves-Ferreira, M., Wellmer, F., Banhara, A., Kumar, V., Riechmann, J.L., and Meyerowitz, E.M. (2007). Global expression profiling applied to the analysis of *Arabidopsis* stamen development. Plant Physiol **145,** 747-762.

Ariizumi, T., Hatakeyama, K., Hinata, K., Sato, S., Kato, T., Tabata, S., and Toriyama, K. (2003). A novel male-sterile mutant of *Arabidopsis thaliana*,

faceless pollen-1, produces pollen with a smooth surface and an acetolysis-sensitive exine. Plant Mol Biol **53,** 107-116.

Ariizumi, T., Kawanabe, T., Hatakeyama, K., Sato, S., Kato, T., Tabata, S., and Toriyama, K. (2008). Ultrastructural characterization of exine development of the transient defective exine 1 mutant suggests the existence of a factor involved in constructing reticulate exine architecture from sporopollenin aggregates. Plant Cell Physiol **49,** 58-67.

Ariizumi, T., Hatakeyama, K., Hinata, K., Inatsugi, R., Nishida, I., Sato, S., Kato, T., Tabata, S., and Toriyama, K. (2004). Disruption of the novel plant protein NEF1 affects lipid accumulation in the plastids of the tapetum and exine formation of pollen, resulting in male sterility in *Arabidopsis thaliana*. Plant J **39,** 170-181.

Atanassov, I., Russinova, E., Antonov, L., and Atanassov, A. (1998). Expression of an anther-specific chalcone synthase-like gene is correlated with uninucleate microspore development in *Nicotiana sylvestris*. Plant Mol Biol **38,** 1169-1178.

Austin, M.B., and Noel, J.P. (2003). The chalcone synthase superfamily of type III polyketide synthases. Nat. Prod. Rep **20,** 79-110.

Baker, M.E., and Blasco, R. (1992). Expansion of the mammalian 3 beta-hydroxysteroid dehydrogenase/plant dihydroflavonol reductase superfamily to include a bacterial cholesterol dehydrogenase, a bacterial UDP-galactose-4-epimerase, and open reading frames in vaccinia virus and fish lymphocystis disease virus. FEBS Lett **301,** 89-93.

Baker, M.E., Luu-The, Y., Simard, J., and Labrie, F. (1990). A common ancestor for mammalian 3 beta-hydroxysteroid dehydrogenase and plant dihydroflavonol reductase. Biochem J **269,** 558-559.

Beisson, F., Li, Y., Bonaventure, G., Pollard, M., and Ohlrogge, J.B. (2007). The Acyltransferase GPAT5 Is Required for the Synthesis of Suberin in Seed Coat and Root of *Arabidopsis*. Plant Cell **19,** 351-368.

Bernhardt, J., Stich, K., Schwarz-Sommer, Z., Saedler, H., and Wienand, U. (1998). SHORT COMMUNICATION Molecular analysis of a second functional A1 gene (dihydroflavonol 4-reductase) in *Zea mays*. The Plant Journal **14,** 483.

Blackmore, S., Wortley, A.H., Skvarla, J.J., and Rowley, J.R. (2007). Pollen wall

development in flowering plants. New Phytol **174,** 483-498.

Blokker, P., Boelen, P., Broekman, R., and Rozema, J. (2006). The Occurrence of p-coumaric Acid and Ferulic Acid in Fossil Plant Materials and their Use as UV-proxy. Plant Ecology **182,** 197-207.

Bohne, G., Richter, E., Woehlecke, H., and Ehwald, R. (2003). Diffusion barriers of tripartite sporopollenin microcapsules prepared from pine pollen. Ann Bot (Lond) **92,** 289-297.

Bowman, J.L., Floyd, S.K., and Sakakibara, K. (2007). Green Genes - Comparative Genomics of the Green Branch of Life. Cell **129,** 229-234.

Brown, R.C., and Lemmon, B.E. (1984). Spore Wall Development in Andreaea (Musci: Andreaeopsida). American Journal of Botany **71,** 412-420.

Bubert, H., Lambert, J., Steuernagel, S., Ahlers, F., and Wiermann, R. (2002). Continuous decomposition of sporopollenin from pollen of *Typha angustifolia* L. by acidic methanolysis. Z Naturforsch C **57,** 1035-1041.

Burbulis, I.E., Iacobucci, M., and Shirley, B.W. (1996). A null mutation in the first enzyme of flavonoid biosynthesis does not affect male fertility in *Arabidopsis*. Plant Cell **8,** 1013-1025.

Chen, X., Goodwin, S.M., Boroff, V.L., Liu, X., and Jenks, M.A. (2003). Cloning and characterization of the *WAX2* gene of *Arabidopsis* involved in cuticle membrane and wax production. Plant Cell **15,** 1170-1185.

Chen, Y.C., and McCormick, S. (1996). Sidecar pollen, an *Arabidopsis thaliana* male gametophytic mutant with aberrant cell divisions during pollen development. Development **122,** 3243-3253.

Cigan, A.M., and Donahue, T.F. (1987). Sequence and structural features associated with translational initiator regions in yeast--a review. Gene **59,** 1-18.

Clough, S.J., and Bent, A.F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*. Plant J **16,** 735-743.

Costa, M.A., Collins, R.E., Anterola, A.M., Cochrane, F.C., Davin, L.B., and Lewis, N.G. (2003). An in silico assessment of gene function and organization of the phenylpropanoid pathway metabolic networks in *Arabidopsis thaliana* and limitations thereof. Phytochemistry **64,** 1097-1112.

Costa, M.A., Bedgar, D.L., Moinuddin, S.G.A., Kim, K.W., Cardenas, C.L., Cochrane, F.C., Shockey, J.M., Helms, G.L., Amakura, Y., and Takahashi, H. (2005). Characterization in vitro and in vivo of the putative multigene 4-coumarate: CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation. Phytochemistry **66,** 2072-2091.

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

de Azevedo Souza, C., Kim, S.S., Koch, S., Kienow, L., Schneider, K., McKim, S.M., Haughn, G.W., Kombrink, E., and Douglas, C.J. (2009). A novel fatty Acyl-CoA Synthetase is required for pollen development and sporopollenin biosynthesis in *Arabidopsis*. Plant Cell **21,** 507-525.

Devic, M., Guilleminot, J., Debeaujon, I., Bechtold, N., Bensaude, E., Koornneef, M., Pelletier, G., and Delseny, M. (1999). The *BANYULS* gene encodes a DFR-like protein and is a marker of early seed coat development. Plant J **19,** 387-398.

Dittrich, F., Zajonc, D., Huhne, K., Hoja, U., Ekici, A., Greiner, E., Klein, H., Hofmann, J., Bessoule, J.J., Sperling, P., and Schweizer, E. (1998). Fatty acid elongation in yeast--biochemical characteristics of the enzyme system and isolation of elongation-defective mutants. Eur J Biochem **252,** 477-485.

Doan, T.T., Carlsson, A.S., Hamberg, M., Bulow, L., Stymne, S., and Olsson, P. (2009). Functional expression of five *Arabidopsis* fatty acyl-CoA reductase genes in *Escherichia coli*. J Plant Physiol **166,** 787-796.

Dobritsa, A.A., Lei, Z., Nishikawa, S., Urbanczyk-Wochniak, E., Huhman, D.V., Preuss, D., and Sumner, L.W. (2010). LAP5 and LAP6 encode anther-specific proteins with similarity to chalcone synthase essential for pollen exine development in *Arabidopsis*. Plant Physiol **153,** 937-955.

Dobritsa, A.A., Shrestha, J., Morant, M., Pinot, F., Matsuno, M., Swanson, R., Moller, B.L., and Preuss, D. (2009). CYP704B1 is a long-chain fatty acid omega-hydroxylase essential for sporopollenin synthesis in pollen of *Arabidopsis*. Plant Physiol **151,** 574-589.

Domínguez, E., Mercado, J.A., Quesada, M.A., and Heredia, A. (1999). Pollen sporopollenin: degradation and structural elucidation. Sexual Plant Reproduction

12, 171-178.

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

Ehlting, J., Buttner, 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. The Plant Journal **19,** 9-20.

Ehlting, J., Mattheus, N., Aeschliman, D.S., Li, E., Hamberger, B., Cullis, I.F., Zhuang, J., Kaneda, M., Mansfield, S.D., and Samuels, L. (2005). Global transcript profiling of primary stems from *Arabidopsis thaliana* identifies candidate genes for missing links in lignin biosynthesis and transcriptional regulators of fiber differentiation. Plant J **42,** 618-640.

Fiebig, A., Mayfield, J.A., Miley, N.L., Chau, S., Fischer, R.L., and Preuss, D. (2000). Alterations in *CER6*, a gene identical to *CUT1*, differentially affect long-chain lipid content on the surface of pollen and stems. Plant Cell **12,** 2001-2008.

Fischer, R., Budde, I., and Hain, R. (1997). Stilbene synthase gene expression causes changes in flower colour and male sterility in tobacco. The Plant Journal **11,** 489-498.

Franke, R., and Schreiber, L. (2007). Suberin - biopolyester forming apoplastic plant interfaces. Current Opinion in Plant Biology **10,** 252-259.

Fulda, M., Shockey, J., Werber, M., Wolter, F.P., and Heinz, E. (2002). Two long-chain acyl-CoA synthetases from *Arabidopsis thaliana* involved in peroxisomal fatty acid beta-oxidation. The Plant Journal **32,** 93-103.

Funa, N., Awakawa, T., and Horinouchi, S. (2007). Pentaketide resorcylic acid synthesis by type III polyketide synthase from *Neurospora crassa*. J Biol Chem **282,** 14476-14481.

Funa, N., Ozawa, H., Hirata, A., and Horinouchi, S. (2006). Phenolic lipid synthesis by type III polyketide synthases is essential for cyst formation in *Azotobacter vinelandii*. Proc Natl Acad Sci U S A **103,** 6356-6361.

Gietz, D., St Jean, A., Woods, R.A., and Schiestl, R.H. (1992). Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res **20,** 1425.

Goldberg, R.B., Beals, T.P., and Sanders, P.M. (1993). Anther development: basic

principles and practical applications. Plant Cell **5,** 1217-1229.

Grienenberger, E., Besseau, S., Geoffroy, P., Debayle, D., Heintz, D., Lapierre, C., Pollet, B., Heitz, T., and Legrand, M. (2009). A BAHD acyltransferase is expressed in the tapetum of *Arabidopsis* anthers and is involved in the synthesis of hydroxycinnamoyl spermidines. Plant J **58,** 246-259.

Grienenberger, E., Kim, S.S., Lallemand, B., Geoffroy, P., Heintz, D., Souza, C.D., Heitz, T., Douglas, C.J., and Legrand, M. (2010). Analysis of TETRAKETIDE α-PYRONE REDUCTASE Function in *Arabidopsis thaliana* Reveals a Previously Unknown, but Conserved, Biochemical Pathway in Sporopollenin Monomer Biosynthesis. Plant Cell **22**: 4067-4083.

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

Hahlbrock, K., and Scheel, D. (1989). Physiology and Molecular Biology of Phenylpropanoid Metabolism. Annual Review of Plant Physiology and Plant Molecular Biology **40,** 347-369.

Hamberger, B., and Hahlbrock, K. (2004). The 4-coumarate: CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proceedings of the National Academy of Sciences **101,** 2209-2214.

Hamberger, B., M, E., M, F., de Azevedo Souza, C., B, B. and Douglas, C. J. (2007). Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the Populus lignin toolbox and conservation and diversification of angiosperm gene families. Can J Bot **85,** 1182-1201.

Hietala, A.M., Eikenes, M., Kvaalen, H., Solheim, H., and Fossdal, C.G. (2003). Multiplex real-time PCR for monitoring Heterobasidion annosum colonization in Norway spruce clones that differ in disease resistance. Appl Environ Microbiol **69,** 4413-4420.

Hoffmann, L., Besseau, S., Geoffroy, P., Ritzenthaler, C., Meyer, D., Lapierre, C., Pollet, B., and Legrand, M. (2004). Silencing of Hydroxycinnamoyl-Coenzyme A Shikimate/Quinate Hydroxycinnamoyltransferase Affects Phenylpropanoid

Biosynthesis. Plant Cell **16,** 1446-1465.

Hofig, K.P., Moller, R., Donaldson, L., Putterill, J., and Walter, C. (2006). Towards male sterility in Pinus radiata--a stilbene synthase approach to genetically engineer nuclear male sterility. Plant Biotechnol J **4,** 333-343.

Hose, E., Clarkson, D.T., Steudle, E., Schreiber, L., and Hartung, W. (2001). The exodermis: a variable apoplastic barrier. Journal of Experimental Botany **52,** 2245.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv Bioinformatics **2008,** 420747.

Ishiguro, S., Kawai-Oda, A., Ueda, J., Nishida, I., and Okada, K. (2001). The DEFECTIVE IN ANTHER DEHISCIENCE gene encodes a novel phospholipase A1 catalyzing the initial step of jasmonic acid biosynthesis, which synchronizes pollen maturation, anther dehiscence, and flower opening in *Arabidopsis*. Plant Cell **13,** 2191-2209.

Ito, T., Nagata, N., Yoshiba, Y., Ohme-Takagi, M., Ma, H., and Shinozaki, K. (2007). *Arabidopsis MALE STERILITY1* encodes a PHD-type transcription factor and regulates pollen and tapetum development. Plant Cell **19,** 3549-3562.

Itoh, Y., Higeta, D., Suzuki, A., Yoshida, H., and Ozeki, Y. (2002). Excision of Transposable Elements from the Chalcone Isomerase and Dihydroflavonol 4-Reductase Genes May Contribute to the Variegation of the Yellow-Flowered Carnation (*Dianthus caryophyllus*). Plant and Cell Physiology **43,** 578-585.

Jiang, C., Schommer, C.K., Kim, S.Y., and Suh, D.Y. (2006). Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*. Phytochemistry **67,** 2531-2540.

Johnson, E.T., Ryu, S., Yi, H., Shin, B., Cheong, H., and Choi, G. (2001). Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. Plant J **25,** 325-333.

Kienow, L., Schneider, K., Bartsch, M., Stuible, H.P., Weng, H., Miersch, O., Wasternack, C., and Kombrink, E. (2008). Jasmonates meet fatty acids: functional analysis of a new acyl-coenzyme A synthetase family from *Arabidopsis*

thaliana. J Exp Bot **59,** 403-419.

Kim, S.S., Grienenberger, E., Lallemand, B., Colpitts, C.C., Kim, S.Y., Souza, C.D., Geoffroy, P., Heintz, D., Krahn, D., Kaiser, M., Kombrink, E., Heitz, T., Suh, D.Y., Legrand, M., and Douglas, C.J. (2010). LAP6/POLYKETIDE SYNTHASE A and LAP5/POLYKETIDE SYNTHASE B Encode Hydroxyalkyl α-Pyrone Synthases Required for Pollen Development and Sporopollenin Biosynthesis in *Arabidopsis thaliana*. Plant Cell **22**: 4045-4066.

Koduri, P.K., Gordon, G.S., Barker, E.I., Colpitts, C.C., Ashton, N.W., and Suh, D.Y. (2010). Genome-wide analysis of the chalcone synthase superfamily genes of Physcomitrella patens. Plant Mol Biol **72,** 247-263.

Kolattukudy, P.E. (2001). Polyesters in higher plants. Adv Biochem Eng Biotechnol **71,** 1-49.

Lacombe, E., Hawkins, S., Van Doorsselaere, J., Piquemal, J., Goffner, D., Poeydomenge, O., Boudet, A.M., and Grima-Pettenati, J. (1997). Cinnamoyl CoA reductase, the first committed enzyme of the lignin branch biosynthetic pathway: cloning, expression and phylogenetic relationships. Plant J **11,** 429-441.

Li, H., Pinot, F., Sauveplane, V., Werck-Reichhart, D., Diehl, P., Schreiber, L., Franke, R., Zhang, P., Chen, L., Gao, Y., Liang, W., and Zhang, D. (2010). Cytochrome P450 family member CYP704B2 catalyzes the ω-hydroxylation of fatty acids and is required for anther cutin biosynthesis and pollen exine formation in rice. Plant Cell **22,** 173-190.

Lynen, F. (1980). On the structure of fatty acid synthetase of yeast. Eur J Biochem **112,** 431-442.

Ma, H. (2005). Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. Annual Review of Plant Biology **56,** 393-434.

McConn, M., and Browse, J. (1996). The Critical Requirement for Linolenic Acid Is Pollen Development, Not Photosynthesis, in an *Arabidopsis* Mutant. Plant Cell **8,** 403-416.

Mizuuchi, Y., Shimokawa, Y., Wanibuchi, K., Noguchi, H., and Abe, I. (2008). Structure function analysis of novel type III polyketide synthases from

Arabidopsis thaliana. Biol Pharm Bull **31,** 2205-2210.

Mo, Y., Nagel, C., and Taylor, L.P. (1992). Biochemical complementation of chalcone synthase mutants defines a role for flavonols in functional pollen. Proc Natl Acad Sci U S A **89,** 7213-7217.

Morant, M., Jorgensen, K., Schaller, H., Pinot, F., Moller, B.L., Werck-Reichhart, D., and Bak, S. (2007). CYP703 is an ancient cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. Plant Cell **19,** 1473-1487.

Nakagawa, T., Kurose, T., Hino, T., Tanaka, K., Kawamukai, M., Niwa, Y., Toyooka, K., Matsuoka, K., Jinbo, T., and Kimura, T. (2007). Development of series of gateway binary vectors, pGWBs, for realizing efficient construction of fusion genes for plant transformation. J Biosci Bioeng **104,** 34-41.

Paxson-Sowders, D.M., Dodrill, C.H., Owen, H.A., and Makaroff, C.A. (2001). DEX1, a novel plant protein, is required for exine pattern formation during pollen development in *Arabidopsis*. Plant Physiol **127,** 1739-1749.

Pichersky, E., and Gang, D.R. (2000). Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends Plant Sci **5,** 439-445.

Piffanelli, P., Ross, J.H.E., and Murphy, D.J. (1998). Biogenesis and function of the lipidic structures of pollen grains. Sexual Plant Reproduction **11,** 65-80.

Pollard, M., Beisson, F., Li, Y., and Ohlrogge, J.B. (2008). Building lipid barriers: biosynthesis of cutin and suberin. Trends Plant Sci **13,** 236-246.

Pompon, D., Louerat, B., Bronine, A., and Urban, P. (1996). Yeast expression of animal and plant P450s in optimized redox environments. Methods Enzymol **272,** 51-64.

Preston, J., Wheeler, J., Heazlewood, J., Li, S.F., and Parish, R.W. (2004). AtMYB32 is required for normal pollen development in *Arabidopsis* thaliana. Plant J **40,** 979-995.

Quilichini, T.D., Friedmann, M.C., Samuels, A.L., and Douglas, C.J. (2010). ATP-binding cassette transporter G26 is required for male fertility and pollen exine formation in *Arabidopsis*. Plant Physiol **154,** 678-690.

Raes, J., Rohde, A., Christensen, J.H., Peer, Y., and Boerjan, W. (2003). Genome-

wide characterization of the lignification toolbox in *Arabidopsis*. Plant Physiology **133,** 1051-1071.

Romanos, M.A., Scorer, C.A., and Clare, J.J. (1992). Foreign gene expression in yeast: a review. Yeast **8,** 423-488.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An *Arabidopsis* thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol Biol **53,** 247-259.

Rowland, O., Lee, R., Franke, R., Schreiber, L., and Kunst, L. (2007). The *CER3* wax biosynthetic gene from *Arabidopsis* thaliana is allelic to *WAX2/YRE/FLP1*. FEBS Lett **581,** 3538-3544.

Rowland, O., Zheng, H., Hepworth, S.R., Lam, P., Jetter, R., and Kunst, L. (2006). *CER4* encodes an alcohol-forming fatty acyl-coenzyme A reductase involved in cuticular wax production in *Arabidopsis*. Plant Physiol **142,** 866-877.

Rozema, J., Broekman, R.A., Blokker, P., Meijkamp, B.B., de Bakker, N., van de Staaij, J., van Beem, A., Ariese, F., and Kars, S.M. (2001). UV-B absorbance and UV-B absorbing compounds (para-coumaric acid) in pollen and sporopollenin: the perspective to track historic UV-B levels. J Photochem Photobiol B **62,** 108-117.

Sanders, P.M., Lee, P.Y., Biesgen, C., Boone, J.D., Beals, T.P., Weiler, E.W., and Goldberg, R.B. (2000). The *Arabidopsis DELAYED DEHISCENCE1* gene encodes an enzyme in the jasmonic acid synthesis pathway. Plant Cell **12,** 1041-1061.

Sanders, P.M., Bui, A.Q., Weterings, K., McIntire, K.N., Hsu, Y.C., Lee, P.Y., Truong, M.T., Beals, T.P., and Goldberg, R.B. (1999). Anther developmental defects in *Arabidopsis thaliana* male-sterile mutants. Sexual Plant Reproduction **11,** 297-322.

Schijlen, E.G., de Vos, C.H., Martens, S., Jonker, H.H., Rosin, F.M., Molthoff, J.W., Tikunov, Y.M., Angenent, G.C., van Tunen, A.J., and Bovy, A.G. (2007). RNA interference silencing of chalcone synthase, the first step in the flavonoid biosynthesis pathway, leads to parthenocarpic tomato fruits. Plant Physiol **144,** 1520-1530.

Schneider, K., Kienow, L., Schmelzer, E., Colby, T., Bartsch, M., Miersch, O., Wasternack, C., Kombrink, E., and Stuible, H.P. (2005). A New Type of Peroxisomal Acyl-Coenzyme A Synthetase from *Arabidopsis thaliana* Has the Catalytic Capacity to Activate Biosynthetic Precursors of Jasmonic Acid. Journal of Biological Chemistry **280,** 13962-13972.

Schnurr, J.A., Shockey, J.M., de Boer, G.J., and Browse, J.A. (2002). Fatty Acid Export from the Chloroplast. Molecular Characterization of a Major Plastidial Acyl-Coenzyme A Synthetase from *Arabidopsis*. Plant Physiology **129,** 1700-1709.

Schweizer, E., and Hofmann, J. (2004). Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. Microbiol Mol Biol Rev **68,** 501-517.

Scott, R.J., Spielman, M., and Dickinson, H.G. (2004). Stamen structure and function. Plant Cell **16 Suppl,** S46-60.

Shimada, S., Takahashi, K., Sato, Y., and Sakuta, M. (2004). Dihydroflavonol 4-reductase cDNA from non-anthocyanin-producing species in the Caryophyllales. Plant Cell Physiol **45,** 1290-1298.

Shirley, B.W., Hanley, S., and Goodman, H.M. (1992). Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis* transparent testa mutations. Plant Cell **4,** 333-347.

Shirley, B.W., Kubasek, W.L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F.M., and Goodman, H.M. (1995). Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. Plant J **8,** 659-671.

Shockey, J.M., Fulda, M.S., and Browse, J.A. (2002). *Arabidopsis* Contains Nine Long-Chain Acyl-Coenzyme A Synthetase Genes That Participate in Fatty Acid and Glycerolipid Metabolism. Plant Physiology **129,** 1710-1722.

Shockey, J.M., Fulda, M.S., and Browse, J. (2003). *Arabidopsis* Contains a Large Superfamily of Acyl-Activating Enzymes. Phylogenetic and Biochemical Analysis Reveals a New Class of Acyl-Coenzyme A Synthetases 1. Plant Physiology **132,** 1065-1076.

Soler, M., Serra, O., Molinas, M., Huguet, G., Fluch, S., and Figueras, M. (2007). A

Genomic Approach to Suberin Biosynthesis and Cork Differentiation. Plant Physiology **144,** 419.

Sorensen, A.M., Krober, S., Unte, U.S., Huijser, P., Dekker, K., and Saedler, H. (2003). The *Arabidopsis ABORTED MICROSPORES (AMS)* gene encodes a MYC class transcription factor. Plant J **33,** 413-423.

Souza Cde, A., Barbazuk, B., Ralph, S.G., Bohlmann, J., Hamberger, B., and Douglas, C.J. (2008). Genome-wide analysis of a land plant-specific *acyl:coenzyme A synthetase (ACS)* gene family in *Arabidopsis*, poplar, rice and Physcomitrella. New Phytol **179,** 987-1003.

Steiner-Lange, S., Unte, U.S., Eckstein, L., Yang, C., Wilson, Z.A., Schmelzer, E., Dekker, K., and Saedler, H. (2003). Disruption of *Arabidopsis thaliana* MYB26 results in male sterility due to non-dehiscent anthers. Plant J **34,** 519-528.

Stuible, H., Buttner, D., Ehlting, J., Hahlbrock, K., and Kombrink, E. (2000). Mutational analysis of 4-coumarate: CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. FEBS Lett **467,** 117-122.

Suzuki, T., Masaoka, K., Nishi, M., Nakamura, K., and Ishiguro, S. (2008). Identification of kaonashi mutants showing abnormal pollen exine structure in *Arabidopsis thaliana*. Plant Cell Physiol **49,** 1465-1477.

Tanaka, Y., Tsuda, S., and Kusumi, T. (1998). Metabolic Engineering to Modify Flower Color. Plant and Cell Physiology **39,** 1119.

Tang, L.K., Chu, H., Yip, W.K., Yeung, E.C., and Lo, C. (2009). An anther-specific dihydroflavonol 4-reductase-like gene (*DRL1*) is essential for male fertility in *Arabidopsis*. New Phytol **181,** 576-587.

Toufighi, K., Brady, S.M., Austin, R., Ly, E., and Provart, N.J. (2005). The Botany Array Resource: e-Northerns, Expression Angling, and promoter analyses. Plant J **43,** 153-163.

Tsai, C.J., Harding, S.A., Tschaplinski, T.J., Lindroth, R.L., and Yuan, Y. (2006). Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in Populus. New Phytol **172,** 47-62.

Tuller, G., Nemec, T., Hrastnik, C., and Daum, G. (1999). Lipid composition of

subcellular membranes of an FY1679-derived haploid yeast wild-type strain grown on different carbon sources. Yeast **15,** 1555-1564.

van der Meer, I.M., Stam, M.E., van Tunen, A.J., Mol, J.N., and Stuitje, A.R. (1992). Antisense inhibition of flavonoid biosynthesis in petunia anthers results in male sterility. Plant Cell **4,** 253-262.

Vizcay-Barrena, G., and Wilson, Z.A. (2006). Altered tapetal PCD and pollen wall development in the *Arabidopsis ms1* mutant. Journal of Experimental Botany **57,** 2709.

Walden, A.R., Walter, C., and Gardner, R.C. (1999). Genes expressed in *Pinus radiata* male cones include homologs to anther-specific and pathogenesis response genes. Plant Physiol **121,** 1103-1116.

Wang, W.-K., Schaal, B.A., Chiou, Y.-M., Murakami, N., Ge, X.-J., Huang, C.-C., and Chiang, T.-Y. (2007). Diverse selective modes among orthologs/paralogs of the chalcone synthase (*Chs*) gene family of *Arabidopsis thaliana* and its relative *A. halleri ssp. gemmifera*. Molecular Phylogenetics and Evolution **44,** 503-520.

Wang, Z., Yeats, T., Han, H., and Jetter, R. (2010). Cloning and characterization of oxidosqualene cyclases from *Kalanchoe daigremontiana*: enzymes catalyzing up to 10 rearrangement steps yielding friedelin and other triterpenoids. J Biol Chem **285,** 29703-29712.

Watanabe, K., Praseuth, A.P., and Wang, C.C. (2007). A comprehensive and engaging overview of the type III family of polyketide synthases. Curr Opin Chem Biol **11,** 279-286.

Wijeratne, A.J., Zhang, W., Sun, Y., Liu, W., Albert, R., Zheng, Z., Oppenheimer, D.G., Zhao, D., and Ma, H. (2007). Differential gene expression in *Arabidopsis* wild-type and mutant anthers: insights into anther cell differentiation and regulatory networks. Plant J **52,** 14-29.

Wilson, Z.A., Morroll, S.M., Dawson, J., Swarup, R., and Tighe, P.J. (2001). The *Arabidopsis MALE STERILITY1* (*MS1*) gene is a transcriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. Plant J **28,** 27-39.

Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics,

biochemistry, cell biology, and biotechnology. Plant Physiol **126,** 485-493.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V., and Provart, N.J. (2007). An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PLoS One **2,** e718.

Wu, S., O'Leary, S.J., Gleddie, S., Eudes, F., Laroche, A., and Robert, L.S. (2008). A chalcone synthase-like gene is highly expressed in the tapetum of both wheat (*Triticum aestivum* L.) and triticale (x *Triticosecale* Wittmack). Plant Cell Rep **27,** 1441-1449.

Xie, D.Y., Sharma, S.B., Paiva, N.L., Ferreira, D., and Dixon, R.A. (2003). Role of anthocyanidin reductase, encoded by *BANYULS* in plant flavonoid biosynthesis. Science **299,** 396-399.

Yang, C., Vizcay-Barrena, G., Conner, K., and Wilson, Z.A. (2007a). MALE STERILITY1 is required for tapetal development and pollen wall biosynthesis. Plant Cell **19,** 3530-3548.

Yang, C., Xu, Z., Song, J., Conner, K., Vizcay Barrena, G., and Wilson, Z.A. (2007b). *Arabidopsis* MYB26/MALE STERILE35 regulates secondary thickening in the endothecium and is essential for anther dehiscence. Plant Cell **19,** 534-548.

Yang, S.L., Xie, L.F., Mao, H.Z., Puah, C.S., Yang, W.C., Jiang, L., Sundaresan, V., and Ye, D. (2003a). Tapetum determinant1 is required for cell specialization in the *Arabidopsis* anther. Plant Cell **15,** 2792-2804.

Yang, W.C., Ye, D., Xu, J., and Sundaresan, V. (1999). The *SPOROCYTELESS* gene of *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. Genes Dev **13,** 2108-2117.

Yang, X., Makaroff, C.A., and Ma, H. (2003b). The *Arabidopsis MALE MEIOCYTE DEATH1* gene encodes a PHD-finger protein that is required for male meiosis. Plant Cell **15,** 1281-1295.

Yau, C., Zhuang, C., Zee, S., and Yip, W. (2005). Expression of a microsporocyte-specific gene encoding dihydroflavonol 4-reductase-like protein is developmentally regulated during early microsporogenesis in rice. Sexual Plant Reproduction **18,** 64-74.

Yuan, T., Fujioka, S., Takatsuto, S., Matsumoto, S., Gou, X., He, K., Russell, S.D.,

and Li, J. (2007). *BEN1*, a gene encoding a dihydroflavonol 4-reductase (DFR)-like protein, regulates the levels of brassinosteroids in *Arabidopsis thaliana*. Plant J **51,** 220-233.

Zhang, X.H., and Chiang, V.L. (1997). Molecular Cloning of 4-Coumarate: Coenzyme A Ligase in Loblolly Pine and the Roles of This Enzyme in the Biosynthesis of Lignin in Compression Wood. Am Soc Plant Biol pp. 65-74.

Zhang, Z.B., Zhu, J., Gao, J.F., Wang, C., Li, H., Zhang, H.Q., Zhang, S., Wang, D.M., Wang, Q.X., Huang, H., Xia, H.J., and Yang, Z.N. (2007). Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in *Arabidopsis*. Plant J **52,** 528-538.

Zhao, D.Z., Wang, G.F., Speal, B., and Ma, H. (2002). The *excess microsporocytes1* gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the *Arabidopsis* anther. Genes Dev **16,** 2021-2031.

Appendices

Appendix A. Amino acid sequences used in construction of the phylogenetic tree shown in Figure 3.2.

Data was reprinted from (Kim et al., 2010) with permission of the American Society of Plant Biologists (Copyright © American Society of Plant Biologists).

>**PKSB**
MGSIDAAVLGSEKKSNPGKATILALGKAFPHQLVMQEYLVDGYFKTTKCDDPELKQKLTRLCKTTTVKTRYVVMSEEILK
KYPELAIEGGSTVTQRLDICNDAVTEMAVEASRACIKNWGRSISDITHVVYVSSSEARLPGGDLYLAKGLGLSPDTHRVL
LYFVGCSGGVAGLRVAKDIAENNPGSRVLLATSETTIIGFKPPSVDRPYDLVGVALFGDGAGAMIIGSDPDPICEKPLFE
LHTAIQNFLPETEKTIDGRLTEQGINFKLSRELPQIIEDNVENFCKKLIGKAGLAHKNYNQMFWAVHPGGPAILNRIEKR
LNLSPEKLSPSRRALMDYGNASSNSIVYVLEYMLEESKKVRNMNEEENEWGLILAFGPGVTFEGIIARNLDV*

>**PpCHS10**
MASRRVEAAFDGQAVELGATIPAANGNGTHQSIKVPGHRQVTPGKTTIMAIGRAVPANTTFNDGLADHYIQEFNLQDPVL
QAKLRRLCETTTVKTRYLVVNKEILDEHPEFLVDGAATVSQRLAITGEAVTQLGHEAATAAIKEWGRPASEITHLVYVSS
SEIRLPGGDLYLAQLLGLRSDVNRVMLYMLGCYGGASGIRVAKDLAENNPGSRVLLITSECTLIGYKSLSPDRPYDLVGA
ALFGDGAAAMIMGKDPIPVLERAFFELDWAGQSFIPGTNKTIDGRLSEEGISFKLGRELPKLIESNIQGFCDPILKRAGG
LKYNDIFWAVHPGGPAILNAVQKQLDLAPEKLQTARQVLRDYGNISSSTCIYVLDYMRHQSLKLKEANDNVNTEPEWGLL
LAFGPGVTIEGALLRNLC*

>**SlChs1**
mgfeniklngmgkkptpgkatvlslgkgfphtlvmqeflvdgyfrntncddpelkqkltrlcktttvktryvvmsdeilk
kcpelamagqatvkqrldicndavtemaidaskacisdwgrpisdithlvyvsssearlpggdlylakglglspetnrvm
lyfsgcsggvagfrvakdiaennpgsrvllatsettiigfkppnpdrpydlvgvalfgdgagamiigsdpnssenplfel
htaiqhflpdtekiidgrlteegisftldralpqiiednieafcdklmssvgltskdyndmfwavhpggpailnrlekrl
dlspdklsasrraltdygnassntivyvmeymieeglkrkngdkndndwglilafgpgl

>**PrCHS1**
msasngtngvvavksrrqhrpgkttamafgrafpdqlvmqeflvdgyfrntncqdpvlrqklerlcktttvktryvvmsd
eilaqhpelavegsatvrqrleisnvavtdmavdacrdclkewgrpvseithlvyvssseirlpggdlylasrlglrsdv
srvmlyflgcyggvtglrvakdlaennpgsrvllatsettilgfrppnperpydlvgaalfgdgaaamvlgtdprpeage
qgfleldwavqqflpdthgtingrlteeginfklgrelpqiiedhiegfcrklmdkagvddynelfwgvhpggpailnrl
ekklslgpeklyysrqaladygnassntivyvldamrqlkggekqspewglilafgpgitfegilarslv

>**OsCHSL2**
MVSTNAGGIASKQASSMAPNPGKATILALGHAFPQQLVMQDYVVDGFMRNTNCDDPELKEKLTRLCTVPDPNLIICSYKY
IYSTIIELACKTTTVKTRYVVMSEEILKSYPELAQEGQPTMKQRLDISNKAVTQMATEASLACVRSWGGALSEITHLVYV
SSSEARFPGGDLHLARALGLSPDVRRVMLAFTGCSGGVAGLRVAKGLAESCPGARVLLATSETTIVGFRPPSPDRPYDLV
GVALFGDGAGAAVVGADPTPVERPLFELHSALQRFLPDTDKTIDGRLTEEGIKFQLGRELPHIIEANVEAFCQKLMQEHP
QAADKLTYGDMFWAVHPGGPAILTKMEGRLGLDGGKLRASRSALRDFGNASSNTIVYVLENMVEETRQRREEAAEEEDCE
WGLILAFGPGITFEGILARNLQARARARD*

>OsCHSL1
MADLGFGDARSGNGSRSQCSRGKAMLLALGKGLPEQVLPQEKVVETYLQDTICDDPATRAKLERLCKTTTVRTRYTVMSK
ELLDEHPELRTEGTPTLTPRLDICNAAVLELGATAARAALGEWGRPAADITHLVYISSSELRLPGGDLFLATRLGLHPNT
VRTSLLFLGCSGGAAALRTAKDIAENNPGSRVLVVAAETTVLGFRPPSPDRPYDLVGAALFGDGASAAIIGAGPIAAEES
PFLELQFSTQEFLPGTDKVIDGKITEEGINFKLGRDLPEKIENRIEGFCRTLMDRVGIKEFNDVFWAVHPGGPAILNRLE
VCLELQPEKLKISRKALMNYGNVSSNTVFYVLEYLRDELKKGMIREEWGLILAFGPGITFEGMLVRGIN*

>**PKSA**
msnsrmngveklsskstrrvanagkatllalgkafpsqvvpqenlvegflrdtkcddafikeklehlcktttvktrytvl
treilakypelttegsptikqrleianeavvemaleaslgcikewgrpvedithivyvssseirlpggdlylsaklglrn
dvnrvmlyflgcyggvtglrvakdiaennpgsrvllttsettilgfrppnkarpydlvgaalfgdgaaaviigadprece
apfmelhyavqqflpgtqnvidgrlteeginfklgrdlpqkieenieefckklmgkagdesmefndmfwavhpggpailn
rletklklekeklessrralvdygnvssntilyvmeymrdelkkkgdaaqewglglafgpgitfegllirsl

>**OsCHS1**
MAAAVTVEEVRRAQRAEGPATVLAIGTATPANCVYQADYPDYYFRITKSEHMVELKEKFKRMCDKSQIRKRYMHLTEEIL
QENPNMCAYMAPSLDARQDIVVVEVPKLGKAAAQKAIKEWGQPRSRITHLVFCTTSGVDMPGADYQLAKMLGLRPNVNRL
MMYQQGCFAGGTVLRVAKDLAENNRGARVLAVCSEITAVTFRGPSESHLDSMVGQALFGDGAAAVIVGSDPDEAVERPLF
QMVSASQTILPDSEGAIDGHLREVGLTFHLLKDVPGLISKNIERALGDAFTPLGISDWNSIFWVAHPGGPAILDQVEAKV
GLDKERMRATRHVLSEYGNMSSACVLFILDEMRKRSAEDGHATTGEGMDWGVLFGFGPGLTVETVVLHSVPITAGAAA*

>**NSCHSLK**
mgkafpaqlvpqdclvegyirdtncqdlaikeklerlcktttvktrytvmskeildkypelategtptikqrleianpav
vemakqasqacikewgrsaeeithivyvssseirlpggdlylatelglrndigrvmlyflgcyggvtglrvakdiaennp
gsrvllttsettilgfrppnkarpydlvgaalfgdgaaaviigtepimgkespfmelnfatqqflpgtnnvidgrlteeg
infklgrdlpekiqdnieefckkiiakadlreakyndlfwavhpggpailnrlentlklqsekldcsrralmdygnvssn
tifyvmeymreelknkknggeewglalafgpgitfegillrsl

>**AtCHSL2**
mlvsarvekqkrvayqgkatvlalgkalpsnvvsqenlveeylreikcdnlsikdklqhlcksttvktrytvmsretlhk
ypelategsptikqrleiandavvqmayeaslvcikewgravedithlvyvsssefrlpggdlylsaqlglsnevqrvml
yflgcygglsglrvakdiaennpgsrvllttsettvlgfrppnkarpynlvgaalfgdgaaaliigadptesespfmelh
camqqflpqtqgvidgrlseegitfklgrdlpqkiednveefckklvakagsgalelndlfwavhpggpailsgletklk
lkpeklecsrralmdygnvssntifyimdkvrdelekkgtegeewglglafgpgitfegflmrnl

>**AtCHS**
MVMAGASSLDEIRQAQRADGPAGILAIGTANPENHVLQAEYPDYYFRITNSEHMTDLKEKFKRMCDKSTIRKRHMHLTEE
FLKENPHMCAYMAPSLDTRQDIVVVEVPKLGKEAAVKAIKEWGQPKSKITHVVFCTTSGVDMPGADYQLTKLLGLRPSVK
RLMMYQQGCFAGGTVLRIAKDLAENNRGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFSDGAAALIVGSDPDTSVGEK
PIFEMVSAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIVKSLDEAFKPLGISDWNSLFWIAHPGGPAILDQVE
IKLGLKEEKMRATRHVLSEYGNMSSACVLFILDEMRRKSAKDGVATTGEGLEWGVLFGFGPGLTVETVVLHSVPL*

>**PpCHS**
MASAGDVTRVALPRGQPRAEGPACVLGIGTAVPPAEFLQSEYPDFFFNITNCGEKEALKAKFKRICDKSGIRKRHMFLTE
EVLKANPGICTYMEPSLNVRHDIVVVQVPKLAAEAAQKAIKEWGGRKSDITHIVFATTSGVNMPGADHALAKLLGLKPTV
KRVMMYQTGCFGGASVLRVAKDLAENNKGARVLAVASEVTAVTYRAPSENHLDGLVGSALFGDGAGVYVVGSDPKPEVEK
PLFEVHWAGETILPESDGAIDGHLTEAGLIFHLMKDVPGLISKNIEKFLNEARKPVGSPAWNEMFWAVHPGGPAILDQVE
AKLKLTKDKMQGSRDILSEFGNMSSASVLFVLDQIRHRSVKMGASTLGEGSEFGFFIGFGPGLTLEVLVLRAAPNSA*

>**OsCHS2**
MVTSTVKLEEVRRMQRAEGMAAVLAIGTATPANCVYQTDYPDYYFRVTNSEHLTNLKERFQRMCESSQIRKRYTHLTEEI
LQENPSMCVFTAPSLDARQDMVVAEVPKLGKAAAEEAIKEWGQPMSRITHLVFCTTNGVDMPGADYQVAKMLGLPTSVKR
LMMYQQGCFAGGTVLRVAKDLAENNRGARVLVVCSEIMAMAFRGPSESHLDSLVGHALFGDGAAAVIVGSDPDEAADERP
LFQIVSASQTILPGTEDAIVGHLREVGLTFHLPKDVPEFISDSVEGALTDAFMPLGVHDWNSIFWVVHPGGPAILDQVEE
KVALHKARMRASRNVLSEYGNMASATVLFVLDEMRKLSADDGHATTGEGMDWGVLFGFGPGLTVETIVLHSVPITAAAPL
IMQ*

>**PoptrCHSL7**
MGYEQIVQGGLTTKANPGKATILALGKAFPHQLVMQEFLVDGYFKNTNCDDLELKQKLTRLCKTTTVKTRYVVMSDEILK
KYPELAIEGLPTVKQRLDICNDAVTRMAIDASRACIKKWGRPVSDITHLVYVSSSEARLPGGDLYLAGGLGLSPETQRVM
LYFAGCSGGVAGLRVAKDIAENNPGSRVLLATSETTIIGFKPPSADRPYDLVGVALFGDGAGAMIVGTDPIPVTESPLFE
LHTAIQNFLPNTEKTIDGRLTEEGISFKLSRELPQIIEDNIEGFCHKLIGNAGLTDKDYNKMFWAVHPGGPAILNRMEKR
FDLLPDKLNASRRALMDYGNASSNTIVYVLEYMIEECRKMNGRL*

>**PoptrCHSL6**
MGSEQIGQGGLTSKASPGKATILALGKAFPHQLVMQEFLVDGYFKNTNCDDPELKQKLTRLCKTTTVKTRYVVMSDEILN
KYPELAIEGIPTIKQRLDICNDAVTQMAIGASRACIKKWGRSVSDITHMVYVSSSEARLPGGDLYLAGGLGLSPETQRVM
LYFSGCSGGVAGLRVAKDIAENNPGSRVLLATSETTIIGFKPPSVDRPYDLVGVALFGDGAGAMVIGTDPVPVTESPLFE
LHTAIQNFLPNTEKTIDGRLTEEGISFKLARELPQIIEDNIEGFCHKLIGVAGLTDKDYNKMFWAVHPGGPAILNRMEKR
LDLLPDKLNASRRALMDYGNASSNTIVYVLEYMIEESRKMKAGAANCDWGLILAFGPGITFEGILARNLTI*

>**PoptrCHS1**
MAPSIEEIRKAQRASGPATILAIGKATPANCVSQADYPDYYFRITNSEHMTELKEKFKRMCDKSMIKKRYMHLTEEILKE
NSSMCEYMAPSLDARQDMVVVEVPKLGKEAAAKAIKEWGQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRSSVKRFMM
YQQGCFAGGTVLRLAKDLAENNKGSRVLVVCSEITAVTFRGPSDTHLDSMVGQALFGDGAAAVIVGADPDTSIERPLFQI
VSAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLVEAFAPIGINDWNSIFWIAHPGGPAILDQVEIKLDL
KEEKLRATRNVLSDYGNMSSACVLFILDEMRNKSLEEGKSTTGEGLEWGVLFGFGPGLTVETVVLHSVPVEQTIYS*

>**PoptrCHS2**
MVTVDEIRKAQRAEGPATILAIGTSTPPNCVDQSTYPDYYFRITNSEHKVELKEKFKRMCEKSMIKKRYMHLTEEILKEN
PSVCEYMAPSLDARQDMVVVEIPKLGKEAAAKAIKEWGQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRSSVKRFMMY
QQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAAIIIGSDPVLGVEKPLFELV
SAAQTILPDSEGAIDGHLREVGLTFHLLKDVPGLISKNVEKSLTEAFKPLGISDWNSLFWIAHPGGPAILDQVEAKLELK
PEKLRATRQVLADYGNMSSACVLFILDEMRKKSAKDGLKSTGEGLEWGVLFGFGPGLTVETVVLHSLPATI*

>**PoptrCHS3**
MVTVDEVRKAQRAEGPAVILAIGTSTPPNCVDQSTYPDYYFRITNSEHKVELKEKFKRMCEKSMIKKRYMHLTEEILKEN
PSVCEYMAPSLDARQDMVVVEVPKLGKEAAAKAIKEWGQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRSSVKRFMMY
QQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAAIIIGSDPVLGVEKPLFELV
SAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNIEKSLTEAFKPLGISDWNSLFWIAHPGGPAILDQVEAKLELK
PEKLRATRQVLADYGNMSSACVLFILDEMRKKSAKDGLKSTGEGLEWGVLFGFGPGLTVETVVLHSVASI*

>**PoptrCHSL4**
MSESDSNGASKHCTTPSRRAPTLGKATLLAIGKAFPSQLIPQECLVEGYIRDTKCDDASIKEKLERLCKTTTVKTRYTVM
SREILDKYPELATEGTPTIRQRLEIANPAVVEMALKASMACINEWGGSVEDITHIVYVSSSEVRLPGGDLYLASQLGLRN
DVGRVMLYFLGCYGGVTGLRVAKDIAENNPGSRVLLTTSETTILGFRPPSKARPYDLVGAALFGDGAAAVIIGANPVIGK
ESPFMELNYSVQQFLPGTQNVIDGRLSEEGIHFKLGRDLPQKIEDNIEEFCNKLMSKAGLTDFNELFWAVHPGGPAILNR
LESKLKLNEEKLECSRRALMDYGNVSSNTIVYVLEYMRDELKRGGGEWGLALAFGPGITFEGILLRSL*

>**PoptrCHSL2**
MALVDEIRKAQRARGPAMVLAIGTAVPVNCFYQADYPDYFFRVTKTENLTELKAKFERICQKSMINKRYMHLTEEMIKEN
PEIGNFMTPSLNVRQDIVLAEVPKLGKEAALKAIQEWGHPMSKITHLVFCTTSGVHMPGADYQLANLLGLSSSIKRLMLY
QQGCYGGGTALRVAKDLAENNAGARVLVVCSEITAITFHAPNEDQLGCLVGQALFGDGAGAAIIGSDPDTLVEKPIFQLV
SAAQIMIPDSEHAIEGHVREMGLLIHLSEDVPKLISDNVEAALREVVTPIGGVLSDWNSLFWAVHAGGRAILDGVEAKLK
LKKEKLGVTRHILREYGNVASACVLFVLDEMRERSVREGKATTGEGLEWGVVIGLGPGLTMETLVLHSVPVAITK*

>**PoptrCHS4**
MVTVDEVRKGQRAEGPATIMAIGTSNPPNCVDQSTYPDYYFRVTNSEHRAELKEKFKRMCEKSMIKKRYIYLTEDMLKEN
PDMRAYMAPSLDARQDMVVVEVPKLGKEAATKAIKEWGQSKSKITHLVFCTTSGVDMPGADYQLTKLLGLRPSVKRLMMY
QQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAIIFRGPNDTHLDSLIGQALFGDGAAAIIIGSDPVVGVEKPLFEIV
SAAQTILPNSAGAIDGHLREAGLTFHLLKDVPGLISNNVEKSLTEAFKPLGISDWNSLFWIAHPGGPAILDQVEAKLGLK
PEKLRATRHVLSEYGNMSSACVLFILDEMRKKSAEDGLQSTGEGLEWGVLFGFGPGLTVETVVLHSVATRV*

>**PoptrCHS5**
MVTVDEIRKTQRAEGPATIMAIGTSTPPNCVDQSAYPDYYFRITNSEHKAELKEKFKRMCEKSMIKKRYMYLTEEILKEN
PSVCEYMAPSLDARQDMVVVEVPRLGKEAATKAIKEWGQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRSSVKRFMMY
QQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAALIIGSDPVIGVEKPLFELV
SAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNVEKSLTEAFKPLGISDWNSLFWIAHPGGPAILDQVEAKLALK
PEKLRATRHVLSEYGNMSSACVLFILDEMRKKSAEDGLQSTGEGLEWGVLFGFGPGLTVETVVLHSVAPTI*

>**PoptrCHS6**
MVTVDEIRKSQRAEGPATIMAIGTSTPPNCVDQSTYPDYYFRITNSEHKAELKEKFKRMCEKSMIKKRYMYLTEEILKEN
PSVCEYMAPSLDARQDMVVVEVPKLGKEAATKAIKEWGQPKSKITHLVFCTTSGVDMPGADYQLTKLLGLRSSVKRFMMY
QQGCFAGGTVLRLAKDLAENNKGARVLVVCSEITAVTFRGPSDTHLDSLVGQALFGDGAAALVIGSDPVIGVEKPLFELV
SAAQTILPDSDGAIDGHLREVGLTFHLLKDVPGLISKNVEKSLTEAFKPLGISDWNSLFWIAHPGGPAILDQVEAKLALK
PEKLRATRHVLSEYGNMSSACVLFILDEMRKKSAEDGLQSTGEGLEWGVLFGFGPGLTVETVVLHSVAPTI*

>**PoptrCHSL3**
MASDQTSQGAQAAQSPATILAIGTANPANFIYQADYPDYYFRVTRSEHMTDLKGKFKRLCEKSEVRKRHFHLTEEILNKN
PTMCTYDGPSLDVRQDVLVTEVPKLGMEAALKAIEEWGRPKSNITHLIFSALAGIDMPGADYQLTRLLGLEPSIKRIMLY
HQGCNIGAATLRIAKDFAENNAGARVLVVSSDLTVGTFRGPSNDNISCLVAQAITGEGAAALIIGADPDMSVERPLFQIL
SASQTIIPDSNDGINGHLREVGLTVHFSRNVPELISRNIGKCLVEAFGPIGVSDWNSLFWIVQPSGAAILNLIEAEVGLA
QEKLSATRHVLSEFGNMGGPTVLFILDEIRRRSLEKRKTTTGEGMEWGVLIGLGAGITVDTVVLHSVPIAEGR*

>**PoptrCHSL1**
MASILAIGTANPPNCFDQADYPDFYFRVTKSEHMTQLKDKFKRICEKSKIRKRYMYITEDTIKKNPSLSTYDAASLDARQ
EILVTEVPKLGKEAALKAIEEWGQPKSKITHLIFCTSSGTHMPGADHELTKLLGLERSVKRFMMYQQGCFTAALALRLSK
DLAENNPGARVLIVCSENMTVCFRAPSETHLDILVGSAIFSDGAAAIIVGADPDTATERPLFQLVSAEQCIVPDSDDGIV
GHIREMGISYYLHKMVPKIVAEGAAQCLVETFNARYGIKDWNSLFYVVHPGGTGVLNKFEEHIGLTKDKLRASRHVLSEY
GNMWGPSMFFVLDEMRRRSAKEGKATTGEGLDLGVLFGFGPGVTIETIVLRSFATD*

>**PoptrCHSL5**
MSKTIGNGASKHYATLTRRSPTPGKATILATGKAFPSQLVPQECLVEGYMRDTKCDDASIKEKLERLCKTTTVKTRYTVM
SKEILEKYPELATEGSPTIKQRLEIANPAVVEMALKASIACINEWGGSVKDITHVVYVSSSEIRLPGGDLYLASQLGLRN
DVGRVMLYFLGCYGGVTGLRVAKDIAENNPGSRILLTTSETTILGFRPPNKARPYDLVGAALFGDGAAAVIIGADPVIGK
ESPFMELSYAVQQFLPGTQNVIDGRLSEEGINFKLGRDLPQKIEDNIEEFCRKLMSKAGLTEFNDLFWAVHPGGPAILNR
LESNLKLNTEKLECSRRALINYGNVSSNTIVYVLEYMKEELKREGGEEWGLALAFGPGITFEGILLRSL*

Appendix B. Alignment of amino acid sequences used to generate the phylogenetic tree shown in Figure 3.2.

Data was reprinted from (Kim et al., 2010) with permission of the American Society of Plant Biologists (Copyright © American Society of Plant Biologists).

```
                ....|....| ....|....| ....|....| ....|....| ....|....|
 ....|....|
                    5           15          25          35          45
55
PpCHS10         MASRRVEAAF DGQAVELGAT IPAANGNGTH QSIKVPGHRQ VTPGKTTIMA
IGRAVPANTT
OsCHSL1         ---------- ---------- --MADLGFGD ARSGNGSRSQ CSRGKAMLLA
LGKGLPEQVL
PrCHS1          ---------- ---------- --MSASNGTN GVVAVKSRRQ HRPGKTTAMA
FGRAFPDQLV
AtCHSL2         ---------- ---------- -------MLV SARVEKQKRV AYQGKATVLA
LGKALPSNVV
PKSA            ---------- ---------- MSNSRMNGVE KLSSKSTRRV ANAGKATLLA
LGKAFPSQVV
NSCHSLK         ---------- ---------- ---------- ---------- ----------
MGKAFPAQLV
PoptrCHSL4      ---------- ---------- MSESDSNGAS KHCTTPSRRA PTLGKATLLA
IGKAFPSQLI
PoptrCHSL5      ---------- ---------- MSKTIGNGAS KHYATLTRRS PTPGKATILA
TGKAFPSQLV
OsCHSL2         ---------- ---------- --MVSTNAGG IASKQASSMA PNPGKATILA
LGHAFPQQLV
SlChsl          ---------- ---------- ------MGFE NIKLNGMGKK PTPGKATVLS
LGKGFPHTLV
PKSB            ---------- ---------- ------MGSI DAAVLGSEKK SNPGKATILA
LGKAFPHQLV
PoptrCHSL7      ---------- ---------- ------MGYE QIVQGGLTTK ANPGKATILA
LGKAFPHQLV
PoptrCHSL6      ---------- ---------- ------MGSE QIGQGGLTSK ASPGKATILA
LGKAFPHQLV
PoptrCHSL1      ---------- ---------- ---------- ---------- ----MASILA
IGTANPPNCF
PpCHS           ---------- ---------- ---MASAGDV TRVALPRGQP RAEGPACVLG
IGTAVPPAEF
PoptrCHSL3      ---------- ---------- ---------M ASDQTSQGAQ AAQSPATILA
IGTANPANFI
PoptrCHSL2      ---------- ---------- ---------M ALVDEIRKAQ RARGPAMVLA
IGTAVPVNCF
OsCHS2          ---------- ---------- -----MVTST VKLEEVRRMQ RAEGMAAVLA
IGTATPANCV
OsCHS1          ---------- ---------- ------MAAA VTVEEVRRAQ RAEGPATVLA
IGTATPANCV
AtCHS           ---------- ---------- ----MVMAGA SSLDEIRQAQ RADGPAGILA
IGTANPENHV
PoptrCHS1       ---------- ---------- --------MA PSIEEIRKAQ RASGPATILA
```

```
IGKATPANCV
PoptrCHS4     ---------- ---------- ---------M VTVDEVRKGQ RAEGPATIMA
IGTSNPPNCV
PoptrCHS5     ---------- ---------- ---------M VTVDEIRKTQ RAEGPATIMA
IGTSTPPNCV
PoptrCHS6     ---------- ---------- ---------M VTVDEIRKSQ RAEGPATIMA
IGTSTPPNCV
PoptrCHS2     ---------- ---------- ---------M VTVDEIRKAQ RAEGPATILA
IGTSTPPNCV
PoptrCHS3     ---------- ---------- ---------M VTVDEVRKAQ RAEGPAVILA
IGTSTPPNCV


              ....|....| ....|....| ....|....| ....|....| ....|....|
 ....|....|
                  65          75          85          95         105
115
PpCHS10       FNDGLADHYI QEFNLQD-PV LQAKLRRL-- ---------- ----------
--CETTTVKT
OsCHSL1       PQEKVVETYL QDTICDD-PA TRAKLERL-- ---------- ----------
--CKTTTVRT
PrCHS1        MQEFLVDGYF RNTNCQD-PV LRQKLERL-- ---------- ----------
--CKTTTVKT
AtCHSL2       SQENLVEEYL REIKCDN-LS IKDKLQHL-- ---------- ----------
--CKSTTVKT
PKSA          PQENLVEGFL RDTKCDD-AF IKEKLEHL-- ---------- ----------
--CKTTTVKT
NSCHSLK       PQDCLVEGYI RDTNCQD-LA IKEKLERL-- ---------- ----------
--CKTTTVKT
PoptrCHSL4    PQECLVEGYI RDTKCDD-AS IKEKLERL-- ---------- ----------
--CKTTTVKT
PoptrCHSL5    PQECLVEGYM RDTKCDD-AS IKEKLERL-- ---------- ----------
--CKTTTVKT
OsCHSL2       MQDYVVDGFM RNTNCDD-PE LKEKLTRLCT VPDPNLIICS YKYIYSTIIE
LACKTTTVKT
SlChsl        MQEFLVDGYF RNTNCDD-PE LKQKLTRL-- ---------- ----------
--CKTTTVKT
PKSB          MQEYLVDGYF KTTKCDD-PE LKQKLTRL-- ---------- ----------
--CKTTTVKT
PoptrCHSL7    MQEFLVDGYF KNTNCDD-LE LKQKLTRL-- ---------- ----------
--CKTTTVKT
PoptrCHSL6    MQEFLVDGYF KNTNCDD-PE LKQKLTRL-- ---------- ----------
--CKTTTVKT
PoptrCHSL1    DQADYPDFYF RVTKSEHMTQ LKDKFKRI-- ---------- ----------
--CEKSKIRK
PpCHS         LQSEYPDFFF NITNCGEKEA LKAKFKRI-- ---------- ----------
--CDKSGIRK
PoptrCHSL3    YQADYPDYYF RVTRSEHMTD LKGKFKRL-- ---------- ----------
--CEKSEVRK
PoptrCHSL2    YQADYPDYFF RVTKTENLTE LKAKFERI-- ---------- ----------
--CQKSMINK
OsCHS2        YQTDYPDYYF RVTNSEHLTN LKERFQRM-- ---------- ----------
--CESSQIRK
OsCHS1        YQADYPDYYF RITKSEHMVE LKEKFKRM-- ---------- ----------
--CDKSQIRK
AtCHS         LQAEYPDYYF RITNSEHMTD LKEKFKRM-- ---------- ----------
--CDKSTIRK
PoptrCHS1     SQADYPDYYF RITNSEHMTE LKEKFKRM-- ---------- ----------
--CDKSMIKK
PoptrCHS4     DQSTYPDYYF RVTNSEHRAE LKEKFKRM-- ---------- ----------
--CEKSMIKK
PoptrCHS5     DQSAYPDYYF RITNSEHKAE LKEKFKRM-- ---------- ----------
```

```
--CEKSMIKK
PoptrCHS6     DQSTYPDYYF RITNSEHKAE LKEKFKRM-- ---------- ----------
--CEKSMIKK
PoptrCHS2     DQSTYPDYYF RITNSEHKVE LKEKFKRM-- ---------- ----------
--CEKSMIKK
PoptrCHS3     DQSTYPDYYF RITNSEHKVE LKEKFKRM-- ---------- ----------
--CEKSMIKK



                ....|....| ....|....| ....|....| ....|....| ....|....|
 ....|....|
                   125         135         145         155         165
175
PpCHS10       RYLVVNKEIL DEHPEFLVDG AATVSQRLAI TGEAVTQLGH EAATAAIKEW
GRPASEITHL
OsCHSL1       RYTVMSKELL DEHPELRTEG TPTLTPRLDI CNAAVLELGA TAARAALGEW
GRPAADITHL
PrCHS1        RYVVMSDEIL AQHPELAVEG SATVRQRLEI SNVAVTDMAV DACRDCLKEW
GRPVSEITHL
AtCHSL2       RYTVMSRETL HKYPELATEG SPTIKQRLEI ANDAVVQMAY EASLVCIKEW
GRAVEDITHL
PKSA          RYTVLTREIL AKYPELTTEG SPTIKQRLEI ANEAVVEMAL EASLGCIKEW
GRPVEDITHI
NSCHSLK       RYTVMSKEIL DKYPELATEG TPTIKQRLEI ANPAVVEMAK QASQACIKEW
GRSAEEITHI
PoptrCHSL4    RYTVMSREIL DKYPELATEG TPTIRQRLEI ANPAVVEMAL KASMACINEW
GGSVEDITHI
PoptrCHSL5    RYTVMSKEIL EKYPELATEG SPTIKQRLEI ANPAVVEMAL KASIACINEW
GGSVKDITHV
OsCHSL2       RYVVMSEEIL KSYPELAQEG QPTMKQRLDI SNKAVTQMAT EASLACVRSW
GGALSEITHL
SlChsl        RYVVMSDEIL KKCPELAMAG QATVKQRLDI CNDAVTEMAI DASKACISDW
GRPISDITHL
PKSB          RYVVMSEEIL KKYPELAIEG GSTVTQRLDI CNDAVTEMAV EASRACIKNW
GRSISDITHV
PoptrCHSL7    RYVVMSDEIL KKYPELAIEG LPTVKQRLDI CNDAVTRMAI DASRACIKKW
GRPVSDITHL
PoptrCHSL6    RYVVMSDEIL NKYPELAIEG IPTIKQRLDI CNDAVTQMAI GASRACIKKW
GRSVSDITHM
PoptrCHSL1    RYMYITEDTI KKNPSLSTYD AASLDARQEI LVTEVPKLGK EAALKAIEEW
GQPKSKITHL
PpCHS         RHMFLTEEVL KANPGICTYM EPSLNVRHDI VVVQVPKLAA EAAQKAIKEW
GGRKSDITHI
PoptrCHSL3    RHFHLTEEIL NKNPTMCTYD GPSLDVRQDV LVTEVPKLGM EAALKAIEEW
GRPKSNITHL
PoptrCHSL2    RYMHLTEEMI KENPEIGNFM TPSLNVRQDI VLAEVPKLGK EAALKAIQEW
GHPMSKITHL
OsCHS2        RYTHLTEEIL QENPSMCVFT APSLDARQDM VVAEVPKLGK AAAEEAIKEW
GQPMSRITHL
OsCHS1        RYMHLTEEIL QENPNMCAYM APSLDARQDI VVVEVPKLGK AAAQKAIKEW
GQPRSRITHL
AtCHS         RHMHLTEEFL KENPHMCAYM APSLDTRQDI VVVEVPKLGK EAAVKAIKEW
GQPKSKITHV
PoptrCHS1     RYMHLTEEIL KENSSMCEYM APSLDARQDM VVVEVPKLGK EAAAKAIKEW
GQPKSKITHL
PoptrCHS4     RYIYLTEDML KENPDMRAYM APSLDARQDM VVVEVPKLGK EAATKAIKEW
GQSKSKITHL
PoptrCHS5     RYMYLTEEIL KENPSVCEYM APSLDARQDM VVVEVPRLGK EAATKAIKEW
GQPKSKITHL
```

```
PoptrCHS6     RYMYLTEEIL KENPSVCEYM APSLDARQDM VVVEVPKLGK EAATKAIKEW
GQPKSKITHL
PoptrCHS2     RYMHLTEEIL KENPSVCEYM APSLDARQDM VVVEIPKLGK EAAAKAIKEW
GQPKSKITHL
PoptrCHS3     RYMHLTEEIL KENPSVCEYM APSLDARQDM VVVEVPKLGK EAAAKAIKEW
GQPKSKITHL



                ....|....| ....|....| ....|....| ....|....| ....|....|
 ....|....|
                   185         195         205         215         225
235
PpCHS10       VYVSSSEIRL PGGDLYLAQL LGLRSDVNRV MLYMLGCYGG ASGIRVAKDL
AENNPGSRVL
OsCHSL1       VYISSSELRL PGGDLFLATR LGLHPNTVRT SLLFLGCSGG AAALRTAKDI
AENNPGSRVL
PrCHS1        VYVSSSEIRL PGGDLYLASR LGLRSDVSRV MLYFLGCYGG VTGLRVAKDL
AENNPGSRVL
AtCHSL2       VYVSSSEFRL PGGDLYLSAQ LGLSNEVQRV MLYFLGCYGG LSGLRVAKDI
AENNPGSRVL
PKSA          VYVSSSEIRL PGGDLYLSAK LGLRNDVNRV MLYFLGCYGG VTGLRVAKDI
AENNPGSRVL
NSCHSLK       VYVSSSEIRL PGGDLYLATE LGLRNDIGRV MLYFLGCYGG VTGLRVAKDI
AENNPGSRVL
PoptrCHSL4    VYVSSSEVRL PGGDLYLASQ LGLRNDVGRV MLYFLGCYGG VTGLRVAKDI
AENNPGSRVL
PoptrCHSL5    VYVSSSEIRL PGGDLYLASQ LGLRNDVGRV MLYFLGCYGG VTGLRVAKDI
AENNPGSRIL
OsCHSL2       VYVSSSEARF PGGDLHLARA LGLSPDVRRV MLAFTGCSGG VAGLRVAKGL
AESCPGARVL
SlChsl        VYVSSSEARL PGGDLYLAKG LGLSPETNRV MLYFSGCSGG VAGFRVAKDI
AENNPGSRVL
PKSB          VYVSSSEARL PGGDLYLAKG LGLSPDTHRV LLYFVGCSGG VAGLRVAKDI
AENNPGSRVL
PoptrCHSL7    VYVSSSEARL PGGDLYLAGG LGLSPETQRV MLYFAGCSGG VAGLRVAKDI
AENNPGSRVL
PoptrCHSL6    VYVSSSEARL PGGDLYLAGG LGLSPETQRV MLYFSGCSGG VAGLRVAKDI
AENNPGSRVL
PoptrCHSL1    IFCTSSGTHM PGADHELTKL LGLERSVKRF MMYQQGCFTA ALALRLSKDL
AENNPGARVL
PpCHS         VFATTSGVNM PGADHALAKL LGLKPTVKRV MMYQTGCFGG ASVLRVAKDL
AENNKGARVL
PoptrCHSL3    IFSALAGIDM PGADYQLTRL LGLEPSIKRI MLYHQGCNIG AATLRIAKDF
AENNAGARVL
PoptrCHSL2    VFCTTSGVHM PGADYQLANL LGLSSSIKRL MLYQQGCYGG GTALRVAKDL
AENNAGARVL
OsCHS2        VFCTTNGVDM PGADYQVAKM LGLPTSVKRL MMYQQGCFAG GTVLRVAKDL
AENNRGARVL
OsCHS1        VFCTTSGVDM PGADYQLAKM LGLRPNVNRL MMYQQGCFAG GTVLRVAKDL
AENNRGARVL
AtCHS         VFCTTSGVDM PGADYQLTKL LGLRPSVKRL MMYQQGCFAG GTVLRIAKDL
AENNRGARVL
PoptrCHS1     VFCTTSGVDM PGADYQLTKL LGLRSSVKRF MMYQQGCFAG GTVLRLAKDL
AENNKGSRVL
PoptrCHS4     VFCTTSGVDM PGADYQLTKL LGLRPSVKRL MMYQQGCFAG GTVLRLAKDL
AENNKGARVL
PoptrCHS5     VFCTTSGVDM PGADYQLTKL LGLRSSVKRF MMYQQGCFAG GTVLRLAKDL
AENNKGARVL
PoptrCHS6     VFCTTSGVDM PGADYQLTKL LGLRSSVKRF MMYQQGCFAG GTVLRLAKDL
AENNKGARVL
```

```
PoptrCHS2     VFCTTSGVDM PGADYQLTKL LGLRSSVKRF MMYQQGCFAG GTVLRLAKDL
AENNKGARVL
PoptrCHS3     VFCTTSGVDM PGADYQLTKL LGLRSSVKRF MMYQQGCFAG GTVLRLAKDL
AENNKGARVL



                ....|....| ....|....| ....|....| ....|....| ....|....|
 ....|....|
                   245         255         265         275         285
295
PpCHS10       LITSECTLIG YKSLSPDRPY DLVGAALFGD GAAAMIMGKD PIPVL-ERAF
FELDWAGQSF
OsCHSL1       VVAAETTVLG FRPPSPDRPY DLVGAALFGD GASAAIIGAG PIAAE-ESPF
LELQFSTQEF
PrCHS1        LATSETTILG FRPPNPERPY DLVGAALFGD GAAAMVLGTD PRPEAGEQGF
LELDWAVQQF
AtCHSL2       LTTSETTVLG FRPPNKARPY NLVGAALFGD GAAALIIGAD PTES--ESPF
MELHCAMQQF
PKSA          LTTSETTILG FRPPNKARPY DLVGAALFGD GAAAVIIGAD PREC--EAPF
MELHYAVQQF
NSCHSLK       LTTSETTILG FRPPNKARPY DLVGAALFGD GAAAVIIGTE PIMGK-ESPF
MELNFATQQF
PoptrCHSL4    LTTSETTILG FRPPSKARPY DLVGAALFGD GAAAVIIGAN PVIGK-ESPF
MELNYSVQQF
PoptrCHSL5    LTTSETTILG FRPPNKARPY DLVGAALFGD GAAAVIIGAD PVIGK-ESPF
MELSYAVQQF
OsCHSL2       LATSETTIVG FRPPSPDRPY DLVGVALFGD GAGAAVVGAD PTPV--ERPL
FELHSALQRF
SlChsl        LATSETTIIG FKPPNPDRPY DLVGVALFGD GAGAMIIGSD PNSS--ENPL
FELHTAIQHF
PKSB          LATSETTIIG FKPPSVDRPY DLVGVALFGD GAGAMIIGSD PDPIC-EKPL
FELHTAIQNF
PoptrCHSL7    LATSETTIIG FKPPSADRPY DLVGVALFGD GAGAMIVGTD PIPVT-ESPL
FELHTAIQNF
PoptrCHSL6    LATSETTIIG FKPPSVDRPY DLVGVALFGD GAGAMVIGTD PVPVT-ESPL
FELHTAIQNF
PoptrCHSL1    IVCSENMTVC FRAPSETHLD ILVGSAIFSD GAAAIIVGAD PDTAT-ERPL
FQLVSAEQCI
PpCHS         AVASEVTAVT YRAPSENHLD GLVGSALFGD GAGVYVVGSD PKPEV-EKPL
FEVHWAGETI
PoptrCHSL3    VVSSDLTVGT FRGPSNDNIS CLVAQAITGE GAAALIIGAD PDMSV-ERPL
FQILSASQTI
PoptrCHSL2    VVCSEITAIT FHAPNEDQLG CLVGQALFGD GAGAAIIGSD PDTLV-EKPI
FQLVSAAQIM
OsCHS2        VVCSEIMAMA FRGPSESHLD SLVGHALFGD GAAAVIVGSD PDEAADERPL
FQIVSASQTI
OsCHS1        AVCSEITAVT FRGPSESHLD SMVGQALFGD GAAAVIVGSD PDEAV-ERPL
FQMVSASQTI
AtCHS         VVCSEITAVT FRGPSDTHLD SLVGQALFSD GAAALIVGSD PDTSVGEKPI
FEMVSAAQTI
PoptrCHS1     VVCSEITAVT FRGPSDTHLD SMVGQALFGD GAAAVIVGAD PDTSI-ERPL
FQIVSAAQTI
PoptrCHS4     VVCSEITAII FRGPNDTHLD SLIGQALFGD GAAAIIIGSD PVVGV-EKPL
FEIVSAAQTI
PoptrCHS5     VVCSEITAVT FRGPSDTHLD SLVGQALFGD GAAALIIGSD PVIGV-EKPL
FELVSAAQTI
PoptrCHS6     VVCSEITAVT FRGPSDTHLD SLVGQALFGD GAAALVIGSD PVIGV-EKPL
FELVSAAQTI
PoptrCHS2     VVCSEITAVT FRGPSDTHLD SLVGQALFGD GAAAIIIGSD PVLGV-EKPL
FELVSAAQTI
```

```
PoptrCHS3     VVCSEITAVT FRGPSDTHLD SLVGQALFGD GAAAIIIGSD PVLGV-EKPL
FELVSAAQTI

                .....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
 .....|.....|
                   305         315         325         335         345
355
PpCHS10       IPGTNKTIDG RLSEEGISFK LGRELPKLIE SNIQGFCDPI L-KRAG--GL
K----YNDI-
OsCHSL1       LPGTDKVIDG KITEEGINFK LGRDLPEKIE NRIEGFCRTL M-DRVG--IK
E----FNDV-
PrCHS1        LPDTHGTING RLTEEGINFK LGRELPQIIE DHIEGFCRKL M-DKAG--VD
D----YNEL-
AtCHSL2       LPQTQGVIDG RLSEEGITFK LGRDLPQKIE DNVEEFCKKL V-AKAG--SG
ALE--LNDL-
PKSA          LPGTQNVIDG RLTEEGINFK LGRDLPQKIE ENIEEFCKKL M-GKAG--DE
SME--FNDM-
NSCHSLK       LPGTNNVIDG RLTEEGINFK LGRDLPEKIQ DNIEEFCKKI I-AKAD--LR
EAK--YNDL-
PoptrCHSL4    LPGTQNVIDG RLSEEGIHFK LGRDLPQKIE DNIEEFCNKL M-SKAG--LT
D----FNEL-
PoptrCHSL5    LPGTQNVIDG RLSEEGINFK LGRDLPQKIE DNIEEFCRKL M-SKAG--LT
E----FNDL-
OsCHSL2       LPDTDKTIDG RLTEEGIKFQ LGRELPHIIE ANVEAFCQKL M-QEHP--QA
ADKLTYGDM-
SlChsl        LPDTEKIIDG RLTEEGISFT LDRALPQIIE DNIEAFCDKL M-SSVG--LT
SKD--YNDM-
PKSB          LPETEKTIDG RLTEQGINFK LSRELPQIIE DNVENFCKKL I-GKAG--LA
HKN--YNQM-
PoptrCHSL7    LPNTEKTIDG RLTEEGISFK LSRELPQIIE DNIEGFCHKL I-GNAG--LT
DKD--YNKM-
PoptrCHSL6    LPNTEKTIDG RLTEEGISFK LARELPQIIE DNIEGFCHKL I-GVAG--LT
DKD--YNKM-
PoptrCHSL1    VPDSDDGIVG HIREMGISYY LHKMVPKIVA EGAAQCLVET FNARYG--IK
D----WNSL-
PpCHS         LPESDGAIDG HLTEAGLIFH LMKDVPGLIS KNIEKFLNEA R-KPVG--SP
A----WNEM-
PoptrCHSL3    IPDSNDGING HLREVGLTVH FSRNVPELIS RNIGKCLVEA F-GPIG--VS
D----WNSL-
PoptrCHSL2    IPDSEHAIEG HVREMGLLIH LSEDVPKLIS DNVEAALREV V-TPIGGVLS
D----WNSL-
OsCHS2        LPGTEDAIVG HLREVGLTFH LPKDVPEFIS DSVEGALTDA F-MPLG--VH
D----WNSI-
OsCHS1        LPDSEGAIDG HLREVGLTFH LLKDVPGLIS KNIERALGDA F-TPLG--IS
D----WNSI-
AtCHS         LPDSDGAIDG HLREVGLTFH LLKDVPGLIS KNIVKSLDEA F-KPLG--IS
D----WNSL-
PoptrCHS1     LPDSDGAIDG HLREVGLTFH LLKDVPGLIS KNIEKSLVEA F-APIG--IN
D----WNSI-
PoptrCHS4     LPNSAGAIDG HLREAGLTFH LLKDVPGLIS NNVEKSLTEA F-KPLG--IS
D----WNSL-
PoptrCHS5     LPDSDGAIDG HLREVGLTFH LLKDVPGLIS KNVEKSLTEA F-KPLG--IS
D----WNSL-
PoptrCHS6     LPDSDGAIDG HLREVGLTFH LLKDVPGLIS KNVEKSLTEA F-KPLG--IS
D----WNSL-
PoptrCHS2     LPDSEGAIDG HLREVGLTFH LLKDVPGLIS KNVEKSLTEA F-KPLG--IS
D----WNSL-
PoptrCHS3     LPDSDGAIDG HLREVGLTFH LLKDVPGLIS KNIEKSLTEA F-KPLG--IS
D----WNSL-
```

```
                 ....|....| ....|....| ....|....| ....|....| ....|....|
 ....|....|
                    365         375         385         395         405
415
PpCHS10          FWAVHPGGPA ILNAVQKQLD LAPEKLQTAR QVLRDYGNIS SSTCIYVLDY
MRHQSLKLKE
OsCHSL1          FWAVHPGGPA ILNRLEVCLE LQPEKLKISR KALMNYGNVS SNTVFYVLEY
LRDELKKGM-
PrCHS1           FWGVHPGGPA ILNRLEKKLS LGPEKLYYSR QALADYGNAS SNTIVYVLDA
MRQLKGGEK-
AtCHSL2          FWAVHPGGPA ILSGLETKLK LKPEKLECSR RALMDYGNVS SNTIFYIMDK
VRDELEKKG-
PKSA             FWAVHPGGPA ILNRLETKLK LEKEKLESSR RALVDYGNVS SNTILYVMEY
MRDELKKKG-
NSCHSLK          FWAVHPGGPA ILNRLENTLK LQSEKLDCSR RALMDYGNVS SNTIFYVMEY
MREELKNKK-
PoptrCHSL4       FWAVHPGGPA ILNRLESKLK LNEEKLECSR RALMDYGNVS SNTIVYVLEY
MRDELKRGG-
PoptrCHSL5       FWAVHPGGPA ILNRLESNLK LNTEKLECSR RALINYGNVS SNTIVYVLEY
MKEELKREG-
OsCHSL2          FWAVHPGGPA ILTKMEGRLG LDGGKLRASR SALRDFGNAS SNTIVYVLEN
MVEETRQRRE
SlChsl           FWAVHPGGPA ILNRLEKRLD LSPDKLSASR RALTDYGNAS SNTIVYVMEY
MIEEGLKRK-
PKSB             FWAVHPGGPA ILNRIEKRLN LSPEKLSPSR RALMDYGNAS SNSIVYVLEY
MLEESKKVR-
PoptrCHSL7       FWAVHPGGPA ILNRMEKRFD LLPDKLNASR RALMDYGNAS SNTIVYVLEY
MIEECRKMN-
PoptrCHSL6       FWAVHPGGPA ILNRMEKRLD LLPDKLNASR RALMDYGNAS SNTIVYVLEY
MIEESRKMK-
PoptrCHSL1       FYVVHPGGTG VLNKFEEHIG LTKDKLRASR HVLSEYGNMW GPSMFFVLDE
MRRRSAKEG-
PpCHS            FWAVHPGGPA ILDQVEAKLK LTKDKMQGSR DILSEFGNMS SASVLFVLDQ
IRHRSVKMG-
PoptrCHSL3       FWIVQPSGAA ILNLIEAEVG LAQEKLSATR HVLSEFGNMG GPTVLFILDE
IRRRSLEKR-
PoptrCHSL2       FWAVHAGGRA ILDGVEAKLK LKKEKLGVTR HILREYGNVA SACVLFVLDE
MRERSVREG-
OsCHS2           FWVVHPGGPA ILDQVEEKVA LHKARMRASR NVLSEYGNMA SATVLFVLDE
MRKLSADDG-
OsCHS1           FWVAHPGGPA ILDQVEAKVG LDKERMRATR HVLSEYGNMS SACVLFILDE
MRKRSAEDG-
AtCHS            FWIAHPGGPA ILDQVEIKLG LKEEKMRATR HVLSEYGNMS SACVLFILDE
MRRKSAKDG-
PoptrCHS1        FWIAHPGGPA ILDQVEIKLD LKEEKLRATR NVLSDYGNMS SACVLFILDE
MRNKSLEEG-
PoptrCHS4        FWIAHPGGPA ILDQVEAKLG LKPEKLRATR HVLSEYGNMS SACVLFILDE
MRKKSAEDG-
PoptrCHS5        FWIAHPGGPA ILDQVEAKLA LKPEKLRATR HVLSEYGNMS SACVLFILDE
MRKKSAEDG-
PoptrCHS6        FWIAHPGGPA ILDQVEAKLA LKPEKLRATR HVLSEYGNMS SACVLFILDE
MRKKSAEDG-
PoptrCHS2        FWIAHPGGPA ILDQVEAKLE LKPEKLRATR QVLADYGNMS SACVLFILDE
MRKKSAKDG-
PoptrCHS3        FWIAHPGGPA ILDQVEAKLE LKPEKLRATR QVLADYGNMS SACVLFILDE
MRKKSAKDG-
```

```
                ....|....| ....|....| ....|....| ....|....| ..
                   425        435        445        455
PpCHS10         ANDNVNTEPE WGLLLAFGPG VTIEGALLRN LC-------- --
OsCHSL1         ------IREE WGLILAFGPG ITFEGMLVRG IN-------- --
PrCHS1          ------QSPE WGLILAFGPG ITFEGILARS LV-------- --
AtCHSL2         -----TEGEE WGLGLAFGPG ITFEGFLMRN L--------- --
PKSA            -----DAAQE WGLGLAFGPG ITFEGLLIRS L--------- --
NSCHSLK         -----NGGEE WGLALAFGPG ITFEGILLRS L--------- --
PoptrCHSL4      --------GE WGLALAFGPG ITFEGILLRS L--------- --
PoptrCHSL5      -------GEE WGLALAFGPG ITFEGILLRS L--------- --
OsCHSL2         -EAAEEEDCE WGLILAFGPG ITFEGILARN LQARARARD- --
SlChsl          -NGD-KNDND WGLILAFGPG L--------- ---------- --
PKSB            --NMNEEENE WGLILAFGPG VTFEGIIARN LDV------- --
PoptrCHSL7      ---------- -GRL------ ---------- ---------- --
PoptrCHSL6      ---AGAANCD WGLILAFGPG ITFEGILARN LTI------- --
PoptrCHSL1      -KATTGEGLD LGVLFGFGPG VTIETIVLRS FATD------ --
PpCHS           -ASTLGEGSE FGFFIGFGPG LTLEVLVLRA APNSA----- --
PoptrCHSL3      -KTTTGEGME WGVLIGLGAG ITVDTVVLHS VPIAEGR--- --
PoptrCHSL2      -KATTGEGLE WGVVIGLGPG LTMETLVLHS VPVAITK--- --
OsCHS2          -HATTGEGMD WGVLFGFGPG LTVETIVLHS VPITAAAPLI MQ
OsCHS1          -HATTGEGMD WGVLFGFGPG LTVETVVLHS VPITAGAAA- --
AtCHS           -VATTGEGLE WGVLFGFGPG LTVETVVLHS VPL------- --
PoptrCHS1       -KSTTGEGLE WGVLFGFGPG LTVETVVLHS VPVEQTIYS- --
PoptrCHS4       -LQSTGEGLE WGVLFGFGPG LTVETVVLHS VATRV----- --
PoptrCHS5       -LQSTGEGLE WGVLFGFGPG LTVETVVLHS VAPTI----- --
PoptrCHS6       -LQSTGEGLE WGVLFGFGPG LTVETVVLHS VAPTI----- --
PoptrCHS2       -LKSTGEGLE WGVLFGFGPG LTVETVVLHS LPATI----- --
PoptrCHS3       -LKSTGEGLE WGVLFGFGPG LTVETVVLHS VASI------ --
```

Appendix C. Amino acid sequences used in construction of the phylogenetic tree shown in Figure 4.14.

>PoptrDFRL4
MKALVTGASGYLGGRLCHGLLKQGHSVRALVRRTSDISELPPPSSGGVFELAYGDITDYQSLLDAFSGCQ
VIFHAAAIVEPWLPDPSKFFSVNVEGLNNVLQAAKETETIEKIIYTSSFFALGSTDGYVADESQVHCEKR
FCTEYERSKMIADKIASQAAAEGVPIVMLYPGVIYGPGKLTTGNIVAQLLIERFAGRLPGYIGYGNDKFS
FCHVDDLVDGHIAAMDKGRQGERYLLTGENASFKLVFDMAAIISETKKPRFSIPLCIIESYGWLLVLVSR
LTGNLPLISPPTVHVLRHQWEYSCEKAKTELGYNPRGLEDGLKEVLPWLKSMGVIKY*

>PoptrDFRL5
MKKIVVTGASGFVGGVLCHTLLKQGHSVRALVRRTSDLSGLPSPSTGENFELAYGDVTDYRSLLDAIFGC
DVIFHAAAAVEPWLPDPSKFFSVNVGGLKNVVQAAKETKMIEKIIYTSSMVALGSTDGYVADESQVHHEK
YFSTEYERSKVAADKVASQAAAEGLPIVTLYPGVVYGPGKLTTGNALAKMLIDRFAGRLPGYIGRGNDRL
SFCHVDDVVGGHIAAMDKGRLGERYLLTGENASFSRVLDIAAIITRTEKPRFSIPLWVIEAYGWLSILIF
HFTGKLPLLCPPSVHVLRHQWEYSCEKARIELDYNPRSLKEGLDELLPWLKSLGAITY*

>At4g33360
MGPKMPNTETENMKILVTGSTGYLGARLCHVLLRRGHSVRALVRRTSDLSDLPPEVE
LAYGDVTDYRSLTDACSGCDIVFHAAALVEPWLPDPSRFISVNVGGLKNVLEAVKET
KTVQKIIYTSSFFALGSTDGSVANENQVHNERFFCTEYERSKAVADKMALNAASEGV
PIILLYPGVIFGPGKLTSANMVARMLIERFNGRLPGYIGSGTDRYSFSHVDDVVEGH
VAAMEKGRLGERYLLTGENASFKLVFDMAALITGTKKPNFSIPLWAINAYGWLSVLI
SRVTGKLPLISPPTVTVLRHQWSYSCDKAKLELGYNPRSLKEGLEEMLPWLKSLGVI
HY

>Os03g08624

MRIAVTGATGYLGSRLCGALADAGHAVRAFALRSAGGGGGGGDVEAGLLPASVELAYGDV
ADVESLAAAFDRCDAVFHVAAAVEAWLPDPSIFITVNVGGLENVLKAARRTPTVKKIVYT
SSFFAIGPTDGYVADETQMHQGKTFCTEYEKSKVLADQIALQAAAEGMPITIVYPGFMYG
PGKLTAGNLVSRILIERFNGRLPGYVGHGHDRESFCHVDDVVAGHVAAMEKGREGERYLL
TGENTSLVQIFDMASRITNTKAPRFHVPLWLLEIYGWISVLVSRITGKLPFISYPAVRVL
RHQWAYSCEKAKKELGYSPRSLTEGLSETLLWLKDSEMIRF*

> PpDFRL4
MRRVMVTGATGYLGGRLCGMLVHAGLTVVALVRKTSQVQELPPEVELVEGDIRDGESVRRAIEGCDYVVHT
AALVGSWLPDSSQFFKVNVEGFKNVIEAVKATPSVKKLIYTSSFFAVGPTDGYIGDETQFHSMKAFYSPYEES
KAFADKLACEAAMEGVPIVSLYPGIIYGPGSMTKGNSLAEMMIERFNGRMPGYVGYKVKKFSFCHIDDVVVAY
LAAIEIGRVGERYMLCGDNMSFHEVFDLAAGLTKTNPAKVTIPMWVLDVAGFLCVQWARFGAWTGISHQIPFI
TTHSVNILKHQWAYSSEKAERELGYKSRPLEEGLLQLLTWLKATGRIKY*

>PoptrDFR1
MGTEAETVCVTGASGFIGSWLIMRLLEKGYAVRATVRDPDNMKKVTHLLELPKASTHLTLWKADLSVEGS
YDEAIQGCTGVFHVATPMDFESKDPENEVIKPTINGVLDIMRACANSKTVRKIVFTSSAGTVDVEEKRKP
VYDESCWSDLDFVQSIKMTGWMYFVSKTLAEQAAWKFAKENNLDFISIIPTLVVGPFIMQSMPPSLLTAL
SLITGNEAHYGILKQGHYVHLDDLCMSHIFLYENPKAEGRYICNSDDANIHDLAKLLREKYPEYNVPAKF
KDIDENLACVAFSSKKLTDLGFEFKYSLEDMFAGAVETCREKGLIPLSHRKQVVEECKENEVVPAS*

>PoptrDFR2
MGVEVETVCVTGASGFIGSWLVMRLLEKGYTVRATVRDPDNIRKVKHLLELPKADTYLTLWKADLSVEGS
FDEAVQGCTGVFHVATPMDFESKDPENEVIKPTINGVLDIMKACAKAKTVRRIVFTSSAGTVDVEEHKKP
VYDESCWSDLEFVQTVKMTGWMYFVSKTLAEQAAWKYAKENNLDFISVIPPLVVGPFIMHSMPPSLITAL
SLITGNEAHYGIIKQGNYVHLDDLCRAHIVLFENPKAEGRYICSSHEATIHDLAKLLREKYPKYNVPAKF
KDIDEDLASVVFSSKKLLDLGFEFKYSLEEMFAGAVETCREKGLIPLSHEK
>AtDFR
MVSQKETVCVTGASGFIGSWLVMRLLERGYFVRATVRDPGNLKKVQHLLDLPNAKTLLTL
WKADLSEEGSYDDAINGCDGVFHVATPMDFESKDPENEVIKPTVNGMLGIMKACVKAKTV
RRFVFTSSAGTVNVEEHQKNVYDENDWSDLEFIMSKKMTGWMYFVSKTLAEKAAWDFAEE
KGLDFISIIPTLVVGPFITTSMPPSLITALSPITRNEAHYSIIRQGQYVHLDDLCNAHIF
LYEQAAAKGRYICSSHDATILTISKFLRPKYPEYNVPSTFEGVDENLKSIEFSSKKLTDM
GFNFKYSLEEMFIESIETCRQKGFLPVSLSYQSISEIKTKNENIDVKTGDGLTDGMKPCN
KTETGITGERTDAPMLAQQMCA

>Os01g44260
MDFESEDPENEVVKPTVEGMLSIMRACRDAGTVKRIVFTSSAGTVNIEERQRPSYDH
DDWSDIDFCRRVKMTGWMYFVSKSLAEKAAMEYAREHGLDLISVIPTLVVGPFISNG
MPPSHVTALALLTGNEAHYSILKQVQFVHLDDLCDAEIFLFESPEARGRYVCSSHDA
TIHGLATMLADMFPEYDVPRSFPGIDADHLQPVHFSSWKLLAHGFRFRYTLEDMFEA
AVRTCREKGLLPPLPPPPTTAVAGGDGSAGVAGEKEPILGRGTGTAVGAETEALVK*

>PoptrDFRL6
TYCVTGANGYIGSWLVKLLLQRGYTVHATLRDLAKSLDLLSSWRGADRLRLFKADLREEGSFDEAVRGCD
GVFHVAASMEFYVAGNEDNENYVQRNIIDPAIEGTLNLLTSCSKSNTVKRVVFTSSISTLTAKDGAGKWR
QVVDETCQTPIDHVWNTKPPGWIYVLSKRLTEEAAFKYAKDNGIDLISVITTTVAGAFLTSSVPSSIRVL
LSPITGDTKFFSILSAVNARMGSIALVHIDDICDAHIFLMEQTRAEGRYICSAHSCVLSQLINHLVEEYP
CSNIQRLAEKQGSISPEISSKKLRDMGFKYKHSIKDIISETI

>At4g27250
MELQGEESKTATYCVTGASGYIGSWLVKSLLQRGYTVHATLRDLAKSEYFQSKWKEN
ERLRLFRADLRDDGSFDDAVKGCDGVFHVAASMEFDISSDHVNLESYVQSKVIEPAL
KGVRNVLSSCLKSKSVKRVVFTSSISTLTAKDENERMRSFVDETCKAHVDHVLKTQA
SGWIYVLSKLVSEEEAFRYAKERGMDLVSVITTTVSGPFLTPFVPSSVQVLLSPITG
DSKLFAILSAVNKRMGSIALVHIEDICRAHLFLMEQPKAKGQYICCVDNIDMHELML
HHFSKDYLCKVQKVNEDEEERECMKPIISSKKLRELGFEYKYGIEEIVDQTIDASIK
IKFPTLNHKLRQ

>Os04g53810
MSSEVERKTVCVTGGNGYVASLLVKMLLEKGYAVQTSVRDPNNPEKVSHFKDMEKLGPLK
VFRANLEDEGSFDEAVAGCHYAFLVAAPVYDKSHKSDDLEKEIVQGGVEGTLNVMRSCAR
AGTVKRVILTSSTAAVSSLRPLEGAGHVLDESSWSDIEYLRSMEKLSPTQAYSISKVLSE

KEATKFAEENGLSLVTLCPVVAVGASPAVRVDTSVPACLSLITGDEEMMNILKGIEKASG
WSMPMVHIEDVCRAEIFVAEEESASGRYICGSLNTTVTEIAGFLAAKYPQYNVRCDCIEE
HHPEKPTISLSSAKLIGEGFEFKYKNLDEMYDDLVAYGKALGLIPN*

>Os04g53800
MSAVERKTACVTGGSGYIASALIKMLLQKGYAVKTTVRNPDDMEKNSHFKELQALGPLKI
FRADLEEEGSFDEAVAGCDYAFLVAAPMNLKSQNPEKELLEAGVQGTLNVLRSCVKAGTV
KRVILTSSAAAVSGQPLQGDGNGSSHVLDESSWSDLDYLRSTNGISPAQAYAIAKVLSEK
EASKLAEENGISLVAVCPVATVGASPAPVANESVANVLSLLSGNEEINTLRMIDQYSGGL
KLVHVDDLCRAEIFLAEKASPSPSGRYICCALNTTMRQIARSLAAKYPHHNVDIDALGGG
LPEKPTILLSSEKLTSEGFEFMYKTVDEMYDDAFVEYGMALGILHY*

>Os04g53850
MSAVERKTACVTGGNGYIASALIKMLLEKGYAVNTTVRNPDDMAKNSHLKDLQALGPLKV
FRADMDEEGSFDDAIAGCDYAFLVAAPMNFNSENPEKDLVEAAVNGTLNAMRSCAKVGTV
KRVIITSSDAAISRRPLQGDGYVLDEESWSDVDYLRTEKPPAWAYSVSKVLLEKAACKFA
EENNMSLVTVFPVFTLGAAPAPVARTSVPGILSLLSGDETHLEVLKPLQWVTGSVSIVHV
DDLCRAEIFLAEKESSSLSSAESSARYICCSFNTTVLALARFMAGRYPQYNVKTDRFDGM
PEKPRVCCSSEKLIREGFEFKYTNMGDILDDLVEYGRALGILPH*

>Os04g53920
MSAVEMKTACVTGGNGYIASALIKMLLQKGYAVNTTVRNPGDDMKKTSHLKDLEALG
PLEVFRADMDEEGSFDDAVAGCDYAFLVAAPVNFQSQNPEKELIEAGVQGTMNVMRS
CVRAGTVKRVILTSSAPAVSGRPLQGDGHVLDEDSWSDVEYLTKEKPPAWAYSVSKV
LMEKAACKLAEENNISLITVFPVFTLGAAPTPTAATSVSAMLSLLSSDETQLKTLKG
LAATGPIPTVHVDDLCRAEVFVAEKESASGRYICSSLSTTVVAFARFVAGKHPRYNV
KTDGFQGFPEKPRVCYSSEKLVREGFEFKWTDLDEVFDDLIEYGNVLGILPQ*

>PoptrANR1
MASQLTKKTACVIGGTGFVASLLVKLLLEKGYAVNTTVRDPDNQKKVAHLIALQNLGDLNIFGADLTDEE
SFNAPIAGCELVFHVATPVNFASEDPENDMIKPAIQGVHNVLKACAKAKTVKRVILTSSAAALSINKLNG
TGLIMDEKNWTDVEFLTSEKPPTWGYPASKTLAEKAAWKFAEENNIDLITVIPSLMTGPSLTLDIPSSVH
LSMSLITGNEFLKNALKGMQMLSGSISITHVEDVCRAHIFLAEKESASGRYICCAVNTSVVELAEFLNKR
YPQYQVPTDFGDFPSKAKLAITSEKLISEGFSFKYGIEEVYDQTVEYFKAKGLLN*

>PoptrANR2
MASQTKKNTACVIGGTGFVASLLIKLLLEKGYAVNTTVRDPDNQKKIAHLIALQNLGDLNIFGADLTNEE
SFNAPIACCDLVFHVATPVNFASEDPENDMIKPAIQGVHNVLKACAKAKTVQRVILTSSAAAVSINKLNG
TGLVMDEKNWTDVEFLTSEKPPTWGYPASKTLAEKAAWKFAEENNIDLITVIPSLMTGPSFTPHIPDSIN
LAMSLITGNKFLINGLKGMQMLSGSISITHVEDVCRAHIFLAEKESASGRYICCGVNTSVVELAKFLNKR
YPQYQVPTDCGDFPSEAKLIITSEKLSSEGFSFKYGIEEIYDQTVEYFKANGLLN*

> AtANR/BAN
MDQTLTHTGSKKACVIGGTGNLASILIKHLLQSGYKVNTTVRDPENEKKIAHLRKLQ
ELGDLKIFKADLTDEDSFESSFSGCEYIFHVATPINFKSEDPEKDMIKPAIQGVINV
LKSCLKSKSVKRVIYTSSAAAVSINNLSGTGIVMNEENWTDVEFLTEEKPFNWGYPI
SKVLAEKTAWEFAKENKINLVTVIPALIAGNSLLSDPPSSLSLSMSFITGKEMHVTG
LKEMQKLSGSISFVHVDDLARAHLFLAEKETASGRYICCAYNTSVPEIADFLIQRYP
KYNVLSEFEEGLSIPKLTLSSQKLINEGFRFEYGINEMYDQMIEYFESKGLIKAK

> PpDFRL1
MGHSTEKSKGTVCVTGATGFVASWLIKCLLQDGYRVRGAVRDPENYEKAAHLWAL
SGAKERLQLVKGDLLVEGSYDAAVAGCEGVFHTAAALVRIKSDPKAEMLDPTILG
TLNVLHSCAKSTTLKRVVLTSSTAAVRFRDDLEQPGAVTYLDEYSWSSIFFCTKY
QIWYSLAKILSEQEAWKFAFLHSIDLVVVLPSFVIGPCLPYPLSKTAQDICDLLN
GLCRNFGIHGRMGYVHVDDVARAHILVYETPSAQGRYICSAQEATPQELVQYLAD
RYPHLQISTKFNDELPKMPYYKLNTTKLQRLGLNCKPLDVMFDDCISFLEEKGLL
KRKPEKTPTSSSTPDEHSKDSVLQNV*

>AtTKPR1
MDQAKGKVCVTGASGFLASWLVKRLLLEGYEVIGTVRDPGNEKKLAHLWKLEGAKERLRL
VKADLMEEGSFDNAIMGCQGVFHTASPVLKPTSNPEEEILRPAIEGTLNVLRSCRKNPSL
KRVVLTSSSSTVRIRDDFDPKIPLDESIWTSVELCKRFQVWYALSKTLAEQAAWKFSEEN

GIDLVTVLPSFLVGPSLPPDLCSTASDVLGLLKGETEKFQWHGQMGYVHIDDVARTHIVV
FEHEAAQGRYICSSNVISLEELVSFLSARYPSLPIPKRFEKLNRLHYDFDTSKIQSLGLK
FKSLEEMFDDCIASLVEQGYLSTVLP

>PoptrDFRL1
MDQIKGRVCVTGASGYLASWLVKRLLLSGYHVTGTVRDPENEKKVAHLWRLEGAKERLRLVKADLMEEGS
FDDAIMECRGVFHTASPAEILEPAIEGTLNVLRSCKRNPSLKRVILTSSSSTLRVRDDFDSNIPLEESSW
SSVELCERLQIWYALSKTLAEKAAWEFCNGNGIDLITVLPSFVIGPSLSPDLCSTATDVLGLLTGESEKF
HWHGRMGYVHIDDVALSHILVYEDETAGGRFLCSSIVLDNDELASFLSQRYPSLPIPKRFEQLKRPYYEF
NTSRLERLGFKFKPIQEMFDDCIASLVEQGHLSSFSLAIN*

>Os08g40440
MENTTKGKVCVTGASGYVASWLVKRLLESGYHVLGTVRDPGNHKKVGHLWNLTGAKE
RLELVRADLLEEGSFDDAVMACEGVFHTASPVITETDSSKAAVLDSAINGTLNVLRS
CKKNPSLKRVVLTSSSSTVRLKDEADLPPNVLLDETSWSSMEFCESLQIWYAIAKTL
AEKAAWEFAKENGIDLVAVLPTFVVGPNLSHELSPTTTDVLGLFQGETTKFTMYGRM
GYVHIDDVASCHILLYETPRAAGSLPCVYGEQTYGFSTAKVRELGMKFRDVEEMFDD
AVDSLRAHGYLLNSVP*

>Os09g32020
MLSRILHGYGGHGGRGFEQTYRCYSAAAFNKPQLEGGDKVIMPASALHRLASLHIDY
PMLFELSHHGDAAAHRVTHCGVLEFVADEGTVIMPRWMMRGMRLDDGGLVVVRSASL
PKGSYAKLQPHTGDFLDTANPKAVLEKTLRSFTCLTTGDTIMVAYNNKEFLIDIVET
KPASAVCIIETDCEVDFAPPLDYKEPEKVQQKPSVPSSKAASEDQDQIKDEPEFRAF
TGSGNRLDGKASKPLAAGISSNPAAASSAISDSNKKVNQETAASGVSNSTRQKKGKL
VFGSNKSSSSSKEPEKAPPVKVDELAKKEEPKFQAFSGTSYSLKRNRDKVSHLWRLP
SAKERLQLEEMLVPAINGTLNVLKSCKKNPFLKRVVLTSSSSTVRIRDESKHPEISL
DETIWSSVALCEKLQLWYALAKISAEKAAWEFAKENNIDLVTVLPSFVIGPSLSHEL
SVTASDILGLLQGDTDRFISYGRMGYVHIDDVASCHILVYEAPQATGRYLCNSVVLD
NNELVALLAKQFPIFPIPRSLRNPYEKQSYELNTSKIQQLGFKFKGVQEMFGDCVES
LKDQGHLLECPL*

>OryzaDFR2
MVISSKGKVCVTGASGFVASWLIKRLLEAGYHVIGTVRDPSNREKVSHLWRLPSAKE
RLQLVRADLMEEGSFDDAVMACEGVFHTASPVLAKSDSNCKEEMLVPAINGTLNVLK
SCKKNPFLKRVVLTSSSSTVRIMDESKHPEISLDETIWSSVALCEKLQLWYALAKIS
AEKAAWEFAKENNIDLVTVLPSFVIGPSLSHELSVTASDILGLLQGDTDRFISYGRM
GYVHIDDVASCHILVYEAPQATGRYLCNSVVLDNNELVALLAKQFPIFPIPRSLRNP
YEKQSYELNTSKIQQLGFKFKGVQEMFGDCVESLKDQGHLLECPL

>PpDFRL2
MDLSKGAEVCVTGGTGYIASCLIQALLQRGYKVRTTARNPDDRAKTGFLWELPGA
TERLEIVGAELLEEGTFDEAVHGVHTVFHTACPVVYDPNGDPEVSMLNPALKGNL
NVLRACTKSHSIQRVVMTSSCSAIRYDHNRRPEDPPLSESVWSSPEYCRDHKMWY
ALAKTLAEKEAFEFAAREGLNLVVICPSFVIGPSLTPIPTSTVFLILDLLRGRAQ
EYPNKRIGFVHIDDVVTAHVLAMEVPEAHGRYICSSDVAHFGDIMSMLKTKYPKL
QTPTRCSDMPPGDDIHHKMDTTKIKKLGLTEFKSIEQMFDDMLRSLHEKHLESL*

>Os01g03670
MPEYCVTGGTGFIASHLIRALLAASHTVRATVRDPEDEAKVGFLWELDGASERLQLV
KADLMVEGSFDDAVRGVDGVFHAASPVVVVGNSSSNNGKPNDDDDEEEVQQRLVEPI
VRGASNVLRSCARASPRPRRVVFTSSCSCVRYGAGAAAALNESHWSDAAYCAAHGLW
YAYAKTLAEREAWRLAKERGLDMVAVNPSFVVGPILSQAPTSTALIVLALLRGELPR
YPNTTVGFVHVDDAVLAHVVAMEDARASGRLICSCHVAHWSEIVGSLRERYPGYPIP
AECGSHKGDDRAHKMDTAKIRALGFPPFLSVQQMFDDCIKSFQDKGLLPPHA*

>PoptrDFRL2
MPEYCVTGGTGFIAAYLVKSLLEKGHRVRTTVRDPGDVGKVGLLREFDGAKERLK
IFKADLLEEGSFDEAIQGVDGVFHTASPVLLPHDDNIQAMLIDPCINGTLNVLNS
CSKANTVKRVVLTSSCSSIRYRDDVQQVSPLNESHWSDPEYCKRYDLWYAYAKTI
GEKEAWRSAKENGIDLVVVNPSFVVGPLLAPQPTSTLLLILAIVKGLRGEYPNMT
IGFVHIDDVVAAHILAMEDKKASGRLVCSGSVAHWSEIIEMLRAKYPSYPYENKC
SSQKGDCNPHSMDTTKIATLGFPPFKTLEEMFDDCIKSFQEKGFL*

>PoptrCCRL1
MPEYCVTGGTGFIAAYLVKSLLEKGHTVRITVRDPGNVRKVGFLQEFNGAKERLK
IFKAELLEEGSFDEAIQGVDGVFHVAAPVLVPYSDRIQETLIDPCIKGTLNVLNS
CLKASSVKRVVFTSSSSTVRYRDDTPQIFSLNESHWSDTEYCKRHNLWYAYAKTV
AEKEAWRVSKENGIDLVSFIPSFVVGPLLAPEPNSTLLLIQSVVKGSRGEYPNMT
VGFTHIDDVVAGNILAMENSEASGRLVCSGPVAHWSQIIKMLRAKYPSYPYENKC
SSQEGDNIPHSMDTTKIAQLGLPPFKTHEQMFDDCIRSLQEKGFL*

>AtTKPR2
mseylvtggtgfiasyiiksllelghtvrttvrnprdeekvgflwefqgakqrlkil
qadltvegsfdeavngvdgvfhtaspvlvpqdhniqetlvdpiikgttnvmsscaks
katlkrivltsscssiryrfdateasplneshwsdpeyckrfnlwygyaktlgerea
wriaeekgldlvvvnpsfvvgpllgpkptstllmilaiakglageypnftvgfvhid
dvvaahvlameepkasgriicsssvahwseiielmrnkypnypfenkcsnkegdnsp
hsmdtrkihelgfgsfkslpemfddciisfqkkgll

>At1g25460
MAEYLVTGGTSFIASHVIKSLLEFGHYVRTTVRDSEDEEKVGFLWDLKGAKERLKIF
EADLTIEGSFDEAVNGVDGVFHIASRVSVRLDNNNLDKFDPNISGTMNVMNSCAKSR
NTVKRIVLTSSSTAIRYRFDATQVSPLNESHWTDLEYCKHFKIWYAYKKTLGEKEAW
RIAADKKLNLVVVIPSFCIGPILSPKPTSSPLIFLSIIKGTRGTYPNFRGGFVHIDD
VVAAQILAMEEPKASGRILCSSSVAHWSEIIEMLRIKYPLYPFETKCGSEEGKDMPH
SLDTTKIHELGFASFKSLTEMFDDCIKCFQDKGLL

>Os09g04050
MPTDETAAAAPATTALSGHGCTVCVTGAGGFIASWLVKRLLEKGYTVRGTVRNPMDPKND
HLRALDGAGERLVLLRADLLDPDSLVAAFTGCEGVFHAASPVTDDPEKMIEPAIRGTRYV
ITAAADTGIKRVVFTSSIGTVYMNPYRDPNKPVDDTCWSDLEYCKRTENWYCYAKTVAEQ
GAWEVARRRGVDLVVVNPVLVLGPLLQATVNASTEHVMKYLTGSAKTYVNAAQAYVHVRD
VAEAHVRVYDCGGARGRYICAESTLHRGDLCRALAKLFPEYPVPSRCKDEAAPPVKGYLF
SNQRLRDLGMDFVPVRQCLYETVRSLQDKGLLPVLPPTADDHHHPSS*

>Os02g08420
MAAAVVCVTGAGGFIGSWIVKLLLARGYAVRGTSRRADDPKNAHLWALDGAAERLTMVSV
DLLDRGSLRAAFAGCHGVIHTASPMHDDPEEIIEPVITGTLNVVEVAADAGVRRVVLSST
IGTMYMDPRRDPDSPLDDSFWSDLDYCKNTKNWYCYAKTIAERKAWEVARGRGVDMAVVI
PVVVLGELLQPGMNTSTKHILKYLTGEAKTYVNESHAYVHVVDAAEAHVRVLEAPGAGGR
RYICAERTLHRGELCRILAGLFPEYPIPTRCRDEINPPKKGYKFTNQPLKDLGIKFTPVH
EYLYEAVKSLEDKGFIKKTSNTKELHRQSSPPQNSPASMLMSKL*

>Os08g34280
MTVIDGAVAADAGGAAAAVVQPGNGQTVCVTGAAGYIASWLVKLLLEKGYTVKGTVRNPD
DPKNAHLKALDGAGERLVLCKADLLDYDAICRAVAGCHGVFHTASPVTDDPEQMVEPAVR
GTEYVINAAAEAGTVRRVVFTSSIGAVTMDPNRGPDVVVDESCWSDLDYCKETRNWYCYG
KAVAEQAAWEAARRRGVELVVVNPVLVIGPLLQPTVNASVAHILKYLDGSASKFANAVQA
YVDVRDVAAAHLLVFESPSAAGRFLCAESVLHREGVVRILAKLFPEYPVPTRCSDEKNPR
KQPYKMSNQKLRDLGLEFRPASQSLYETVKCLQEKGHLPVLAAEKTEEEAGEVQGGIAIR
A*

>Os09g25150
MTVVVVADDAAAAAAAAQQQEELPPGHGQTVCVTGAAGYIASWLVKLLLERGYTVKGTVR
NPDDPKNAHLKALDGADERLVLCKADLLDYDSIRAAVDGCHGVFHTASPVTDDPEQMVEP
AVRGTEYVIKAAAEAGTVRRVVFTSSIGAVTMDPNRGPDVVVDESCWSDLEFCKKTKNWY
CYGKAVAEQEACKAAEERGVDLVVVSPVLVVGPLLQPTVNASAVHILKYLDGSAKKYANA
VQAYVDVRDVAAAHVRVFEAPEASGRHLCAERVLHREDVVHILGKLFPEYPVPTR*

>PoptrCCR2
MPVDASSLSGQGQTICVTGAGGFIASWMVKLLLDKGYTVRGTARNPADPKNSHLRGLEGAEERLTLCKAD
LLDYESLKEAIQGCDGVFHTASPVTDDPEEMVEPAVNGTKNVIIAAAEAKVRRVVFTSSIGAVYMDPNKG
PDVVIDESCWSDLEFCKNTKNWYCYGKAVAEQAAWDMAKEKGVDLVVVNPVLVLGPLLQPTVNASITHIL
KYLTGSAKTYANSVQAYVHVRDVALAHILVFETPSASGRYLCSESVLHRGEVVEILAKFFPEYPIPTKCS
DEKNPRKQPYKFSNQKLRDLGFEFTPVKQCLYETVKSLQEKGHLPIPKQAAEESLKIQ*

>AtCCR1
MPVDVASPAGKTVCVTGAGGYIASWIVKILLERGYTVKGTVRNPDDPKNTHLRELEGGKERLILCKADLQ
DYEALKAAIDGCDGVFHTASPVTDDPEQMVEPAVNGAKFVINAAAEAKVKRVVITSSIGAVYMDPNRDPE
AVVDESCWSDLDFCKNTKNWYCYGKMVAEQAAWETAKEKGVDLVVLNPVLVLGPPLQPTINASLYHVLKY
LTGSAKTYANLTQAYVDVRDVALAHVLVYEAPSASGRYLLAESARHRGEVVEILAKLFPEYPLPTKCKDE
KNPRAKPYKFTNQKIKDLGLEFTSTKQSLYDTVKSLQEKGHLAPPPPPPSASQESVENGIKIGS

>AtCCR2
MLVDGKLVCVTGAGGYIASWIVKLLLERGYTVRGTVRNPTDPKNNHLRELQGAKERLTLHSADLLDYEAL
CATIDGCDGVFHTASPMTDDPETMLEPAVNGAKFVIDAAAKAKVKRVVFTSSIGAVYMNPNRDTQAIVDE
NCWSDLDFCKNTKNWYCYGKMLAEQSAWETAKAKGVDLVVLNPVLVLGPPLQSAINASLVHILKYLTGSA
KTYANLTQVYVDVRDVALGHVLVYEAPSASGRYILAETALHRGEVVEILAKFFPEYPLPTKCSDEKNPRA
KPYKFTTQKIKDLGLEFKPIKQSLYESVKSLQEKGHLPLPQDSNQNEVIIES

> PpDFRL3
MANGQVVCVTGANGFIASWLVKSLLERGYTVRGTVRNPEKSKHLLNLPGANERLELIEADLLAPEAFDSAVH
GCHGVFHTASPFHFNITDPDSQLIEPAVKGTLNVLESCAKAGTKKIVLTSSVAAVAYSPKRAGASVVDETFFS
DPEFCQKEQRWYVLSKTLAESAAWEFVKEHNLNMVAINPTMVIGPLLQSSMNTSNELLLGFLNGTAKSFPNQ
AVGWVSVKDVAMAHILAYEKPEAEGRYIINERLIHYGEMVSLLMNRYPQYPIVAKDADDSTRLPSYNLSNEKIK
KLGLTFQPLEEALDETVACFKELKLLD*

>PoptrCCRL4
MSSLVSYRAVATGTERMSRGGDGKVVCVTGGSGYIASWLVKLLLQRGYTVKTTVR
DPNDPKKTEHLLALEGAKERLHLFKANLLEEGAFDPIVDGCEGVFHTASPVSFSP
TDDPQVDLIDPALKGTLNVLRSCAKVHSIRRVVLTSSAAACIYSGKPLNHDVVID
ETWYSDPAICKELKAWYALSKTLAEEAAWNFAKENATDLVTVHPSFVIGPLLQPT
LNLSVEMILDLVNGAETYPNGYYRCIDVRDVANAHIQAFEIPSASGRYVLTAYVT
TFSEVLKIIRENYPTLRLPEKSTESMFKPYQVSKEKAKTLGINFTPLDLSLVDTI
ESLKEKGFLKI*

>PoptrCCRL3
MSGEGKVVSVTGASGYIASWLVKLLLERGYTVKASVRDPNDAKKTEHLLALDGAK
ERLQLFKADLLDEGSFDPVVEGCECVFHTASPFYFTVNDPQAELVDPALKGTVNV
LRSCTKIPSIKRVVITSSMAAVVFNGKSLAPDVVVDETWFSDSDFCEKSKLWYHL
SKTLAEEAAWKFTKENGIDMVTLNPGLVIGPLLQPTLNQSAESVLDLINGAKSYP
NTTYRWVDVRDVANAHIYALENPSANGRYCLVGTVIHSSEAVKILSKLYPDLTIP
KQCADDKPPMPKYQVSKERAASLGVKYTPLEASLKDTIESLKEKNFVSF*

>Os01g34480
MSSESEAAPGTGKLVCVTGASGYIASWLVRLLLARGYTVRATIRDTSDPKKTLHLRA
LDGANERLHLFEANLLEEGSFDAAVNGCDCVFHTASPFYHNVKDPKAELLDPAVKGT
LNVLGSCKKASIRRVIVTSSMAAVAYNGKPRTPDVVVDETWFSVPEICEKHQQWYVL
SKTLAEEAAWKFSKDNGFEIVTVNPAMVIGPLLQPSLNTSAEAILKLINGSSSTYPN
FSFGWINVKDVALAHILAYEVPSANGRYCMVERVAHYSELVQIIREMYPNIPLPDKC
ADDKPSVPIYQVSKEKIKSLGLELTPLHTSIKETIESLKEKGFVTFDSSNL*

>PoptrCCRL1
MSTGAGKIVCVTGASGYIASWIVKLLLSRGYTVKASVRDPNDPKKTQHLRALRGA
QERLELVKANLLEEGSFDSIVEGCEGVFHTASPFYHDVKDPQAELLDPAVKGTLN
VLGSCARHPSIKRVVLTSSMAAVAYNRKPRTPDVVVDETWFSDPELCRESKLWYV
LSKTLAEDAAWKFAKEKGMDMVAINPSMVIGPLLQPTLNTSAAAILSLIKGAQTF
SNASFGWINVKDVANAHIQAFELSSASGRYCLVERVAHHSEVVKILRELYPDLQL
PEKCADDKPYVPIYQVSKEKAKSLGIEFIPLEASIKETVESLKEKGFVSF*

>PoptrCCRL2
MSSGAGKIVCVTGASGYIASWLVKLLLSRGYTVKASVRDPNDPKKTEHLRALNGA
QERLQLFKANLLEEGSFDSIVEGCEGVFHTASPFYHDVKDPQVELLDPAVKGTLN
VLGSCAKHPSIRRVVLTSSVAAVAYNGKPRTPDVVVDETWFSDPNLCRESKVWYV
LSKTLAEDAAWKFAKEKDMDMVAINPAMVIGPLLQPTLNTSAAAILSLIKGAQTF
PNASFGWINVKDVANAHIQAFELSSASGRYCLVERVAHYSEVVKILHELYPDLQL
PEKCADDKPYVPIYQVSKEKAKSLGVEFIPLEASVKETVESLKEKGFVSF*

>AtCCRL14
mansgegkvvcvtgasgyiaswlvkfllsrgytvkasvrdpsdpkktqhlvslegak
erlhlfkadlleqgsfdsaidgchgvfhtaspffndakdpqaelidpavkgtlnvln
scakassvkrvvvtssmaavgyngkprtpdvtvdetwfsdpelceaskmwyvlsktl
aedaawklakekgldivtinpamvigpllqptlntsaaailnlingaktfpnlsfgw
vnvkdvanahiqafevpsangryclvervvhhseivnilrelypnlplpercvdenp
yvptyqvskdktrslgidyiplkvsiketveslkekgfaqf

Appendix D. Alignment of amino acid sequences used to generate the phylogenetic tree shown in Figure 4.14.

```
           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              5        15        25        35        45        55
PpDFRL4     ---------- ---------- ---------- ---------- ---------- ----------
Os03g08624  ---------- ---------- ---------- ---------- ---------- ----------
At4g33360   ---------- ---------- ---------- ---------- ---------- ---------M
PoptrDFRL5  ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFRL4  ---------- ---------- ---------- ---------- ---------- ----------
Os02g08420  ---------- ---------- ---------- ---------- ---------- ----------
Os09g04050  ---------- ---------- ---------- ---------- ---------- -MPTDETAAA
Os08g34280  ---------- ---------- ---------- ---------- ------MTVI DGAVAADAGG
Os09g25150  ---------- ---------- ---------- ---------- ---MTVVVVA DDAAAAAAAA
PoptrCCR2   ---------- ---------- ---------- ---------- ---------- ---------M
AtCCR1      ---------- ---------- ---------- ---------- ---------- ----------
AtCCR2      ---------- ---------- ---------- ---------- ---------- ----------
PpDFRL3     ---------- ---------- ---------- ---------- ---------- ----------
PoptrCCRL4  ---------- ---------- ---------- ---------- --------MS SLVSYRAVAT
PoptrCCRL3  ---------- ---------- ---------- ---------- ---------- ----------
Os01g34480  ---------- ---------- ---------- ---------- ---------- ---------M
AtCCRL14    ---------- ---------- ---------- ---------- ---------- ----------
PoptrCCRL1  ---------- ---------- ---------- ---------- ---------- ----------
PoptrCCRL2  ---------- ---------- ---------- ---------- ---------- ----------
PpDFRL1     ---------- ---------- ---------- ---------- ---------- ----------
AtTKPR1     ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFRL1  ---------- ---------- ---------- ---------- ---------- ----------
Os08g40440  ---------- ---------- ---------- ---------- ---------- ----------
Os09g32020  MLSRILHGYG GHGGRGFEQT YRCYSAAAFN KPQLEGGDKV IMPASALHRL ASLHIDYPML
OryzaDFR2   ---------- ---------- ---------- ---------- ---------- ----------
PpDFRL2     ---------- ---------- ---------- ---------- ---------- ----------
Os01g03670  ---------- ---------- ---------- ---------- ---------- ----------
At1g25460   ---------- ---------- ---------- ---------- ---------- ----------
AtTKPR2     ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFRL2  ---------- ---------- ---------- ---------- ---------- ----------
PoptrCCRL1  ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFRL6  ---------- ---------- ---------- ---------- ---------- ----------
At4g27250   ---------- ---------- ---------- ---------- ---------- ----------
Os01g44260  ---------- ---------- ---------- ---------- ---------- ----------
AtDFR       ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFR1   ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFR2   ---------- ---------- ---------- ---------- ---------- ----------
AtANR/BAN   ---------- ---------- ---------- ---------- ---------- ----------
PoptrANR1   ---------- ---------- ---------- ---------- ---------- ----------
PoptrANR2   ---------- ---------- ---------- ---------- ---------- ----------
Os04g53810  ---------- ---------- ---------- ---------- ---------- ----------
Os04g53800  ---------- ---------- ---------- ---------- ---------- ----------
Os04g53850  ---------- ---------- ---------- ---------- ---------- ----------
Os04g53920  ---------- ---------- ---------- ---------- ---------- ----------

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              65        75        85        95       105       115
```

```
PpDFRL4     ---------- MRRVMVTGAT GYLGG----- ---RLCGML- VHAGLTV--- ----------
Os03g08624  ---------- -MRIAVTGAT GYLGS----- ---RLCGAL- ADAGHAV--- ----------
At4g33360   GPKMPNTETE NMKILVTGST GYLGA----- ---RLCHVL- LRRGHSV--- ----------
PoptrDFRL5  ---------- MKKIVVTGAS GFVGG----- ---VLCHTL- LKQGHSV--- ----------
PoptrDFRL4  ---------- -MKALVTGAS GYLGG----- ---RLCHGL- LKQGHSV--- ----------
Os02g08420  --------MA AAVVCVTGAG GFIGS----- ---WIVKLL- LARGYAV--- ----------
Os09g04050  APATTALSGH GCTVCVTGAG GFIAS----- ---WLVKRL- LEKGYTV--- ----------
Os08g34280  AAAAVVQPGN GQTVCVTGAA GYIAS----- ---WLVKLL- LEKGYTV--- ----------
Os09g25150  QQQEELPPGH GQTVCVTGAA GYIAS----- ---WLVKLL- LERGYTV--- ----------
PoptrCCR2   PVDASSLSGQ GQTICVTGAG GFIAS----- ---WMVKLL- LDKGYTV--- ----------
AtCCR1      -MPVDVASPA GKTVCVTGAG GYIAS----- ---WIVKIL- LERGYTV--- ----------
AtCCR2      ------MLVD GKLVCVTGAG GYIAS----- ---WIVKLL- LERGYTV--- ----------
PpDFRL3     -------MAN GQVVCVTGAN GFIAS----- ---WLVKSL- LERGYTV--- ----------
PoptrCCRL4  GTERMSRGGD GKVVCVTGGS GYIAS----- ---WLVKLL- LQRGYTV--- ----------
PoptrCCRL3  ------MSGE GKVVSVTGAS GYIAS----- ---WLVKLL- LERGYTV--- ----------
Os01g34480  SSESEAAPGT GKLVCVTGAS GYIAS----- ---WLVRLL- LARGYTV--- ----------
AtCCRL14    ----MANSGE GKVVCVTGAS GYIAS----- ---WLVKFL- LSRGYTV--- ----------
PoptrCCRL1  -----MSTGA GKIVCVTGAS GYIAS----- ---WIVKLL- LSRGYTV--- ----------
PoptrCCRL2  -----MSSGA GKIVCVTGAS GYIAS----- ---WLVKLL- LSRGYTV--- ----------
PpDFRL1     --MGHSTEKS KGTVCVTGAT GFVAS----- ---WLIKCL- LQDGYRV--- ----------
AtTKPR1     ------MDQA KGKVCVTGAS GFLAS----- ---WLVKRL- LLEGYEV--- ----------
PoptrDFRL1  ------MDQI KGRVCVTGAS GYLAS----- ---WLVKRL- LLSGYHV--- ----------
Os08g40440  -----MENTT KGKVCVTGAS GYVAS----- ---WLVKRL- LESGYHV--- ----------
Os09g32020  FELSHHGDAA AHRVTHCGVL EFVADEGTVI MPRWMMRGMR LDDGGLVVVR SASLPKGSYA
OryzaDFR2   -----MVISS KGKVCVTGAS GFVAS----- ---WLIKRL- LEAGYHV--- ----------
PpDFRL2     -----MDLSK GAEVCVTGGT GYIAS----- ---CLIQAL- LQRGYKV--- ----------
Os01g03670  ---------- MPEYCVTGGT GFIAS----- ---HLIRAL- LAASHTV--- ----------
At1g25460   ---------- MAEYLVTGGT SFIAS----- ---HVIKSL- LEFGHYV--- ----------
AtTKPR2     ---------- MSEYLVTGGT GFIAS----- ---YIIKSL- LELGHTV--- ----------
PoptrDFRL2  ---------- MPEYCVTGGT GFIAA----- ---YLVKSL- LEKGHRV--- ----------
PoptrCCRL1  ---------- MPEYCVTGGT GFIAA----- ---YLVKSL- LEKGHTV--- ----------
PoptrDFRL6  ---------- --TYCVTGAN GYIGS----- ---WLVKLL- LQRGYTV--- ----------
At4g27250   -MELQGEESK TATYCVTGAS GYIGS----- ---WLVKSL- LQRGYTV--- ----------
Os01g44260  ---------- ---------- ---------- ---------- ---------- ----------
AtDFR       ------MVSQ KETVCVTGAS GFIGS----- ---WLVMRL- LERGYFV--- ----------
PoptrDFR1   ------MGTE AETVCVTGAS GFIGS----- ---WLIMRL- LEKGYAV--- ----------
PoptrDFR2   ------MGVE VETVCVTGAS GFIGS----- ---WLVMRL- LEKGYTV--- ----------
AtANR/BAN   -MDQTLTHTG SKKACVIGGT GNLAS----- ---ILIKHL- LQSGYKV--- ----------
PoptrANR1   ----MASQLT KKTACVIGGT GFVAS----- ---LLVKLL- LEKGYAV--- ----------
PoptrANR2   ----MASQTK KNTACVIGGT GFVAS----- ---LLIKLL- LEKGYAV--- ----------
Os04g53810  ----MSSEVE RKTVCVTGGN GYVAS----- ---LLVKML- LEKGYAV--- ----------
Os04g53800  -----MSAVE RKTACVTGGS GYIAS----- ---ALIKML- LQKGYAV--- ----------
Os04g53850  -----MSAVE RKTACVTGGN GYIAS----- ---ALIKML- LEKGYAV--- ----------
Os04g53920  -----MSAVE MKTACVTGGN GYIAS----- ---ALIKML- LQKGYAV--- ----------

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              125       135       145       155       165       175
PpDFRL4     ---------- ---------- ---------- ---------- ---------- ----------
Os03g08624  ---------- ---------- ---------- ---------- ---------- ----------
At4g33360   ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFRL5  ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFRL4  ---------- ---------- ---------- ---------- ---------- ----------
Os02g08420  ---------- ---------- ---------- ---------- ---------- ----------
Os09g04050  ---------- ---------- ---------- ---------- ---------- ----------
Os08g34280  ---------- ---------- ---------- ---------- ---------- ----------
Os09g25150  ---------- ---------- ---------- ---------- ---------- ----------
PoptrCCR2   ---------- ---------- ---------- ---------- ---------- ----------
AtCCR1      ---------- ---------- ---------- ---------- ---------- ----------
AtCCR2      ---------- ---------- ---------- ---------- ---------- ----------
PpDFRL3     ---------- ---------- ---------- ---------- ---------- ----------
PoptrCCRL4  ---------- ---------- ---------- ---------- ---------- ----------
PoptrCCRL3  ---------- ---------- ---------- ---------- ---------- ----------
Os01g34480  ---------- ---------- ---------- ---------- ---------- ----------
```

```
AtCCRL14    ---------- ---------- ---------- ---------- ---------- ----------
PoptrCCRL1  ---------- ---------- ---------- ---------- ---------- ----------
PoptrCCRL2  ---------- ---------- ---------- ---------- ---------- ----------
PpDFRL1     ---------- ---------- ---------- ---------- ---------- ----------
AtTKPR1     ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFRL1  ---------- ---------- ---------- ---------- ---------- ----------
Os08g40440  ---------- ---------- ---------- ---------- ---------- ----------
Os09g32020  KLQPHTGDFL DTANPKAVLE KTLRSFTCLT TGDTIMVAYN NKEFLIDIVE TKPASAVCII
OryzaDFR2   ---------- ---------- ---------- ---------- ---------- ----------
PpDFRL2     ---------- ---------- ---------- ---------- ---------- ----------
Os01g03670  ---------- ---------- ---------- ---------- ---------- ----------
At1g25460   ---------- ---------- ---------- ---------- ---------- ----------
AtTKPR2     ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFRL2  ---------- ---------- ---------- ---------- ---------- ----------
PoptrCCRL1  ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFRL6  ---------- ---------- ---------- ---------- ---------- ----------
At4g27250   ---------- ---------- ---------- ---------- ---------- ----------
Os01g44260  ---------- ---------- ---------- ---------- ---------- ----------
AtDFR       ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFR1   ---------- ---------- ---------- ---------- ---------- ----------
PoptrDFR2   ---------- ---------- ---------- ---------- ---------- ----------
AtANR/BAN   ---------- ---------- ---------- ---------- ---------- ----------
PoptrANR1   ---------- ---------- ---------- ---------- ---------- ----------
PoptrANR2   ---------- ---------- ---------- ---------- ---------- ----------
Os04g53810  ---------- ---------- ---------- ---------- ---------- ----------
Os04g53800  ---------- ---------- ---------- ---------- ---------- ----------
Os04g53850  ---------- ---------- ---------- ---------- ---------- ----------
Os04g53920  ---------- ---------- ---------- ---------- ---------- ----------

          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
            185       195       205       215       225       235
PpDFRL4     ---------- ---------- ---------V ALVRKT---- ------SQVQ ELPPEVELVE
Os03g08624  ---------- ---------- ---------R AFALRS-AGG GGGGGDVEAG LLPASVELAY
At4g33360   ---------- ---------- ---------R ALVRRT---- ------SDLS DLPPEVELAY
PoptrDFRL5  ---------- ---------- ---------R ALVRRT-SDL SG----LPSP STGENFELAY
PoptrDFRL4  ---------- ---------- ---------R ALVRRT-SDI SE----LPPP SSGGVFELAY
Os02g08420  ---------- ---------- ---------R GTSRRA-DDP KN-AHLWALD GAAERLTMVS
Os09g04050  ---------- ---------- ---------R GTVRNP-MDP KN-DHLRALD GAGERLVLLR
Os08g34280  ---------- ---------- ---------K GTVRNP-DDP KN-AHLKALD GAGERLVLCK
Os09g25150  ---------- ---------- ---------K GTVRNP-DDP KN-AHLKALD GADERLVLCK
PoptrCCR2   ---------- ---------- ---------R GTARNP-ADP KN-SHLRGLE GAEERLTLCK
AtCCR1      ---------- ---------- ---------K GTVRNP-DDP KN-THLRELE GGKERLILCK
AtCCR2      ---------- ---------- ---------R GTVRNP-TDP KN-NHLRELQ GAKERLTLHS
PpDFRL3     ---------- ---------- ---------R GTVRNP---- EKSKHLLNLP GANERLELIE
PoptrCCRL4  ---------- ---------- ---------K TTVRDP-NDP KKTEHLLALE GAKERLHLFK
PoptrCCRL3  ---------- ---------- ---------K ASVRDP-NDA KKTEHLLALD GAKERLQLFK
Os01g34480  ---------- ---------- ---------R ATIRDT-SDP KKTLHLRALD GANERLHLFE
AtCCRL14    ---------- ---------- ---------K ASVRDP-SDP KKTQHLVSLE GAKERLHLFK
PoptrCCRL1  ---------- ---------- ---------K ASVRDP-NDP KKTQHLRALR GAQERLELVK
PoptrCCRL2  ---------- ---------- ---------K ASVRDP-NDP KKTEHLRALN GAQERLQLFK
PpDFRL1     ---------- ---------- ---------R GAVRDP-ENY EKAAHLWALS GAKERLQLVK
AtTKPR1     ---------- ---------- ---------I GTVRDP-GNE KKLAHLWKLE GAKERLRLVK
PoptrDFRL1  ---------- ---------- ---------T GTVRDP-ENE KKVAHLWRLE GAKERLRLVK
Os08g40440  ---------- ---------- ---------L GTVRDP-GNH KKVGHLWNLT GAKERLELVR
Os09g32020  ETDCEVDFAP PLDYKEPEKV QQKPSVPSSK AASEDQ-DQI KDEPEFRAFT GSGNRLDGKA
OryzaDFR2   ---------- ---------- ---------I GTVRDP-SNR EKVSHLWRLP SAKERLQLVR
PpDFRL2     ---------- ---------- ---------R TTARNP-DDR AKTGFLWELP GATERLEIVG
Os01g03670  ---------- ---------- ---------R ATVRDP-EDE AKVGFLWELD GASERLQLVK
At1g25460   ---------- ---------- ---------R TTVRDS-EDE EKVGFLWDLK GAKERLKIFE
AtTKPR2     ---------- ---------- ---------R TTVRNP-RDE EKVGFLWEFQ GAKQRLKILQ
PoptrDFRL2  ---------- ---------- ---------R TTVRDP-GDV GKVGLLREFD GAKERLKIFK
PoptrCCRL1  ---------- ---------- ---------R ITVRDP-GNV RKVGFLQEFN GAKERLKIFK
PoptrDFRL6  ---------- ---------- ---------H ATLRDL---A KSLDLLSSWR GA-DRLRLFK
```

```
At4g27250   ---------- ---------- ---------H ATLRDL-AKS EYFQSKWKEN ---ERLRLFR
Os01g44260  ---------- ---------- ---------- ---------- ---------- ----------
AtDFR       ---------- ---------- ---------R ATVRDP-GNL KKVQHLLDLP NAKTLLTLWK
PoptrDFR1   ---------- ---------- ---------R ATVRDP-DNM KKVTHLLELP KASTHLTLWK
PoptrDFR2   ---------- ---------- ---------R ATVRDP-DNI RKVKHLLELP KADTYLTLWK
AtANR/BAN   ---------- ---------- ---------N TTVRDP-ENE KKIAHLRKLQ ELGD-LKIFK
PoptrANR1   ---------- ---------- ---------N TTVRDP-DNQ KKVAHLIALQ NLGD-LNIFG
PoptrANR2   ---------- ---------- ---------N TTVRDP-DNQ KKIAHLIALQ NLGD-LNIFG
Os04g53810  ---------- ---------- ---------Q TSVRDP-NNP EKVSHFKDME KLGP-LKVFR
Os04g53800  ---------- ---------- ---------K TTVRNP-DDM EKNSHFKELQ ALGP-LKIFR
Os04g53850  ---------- ---------- ---------N TTVRNP-DDM AKNSHLKDLQ ALGP-LKVFR
Os04g53920  ---------- ---------- ---------N TTVRNPGDDM KKTSHLKDLE ALGP-LEVFR

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              245       255       265       275       285       295
PpDFRL4     GDIRDGESVR RAIEGCDYVV HT-------- ---------- ---------- ----------
Os03g08624  GDVADVESLA AAFDRCDAVF HV-------- ---------- ---------- ----------
At4g33360   GDVTDYRSLT DACSGCDIVF HA-------- ---------- ---------- ----------
PoptrDFRL5  GDVTDYRSLL DAIFGCDVIF HA-------- ---------- ---------- ----------
PoptrDFRL4  GDITDYQSLL DAFSGCQVIF HA-------- ---------- ---------- ----------
Os02g08420  VDLLDRGSLR AAFAGCHGVI HT-------- ---------- ---------- ----------
Os09g04050  ADLLDPDSLV AAFTGCEGVF HA-------- ---------- ---------- ----------
Os08g34280  ADLLDYDAIC RAVAGCHGVF HT-------- ---------- ---------- ----------
Os09g25150  ADLLDYDSIR AAVDGCHGVF HT-------- ---------- ---------- ----------
PoptrCCR2   ADLLDYESLK EAIQGCDGVF HT-------- ---------- ---------- ----------
AtCCR1      ADLQDYEALK AAIDGCDGVF HT-------- ---------- ---------- ----------
AtCCR2      ADLLDYEALC ATIDGCDGVF HT-------- ---------- ---------- ----------
PpDFRL3     ADLLAPEAFD SAVHGCHGVF HT-------- ---------- ---------- ----------
PoptrCCRL4  ANLLEEGAFD PIVDGCEGVF HT-------- ---------- ---------- ----------
PoptrCCRL3  ADLLDEGSFD PVVEGCECVF HT-------- ---------- ---------- ----------
Os01g34480  ANLLEEGSFD AAVNGCDCVF HT-------- ---------- ---------- ----------
AtCCRL14    ADLLEQGSFD SAIDGCHGVF HT-------- ---------- ---------- ----------
PoptrCCRL1  ANLLEEGSFD SIVEGCEGVF HT-------- ---------- ---------- ----------
PoptrCCRL2  ANLLEEGSFD SIVEGCEGVF HT-------- ---------- ---------- ----------
PpDFRL1     GDLLVEGSYD AAVAGCEGVF HT-------- ---------- ---------- ----------
AtTKPR1     ADLMEEGSFD NAIMGCQGVF HT-------- ---------- ---------- ----------
PoptrDFRL1  ADLMEEGSFD DAIMECRGVF HT-------- ---------- ---------- ----------
Os08g40440  ADLLEEGSFD DAVMACEGVF HT-------- ---------- ---------- ----------
Os09g32020  SKPLAAGISS NPAAASSAIS DSNKKVNQET AASGVSNSTR QKKGKLVFGS NKSSSSSKEP
OryzaDFR2   ADLMEEGSFD DAVMACEGVF HT-------- ---------- ---------- ----------
PpDFRL2     AELLEEGTFD EAVHGVHTVF HT-------- ---------- ---------- ----------
Os01g03670  ADLMVEGSFD DAVRGVDGVF HA-------- ---------- ---------- ----------
At1g25460   ADLTIEGSFD EAVNGVDGVF HI-------- ---------- ---------- ----------
AtTKPR2     ADLTVEGSFD EAVNGVDGVF HT-------- ---------- ---------- ----------
PoptrDFRL2  ADLLEEGSFD EAIQGVDGVF HT-------- ---------- ---------- ----------
PoptrCCRL1  AELLEEGSFD EAIQGVDGVF HV-------- ---------- ---------- ----------
PoptrDFRL6  ADLREEGSFD EAVRGCDGVF HV-------- ---------- ---------- ----------
At4g27250   ADLRDDGSFD DAVKGCDGVF HV-------- ---------- ---------- ----------
Os01g44260  ---------- ---------- ---------- ---------- ---------- ----------
AtDFR       ADLSEEGSYD DAINGCDGVF HV-------- ---------- ---------- ----------
PoptrDFR1   ADLSVEGSYD EAIQGCTGVF HV-------- ---------- ---------- ----------
PoptrDFR2   ADLSVEGSFD EAVQGCTGVF HV-------- ---------- ---------- ----------
AtANR/BAN   ADLTDEDSFE SSFSGCEYIF HV-------- ---------- ---------- ----------
PoptrANR1   ADLTDEESFN APIAGCELVF HV-------- ---------- ---------- ----------
PoptrANR2   ADLTNEESFN APIACCDLVF HV-------- ---------- ---------- ----------
Os04g53810  ANLEDEGSFD EAVAGCHYAF LV-------- ---------- ---------- ----------
Os04g53800  ADLEEEGSFD EAVAGCDYAF LV-------- ---------- ---------- ----------
Os04g53850  ADMDEEGSFD DAIAGCDYAF LV-------- ---------- ---------- ----------
Os04g53920  ADMDEEGSFD DAVAGCDYAF LV-------- ---------- ---------- ----------

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              305       315       325       335       345       355
PpDFRL4     --AALV---- ------G-SW LPDS------ ---------- ---------- --------SQ
```

```
Os03g08624   --AAAV---- ------E-AW LPDP------ ---------- ---------- --------SI
At4g33360    --AALV---- ------E-PW LPDP------ ---------- ---------- --------SR
PoptrDFRL5   --AAAV---- ------E-PW LPDP------ ---------- ---------- --------SK
PoptrDFRL4   --AAIV---- ------E-PW LPDP------ ---------- ---------- --------SK
Os02g08420   --ASPM---- ---------- HDDP------ ---------- ---------- --------EE
Os09g04050   --ASPV---- ---------- TDDP------ ---------- ---------- --------EK
Os08g34280   --ASPV---- ---------- TDDP------ ---------- ---------- --------EQ
Os09g25150   --ASPV---- ---------- TDDP------ ---------- ---------- --------EQ
PoptrCCR2    --ASPV---- ---------- TDDP------ ---------- ---------- --------EE
AtCCR1      --ASPV---- ---------- TDDP------ ---------- ---------- --------EQ
AtCCR2      --ASPM---- ---------- TDDP------ ---------- ---------- --------ET
PpDFRL3     --ASPF---- ------H-FN ITDP------ ---------- ---------- ------D-SQ
PoptrCCRL4   --ASPV---- ------SFSP TDDP------ ---------- ---------- ------Q-VD
PoptrCCRL3   --ASPF---- ------Y-FT VNDP------ ---------- ---------- ------Q-AE
Os01g34480   --ASPF---- ------Y-HN VKDP------ ---------- ---------- ------K-AE
AtCCRL14    --ASPF---- ------F-ND AKDP------ ---------- ---------- ------Q-AE
PoptrCCRL1   --ASPF---- ------Y-HD VKDP------ ---------- ---------- ------Q-AE
PoptrCCRL2   --ASPF---- ------Y-HD VKDP------ ---------- ---------- ------Q-VE
PpDFRL1     --AAAL---- ------V-RI KSDP------ ---------- ---------- ------K-AE
AtTKPR1     --ASPV---- ------L-KP TSNP------ ---------- ---------- ------E-EE
PoptrDFRL1   --ASPA---- ---------- ---------- ---------- ---------- ---------E
Os08g40440   --ASPV---- --------IT ETDS------ ---------- ---------- ------SKAA
Os09g32020   EKAPPV---- ---KVDELAK KEEPKFQAFS GTSYSLKRNR DKVSHLWRLP SAKERLQLEE
OryzaDFR2    --ASPV---- ------L-AK SDSN------ ---------- ---------- -----CK-EE
PpDFRL2     --ACPV---- ------VYDP NGDP------ ---------- ---------- ------E-VS
Os01g03670   --ASPVVVVG NSSSNNG-KP NDDDD----- ---------- ---------- --EEEVQ-QR
At1g25460    --ASRV---- ------S-VR LDNN------ ---------- ---------- ------N-LD
AtTKPR2     --ASPV---- --------LV PQDH------ ---------- ---------- ------NIQE
PoptrDFRL2   --ASPV---- ------L-LP HDDN------ ---------- ---------- -----IQ-AM
PoptrCCRL1   --AAPV---- ------L-VP YSDR------ ---------- ---------- -----IQ-ET
PoptrDFRL6   --AASM---- ------EFYV AGNE------ ---------- ---------- --------DN
At4g27250    --AASM---- ------EFDI SSDH------ ---------- ---------- ------V-NL
Os01g44260   -----M---- ------D-FE SEDP------ ---------- ---------- ------E-NE
AtDFR       --ATPM---- ------D-FE SKDP------ ---------- ---------- ------E-NE
PoptrDFR1    --ATPM---- ------D-FE SKDP------ ---------- ---------- ------E-NE
PoptrDFR2    --ATPM---- ------D-FE SKDP------ ---------- ---------- ------E-NE
AtANR/BAN    --ATPI---- ------N-FK SEDP------ ---------- ---------- ------E-KD
PoptrANR1    --ATPV---- ------N-FA SEDP------ ---------- ---------- ------E-ND
PoptrANR2    --ATPV---- ------N-FA SEDP------ ---------- ---------- ------E-ND
Os04g53810   --AAPV---- ---YDKS-HK SDDL------ ---------- ---------- ------E-KE
Os04g53800   --AAPM---- ------N-LK SQNP------ ---------- ---------- ------E-KE
Os04g53850   --AAPM---- ------N-FN SENP------ ---------- ---------- ------E-KD
Os04g53920   --AAPV---- ------N-FQ SQNP------ ---------- ---------- ------E-KE

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
             365       375       385       395       405       415
PpDFRL4     -------FFK VNVEGFKNVI EAVKAT-PSV KKLIYTSSFF AVGPTDG--- -----YIGDE
Os03g08624   -------FIT VNVGGLENVL KAARRT-PTV KKIVYTSSFF AIGPTDG--- -----YVADE
At4g33360    -------FIS VNVGGLKNVL EAVKET-KTV QKIIYTSSFF ALGSTDG--- -----SVANE
PoptrDFRL5   -------FFS VNVGGLKNVV QAAKET-KMI EKIIYTSSMV ALGSTDG--- -----YVADE
PoptrDFRL4   -------FFS VNVEGLNNVL QAAKET-ETI EKIIYTSSFF ALGSTDG--- -----YVADE
Os02g08420   -------IIE PVITGTLNVV EVAADA--GV RRVVLSSTIG TMYMDPRR-- DPD--SPLDD
Os09g04050   -------MIE PAIRGTRYVI TAAADT--GI KRVVFTSSIG TVYMNPYR-- DPN--KPVDD
Os08g34280   -------MVE PAVRGTEYVI NAAAEA-GTV RRVVFTSSIG AVTMDPNR-- GPD--VVVDE
Os09g25150   -------MVE PAVRGTEYVI KAAAEA-GTV RRVVFTSSIG AVTMDPNR-- GPD--VVVDE
PoptrCCR2    -------MVE PAVNGTKNVI IAAAEA--KV RRVVFTSSIG AVYMDPNK-- GPD--VVIDE
AtCCR1      -------MVE PAVNGAKFVI NAAAEA--KV KRVVITSSIG AVYMDPNR-- DPE--AVVDE
AtCCR2      -------MLE PAVNGAKFVI DAAAKA--KV KRVVFTSSIG AVYMNPNR-- DTQ--AIVDE
PpDFRL3     -------LIE PAVKGTLNVL ESCAKA--GT KKIVLTSSVA AVAYSPK-R- AGA--SVVDE
PoptrCCRL4   -------LID PALKGTLNVL RSCAKV-HSI RRVVLTSSAA ACIYSGKPL- NHD--VVIDE
PoptrCCRL3   -------LVD PALKGTVNVL RSCTKI-PSI KRVVITSSMA AVVFNGKSL- APD--VVVDE
Os01g34480   -------LLD PAVKGTLNVL GSCKKA--SI RRVIVTSSMA AVAYNGKPR- TPD--VVVDE
AtCCRL14    -------LID PAVKGTLNVL NSCAKA-SSV KRVVVTSSMA AVGYNGKPR- TPD--VTVDE
```

```
PoptrCCRL1   -------LLD PAVKGTLNVL GSCARH-PSI KRVVLTSSMA AVAYNRKPR- TPD--VVVDE
PoptrCCRL2   -------LLD PAVKGTLNVL GSCAKH-PSI RRVVLTSSVA AVAYNGKPR- TPD--VVVDE
PpDFRL1      -------MLD PTILGTLNVL HSCAKS-TTL KRVVLTSSTA AVRFRDDLE- QPGAVTYLDE
AtTKPR1      -------ILR PAIEGTLNVL RSCRKN-PSL KRVVLTSSSS TVRIRDDF-- DPK--IPLDE
PoptrDFRL1   -------ILE PAIEGTLNVL RSCKRN-PSL KRVILTSSSS TLRVRDDF-- DSN--IPLEE
Os08g40440   -------VLD SAINGTLNVL RSCKKN-PSL KRVVLTSSSS TVRLKDEADL PPN--VLLDE
Os09g32020   -------MLV PAINGTLNVL KSCKKN-PFL KRVVLTSSSS TVRIRDESK- HPE--ISLDE
OryzaDFR2    -------MLV PAINGTLNVL KSCKKN-PFL KRVVLTSSSS TVRIMDESK- HPE--ISLDE
PpDFRL2      -------MLN PALKGNLNVL RACTKS-HSI QRVVMTSSCS AIRYDHNRR- PED--PPLSE
Os01g03670   -------LVE PIVRGASNVL RSCARASPRP RRVVFTSSCS CVRYGAGAA- -----AALNE
At1g25460    -------KFD PNISGTMNVM NSCAKSRNTV KRIVLTSSST AIRYRFDA-- TQV--SPLNE
AtTKPR2     T------LVD PIIKGTTNVM SSCAKSKATL KRIVLTSSCS SIRYRFDA-- TEA--SPLNE
PoptrDFRL2   -------LID PCINGTLNVL NSCSKA-NTV KRVVLTSSCS SIRYRDDV-- QQV--SPLNE
PoptrCCRL1   -------LID PCIKGTLNVL NSCLKA-SSV KRVVFTSSSS TVRYRDDT-- PQI--FSLNE
PoptrDFRL6  ENYVQRNIID PAIEGTLNLL TSCSKS-NTV KRVVFTSSIS TLTAKDGAG- KWR--QVVDE
At4g27250   ESYVQSKVIE PALKGVRNVL SSCLKS-KSV KRVVFTSSIS TLTAKDENE- RMR--SFVDE
Os01g44260   -------VVK PTVEGMLSIM RACRDA-GTV KRIVFTSSAG TVNIEERQR- -----PSYDH
AtDFR       -------VIK PTVNGMLGIM KACVKA-KTV RRFVFTSSAG TVNVEEHQK- -----NVYDE
PoptrDFR1    -------VIK PTINGVLDIM RACANS-KTV RKIVFTSSAG TVDVEEKRK- -----PVYDE
PoptrDFR2    -------VIK PTINGVLDIM KACAKA-KTV RRIVFTSSAG TVDVEEHKK- -----PVYDE
AtANR/BAN    -------MIK PAIQGVINVL KSCLKS-KSV KRVIYTSSAA AVSINNLS-- GTG--IVMNE
PoptrANR1    -------MIK PAIQGVHNVL KACAKA-KTV KRVILTSSAA ALSINKLN-- GTG--LIMDE
PoptrANR2    -------MIK PAIQGVHNVL KACAKA-KTV QRVILTSSAA AVSINKLN-- GTG--LVMDE
Os04g53810   -------IVQ GGVEGTLNVM RSCARA-GTV KRVILTSSTA AVSSLRPLE- GAG--HVLDE
Os04g53800   -------LLE AGVQGTLNVL RSCVKA-GTV KRVILTSSAA AVSGQPLQGD GNGSSHVLDE
Os04g53850   -------LVE AAVNGTLNAM RSCAKV-GTV KRVIITSSDA AISRRPLQ-- GDG--YVLDE
Os04g53920   -------LIE AGVQGTMNVM RSCVRA-GTV KRVILTSSAP AVSGRPLQ-- GDG--HVLDE

           ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
             425       435       445       455       465       475
PpDFRL4     TQFHSMKAFY SP-------- -YEESKAFAD KLACEAAME- GVPIVSLYPG IIYGPGSMTK
Os03g08624   TQMHQGKTFC TE-------- -YEKSKVLAD QIALQAAAE- GMPITIVYPG FMYGPGKLTA
At4g33360    NQVHNERFFC TE-------- -YERSKAVAD KMALNAASE- GVPIILLYPG VIFGPGKLTS
PoptrDFRL5   SQVHHEKYFS TE-------- -YERSKVAAD KVASQAAAE- GLPIVTLYPG VVYGPGKLTT
PoptrDFRL4   SQVHCEKRFC TE-------- -YERSKMIAD KIASQAAAE- GVPIVMLYPG VIYGPGKLTT
Os02g08420   SFWSDLD-YC KNTK----NW -YCYAKTIAE RKAWEVARGR GVDMAVVIPV VVLGELLQPG
Os09g04050   TCWSDLE-YC KRTE----NW -YCYAKTVAE QGAWEVARRR GVDLVVVNPV LVLGPLLQAT
Os08g34280   SCWSDLD-YC KETR----NW -YCYGKAVAE QAAWEAARRR GVELVVVNPV LVIGPLLQPT
Os09g25150   SCWSDLE-FC KKTK----NW -YCYGKAVAE QEACKAAEER GVDLVVVSPV LVVGPLLQPT
PoptrCCR2    SCWSDLE-FC KNTK----NW -YCYGKAVAE QAAWDMAKEK GVDLVVVNPV LVLGPLLQPT
AtCCR1      SCWSDLD-FC KNTK----NW -YCYGKMVAE QAAWETAKEK GVDLVVLNPV LVLGPPLQPT
AtCCR2      NCWSDLD-FC KNTK----NW -YCYGKMLAE QSAWETAKAK GVDLVVLNPV LVLGPPLQSA
PpDFRL3     TFFSDPE-FC QKEQ----RW -YVLSKTLAE SAAWEFVKEH NLNMVAINPT MVIGPLLQSS
PoptrCCRL4   TWYSDPA-IC KELK----AW -YALSKTLAE EAAWNFAKEN ATDLVTVHPS FVIGPLLQPT
PoptrCCRL3   TWFSDSD-FC EKSK----LW -YHLSKTLAE EAAWKFTKEN GIDMVTLNPG LVIGPLLQPT
Os01g34480   TWFSVPE-IC EKHQ----QW -YVLSKTLAE EAAWKFSKDN GFEIVTVNPA MVIGPLLQPS
AtCCRL14    TWFSDPE-LC EASK----MW -YVLSKTLAE DAAWKLAKEK GLDIVTINPA MVIGPLLQPT
PoptrCCRL1   TWFSDPE-LC RESK----LW -YVLSKTLAE DAAWKFAKEK GMDMVAINPS MVIGPLLQPT
PoptrCCRL2   TWFSDPN-LC RESK----VW -YVLSKTLAE DAAWKFAKEK DMDMVAINPA MVIGPLLQPT
PpDFRL1     YSWSSIF-FC TKYQ----IW -YSLAKILSE QEAWKFAFLH SIDLVVVLPS FVIGPCLPYP
AtTKPR1     SIWTSVE-LC KRFQ----VW -YALSKTLAE QAAWKFSEEN GIDLVTVLPS FLVGPSLPPD
PoptrDFRL1   SSWSSVE-LC ERLQ----IW -YALSKTLAE KAAWEFCNGN GIDLITVLPS FVIGPSLSPD
Os08g40440   TSWSSME-FC ESLQ----IW -YAIAKTLAE KAAWEFAKEN GIDLVAVLPT FVVGPNLSHE
Os09g32020   TIWSSVA-LC EKLQ----LW -YALAKISAE KAAWEFAKEN NIDLVTVLPS FVIGPSLSHE
OryzaDFR2    TIWSSVA-LC EKLQ----LW -YALAKISAE KAAWEFAKEN NIDLVTVLPS FVIGPSLSHE
PpDFRL2     SVWSSPE-YC RDHK----MW -YALAKTLAE KEAFEFAARE GLNLVVICPS FVIGPSLTPI
Os01g03670   SHWSDAA-YC AAHG----LW -YAYAKTLAE REAWRLAKER GLDMVAVNPS FVVGPILSQA
At1g25460    SHWTDLE-YC KHFK----IW -YAYKKTLGE KEAWRIAADK KLNLVVVIPS FCIGPILSPK
AtTKPR2     SHWSDPE-YC KRFN----LW -YGYAKTLGE REAWRIAEEK GLDLVVVNPS FVVGPLLGPK
PoptrDFRL2   SHWSDPE-YC KRYD----LW -YAYAKTIGE KEAWRSAKEN GIDLVVVNPS FVVGPLLAPQ
PoptrCCRL1   SHWSDTE-YC KRHN----LW -YAYAKTVAE KEAWRVSKEN GIDLVSFIPS FVVGPLLAPE
PoptrDFRL6   TCQTPID-HV WNTK--PPGW IYVLSKRLTE EAAFKYAKDN GIDLISVITT TVAGAFLTSS
At4g27250    TCKAHVD-HV LKTQ--ASGW IYVLSKLVSE EEAFRYAKER GMDLVSVITT TVSGPFLTPF
```

```
Os01g44260   DDWSDID-FC RRVK--MTGW MYFVSKSLAE KAAMEYAREH GLDLISVIPT LVVGPFISNG
AtDFR       NDWSDLE-FI MSKK--MTGW MYFVSKTLAE KAAWDFAEEK GLDFISIIPT LVVGPFITTS
PoptrDFR1    SCWSDLD-FV QSIK--MTGW MYFVSKTLAE QAAWKFAKEN NLDFISIIPT LVVGPFIMQS
PoptrDFR2    SCWSDLE-FV QTVK--MTGW MYFVSKTLAE QAAWKYAKEN NLDFISVIPP LVVGPFIMHS
AtANR/BAN    ENWTDVE-FL TEEK--PFNW GYPISKVLAE KTAWEFAKEN KINLVTVIPA LIAGNSLLSD
PoptrANR1    KNWTDVE-FL TSEK--PPTW GYPASKTLAE KAAWKFAEEN NIDLITVIPS LMTGPSLTLD
PoptrANR2    KNWTDVE-FL TSEK--PPTW GYPASKTLAE KAAWKFAEEN NIDLITVIPS LMTGPSFTPH
Os04g53810   SSWSDIE-YL RSMEKLSPTQ AYSISKVLSE KEATKFAEEN GLSLVTLCPV VAVGASPAVR
Os04g53800   SSWSDLD-YL RSTNGISPAQ AYAIAKVLSE KEASKLAEEN GISLVAVCPV ATVGASPAPV
Os04g53850   ESWSDVD-YL RTEK--PPAW AYSVSKVLLE KAACKFAEEN NMSLVTVFPV FTLGAAPAPV
Os04g53920   DSWSDVE-YL TKEK--PPAW AYSVSKVLME KAACKLAEEN NISLITVFPV FTLGAAPTPT

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               485       495       505       515       525       535
PpDFRL4     GNSLAEMMIE RFNGRMPGYV GYKVK----- -KFSFCHIDD VVVAYLAAIE IGR-------
Os03g08624   GNLVSRILIE RFNGRLPGYV GHGHD----- -RESFCHVDD VVAGHVAAME KGR-------
At4g33360    ANMVARMLIE RFNGRLPGYI GSGTD----- -RYSFSHVDD VVEGHVAAME KGR-------
PoptrDFRL5   GNALAKMLID RFAGRLPGYI GRGND----- -RLSFCHVDD VVGGHIAAMD KGR-------
PoptrDFRL4   GNIVAQLLIE RFAGRLPGYI GYGND----- -KFSFCHVDD LVDGHIAAMD KGR-------
Os02g08420   MNTSTKHILK YLTGEAKTYV NE-------- -SHAYVHVVD AAEAHVRVLE APG-------
Os09g04050   VNASTEHVMK YLTGSAKTYV NA-------- -AQAYVHVRD VAEAHVRVYD CGG-------
Os08g34280   VNASVAHILK YLDGSASKFA NA-------- -VQAYVDVRD VAAAHLLVFE SPS-------
Os09g25150   VNASAVHILK YLDGSAKKYA NA-------- -VQAYVDVRD VAAAHVRVFE APE-------
PoptrCCR2    VNASITHILK YLTGSAKTYA NS-------- -VQAYVHVRD VALAHILVFE TPS-------
AtCCR1      INASLYHVLK YLTGSAKTYA NL-------- -TQAYVDVRD VALAHVLVYE APS-------
AtCCR2      INASLVHILK YLTGSAKTYA NL-------- -TQVYVDVRD VALGHVLVYE APS-------
PpDFRL3     MNTSNELLLG FLNGTAKSFP NQ-------- -AVGWVSVKD VAMAHILAYE KPE-------
PoptrCCRL4   LNLSVEMILD LVNG-AETYP NG-------- -YYRCIDVRD VANAHIQAFE IPS-------
PoptrCCRL3   LNQSAESVLD LING-AKSYP NT-------- -TYRWVDVRD VANAHIYALE NPS-------
Os01g34480   LNTSAEAILK LINGSSSTYP NF-------- -SFGWINVKD VALAHILAYE VPS-------
AtCCRL14    LNTSAAAILN LING-AKTFP NL-------- -SFGWVNVKD VANAHIQAFE VPS-------
PoptrCCRL1   LNTSAAAILS LIKG-AQTFS NA-------- -SFGWINVKD VANAHIQAFE LSS-------
PoptrCCRL2   LNTSAAAILS LIKG-AQTFP NA-------- -SFGWINVKD VANAHIQAFE LSS-------
PpDFRL1     LSKTAQDICD LLNGLCRNFG IHG------- -RMGYVHVDD VARAHILVYE TPS-------
AtTKPR1     LCSTASDVLG LLKGETEKFQ WHG------- -QMGYVHIDD VARTHIVVFE HEA-------
PoptrDFRL1   LCSTATDVLG LLTGESEKFH WHG------- -RMGYVHIDD VALSHILVYE DET-------
Os08g40440   LSPTTTDVLG LFQGETTKFT MYG------- -RMGYVHIDD VASCHILLYE TPR-------
Os09g32020   LSVTASDILG LLQGDTDRFI SYG------- -RMGYVHIDD VASCHILVYE APQ-------
OryzaDFR2    LSVTASDILG LLQGDTDRFI SYG------- -RMGYVHIDD VASCHILVYE APQ-------
PpDFRL2     PTSTVFLILD LLRGRAQEYP NK-------- -RIGFVHIDD VVTAHVLAME VPE-------
Os01g03670   PTSTALIVLA LLRGELPRYP NT-------- -TVGFVHVDD AVLAHVVAME DAR-------
At1g25460    PTSSPLIFLS IIKGTRGTYP NF-------- -RGGFVHIDD VVAAQILAME EPK-------
AtTKPR2     PTSTLLMILA IAKGLAGEYP NF-------- -TVGFVHIDD VVAAHVLAME EPK-------
PoptrDFRL2   PTSTLLLILA IVKGLRGEYP NM-------- -TIGFVHIDD VVAAHILAME DKK-------
PoptrCCRL1   PNSTLLLIQS VVKGSRGEYP NM-------- -TVGFTHIDD VVAGNILAME NSE-------
PoptrDFRL6   VPSSIRVLLS PITGDTKFFS ILSAVNARMG -SIALVHIDD ICDAHIFLME QTR-------
At4g27250    VPSSVQVLLS PITGDSKLFA ILSAVNKRMG -SIALVHIED ICRAHLFLME QPK-------
Os01g44260   MPPSHVTALA LLTGNEAHYS ILK------- -QVQFVHLDD LCDAEIFLFE SPE-------
AtDFR       MPPSLITALS PITRNEAHYS IIR------- -QGQYVHLDD LCNAHIFLYE QAA-------
PoptrDFR1    MPPSLLTALS LITGNEAHYG ILK------- -QGHYVHLDD LCMSHIFLYE NPK-------
PoptrDFR2    MPPSLITALS LITGNEAHYG IIK------- -QGNYVHLDD LCRAHIVLFE NPK-------
AtANR/BAN    PPSSLSLSMS FITGKEMHVT GLKEMQKLSG -SISFVHVDD LARAHLFLAE KET-------
PoptrANR1    IPSSVHLSMS LITGNEFLKN ALKGMQMLSG -SISITHVED VCRAHIFLAE KES-------
PoptrANR2    IPDSINLAMS LITGNKFLIN GLKGMQMLSG -SISITHVED VCRAHIFLAE KES-------
Os04g53810   VDTSVPACLS LITGDEEMMN ILKGIEKASG WSMPMVHIED VCRAEIFVAE EES-------
Os04g53800   ANESVANVLS LLSGNEEINT LRMIDQYSG- -GLKLVHVDD LCRAEIFLAE KASPS-----
Os04g53850   ARTSVPGILS LLSGDETHLE VLKPLQWVTG -SVSIVHVDD LCRAEIFLAE KESSSLSSAE
Os04g53920   AATSVSAMLS LLSSDETQLK TLKGLA-ATG -PIPTVHVDD LCRAEVFVAE KES-------

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
               545       555       565       575       585       595
PpDFRL4     VGE-RYMLCG DNMSFHEVFD LAAGLTKTNP AKVTIPMWVL DVAGFLCVQW ARFGAWTGIS
Os03g08624   EGE-RYLLTG ENTSLVQIFD MASRITNTKA PRFHVPLWLL EIYGWISVLV SRI------T
```

```
At4g33360   LGE-RYLLTG ENASFKLVFD MAALITGTKK PNFSIPLWAI NAYGWLSVLI SRV------T
PoptrDFRL5  LGE-RYLLTG ENASFSRVLD IAAIITRTEK PRFSIPLWVI EAYGWLSILI FHF------T
PoptrDFRL4  QGE-RYLLTG ENASFKLVFD MAAIISETKK PRFSIPLCII ESYGWLLVLV SRL------T
Os02g08420  AGGRRYICAE RTLHRG---E L-CRILAGLF PEYPIPTRCR DEI------- ---------N
Os09g04050  ARG-RYICAE STLHRG---D L-CRALAKLF PEYPVPSRCK DEA------- ---------A
Os08g34280  AAG-RFLCAE SVLHRE---G V-VRILAKLF PEYPVPTRCS DEK------- ---------N
Os09g25150  ASG-RHLCAE RVLHRE---D V-VHILGKLF PEYPVPTR-- ---------- ----------
PoptrCCR2   ASG-RYLCSE SVLHRG---E V-VEILAKFF PEYPIPTKCS DEK------- ---------N
AtCCR1      ASG-RYLLAE SARHRG---E V-VEILAKLF PEYPLPTKCK DEK------- ---------N
AtCCR2      ASG-RYILAE TALHRG---E V-VEILAKFF PEYPLPTKCS DEK------- ---------N
PpDFRL3     AEG-RYIINE RLIHYG---E M-VSLLMNRY PQYPIVAKDA DDS------- ----------
PoptrCCRL4  ASG-RYVLTA YVTTFS---E V-LKIIRENY PTLRLPEKST ESM------- ----------
PoptrCCRL3  ANG-RYCLVG TVIHSS---E A-VKILSKLY PDLTIPKQCA DDK------- ----------
Os01g34480  ANG-RYCMVE RVAHYS---E L-VQIIREMY PNIPLPDKCA DDK------- ----------
AtCCRL14    ANG-RYCLVE RVVHHS---E I-VNILRELY PNLPLPERCV DEN------- ----------
PoptrCCRL1  ASG-RYCLVE RVAHHS---E V-VKILRELY PDLQLPEKCA DDK------- ----------
PoptrCCRL2  ASG-RYCLVE RVAHYS---E V-VKILHELY PDLQLPEKCA DDK------- ----------
PpDFRL1     AQG-RYICSA QEATPQ---E L-VQYLADRY PHLQISTKFN DEL------- ----------
AtTKPR1     AQG-RYICSS NVISLE---E L-VSFLSARY PSLPIPKRFE KLN------- ----------
PoptrDFRL1  AGG-RFLCSS IVLDND---E L-ASFLSQRY PSLPIPKRFE QLK------- ----------
Os08g40440  AAG------- ---------- ---------- ---SLPCVYG EQT------- ----------
Os09g32020  ATG-RYLCNS VVLDNN---E L-VALLAKQF PIFPIPRSLR NPY------- ----------
OryzaDFR2   ATG-RYLCNS VVLDNN---E L-VALLAKQF PIFPIPRSLR NPY------- ----------
PpDFRL2     AHG-RYICSS DVAHFG---D I-MSMLKTKY PKLQTPTRCS DMP------- ---------P
Os01g03670  ASG-RLICSC HVAHWS---E I-VGSLRERY PGYPIPAECG SHK------- ----------
At1g25460   ASG-RILCSS SVAHWS---E I-IEMLRIKY PLYPFETKCG SEE------- ----------
AtTKPR2     ASG-RIICSS SVAHWS---E I-IELMRNKY PNYPFENKCS NKE------- ----------
PoptrDFRL2  ASG-RLVCSG SVAHWS---E I-IEMLRAKY PSYPYENKCS SQK------- ----------
PoptrCCRL1  ASG-RLVCSG PVAHWS---Q I-IKMLRAKY PSYPYENKCS SQE------- ----------
PoptrDFRL6  AEG-RYICSA HSCVLS---Q L-INHLVEEY PCSNIQRLAE KQG------- ----------
At4g27250   AKG-QYICCV DNIDMH---E LMLHHFSKDY LCKVQKVNED EEE------- ----------
Os01g44260  ARG-RYVCSS HDATIH---G L-ATMLADMF PEYDVPRSFP GID------- ---------A
AtDFR       AKG-RYICSS HDATIL---T I-SKFLRPKY PEYNVPSTFE GVD------- ----------
PoptrDFR1   AEG-RYICNS DDANIH---D L-AKLLREKY PEYNVPAKFK DID------- ----------
PoptrDFR2   AEG-RYICSS HEATIH---D L-AKLLREKY PKYNVPAKFK DID------- ----------
AtANR/BAN   ASG-RYICCA YNTSVP---E I-ADFLIQRY PKYNVLSEFE EGL------- ----------
PoptrANR1   ASG-RYICCA VNTSVV---E L-AEFLNKRY PQYQVPTDFG DF-------- ----------
PoptrANR2   ASG-RYICCG VNTSVV---E L-AKFLNKRY PQYQVPTDCG DF-------- ----------
Os04g53810  ASG-RYICGS LNTTVT---E I-AGFLAAKY PQYNVRCDCI EEH------- ---------H
Os04g53800  PSG-RYICCA LNTTMR---Q I-ARSLAAKY PHHNVDIDAL GGG------- ---------L
Os04g53850  SSA-RYICCS FNTTVL---A L-ARFMAGRY PQYNVKTDRF DGM------- ----------
Os04g53920  ASG-RYICSS LSTTVV---A F-ARFVAGKH PRYNVKTDGF QGF------- ----------

            ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
              605       615       625       635       645       655
PpDFRL4     HQIPFITTHS VNILKHQWAY SSEKAERELG Y-KSR--PLE EGLLQL-LTW LKATGRIKY-
Os03g08624  GKLPFISYPA VRVLRHQWAY SCEKAKKELG Y-SPR--SLT EGLSET-LLW LKDSEMIRF-
At4g33360   GKLPLISPPT VTVLRHQWSY SCDKAKLELG Y-NPR--SLK EGLEEM-LPW LKSLGVIHY-
PoptrDFRL5  GKLPLLCPPS VHVLRHQWEY SCEKARIELD Y-NPR--SLK EGLDEL-LPW LKSLGAITY-
PoptrDFRL4  GNLPLISPPT VHVLRHQWEY SCEKAKTELG Y-NPR--GLE DGLKEV-LPW LKSMGVIKY-
Os02g08420  PPKKGYKFTN QPL------- ------KDLG I-KFT--PVH EYLYEA-VKS LEDKGFIKKT
Os09g04050  PPVKGYLFSN QRL------- ------RDLG M-DFV--PVR QCLYET-VRS LQDKGLLPVL
Os08g34280  PRKQPYKMSN QKL------- ------RDLG L-EFR--PAS QSLYET-VKC LQEKGHLPVL
Os09g25150  ---------- ---------- ---------- ---------- ---------- ----------
PoptrCCR2   PRKQPYKFSN QKL------- ------RDLG F-EFT--PVK QCLYET-VKS LQEKGHLPIP
AtCCR1      PRAKPYKFTN QKI------- ------KDLG L-EFT--STK QSLYDT-VKS LQEKGHLAPP
AtCCR2      PRAKPYKFTT QKI------- ------KDLG L-EFK--PIK QSLYES-VKS LQEKGHLPLP
PpDFRL3     TRLPSYNLSN EKI------- ------KKLG L-TFQ--PLE EALDET-VAC FKELKLLD--
PoptrCCRL4  --FKPYQVSK EKA------- ------KTLG I-NFT--PLD LSLVDT-IES LKEKGFLKI-
PoptrCCRL3  PPMPKYQVSK ERA------- ------ASLG V-KYT--PLE ASLKDT-IES LKEKNFVSF-
Os01g34480  PSVPIYQVSK EKI------- ------KSLG L-ELT--PLH TSIKET-IES LKEKGFVTFD
AtCCRL14    PYVPTYQVSK DKT------- ------RSLG I-DYI--PLK VSIKET-VES LKEKGFAQF-
PoptrCCRL1  PYVPIYQVSK EKA------- ------KSLG I-EFI--PLE ASIKET-VES LKEKGFVSF-
```

```
PoptrCCRL2   PYVPIYQVSK EKA------- ------KSLG V-EFI--PLE ASVKET-VES LKEKGFVSF-
PpDFRL1     PKMPYYKLNT TKL------- ------QRLG L-NCK--PLD VMFDDC-ISF LEEKGLLKRK
AtTKPR1     --RLHYDFDT SKI------- ------QSLG L-KFK--SLE EMFDDC-IAS LVEQGYLSTV
PoptrDFRL1   --RPYYEFNT SRL------- ------ERLG F-KFK--PIQ EMFDDC-IAS LVEQGHLSSF
Os08g40440   -----YGFST AKV------- ------RELG M-KFR--DVE EMFDDA-VDS LRAHGYLLNS
Os09g32020   -EKQSYELNT SKI------- ------QQLG F-KFK--GVQ EMFGDC-VES LKDQGHLLEC
OryzaDFR2    -EKQSYELNT SKI------- ------QQLG F-KFK--GVQ EMFGDC-VES LKDQGHLLEC
PpDFRL2     GDDIHHKMDT TKI------- ------KKLG LTEFK--SIE QMFDDM-LRS LHEKHLESL-
Os01g03670   GDDRAHKMDT AKI------- ------RALG FPPFL--SVQ QMFDDC-IKS FQDKGLLPPH
At1g25460    GKDMPHSLDT TKI------- ------HELG FASFK--SLT EMFDDC-IKC FQDKGLL---
AtTKPR2     GDNSPHSMDT RKI------- ------HELG FGSFK--SLP EMFDDC-IIS FQKKGLL---
PoptrDFRL2   GDCNPHSMDT TKI------- ------ATLG FPPFK--TLE EMFDDC-IKS FQEKGFL---
PoptrCCRL1   GDNIPHSMDT TKI------- ------AQLG LPPFK--THE QMFDDC-IRS LQEKGFL---
PoptrDFRL6   --SISPEISS KKL------- ------RDMG F-KYKH-SIK DIISET-I-- ----------
At4g27250    RECMKPIISS KKL------- ------RELG F-EYKY-GIE EIVDQT-IDA SIKIKFPTLN
Os01g44260   DHLQPVHFSS WKL------- ------LAHG F-RFRY-TLE DMFEAA-VRT CREKGLLPPL
AtDFR       ENLKSIEFSS KKL------- ------TDMG F-NFKY-SLE EMFIES-IET CRQKGFLPVS
PoptrDFR1    ENLACVAFSS KKL------- ------TDLG F-EFKY-SLE DMFAGA-VET CREKGLIPLS
PoptrDFR2    EDLASVVFSS KKL------- ------LDLG F-EFKY-SLE EMFAGA-VET CREKGLIPLS
AtANR/BAN    -SIPKLTLSS QKL------- ------INEG F-RFEY-GIN EMYDQM-IEY FESKGLIKAK
PoptrANR1    PSKAKLAITS EKL------- ------ISEG F-SFKY-GIE EVYDQT-VEY FKAKGLLN--
PoptrANR2    PSEAKLIITS EKL------- ------SSEG F-SFKY-GIE EIYDQT-VEY FKANGLLN--
Os04g53810   PEKPTISLSS AKL------- ------IGEG F-EFKYKNLD EMYDDL-VAY GKALGLIPN-
Os04g53800   PEKPTILLSS EKL------- ------TSEG F-EFMYKTVD EMYDDAFVEY GMALGILHY-
Os04g53850   PEKPRVCCSS EKL------- ------IREG F-EFKYTNMG DILDDL-VEY GRALGILPH-
Os04g53920   PEKPRVCYSS EKL------- ------VREG F-EFKWTDLD EVFDDL-IEY GNVLGILPQ-

           ....|....| ....|....| ....|....| ....|....| ....|....| ....
             665       675       685       695       705
PpDFRL4     ---------- ---------- ---------- ---------- ---------- ----
Os03g08624   ---------- ---------- ---------- ---------- ---------- ----
At4g33360    ---------- ---------- ---------- ---------- ---------- ----
PoptrDFRL5   ---------- ---------- ---------- ---------- ---------- ----
PoptrDFRL4   ---------- ---------- ---------- ---------- ---------- ----
Os02g08420   SNTKELHRQS SPPQNSPASM LMSKL----- ---------- ---------- ----
Os09g04050   PPTADDHHHP SS-------- ---------- ---------- ---------- ----
Os08g34280   AAEKTEEEAG EVQGGIAIRA ---------- ---------- ---------- ----
Os09g25150   ---------- ---------- ---------- ---------- ---------- ----
PoptrCCR2    KQAAEESLKI Q--------- ---------- ---------- ---------- ----
AtCCR1      PPPPSASQES VENGIKIGS- ---------- ---------- ---------- ----
AtCCR2      QDSNQNEVII ES-------- ---------- ---------- ---------- ----
PpDFRL3     ---------- ---------- ---------- ---------- ---------- ----
PoptrCCRL4   ---------- ---------- ---------- ---------- ---------- ----
PoptrCCRL3   ---------- ---------- ---------- ---------- ---------- ----
Os01g34480   SSNL------ ---------- ---------- ---------- ---------- ----
AtCCRL14     ---------- ---------- ---------- ---------- ---------- ----
PoptrCCRL1   ---------- ---------- ---------- ---------- ---------- ----
PoptrCCRL2   ---------- ---------- ---------- ---------- ---------- ----
PpDFRL1     PEKTPTSSST PDEHSKDSVL QNV------- ---------- ---------- ----
AtTKPR1     LP-------- ---------- ---------- ---------- ---------- ----
PoptrDFRL1   SLAIN----- ---------- ---------- ---------- ---------- ----
Os08g40440   VP-------- ---------- ---------- ---------- ---------- ----
Os09g32020   PL-------- ---------- ---------- ---------- ---------- ----
OryzaDFR2    PL-------- ---------- ---------- ---------- ---------- ----
PpDFRL2     ---------- ---------- ---------- ---------- ---------- ----
Os01g03670   A--------- ---------- ---------- ---------- ---------- ----
At1g25460    ---------- ---------- ---------- ---------- ---------- ----
AtTKPR2     ---------- ---------- ---------- ---------- ---------- ----
PoptrDFRL2   ---------- ---------- ---------- ---------- ---------- ----
PoptrCCRL1   ---------- ---------- ---------- ---------- ---------- ----
PoptrDFRL6   ---------- ---------- ---------- ---------- ---------- ----
At4g27250    HKLRQ----- ---------- ---------- ---------- ---------- ----
Os01g44260   PPPPTTAVAG GDGSAGVAGE KEPILGRGTG TAVGAETEAL VK-------- ----
```

```
AtDFR       LSYQSISEIK TKNENIDVKT GDGLTDGMKP CNKTETGITG ERTDAPMLAQ QMCA
PoptrDFR1    HRKQVVEECK ENEVVPAS-- ---------- ---------- ---------- ----
PoptrDFR2    HEK------- ---------- ---------- ---------- ---------- ----
AtANR/BAN    ---------- ---------- ---------- ---------- ---------- ----
PoptrANR1    ---------- ---------- ---------- ---------- ---------- ----
PoptrANR2    ---------- ---------- ---------- ---------- ---------- ----
Os04g53810   ---------- ---------- ---------- ---------- ---------- ----
Os04g53800   ---------- ---------- ---------- ---------- ---------- ----
Os04g53850   ---------- ---------- ---------- ---------- ---------- ----
Os04g53920   ---------- ---------- ---------- ---------- ---------- ----
```