THE ACYL-COA LIGASE–LIKE (*ACLL*) GENE FAMILY IN ARABIDOPSIS AND POPLAR

by

CLARICE DE AZEVEDO SOUZA

B.Sc., Universidade de Sao Paulo, Sao Paulo – Brazil, 2001

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

August, 2007

ABSTRACT

Many genes of unknown function have been annotated in plant genome projects, and many of these may encode undiscovered enzymes. For example, completion of the *Arabidopsis thaliana* genome sequence revealed large families of phenylpropanoid-like enzymes of unknown functions. Using an *in silico* similarity search based on the amino-acid sequences of known Arabidopsis genes encoding 4-coumarate:CoA ligase (4CL), I identified nine putative genes as members of the Arabidopsis acyl-CoA ligase-like (*ACLL*) gene superfamily which encode a plant-specific clade of enzymes closely related to true 4CLs. I also identified all ACLLs in the fully sequenced poplar and rice genomes. Phylogenetic analysis of amino-acid sequences revealed five ACLL clades, each containing at least one ACLL member from each species, suggesting conserved biochemical functions for ACLL enzymes. In four of five clades, most of the ACLL representatives have the PTS1 peroxisomal target sequence, indicating a likely function in that organelle. I established tissue expression profiles and the wound and herbivory responsiveness of Arabidopsis and poplar *ACLL* genes, and this revealed similar expression patterns for potentially orthologous genes. Finally, I mined publicly available microarray databases for co-expressed Arabidopsis genes, and this data provides clues for potential ACLL biochemical functions. The only non-peroxisomal clade is the one most closely related to true 4CLs and contains a single copy gene in Arabidopsis (*ACLL5*) and poplar (*ACLL13*). These genes are flower and anther-preferred in expression, and because of the apparent conservation in sequence and in expression, were chosen for functional analysis. *ACLL5* is transiently expressed in tapetum cells just prior to release of microspores from tetrads, suggesting a role in pollen wall and/or sporopollenenin formation. In support of this, an *acll5* transposon insertion mutant is male sterile and fails to produce pollen grains. These data suggest that ACLL5 and similar enzymes from other species, produce CoA ester intermediates used in an unknown pathway required for pollen wall formation. *In silico* co-expression analysis in Arabidopsis has revealed potential other members of this pathway, also conserved across angiosperms. This work highlights the utility of the Arabidopsis model system in the discovery of genes in other plant species with genome sequence information.

TABLE OF CONTENTS

ABSTRACT II

TABLE OF CONTENTS III

LIST OF TABLES VI

LIST OF FIGURES VII

LIST OF ABBREVIATIONS IX

AGNOWLEDGEMENTS XI

CO-AUTHORSHIP STATEMENT XII

CHAPTER 1 - INTRODUCTION 1

1.1 NATURAL PRODUCT DIVERSITY AND PLANT GENOMICS 1
1.1.1 FULLY SEQUENCED PLANT GENOMES AND LESSONS LEARNED FROM THEM 5
1.2 PHENYLPROPANOID METABOLISM AND ADENYLATE-FORMING ENZYMES 8
1.2.1 PHENYLPROPANOID METABOLISM 8
1.2.2 ADENYLATE-FORMING ENZYMES 11
1.2.3 *4CL* GENE FAMILIES 13
1.2.4 *4CL* SUBFUNCTIONALIZATION 13
1.2.5 IDENTIFICATION OF *4CL-LIKE* GENES (*4CLLS*) AND PHYLOGENETIC RELATIONSHIPS TO OTHER ADENYLATE-FORMING ENZYMES 17
1.2.6 4CL PROTEIN STRUCTURE 18
1.3 THE ACLLS AND THESIS OBJECTIVES 19

CHAPTER 2 - MATERIAL AND METHODS 22

2.1 NUCLEIC ACID METHODS 22
2.1.1 GENOMIC DNA AND TOTAL RNA ISOLATION 22
2.1.2 DNA EXTRACTION FROM AGAROSE GELS 22
2.1.3 PLASMID DNA PREPARATION AND SEQUENCING 22
2.1.4 SEQUENCE ALIGNMENT AND EDITING 23
2.1.5 REVERSE TRANSCRIPTION FOR CDNA SYNTHESIS 23
2.1.6 GENERAL RECOMBINANT DNA METHODS 23
2.2 PLANT GROWTH AND MAINTENANCE 26
2.2.1 SEED HARVESTING AND SOWING 26
2.2.2 PLANT GROWTH CONDITIONS 26
2.2.3 MECHANICAL WOUND AND HERBIVORY 27
2.3 GENERAL TRANSFORMATION PROCEDURES 27
2.3.1 BACTERIA TRANSFORMATION 27

2.3.2 ARABIDOPSIS TRANSFORMATION 27
2.3.3 TOBACCO TRANSFORMATION (DONE BY K. TURNER, BC INSTITUTE OF TECHNOLOGY) 28
2.4 GENERAL BIOINFORMATICS PROCEDURES **29**
2.4.1 SEQUENCE SELECTION AND PHYLOGENETIC TREE CONSTRUCTION 29
2.4.2 IDENTIFICATION OF *CIS*-ACTING PROMOTER ELEMENTS 30
2.4.3 SEARCH OF CO-REGULATED GENES IN PUBLIC ARABIDOPSIS MICROARRAY DATABASE 30
2.6 GENE EXPRESSION ANALYSIS **30**
2.6.1 QUANTITATIVE REAL-TIME PCR 30
2.6.2 SEMI-QUANTITATIVE RT-PCR 34
2.6.3 GUS HISTOCHEMICAL ASSAY 34
2.7 SUB-CELLULAR LOCALIZATION OF ARATHACLL4 AND POPTRACLL5 ..... **35**
2.8 IDENTIFICATION AND CHARACTERIZATION OF AN *ACLL5* INSERTION MUTANT **35**
2.8.1 GENETIC METHODS 35
2.8.2 PHENOTYPIC ANALYSIS OF THE *ACLL5-1* MUTANT 36
2.8.3 *IN SITU* HYBRIDIZATION (EXPERIMENT PERFORMED BY S.MCKIM, UBC) 37

CHAPTER 3 - GENOME-WIDE PHYLOGENETIC ANALYSIS AND COMPARATIVE GENOMICS OF THE PLANT-SPECIFIC ACYL:COENZYMEA LIGASE-LIKE (*ACLL*) GENE FAMILY IN ARABIDOPSIS AND POPLAR **38**

3.1 INTRODUCTION **38**
3.2 RESULTS **42**
3.2.1 PHYLOGENETIC ANALYSIS OF 100 ACLLS FROM PLANTS AND MICROORGANISMS 42
3.2.2 MOST ACLLS CONTAIN THE PTS1 (PEROXISOMAL TARGET SIGNAL 1) 45
3.2.3 SPECIES-SPECIFIC *ACLL* GENE FAMILY EVOLUTION 45
3.2.4 COMPARATIVE ANALYSIS OF ARABIDOPSIS AND POPLAR 4CL AND ACLL PROTEINS 50
3.2.5 COMPARATIVE EXPRESSION ANALYSIS OF ARABIDOPSIS AND POPLAR GENES 52
3.2.7 *IN SILICO* CO-EXPRESSION ANALYSIS OF ARABIDOPSIS *ACLL* GENES 60
3.2.8 POPLAR GENES ACTIVATED BY ADDITIONAL STRESS TREATMENTS 63
3.2.9 SUB-CELLULAR LOCALIZATION OF POPTRACLL5 67
3.3 DISCUSSION **69**
3.3.1 *4CL* AND *ACLL* GENE EVOLUTION 69
3.3.2 *ACLL* GENE FAMILY STRUCTURE AND EXPRESSION PATTERNS 71
3.3.3 SUMMARY AND COMBINING DATA TO MAKE FUNCTIONAL INFERENCES 77

CHAPTER 4 – THE *ARABIDOPSIS THALIANA* FLOWER-SPECIFIC ACYL-COENZYMEA LIGASE GENE *ACLL5* IS CONSERVED IN ANGIOSPERMS AND IS REQUIRED FOR MALE FERTILITY **81**

4.1 INTRODUCTION **81**
4.2 RESULTS **86**
4.2.1 IDENTIFICATION OF AN *ACLL5* INSERTION MUTANT 86
4.2.2 *ACLL5* MUTATION IS CORRELATED WITH MALE STERILITY AND ABSENCE OF POLLEN GRAINS 88
4.2.3 GENETIC CHARACTERIZATION OF THE *ACLL5* MUTANT 90
4.2.4 PHENOTYPIC ANALYSIS OF ANTHER DEVELOPMENT IN THE *ACLL5-1* MUTANT 91
4.2.5 *IN SITU* HYBRIDIZATION ANALYSIS OF *ACLL5* EXPRESSION IN DEVELOPING ANTHERS 94
4.2.6 CO-EXPRESSION ANALYSIS OF *ACLL5* IN *ARABIDOPSIS* 96
4.3 DISCUSSION **102**

CHAPTER 5- CONCLUSIONS AND FUTURE DIRECTIONS 112

5.1 A TIMELINE OF DISCOVERIES 112
5.2 FUTURE DIRECTIONS 114
5.2.1 MUTANT ANALYSIS 114
5.2.2 *ACLL5* MUTANT COMPLEMENTATION 117
5.2.3 MUTANT STUDIES IN POPLAR AND OTHER PLANT SPECIES 119
5.2.4 BIOCHEMICAL CHARACTERIZATION POPLAR CLADE D ACLLS 120
5.2.5 4CL/ACLL STRUCTURAL INFORMATION AND IDENTIFICATION OF SUBSTRATES 121
5.2.6 CONTINUOUS MINING OF DATA 124
5.3 FINAL REMARKS 126

REFERENCES 128

APPENDIX 1 140

LIST OF TABLES

TABLE 1.1: KINETIC PROPERTIES OF RECOMBINANT ARABIDOPSIS 4CLS..... 15

TABLE 2.1: PRIMERS USED FOR *ACLL* PROMOTER AMPLIFICATION AND CLONING..... 25

TABLE 2.2: PRIMERS USED FOR GENERATING GFP FUSIONS OF ARATHACLL4 AND POPTRACLL5..... 26

TABLE 2.3: PRIMERS USED FOR QUANTITATIVE AND SEMI-QUANTITATIVE RT- PCR..... 33

TABLE 2.4: PRIMERS USED FOR GENOTYPING *ACLL5* TRANSPOSON INSERTION LINES..... 36

TABLE 3.1: ANNOTATION OF *POPULUS TRICHOCARPA* AND *ORYZA SATIVA ACLL* GENES..... 47

TABLE 3.2: AMINO ACID IDENTITY COMPARISON OF FULL-LENGTH AMINO ACID SEQUENCES OF 4CLS AND ACLLS IN THE DIFFERENT CLADES..... 52

TABLE 4.1: CONSENSUS *CIS* ELEMENT MATCHES IN *ACLL5* AND CO-EXPRESSED PHENYLPROPANOID-LIKE GENE PROMOTER REGIONS..... 106

LIST OF FIGURES

FIGURE 1.1: SCHEMATIC VIEW OF THE RELATIONSHIP OF PHENYLPROPANOID METABOLISM TO PRIMARY METABOLISM.....9

FIGURE 1.2: A SIMPLIFIED SCHEME OF THE PHENYLPROPANOID PATHWAY.....11

FIGURE 1.3: THE TWO-STEP MECHANISM OF THE COA-LIGASE REACTION..... 12

FIGURE 2.1: PCAMBIA VECTOR 1305.1 USED FOR PROMOTER::*GUS* CLONING..... 24

FIGURE 3.1: PHYLOGENETIC TREE OF 100 ACLLS FROM VARIOUS ORGANISMS.....43

FIGURE 3.2: PHYLOGENETIC RELATIONSHIP OF PLANT-SPECIFIC ACLLS, FROM ARABIDOPSIS, POPLAR AND RICE..... 46

FIGURE 3.3: SCHEMATIC REPRESENTATION OF THE *ARABIDOPSIS THALIANA* GENOME, SHOWING THE LOCATION OF ACLL GENES WITH THE RESPECTIVE CLADES..... 49

FIGURE 3.4: TISSUE EXPRESSION PROFILE OF ACLLS IN ARABIDOPSIS AND POPLAR..... 55

FIGURE 3.5: EFFECT OF WOUND STRESS ON PEROXISOMAL ACLLS.....58

FIGURE 3.6: PAJEK CO-EXPRESSION NETWORKS GENERATED FROM PRIME CORRELATED GENE SEARCH TOOL DATA (HTTP://PRIME.PSC.RIKEN.JP). (**A**) CO-EXPRESSION NETWORK OF ARATHACLL3. (**B**) CO-EXPRESSION NETWORK OF ARATHACLL4.....61

FIGURE 3.7: REAL-TIME PCR DATA SHOWED EXPRESSION OF POPLAR *ACLLS* AFTER SIMULATED HERBIVORY, HERBIVORY BY THE FOREST TENT CATERPILLAR (*MALACOSOMA DISSTRIA*), AND EXPOSURE TO MEJA..... 66

FIGURE 3.8: CONFOCAL MICROSCOPY IMAGE SHOWING SUB-CELLULAR LOCALIZATION OF ARATHACLL4 (OPDA/OPC8-COA LIGASE) AND THE POPLAR HOMOLOGUE POPTRACLL5 IN GUARD CELLS OF TRANSGENIC TOBACCO LINES.....68

FIGURE 4.1: (**A**) SCHEMATIC REPRESENTATION OF THE ACLL5 (AT1G62940) GENE, SHOWING THE LOCATION OF THE TRANSPOSON INSERTION. (**B**) ACLL5 EXPRESSION IN WILD-TYPE (WT) AND ACLL5-1 HOMOZYGOTE LINES.....86

FIGURE 4.2: PHENOTYPIC CHARACTERIZATION *ACLL5*-1 HOMOZYGOUS PLANTS.....88

FIGURE 4.3: SEM OF WILD TYPE (WT) AND HOMOZYGOUS *ACLL5*-1 ANTHERS..... 89

FIGURE 4.4: ANTHER CROSS-SECTIONS (1μM) OF WILD TYPE AND HOMOZYGOUS *ACLL5*-1 MUTANT. DEVELOPMENTAL STAGES ARE ACCORDING TO SANDERS *ET AL.* 1999.....92

FIGURE 4.5: *IN SITU* HYBRIDIZATION ANALYSIS IN DEVELOPING WILD-TYPE FLOWERS...95

FIGURE 4.6: (**A**) SELECTED GENES WITH HIGH CO-EXPRESSION COEFFICIENTS WITH *ACLL5*/AT1G62940 INVOLVED IN LIPID AND PHENYLPROPANIOD-LIKE METABOLISM (**B**) PUTATIVE DUPLICATED PHENYLPROPANOID PATHWAY..... 97

FIGURE 4.7: PHYLOGENETIC ANALYSES OF PROTEIN SEQUENCES OF ARABIDOPSIS, POPLAR, RICE, PHYSCOMITRELLA AND OTHER SPECIES GENES EXPRESSED DURING MICROSPORE DEVELOPMENT. (**A**) CHALCONE SYNTHASE (CHS) AND CHALCONE SYNTHASE-LIKES (CHSL). (**B**) DEHYDROFLAVONOL REDUCTASE (DFR) AND DEHYDROFLAVONOL REDUCTASE LIKE FAMILIES. (**C**) 4-COUMARATE-COA LIGASE (4CL) AND ACYL-COA LIGASE 5 AND HOMOLOGUES (ACLL).....101

LIST OF ABBREVIATIONS

4CL	4-Coumarate:CoA Ligase
AAE	Acyl-Activating Enzyme
ACLL	Acyl-CoA Ligase Like
AOC	Allene Oxide Cyclase
AOS	Allene Oxide Synthase
C3H	Coumaroyl-Shikimate 3'-Hydroxylase
C4H	Cinnamate 4-Hydroxylase
CAD	Cinnamyl Alcohol:NADP+ Dehydrogenase
CCOMT	Caffeoyl-CoA O-Methyltransferase Enzyme
CCR	Cinnamyl-CoA Reductase
CHS	Chalcone Synthase
COMT	Caffeic Acid O-Methyltransferase
CPR	Cytochrome P450 Reductase
CT	Threshold Cycles
dex1	Defective in Exine Formation
DFR	Dihydroflavonol Reductase
dsRNA	Double Stranded RNA
dyt1	Dysfunctional Tapetum 1
F5H	Ferulate-5-Hydroxylase
HAL	Histidine-Ammonia Lyase
HCT	Hydroxycinnamyl CoA Shikimate/Quinate Hydroxycinnamyltransferase
HMGR2	3-Hydroxy-3-Methylglutaryl-CoA Reductase 2
IAA	Indole-3-Acetic Acid
IPP	Isopentenyl Diphosphate
JA	Jasmonic Acid
LOX	Lipoxygenase
LTP	Lipid Transfer Proteins
MeJA	Methyjasmonate
MEP	Methylerythritol Phosphate
ML	Maximum-Likelihood
MPSS	Massively Parallel Signature Sequencing
ms1	Male Sterile 1
ms2	Male Sterile 2
NASC	The European Arabidopsis Stock Centre
OPC8	3-oxo-2(2'[Z]-Pentenyl)-Cyclopentane-1-Octanoic Acid
OPDA	12-oxo-Phytodienoic Acid
OPR3	OPDA Reductase
OS	Overall Sequence
PAL	Phenylalanine-Ammonia Lyase
PCD	Programmed Cell Death
PTS1	Peroxisomal Target Signal 1
PTS2	Peroxisomal Target Signal 2
SBD	Substrate Binding Domains

SEM	Scanning Electron Microscopy
tt4	Transparent Testa 4
WT	Wild-Type

AGNOWLEDGEMENTS

The past 6 years that I've been at UBC were probably the most enriching and challenging years of my life. It's with joy and a little sadness that I will complete my first life-long goal by finishing this document. While I struggle to think about how to start writing this small section after writing an entire thesis, I realize that maybe it's because I just have way too much to say and only one page to say it all...

I will start with my very thanks to Juergen Ehlting. As a freshly titled B.Sc. in Biology, I was avidly looking for an internship abroad to study genetics of plant secondary metabolism. Juergen not only made that possible, but he captivated me with his passion for Science. Thanks Juergen, your dedication really changed my life.

My supervisor Carl Douglas was a constant source of support. Thank you Carl for directing my project while giving me the freedom to learn and develop. I am extremely grateful for your enormous patience and understanding in various moments throughout my degree. I'll carry with me the knowledge and experience that you shared to the next steps of my scientific career.

My deepest thanks to my committee members Joerg Bohlmann, Brian Ellis and Ljerka Kunst for their time and effort in giving valuable feedback on my work and allowing me to use their teams expertise for advice and lab equipment.

My special thanks to past and present Douglas Lab members: Dae Kyun Ro, Lee Johnson, Qing Wang, Barham Soltani, Tamara Allen, Shana Gutman, David Johnston, Eryang Li, June Kim, Tom Liu, SongSoo Kim and specially Bjoern Hamberger. From outside the Douglas Lab, I am thankful for the help of so many talented and friendly people, in particular: Barbara Ehlting, Sandra Goritschnig, Sarah McKim, Jamie Pighin, Hardy Hall, Nathalie Mathews, Britta Hamberger, the EM lab crew and the very efficient Botany office staff.

Of course, I thank my beautiful host country Canada. I could not have picked a better place to live and I know I will miss Vancouver! To all the friends I made in Canada from 2001 to 2007: you were a crucial support network that kept me out of trouble, and helped me a great deal in and outside the lab. I would have not made it if it weren't for you. You know who you are and I thank you all very much. To my dear friends back home and my boyfriend Ryan: thanks for the great times together, the countless emails and phone calls - very necessary after long days (and long nights) in the lab!

Finally, I thank for the enormous encouragement from my family: my brother Olavo and my sister Lulu for always setting the bar high... and my mom and dad, Beatriz and Leopoldo, for always pushing me to be the best I can be, for the constant source of emotional and (very appreciated) financial support. I owe my success to you.

Thanks to all above who allowed me to fill this page with words. I am one lucky person.

CO-AUTHORSHIP STATEMENT

The following people have contributed significantly to experimental data collection presented in this thesis:

- Dr. Brad Barbazuk from Donald Danforth Plant Science Center, St. Louis, Missouri, provided the maize *ACLL* nucleotide sequences. The candidate used this data for the phylogenetic analysis depicted on Figure 3.1.
- Dr. Steven Ralph from UBC Michael Smith Laboratories-Genome BC Treenomix project donated all the cDNA used for all gene expression studies in poplar presented in Chapter 3. The candidate performed all the experiments using this material, interpreted the data and prepared Figures 3.4, 3.5 and 3.7.
- Dr. Keith Turner from BC Institute of Technology generated the transgenic tobacco plants presented in Chapter 3 using transgenic *Agrobacterium* strains provided by the candidate. The candidate also photographed and processed the images presented in Figure 3.8.
- Sarah McKim from UBC Department of Botany did the *in situ* hybridization experiment presented in Chapter 4. The candidate analyzed the results of this experiment and prepared the images shown in Figure 4.5.

CHAPTER 1 - INTRODUCTION

1.1 Natural product diversity and plant genomics

Plants are a highly diverse group of sessile organisms that have evolved incredible ecological adaptations to occupy a wide range of terrestrial and aquatic habitats. With over 250,000 species, the plant kingdom represents a variety of life forms and survival strategies that reflect the evolutionary pressures driving speciation (Raven *et al.*, 1996). Our presence on Earth relies on the very basic resources and processes that plants generate, such as food, oxygen, fuel, lumber and fibers. Plants have the ability to synthesize an enormous variety of structurally diverse natural products that are important for their fitness in the natural habitat, and also important for human use. These specialized organic compounds do not appear to be directly related to growth and development, and for this reason, are traditionally referred to as products of secondary metabolism (Croteau *et al.*, 2000). Major classes of natural products, or secondary metabolites, are terpenoids, alkaloids and phenylpropanoids.

Terpenoids are the most structurally diverse class of natural products and are derived from repetitive fusions of five-carbon isopentane units (also called isoprene units) (Croteau *et al.*, 2000). Usually the production of terpenoids is located within specialized anatomical structures, such as glandular trichomes (Turner *et al.*, 1999) and resin ducts (Martin *et al.*, 2002). Terpenoids are unified based on the common biosynthetic origin from the fundamental precursor isopentenyl diphosphate (IPP), which is synthesized via the acetate/mevalonate pathway in the cytosol and ER and via the methylerythritol

phosphate (MEP) pathway in the plastids. Examples of terpenoids are those found as components of essential oils and vitamin precursors, like β-carotene. Stress-induced terpenoids can play important roles in plant defense (Croteau *et al.*, 2000).

About 20% of flowering plants produce alkaloids, which are low molecular weight, nitrogen containing molecules, usually pharmacologically active (Facchini and St-Pierre, 2005). The human use of plant-derived alkaloids has been documented since our ancient history. For example, in 399 B.C. the Greek philosopher Socrates was executed by consuming an extract of coniine-containing hemlock. The study of alkaloids has been primary driven due to its importance in medicine, such as the potent analgesics codeine and morphine, derived from opium poppy (*Papaver somniferum*). Today, it is known that alkaloids derive in most cases from amino acids, and over 12000 alkaloids have been isolated (Facchini and St-Pierre, 2005). In plants, alkaloids are associated with chemical defense such as feeding deterrence against herbivores, and therefore play important roles in ecological interactions (Steppuhn *et al.*, 2004).

About 40% of the organic carbon in the biosphere is found in the form of plant phenolic compounds, such as lignin (Croteau *et al.*, 2000). Phenylpropanoid metabolism, from which most of the phenolic compounds derive, is discussed in detail in the later sections of this chapter.

Plant hormones, such as auxins, salicylic acid and jasmonates, constitute an array of plant metabolic products that add to the repertoire of plant chemicals. The biosynthesis of some

of these compounds share precursors with the biosynthesis of some secondary metabolites, as it is the case for gibberellic acid, which derives from IPP and its conversion to geranyl diphosphate via the terpenoid pathway (Croteau *et al.*, 2000). This is an example of the fine line dividing primary and secondary metabolism, and an indication of the origin of plants' specialized metabolism deriving from ubiquitous biochemical pathways.

Over the last two decades, there have been efforts using molecular tools for the study of plant natural products. As one example, these approaches were useful in the identification of the set of structural genes and regulatory elements necessary for phenylpropanoid metabolism, which is thought to be a crucial biochemical pathway for the colonization of the land environment (Douglas, 1996). The early studies with parsley cell cultures such as enzyme activity studies with PAL and 4CL (Hahlbrock *et al.*, 1981), and the cloning and expression of genes coding PAL (Hahlbrock and Scheel, 1989), C4H (Koopmann *et al.*, 1999) and 4CL (Douglas *et al.*, 1987) enzymes, provided the initial picture of the molecular toolbox defining the phenylpropanoid pathway. With the completion of genome sequences of plant species of different angiosperm lineages and ecological adaptations, starting with the Arabidopsis genome (The Arabidopsis Genome Initiative, 2000), and the more recent additions of the rice (Yuan *et al.*, 2003) and poplar (Tuskan *et al.*, 2006) genomes, as well as generation of significant genome sequence information from crop plants (Paterson, 2006) and basal land plant lineages like the moss *Physcomitrella patens* (Nishiyama *et al.*, 2003), the era of plant genomics is bringing about new possibilities to study plant metabolism in the context of land plant evolution.

For example, the picture of the diversity of genes encoding phenylpropanoid pathway enzymes has become more clear, and insights can be made correlating metabolic diversity to genome evolution and ecological adaptations (Hamberger *et al.*, submitted; Tsai *et al.*, 2006; Tuskan *et al.*, 2006).

As introduced above, a characteristic of plants is their metabolic diversity, but the origins of this diversity are not well understood. A common mechanism for generating metabolic diversity is the recruitment of enzymes from pre-existing biochemical pathways in a given organism (Austin and Noel, 2003; Ritter and Schulz, 2004). One interesting example that may be key to the evolution of phenylpropanoid metabolism is the evolutionary history of the enzyme phenylalanine-ammonia lyase (PAL). The similarity of the phenylpropanoid enzyme PAL to histidine-ammonia lyase (HAL) in the His degradation pathway has been noted (Ritter and Schulz, 2004). This study demonstrated, via sequence and structural similarities between PAL and HAL, that PAL, involved in specialized plant metabolism, was likely recruited from the central metabolic pathway of amino acid degradation. Chalcone synthase (CHS), a key enzyme of the flavonoid branch pathway, and related enzymes were also likely recruited from enzymes functioning in primary metabolism. Although their origin is not fully elucidated, evidence suggests that chalcone synthases evolved from ketoacyl synthase III enzymes, involved in fatty acid biosynthesis (Austin and Noel, 2003). Extensive gene duplication and subsequent genetic variation gave rise to most or all of the diversity in the CHS family seen today (Austin and Noel, 2003). As more genomes are sequenced, comparative genomics approaches will make it increasingly possible to follow the evolutionary history of gene families.

1.1.1 Fully sequenced plant genomes and lessons learned from them

Arabidopsis thaliana was the first plant with a fully sequenced genome, providing a foundation for functional characterization of plant genes as well as a platform for development of tools relevant to evolutionary biology, agriculture, bioinformatics, and comparative genomics (The Arabidopsis Genome Initiative, 2000). Much was learned with the completion of the Arabidopsis genome in terms of genome structure and evolution, when compared to other available genomes. At the time the Arabidopsis genome was completed, whole genome comparisons to other eukaryotes were only possible between Arabidopsis and the yeast *Saccharomyces cerevisiae*, the fruit fly *Drosophila melanogaster* and the worm *Caenorhabditis elegans*. Results of those analyses demonstrated conservation of protein families among all eukaryotes, and another ~150 protein families unique to plants, most of unknown function (The Arabidopsis Genome Initiative, 2000), highlighting our lack of knowledge about unique aspects of plant development and metabolism.

Large-scale analysis of the *Arabidopsis* genome revealed that it has undergone at least two whole genome duplications in its evolutionary history, in addition to numerous tandem duplications and further reshuffling of chromosome segments. In fact, it has been estimated that about 90% of loci in Arabidopsis are duplicated, 17% of which are arranged in tandem arrays (Moore and Purugganan, 2005; The Arabidopsis Genome Initiative, 2000). One study suggests that the most recent whole genome duplication occurred 24 to 40 Mya, during the early emergence of the crucifer (mustard) family (Blanc *et al.*, 2003). The evolutionary fate of duplicated genes has been long debated and

it is generally accepted that duplicated genes are a source of raw material for evolutionary novelty to develop. A study of regulatory genes in Arabidopsis showed evidence that many gene families have expanded and diversified over the course of evolution, as a result of gene duplication and divergence (Duarte *et al.*, 2006). In many cases, retention of duplicated genes is accompanied by either changes in gene expression patterns (subfunctionalization) or by changes in protein function (neofunctionalization) suggesting that changes in gene number, sequence, and expression can give rise to phenotypic variation (Moore and Purugganan, 2005).

A first comparative genomics approach between plants became possible with the completion of the rice genome (International Rice Genome Sequencing Project, 2005; Itoh *et al.*, 2007; Yuan *et al.*, 2003). The rice genome contains about the same number of genes as Arabidopsis, but only about one third of the protein-coding sequences in rice have putative orthologues in Arabidopsis. There were many species-specific gene families discovered that could account for the phenotypic differences evident between these species (Itoh *et al.*, 2007).

The completion of the first tree genome sequence (*Populus trichocarpa* or poplar; (Tuskan *et al.*, 2006) allows for interesting comparisons between the complete gene sets of two plant species with very different life histories: the herbaceous annual Arabidopsis and the tree *Populus*. The poplar genome sequencing and genome annotation effort identified more than 45,000 putative protein-coding genes, with an average of 1.5 putative poplar homologues for every Arabidopsis gene. The poplar genome sequence

revealed that there has been a recent whole genome duplication event, referred to as the salicoid duplication event, evidenced by the identification of blocks of genes with conserved synteny located on different chromosomes. In addition, comparative genomics approaches showed that both *Arabidopsis* and *Populus* lineages share an ancient genome duplication, the eurosid duplication event, and that tandem duplications appear to be relatively more common in Arabidopsis than in poplar (Tuskan *et al.*, 2006).

Genes overrepresented in one or the other organism might be correlated with adaptations that could lead to speciation events (Itoh *et al.*, 2007). The whole genome duplication in the salicoid lineage, including poplar, would have provided a wealth of new genes, and it is interesting that the emergence of the *Populus* genus in the fossil record coincides with the salicoid whole genome duplication (Tuskan *et al.*, 2006). On the other hand, genes conserved between lineages may encode proteins with common, conserved functions. It has been suggested that gene expression differences account for a large share of the phenotypic variability seen between species (Nielsen, 2006). With regard to evolution of metabolic pathways, even small changes in expression of enzyme encoding genes could lead to changes in developmental and environmentally specified enzyme levels and such changes could cause dramatic changes in flux through metabolic pathways (Nielsen, 2006). Thus, since enzymes of conserved sequence may share the same biochemical function, but not necessarily the same biological functions due to expression differences, information about gene expression is important when assessing homologous pairs of genes for conserved function. However, with the limited gene expression information

available today for plant species other that Arabidopsis, gene sequences alone still provide a good starting point for creating hypotheses about gene origin and function.

1.2 Phenylpropanoid metabolism and adenylate-forming enzymes

1.2.1 Phenylpropanoid metabolism

Plants can efficiently channel carbon from primary metabolism to the phenylpropanoid metabolism mostly via the amino acid phenylalanine (Figure1.1) (Douglas, 1996). Phenylalanine-derived natural products are crucial compounds in plants, with a variety of functions ranging from UV-protection, inter-species signaling and antimicrobial activity, to structural composition of the cell wall (Dixon and Paiva, 1995; Hahlbrock and Scheel, 1989). Protection against UV-irradiation conferred by flavonoids as well as mechanical support and water impermeability for water transport provided by lignin suggest that the evolution of phenylpropanoid metabolism was likely of fundamental importance in the ability of plants to colonize land (Douglas, 1996).

General phenylpropanoid metabolism in plants is composed of three main enzymatic steps (Figure 1.2). Phenylalanine ammonia lyase (PAL) is the first enzyme of the pathway and catalyzes the conversion of the amino acid phenylalanine into *trans*-cinnamic acid. Subsequently, the cytochrome P450-dependant enzyme cinnamate 4-hydroxylase (C4H), in conjunction with cytochrome P450 reductase (CPR), hydroxylates *trans*-cinnamic acid yielding *p*-coumaric acid. Finally, 4-coumarate:CoA ligase (4CL) generates CoA esters of *p*-coumarate and its derivatives in a two step reaction (Figure

1.3) (Hahlbrock and Scheel, 1989). This two-step reaction, in which an adenylate-substrate intermediate is formed in the presence of ATP and Mg^{2+}, is a common mechanism shared with other adenylate-forming enzymes, such as firefly luciferases and non-ribosomal peptide synthases (Becker-Andre *et al.*, 1991), as discussed in detail below. Thioesters activated by 4CL are then used as precursors for the downstream branches of the phenylpropanoid pathway, for the production of a variety of plant natural products, including lignin and flavonoids (Figure 1.2) (Hahlbrock and Scheel, 1989).

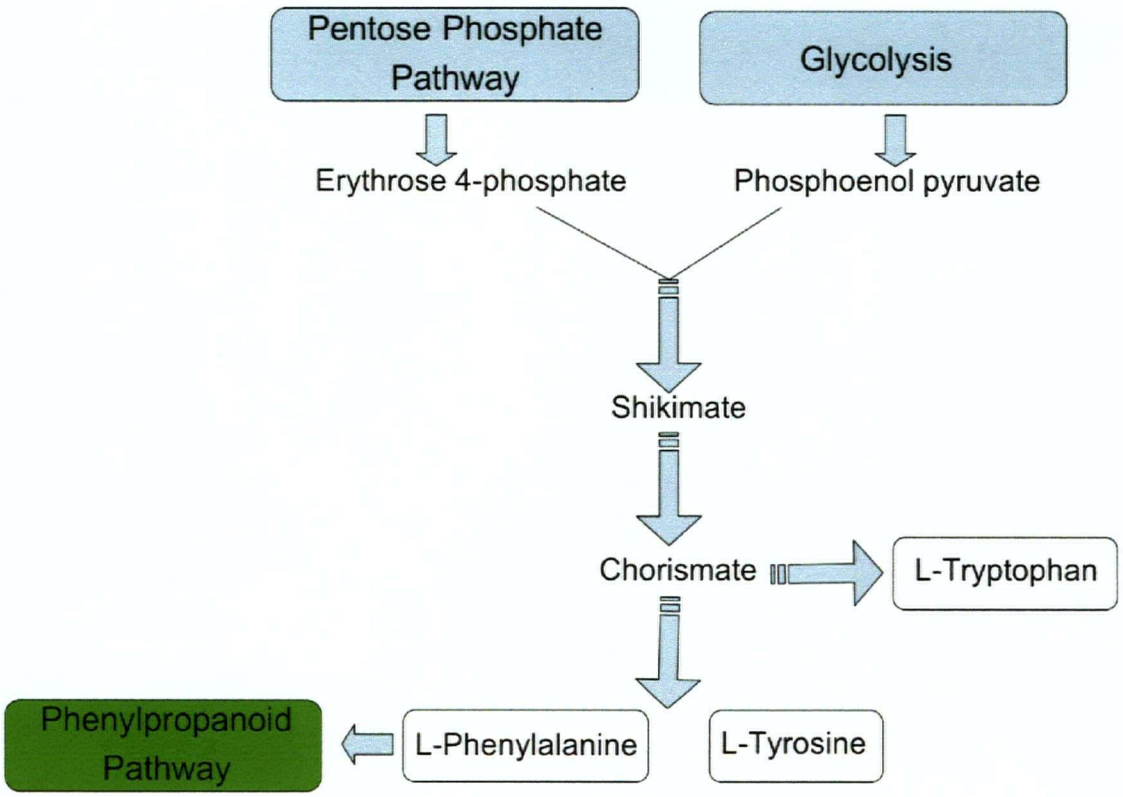


Figure 1.1: Schematic view of the relationship of phenylpropanoid metabolism to primary metabolism. The shikimate pathway leads to biosynthesis of aromatic amino acids including phenylalanine, starting with metabolic precursors. Phenylpropanoid metabolism, branches from the shikimate pathway mostly via phenylalanine.

The downstream reactions following the general phenylpropanoid pathway are quite diverse and much is still to be learned about them. The known general steps for the monolignol biosynthesis branch pathway include the donation of the acyl group of the CoA ester formed by the 4CL reaction, by the enyzyme Hydroxycinnamyl CoA shikimate/quinate hydroxycinnamyltransferase (HCT), to a shikimic acid or quinic acid acceptor, yielding the corresponding shikimate or quinate esters. The phenolic rings of these esters are then adorned at position 3 with a hydroxyl group by action of the P450 enzyme coumaroyl-shikimate 3'-hydroxylase (C3H). The 3' hydroxyl group is then substituted by a methyl group by caffeoyl-CoA *O*-methyltransferase enzyme (CCOMT), and the 5' hydroxyl group may be further introduced by the action of enzymes originally characterized as ferulate-5-hydroxylase (F5H) and methoxylated by caffeic acid *O*-methyltransferase (COMT), although these reactions likely occur *in vivo* at the level of aldehydes (Humphreys and Chapple, 2002). The resulting methylated CoA esters are then reduced to the respective aldehydes by action of cinnamyl-CoA reductase (CCR), and are finally further reduced to the corresponding monolignol alcohols by cinnamyl alcohol:NADP+ dehydrogenase (CAD) (Costa *et al.*, 2003; Hamberger *et al.*, submitted; Humphreys and Chapple, 2002). The first enzyme on the flavonoid branch pathway is chalcone synthase (CHS), which catalyzes the condensation of the product of the 4CL reaction, *p*-coumaroyl-CoA, with three molecules of malonyl-CoA to generate a C15 skeleton, which is the backbone for a variety of elaborations by downstream enzymes that generate flavonoid diversity (Noel *et al.*, 2005).

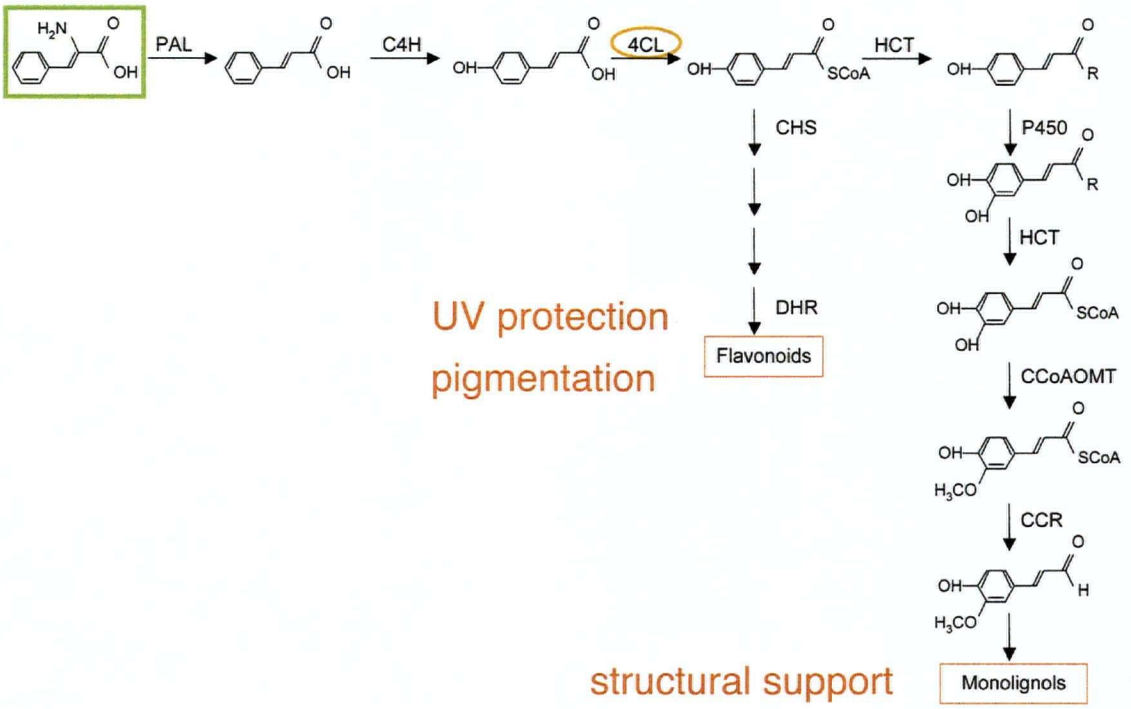


Figure 1.2: A simplified scheme of the phenylpropanoid pathway, showing the two main branch pathways deriving from the product of the 4CL enzyme and the resulting final products.

1.2.2 Adenylate-forming enzymes

Adenylate-forming enzymes constitute a large class of enzymes that catalyze diverse reactions, all characterized by two-step reaction mechanism requiring Mg^{++} and involving pyrophosphorylysis of ATP and formation of an enzyme-bound AMP-substrate intermediate (adenylate). This reaction mechanism is used, for example, in the formation of non-ribosomal bacterial peptide synthases (Conti *et al.*, 1997) and by the firefly luciferase enzyme (Deluca, 1976). The final acyl acceptor varies considerably depending on the type of enzyme (Shockey *et al.*, 2003). In the CoA ligase reaction, the AMP is released after nucleophilic attack of the carbonyl carbon of the adenylate by the free

electrons of the thiol group of the CoA acyl acceptor, forming the final CoA thioester (Figure 1.3) (Shockey *et al.*, 2003). Adenylate-forming enzymes are involved in carboxylic acid activation by formation of CoA esters, reactions that play vital roles in all living organisms, providing precursors for biosynthesis or breakdown pathways of many important metabolites (Shockey *et al.*, 2003). Known functions of adenylate-forming enzymes in plants include long chain fatty-acid activation (Shockey *et al.*, 2002), synthesis of acetyl-CoA important for lipid accumulation in developing seeds (Ke *et al.*, 2000), biosynthesis of molecules such as jasmonic acid, indole-3-acetic acid (IAA) and salicylic acid (Staswick *et al.*, 2002), and phenolic acid activation in phenylpropanoid metabolism.

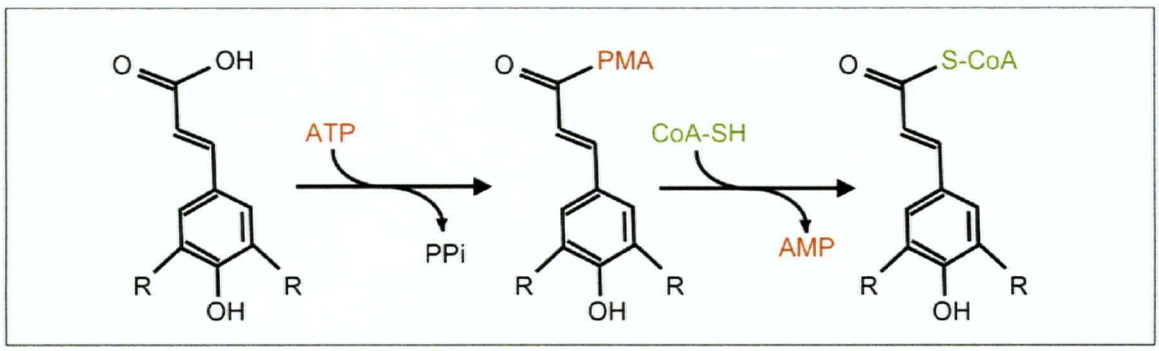


Figure 1.3: The two-step mechanism of the CoA-ligase reaction. The first step uses ATP to generate an adenylate intermediate. The second step is characterized by the transfer of the acyl molecule to the CoA acceptor forming a thioester bond, and release of AMP.

1.2.3 *4CL* gene families

The first *4CL* gene cloned was derived from parsley (Douglas *et al.*, 1987) and activities of recombinant parsley 4CL enzymes were demonstrated (Lozoya *et al.*, 1988). Since then, *4CL* genes have been isolated from several plants, such as tobacco (Lee and Douglas, 1996), loblolly pine (Zhang and Chiang, 1997), poplar (Allina *et al.*, 1998), aspen (Hu *et al.*, 1998), *Arabidopsis* (Ehlting *et al.*, 1999) and soybean (Lindermayr *et al.*, 2002), and in these cases these genes were also shown to encode *bona fide* 4CL enzymes by expression of recombinant proteins. In all angiosperms examined, 4CL is encoded by multi-gene families. The *4CL* gene family in *Arabidopsis thaliana* is comprised of four genes, *4CL1*, *4CL2*, *4CL3* (Ehlting *et al.*, 1999) and *4CL4* (Hamberger and Hahlbrock, 2004). Phylogenetic analysis of all known plant *4CL* genes, including Arabidopsis *4CLs*, showed that they fall into two classes, which likely arose early in angiosperm evolution (Cukovic *et al.*, 2001; Ehlting *et al.*, 1999). Class I *4CLs* appear to be associated with the biosynthesis of lignin and other phenylpropanoids, while class II *4CLs* are associated with flavonoid biosynthesis (Ehlting *et al.*, 1999).

1.2.4 *4CL* subfunctionalization

It has been demonstrated that 4CL enzyme isoforms encoded by different gene family members have the capacity to convert different substrates, thus directing the flux of the general phenylpropanoid metabolism into the major branch pathways of flavonoid or monolignol biosynthesis. In aspen (*Populus tremuloides*), *4CL1* and *4CL2* are differentially expressed and exhibit highly divergent substrate preference associated with their respective functions of lignin biosynthesis in developing xylem tissues, and

biosynthesis of other phenolics, such as flavonoids, in epidermal cells (Hu *et al.*, 1998). In soybean (*Glycine max*) the three structurally and functionally distinct cDNAs encoding 4CL enzymes were also shown to be divergent at the levels of catalytic specificity and expression (Lindermayr *et al.*, 2002). Interesting data came from the study of loblolly pine (*Pinus taeda*), in which a single 4CL protein that exhibits broad substrate specificity has been described (Harding *et al.*, 2002; Zhang and Chiang, 1997). Those data, together with phylogenetic data, suggest that subfunctionalization in terms of substrate preference and gene expression, apparent in angiosperm lineages, may have originated after the divergence of angiosperms from gymnosperms (Hamberger *et al.*, submitted).

The *4CL* gene family has been particularly well studied in Arabidopsis, where not only gene expression and substrate specificity have been characterized (Ehlting *et al.*, 1999; Hamberger and Hahlbrock, 2004; Soltani *et al.*, 2006) but phenotypic effects of *4CL* knock-out mutations have been observed, as discussed below (Hamberger B. and Douglas C., unpublished). All Arabidopsis 4CL proteins expressed heterologously in *E. coli* have activity towards 4-coumarate and therefore are *bona fide* 4CLs, although substrate preference is largely complementary among the four 4CL enzymes (Table 1.1). 4CL1 has highest activity with *p*-coumarate and caffeate, overlapping with the highest activities of 4CL2 (caffeate) and 4CL3 (4-coumarate). 4CL4 is unique in the ability to convert very efficiently ferulate and sinnapate, and its been suggested that this enzyme could play a role in the biosynthesis of soluble sinapate-containing phenolics, alternative to or in addition to a role in lignin biosynthesis (Hamberger and Hahlbrock, 2004). Interestingly, a sinapate-utilizing 4CL isoform has also been identified in soybean

(Lindermayr *et al.*, 2002), but otherwise this activity does not appear to be widespread among 4CL enzymes (Allina *et al.*, 1998; Hu *et al.*, 1998).

Table 1.1: Kinetic properties of recombinant Arabidopsis 4CLs (Data from: Ehlting *et al.*, 1999; Hamberger and Hallbrock 2004)

Enzyme	Substrate	Km (µM)	Vmax (% coumarate)	Vmax/Km
4CL1	Cinnamate	6320	103	0.02
	4-Coumarate	38	100	2.6
	Caffeate	11	27	2.5
	Ferulate	199	53	0.26
	Sinapate	n.c.	-	-
4CL2	Cinnamate	6630	21	0.003
	4-Coumarate	252	100	0.39
	Caffeate	20	74	3.7
	Ferulate	n.c.	-	-
	Sinapate	n.c.	-	-
4CL3	Cinnamate	2070	164	0.08
	4-Coumarate	23	100	4.4
	Caffeate	374	129	0.35
	Ferulate	166	86	0.52
	Sinapate	n.c.	-	-
4CL4	Cinnamate			
	4-Coumarate	432	100	0.3
	Caffeate	186	187	1.1
	Ferulate	26	153	6.6
	Sinapate	20	105	6.7

Biochemical data demonstrating substrate preference suggests that 4CL1 could potentially complement the activity of 4CL2 and 3 *in vitro* (Table1.1). In addition, the gene expression patterns of Arabidopsis *4CL* genes show differential regulation and expression subfunctionalization of family members. Expression of *4CL1* is higher in seedling roots and in bolting stems of mature plants, *4CL2* is most highly expressed in roots and *4CL3* is most highly expressed in flowers (Ehlting *et al.*, 1999). Examination of *4CL* promoter-GUS fusion expression suggests that *4CL1* and *4CL2* promoter activity is confined primarily to the vasculature of aerial organs, with some broader expression in roots (Soltani *et al.*, 2006). In contrast, the *4CL3* promoter drives *GUS* expression in epidermal cells with no preferential expression in vascular tissues, and the *4CL4*

promoter exhibits low overall activity but is wound-induced (Soltani *et al.*, 2006). These differential expression patterns combined with the biochemical data suggest that subfunctionalization at the levels of enzymatic properties and gene expression has occurred in the evolution of the *4CL* gene family in Arabidopsis, such that specific genes and enzymes are specialized for the production of phenylpropanoids required for specialized organs and tissues.

It is known that phenylpropanoids participate in defense mechanisms, and evidence for that has been gathered at the chemical, biochemical and genetic levels (Dixon and Paiva, 1995). For example, flavonoids were shown to accumulate in bean leaves upon UV radiation treatment (Beggs *et al.*, 1985), studies on parsley suspension culture cells demonstrated increases in PAL and 4CL activity after elicitor treatment (Hahlbrock *et al.*, 1981), and finally studies on parsley *4CL* genes demonstrated gene up-regulation upon elicitor and UV treatment (Douglas *et al.*, 1987). In Arabidopsis, *4CL1*, *4CL2* and *4CL4* expression is wound inducible (Ehlting *et al.*, 1999; Soltani *et al.*, 2006), consistent with the role of these genes in phenylpropanoid biosynthesis for defense purposes, while *4CL3* shows no wound inducibility but is up-regulated after UV stress, consistent with the role of this gene in flavonoid biosynthesis (Ehlting *et al.*, 1999). These data lend further support to the subfunctionalization of *4CL* gene expression in Arabidopsis.

Reverse genetic approaches in Arabidopsis provides further evidence for specialized functions of *4CL* genes in the production of lignin and flavonoids. Although single *4CL1* or *4CL2* insertion mutants have no visible phenotypes, and a single *4CL3* mutant has a

subtle UV-induced flavonoid-deficiency, double mutants of *4CL1/2* and *4CL1/3* have strong phenotypes: the *4cl1/2* double mutant is deficient in lignin biosynthesis and is a dwarf at maturity, while the *4cl2/3* double mutant is severely deficient in developmental soluble flavonoid and anthocyanin production (Hamberger B. and Douglas C., unpublished). These data support subfunctionalization of 4CL enzyme activity in Arabidopsis under normal laboratory conditions. However, there is clearly a functional overlap, since loss of function of a given isoform is largely silent, presumably due to partially redundant function provided by a second isoform that is sufficient to fulfill the biochemical needs of the plant in growth chamber conditions.

1.2.5 Identification of *4CL-like* genes (*4CLLs*) and phylogenetic relationships to other adenylate-forming enzymes

With the sequencing of the Arabidopsis genome and with sequence data from other plants, it became apparent that genes encoding enzymes related to, but phylogenetically distinct from true *4CL* genes exist (Cukovic *et al.*, 2001). With the completion of the Arabidipsis genome, a complete set of "*4CL-like* genes" was identified based on sequence similarity and phylogenetic relationships to known *4CL* genes (Costa *et al.*, 2003; Shockey *et al.*, 2003; this study). In the annotation of *4CL-like* genes, reported by Ehlting *et al.* (2005), FASTA searches were carried out using 4CL1, 4CL2, 4CL3 and 4CL4 amino acid sequences against all annotated Arabidopsis genes (provided by MATDB, http://mips.gsf.de/proj/thal/db/index.html), and putative genes that displayed more than 30% identity on the amino acid level to at least one of the 4CL proteins over a stretch of more than 300 amino acids were selected. Initial phylogenetic analysis revealed those most closely related to true 4CLs and became the focus of further study. While

many genes of this class have been termed "4CL" (Costa *et al.*, 2003), "4CL-like" (Ehlting *et al.*, 2005), and "AAE" for Acyl-Activating enzyme (Shockey *et al.*, 2003), I have subsequently used the term ACLL for Acyl-CoA Ligase Like. The annotation of these genes as ACLLs is based on sequence similarity only, and the biochemical and biological functions of the ACLL proteins were still unresolved at the onset of this thesis.

1.2.6 4CL protein structure

Since the substrate utilization profiles of recombinant Arabidopsis 4CL1 and 4CL2 proteins differ markedly (Table 1.1; Ehlting *et al.*, 1999, Hamberger and Hahlbrock, 2004), it was possible to define 4CL substrate recognition domains based on the activity of chimeric proteins. These were localized between two highly conserved regions, LPFSSGTTGLPKG (box I) and GEICIRG (box II), and are approximately 100 amino acids long (Ehlting *et al.*, 2001). Point mutations in the putative substrate binding domains (SBD) of different 4CLs can cause changes in substrate recognition, giving rise to the concept of a substrate binding pocket with limited number of contact residues involved in substrate recognition (Stuible and Kombrink, 2001). It is known that adenylate-forming enzymes utilize very diverse substrates, so amino acid sequence information alone is not sufficient to provide definitive information on substrate usage by 4CL and related enzymes. A crystal structure for 4CL would allow more definitive identification of key functional amino acids in the 4CL SBD, and could help with predictions of the substrates of ACLL enzymes.

A model for identifying the important amino acid residues responsible for the substrate recognition in 4CL2 has been proposed. This "specificity code" is composed of 12 amino acid residues (Schneider *et al.*, 2003), based on homology modeling of 4CL2 to the known structure of the bacterial phenylalanine-activating domain of gramicidin S syntethase (PheA). Together with previous mutation analysis of 4CL enzymes, this study introduced the concept that the specificity of 4CL isoforms towards their hydroxycinnamic acid substrates is due to size exclusion controlled by four amino acids in the putative substrate-binding pocket. In addition, increasing hydrophobicity of specific residues this region resulted in variants of 4CL2 with enhanced conversion of cinnamic acid (Stuible and Kombrink, 2001). In the ACLLs, four of the 12 amino acids corresponding to those responsible for the "specificity code" in 4CL2 are conserved, whereas the other eight, including the ones causing steric hindrance of the substrate are not conserved. Knowledge of the types of substrates ACLLs convert may allow future use of this type of modeling approach to identify amino acids within the SBD of ACLLs that are responsible for substrate recognition, and might be used to predict potential substrates of other ACLLs.

1.3 The ACLLs and thesis objectives

4CL enzymes have been extensively studied for three decades due to their central role in the general phenylpropanoid pathway. Therefore, upon completion of the Arabidopsis genome and subsequent annotation of "*4CL-like*" *ACLL* genes, ACLLs were an obvious target for identifying new enzymatic functions and novel pathways that could be related

to the general phenylpropanoid pathway or other metabolic pathways, therefore these became targets for functional analysis by several groups.

The objectives of my thesis were to:

1. Provide a general overview of Arabidopsis *ACLL* gene repertoire and gene family structure using the known Arabidopsis *4CL* genes as a guiding tool.

2. Use expression analysis, reverse genetics, and bioinformatic analyses to characterize their evolution, biological and biochemical functions.

3. Provide a comparison of the *ACLL* gene family in the three sequenced plant genomes (Arabidopsis, rice, and poplar) and obtain poplar ACLL expression data to complement data obtained in Arabidopsis.

Several approaches were used to collect such information. Using bioinformatics tools, sequence data from various organisms was assembled to further reconstruct the evolution of the *ACLL* gene family, and *ACLL* homologues were discovered in other plant species, such as *Physcomitrella patens*, rice and poplar. I assessed expression patterns to verify subfunctionalization or evidence of overlapping functions of *ACLL* genes in Arabidopsis and poplar. Reverse genetics approaches such as analysis of DNA insertion mutants were used to find clues for *ACLL* biological function. Finally, I identified co-expressed genes using public microarray data and identified common regulatory elements in the promoter regions of co-expressed genes that have allowed me to pinpoint known and putative novel biochemical pathways in which ACLL enzymes may be participating. This work has allowed the biological and biochemical functions of some *ACLL* genes to be deduced,

and has generated hypotheses regarding the functions of others, which will require further experimentation.

Identification of additional characteristics of the ACLLs would pave the way for biological and biochemical characterizations, possibly including identification of substrates for ACLL proteins, which could lead to the identification of new biochemical pathways that require adenylate-forming enzymes.

CHAPTER 2 - MATERIAL AND METHODS

2.1 Nucleic acid methods

2.1.1 Genomic DNA and total RNA isolation

Total Arabidopsis genomic DNA extraction was carried out using young leaf tissue (Columbia ecotype), ground in a bead beater at 4°C, with the use of Nucleon PhytoPure kit (Amersham-Pharmacia) according to manufacturer's instructions. Arabidopsis RNA was isolated from the specified tissues frozen in liquid nitrogen, and ground to a fine powder using Trizol reagent (Gibco-BRL), following manufacturer's instructions.

2.1.2 DNA extraction from agarose gels

DNA bands were cut from 1% agarose gels and the DNA extracted using the QiaQuick Gel Extraction Kit (Qiagen) according to manufacturer's instructions.

2.1.3 Plasmid DNA preparation and sequencing

Plasmid DNA was prepared with the use of Qiagen spin Miniprep and Midiprep kits, following the manufacturer's instructions (Qiagen). DNA sequencing was performed by the University of British Columbia Nucleic Acid and Protein Service unit, using BigDye 3.0 (Applied Biosystems) and a Prism Sequencer (Applied Biosystems).

2.1.4 Sequence alignment and editing

Multiple sequence alignment was done using the Genomatix DiAlign software (http://www.genomatix.de/cgi-bin/dialign/dialign.pl). Sequence editing and restriction mapping was done using the SeqPup software (http://iubio.bio.indiana.edu/soft/molbio/seqpup/java/seqpup-doc.html).

2.1.5 Reverse transcription for cDNA synthesis

RNA samples were isolated and quality assessed by visual inspection of rRNA on a 1% agarose gel. RNA samples were then quantified spectrophotometrically and 2μg RNA/ 20μl reaction was used to generate first strand cDNA using Superscript II Reverse Transcriptase (Invitrogen) following the manufacturer's protocol.

2.1.6 General recombinant DNA methods

DNA PCR amplified with a proof-reading enzyme (indicated below) was digested with the appropriate restriction enzymes (Invitrogen, Roche and New England BioLabs). Restriction enzyme digests were performed at 40μl volume, following the manufacturers' protocol for each enzyme. Digested DNA was purified using the method described on section 2.1.2. Ligation reactions were performed using T4 ligase (Roche) at 10μl volume, following the manufactures' protocol. All clones used for *Agrobacterium*-mediated transformation were sequenced before further use.

2.1.5.1 Promoter::GUS constructs

Arabidopsis *ACLL* promoter regions were amplified from genomic DNA using pwo enzyme (Roche) and cloned into the pCambia 1305.1 vector (Figure 2.1; http://www.bioforge.net/forge/entry.jspa?externalID=41&categoryID=3) containing the *GUS* (beta-glucoronidase) reporter gene (Jefferson *et al.*, 1987). The multiple cloning site of the vector (5'end) and *Nco*1 sites at *ACLL* start codons (3'end) were used to generate in-frame insertions of the PCR fragment to the *GUS* gene. Primers contained a 3' "tail" to introduce a compatible restriction site sequence and are given in Table 2.1.

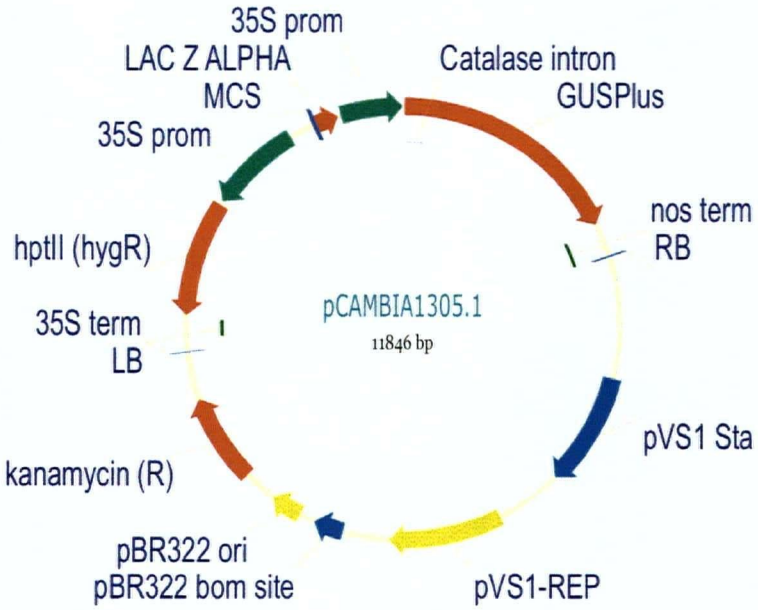


Figure 2.1: pCambia vector 1305.1 used for promoter::*GUS* cloning.

Table 2.1: Primers used for *ACLL* promoter amplification and cloning. Restriction sites are underlined and mismatches are in grey.

Gene	Primer	Sequence (5' - 3')	RE	size (bp)
ACLL1	new3F	GCTTTACTCGAGGGGAACCACAGGG	Xho I	2000
	3R	CGCCATGGTAAAGGACTTTGGTTGTATC	Nco I	
ACLL2	12F prom 1500	AAATGTCGACGTTGTTGGATCGATTGAAAGAAC	Sal I	1500
	12R prom 1500	CTCAGGATACGCCATGGTTCTAATATG	Nco I	
ACLL3	new7F	GTTTCTCGAGGTTATTTCAGCAATGAGGAAGC	Xho I	1900
	7R	CTCAGGATACGCCATGGTTTCTCTATACG	Nco I	
ACLL4	new6F	GAAACTCGAGCTGTCCAGAAGCAAGAGAGTCTC	Xho I	1700
	6R	GAAGCCATGGATTTTGTTTCTATGTAACTTGAC	Nco I	
ACLL5	new4F	GTTTCTCGAGGATTTATCACCTGAATAGTATTCTTCCAGATTGG	Xho I	1950
	4R	CTTTTGACTCTCCATGGTTAAACGAATTGAATTTGATTTATG	Nco I	
ACLL6	new1F	GTTTCTCGAGATGGAATGAAAACACCCGGTCCGGTTC	Xho I	1900
	new1R	GGATTTCTCCATGGTTTCCGATCTCG	Nco I	
ACLL7	new2F	GTTTCTCGAGCATTTGGCCGGCGATAACATCAGAG	Xho I	2000
	2R	GCCGCCATGGGAGAGAAGCAGAGTTTAAG	Nco I	
ACLL8	8F prom 1500	AAATCCTCGACCAAGAGACGGTCGAATGGC	Xho I	1500
	8R	GAATTCGCCATGGTTCTTTGGTTGGATTAG	Nco I	
ACLL9	5F	GTTTCCGCGGCCCAATGGTGAAGGATACAAGCC	Sac II	2100
	new5R	GTCACCATGGTGGAAGACGATTAGAGATAC	Nco I	
pCambia	pCambiaF	GCGGATAACAATTTCACACAGGAAAC		
	pCambiaR	GGGTCCTAACCAAGAAAATGAAGG		

2.1.5.2 ArathACLL4 and PoprtACLL5 GFP fusions

The coding sequences of both *ArathACLL4* and *PoptrACLL5* were amplified from cDNA made from RNA extracted from organs where these genes were most highly expressed, according to the results depicted on Figure 3.4 and 3.7 (flowers and MeJA treated leaves, respectively). The PCR amplification reactions (primers given in Table 2.2) were performed with Phusion high fidelity enzyme (Finnzymes) according to the manufacture's protocol. The PCR products were cloned into a Gateway (Invitrogen) compatible entry vector using TOPO-TA cloning kit (Invitrogen) and subsequently recombined into the destination vector using LR Clonase II enzyme mix according to the manufacturer's instructions (Invitrogen). PCR products were cloned in-frame N-terminal to the *GFP* gene driven by the CMV 35S promoter.

Table 2.2: Primers used for generating GFP fusions of ArathACLL4 and PoptrACLL5

Gene	Primer	sequence (5'-3')
ArathACLL4	CFP-ArathACLL4-F	ATGGCTTCAGTGAATTCTCGA
	CFP-ArathACLL4-R	TCAAAGCTTGGAGTTGGAAGT
PoptrACLL5	CFP-PoptrACLL5-F	ATGGCAGACAACAACAACCTCACA
	CFP-PoptrACLL5-R	TCAGAGCTTGGAGGTTGCGAG

2.2 Plant growth and maintenance

2.2.1 Seed harvesting and sowing

Mature *Arabidopsis thaliana* (Arabidopsis) plants were let to dry at room temperature. Fully dried siliques were harvested and seeds were separated from plant debris. Seeds were sterilized in 70% EtOH for 2min followed by 100% EtOH for 2 mins. Seeds were dried and sown on Petri plates containing ½ MS (Murashige and Skoog) salts (Sigma Aldrich), supplemented with 1% sucrose and 0.6% agar medium. Plates were placed at 4°C for 2 to 3 days for seed stratification and then transferred to a growth chamber at 20°C under continuous light until first cotyledons were developed.

2.2.2 Plant growth conditions

Arabidopsis seedlings were transferred from plates to pots containing moist soil (Sunshine mix 5, Sungrow Horticulture, Saba Beach, Alberta) and pots were covered with Saran wrap to prevent dehydration of the soil during establishment of seedlings. After 3 days the plastic wrap was cut with razor blade and finally removed after another 3 days. Plants were watered as needed and kept in the growth chamber at 20°C under long day conditions (18h light) until maturity. Material from poplar plants was obtained from S. Ralph (Genome BC). Poplar growth conditions are described in Ralph *et al.* (2006).

2.2.3 Mechanical wound and herbivory

Arabidopsis plants, grown as described above, were wounded with pliers on the full surface of leaf blade and harvested after 1h, 4h, and 24h. Wounded and unwounded control plants were harvested at the same time and placed immediately in liquid nitrogen. Poplar stress experiments were done on leaves of *Populus trichocarpa* X *P. deltoides* clone H11, during time course of 2h, 6h, and 24h. Mechanical wounding, herbivory, regurgitant, and methyl jasmonate treatments are described in Ralph *et al.* (2006) and in Hamberger *et al.* (2007).

2.3 General transformation procedures

2.3.1 Bacteria transformation

Competent *E. coli* (DH5α) or *Agrobacterium* (GV3303) cells were kept at -80°C until ready to use. Cells were thawed on ice and 10μl ligation reaction mixture or 100ng plasmid DNA was added to the bacteria. The bacteria was transformed by heat shock (42°C for 45 secs for *E. coli* and 37°C for 2mins for *Agrobacterium*) and placed on agar LB media Petri plates containing the appropriate selection antibiotic. Single colonies were picked and cultures grown on liquid LB media. Plasmids were isolated from cultures (section 2.1.3) and the presence of transgenes was tested by test digestion of the plasmid.

2.3.2 Arabidopsis transformation

Reproductive Arabidopsis plants containing many unopened flowers were used for

transformation by the floral dip protocol (Clough and Bent, 1998). After dipping, pots were kept overnight in dim light covered by a plastic bag to maintain high humidity. Two days later the pots were placed back into the growth chamber. Mature plants were harvested for seeds, and seeds were sown in appropriate selection media. Transformant seedlings (T1) were tested for the presence of the transgene by PCR using at least one plasmid specific primer. The next generation derived from self-crosses (T2) was grown and seeds (T3) were collected from individual plants. A subset of T3 seeds from each T2 plant was placed on selective media for germination to identify homozygous lines. Another subset of seeds from the same individual line was grown in parallel on non-selective media and the lines that were homozygous (based on antibiotic selection) were used for experimentation.

2.3.3 Tobacco transformation (done by K. Turner, BC Institute of Technology)

Leaves disks from sterile tobacco plants grown *in vitro* (1 cm^2 pieces) were placed in a 1/10 dilution of transformed *Agrobacterium* cultures. Leaves disks were blot dried on sterile paper towel and cultured on MS for 48 hours and then transferred to the regeneration medium. The regeneration medium was MS with 1% sucrose, 1.0 mg/l BAP and 0.1 mg/l NAA supplemented with the appropriate antibiotic. Leaf disks were grown at room temp with 16 hours days. Shoots emerging from the leaf disks were transferred into sterile universal jars containing MS medium without hormones to induce root formation. All shoots generated by tissue culture derived from an independent transformation event.

2.4 General bioinformatics procedures

2.4.1 Sequence selection and phylogenetic tree construction

The set of Arabidopsis genes characterized as encoding 4CL enzymes (Ehlting *et al.*, 1999) was used in homology searches to identify potential *4CL-like /ACLL* genes the Arabidopsis genome, using the database maintained at the Arabidopsis Information Resource (http//www.arabidopsis.org). Poplar homologues were identified by reciprocal BLAST searches of the poplar genome assembly (Joint Genomics Institute, Populus trichocarpa v.1.1; http://genome.jgi-psf.org/Poptr1/Poptr1.home.html) using *4CL* and *ArathACLL* sequences as queries. The poplar gene models (from automated *ab initio* gene-calling programs; Tuskan *et al.*, 2006) assigned for a given locus were evaluated, annotated manually and revised as necessary (Table 3.1) All annotated candidates corresponded to loci anchored to poplar linkage groups or to sequence scaffolds, as described in Tuskan *et al.* (2006). Corresponding rice homologues (Table 3.1) were identified in the rice genome using BLAST searches of the rice genome annotation at The Institute for Genome Research (TIGR; http://www.tigr.org/tdb/e2k1/osa1/). *Physcomitrella* sequences were selected in the same fashion from the JGI website. Selected microorganism sequences were obtained by BLAST searches using *4CL* and *ACLL* sequences as queries at the NCBI website (http://130.14.29.110/BLAST/). Protein sequences for were aligned using the Genomatix Dialign program (http://www.genomatix.de/cgi-bin/dialign/dialign.pl) and the multiple protein sequence alignments were manually optimized. To reconstruct phylogenetic trees, maximum likelihood analyses with 1000 bootstrap replicates were carried out using PhyML v2.4.4 (Guindon and Gascuel, 2003) with the JTT model of amino acid substitution.

2.4.2 Identification of *cis*-acting promoter elements

Promoters were defined as approximately 2Kb upstream region from the start codon. We used the PLACE online tool (http://bbc.botany.utoronto.ca/ntools/cgi-bin/BAR_Promomer.cgi) for identification of common elements in the input list of co-expressed genes (Chapter 4; Table 4.1).

2.4.3 Search of co-regulated genes in public Arabidopsis microarray database

Genes co-expressed with *ArathACLLs* were identified using the Platform for Riken Metabolomics (PRIMe - http://prime.psc.riken.jp/?action=coexpression_index) Correlated Gene Search tool, using the union of sets method. Input data was each *ArathACLL* locus ID. All data matrices available were analyzed and those displaying co-expressed genes with highest Pearson coefficients were selected. For each *ArathACLL* search the top 100 genes and/or over 0.6 Pearson coefficient values are shown in Appendix 1. Pajek data output file was used for generating co-expression networks (V. Batagelj *et al.*, 2003).

2.6 Gene expression analysis

2.6.1 Quantitative Real-Time PCR

For the real-time quantitative RT-PCR described in Chapter 3 (Figures 3.4 and 3.7), total Arabidopsis RNA extracted from different tissues (10 μg) was first digested with 15U DNAse in 1x buffer (Invitrogen) for 15 min at room temperature. The reaction was stopped with EDTA (2.5 mM final concentration) and heat-inactivated (65°C, 10 min).

RNA was precipitated with 1 volume of isopropanol and a 1/10 volume of 3M sodium acetate at - 80°C for at least 30 min, and subsequently pelleted at 14,000 rpm in an Eppendorf 3415C microcentrifuge for 40 min at 4°C. The precipitate was washed with 70% ethanol, re-centrifuged, air dried and re-suspended in RNAse free water to an approximate concentration of 0.5 μg/μl. Concentrations were determined spectrophotometrically. 10 μg total RNA was used for reverse transcription with 0.27 μM oligoDT primer, 0.15 mM dNTP's, 40 U RNAseOut, and 400 U SuperscriptII (Invitrogen) in 10 mM DTT and 1 x first strand buffer in a total volume of 40 μl. Prior to addition of enzymes the solution was heated to 65°C for 5 min and for primer annealing cooled to 42°C. Following an incubation at 42°C for 50 min, the reaction was inactivated by heating at 70°C for 15 min. Based on A_{260} concentrations determined for the DNAse treated total RNA samples, cDNA samples were diluted to a concentration of 1ng/μl. Poplar cDNA samples were obtained from S. Ralph (UBC Michael Smith Laboratories; Genome BC Treenomix project) at 1.67ng/μl for use in real-time PCR. For quantitative PCR reactions, 10ng of cDNA was incubated with 10μl QuantiTect SYBR Green PCR mastermix (Qiagen) and 30nmole of each a forward and a reverse primer in a total volume of 20μl. After an initial denaturation step at 95°C for 15 min, 40 cycles at (95°C for 15 sec, 55°C for 30 sec, and 68°C for 45 sec) followed by a fluorescence reading were performed. After a final incubation at 68°C for 5 min, a melting curve was generated ranging from 90°C to 60°C. Threshold cycles were adjusted manually, and the resulting threshold cycles (CT) were subtracted from CT values obtained for a housekeeping control amplified in parallel on each plate 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, and relative quantities were calculated using the ΔΔCT method as described in (Hietala *et al.*, 2003). ΔCT were calculated after normalization using the following control genes: Arabidopsis adenine phosphoribosyltransferase (APT1 - At1g27450) for all Arabidopsis expression experiments and poplar eukaryotic translation initiation factor 5A-1 / eIF-5A 1 (c672 - closest homologue to At1g13950) for all poplar experiments. ΔΔCT was calculated using the following reference tissues: of the highest expressing tissue for developmental expression analysis (Figure 3.4), unwounded leaf tissue (Figure 3.5) and unstressed leaf tissue (Figure 3.7). Only intron-spanning primers were used (Table 2.3). Selected reactions were sequenced for quality control.

Table 2.3: Primers used for quantitative and semi-quantitative RT- PCR

Gene	Primer	Sequence (5'-3')
Arabidopsis		
ACLL1	RT-CLL3F	GAAGTCCTACTGTGATGAAAGG
	RT-CLL3R	AGCTTTCATGTCAGGGATGGG
ACLL2	RT-CLL12F	CAAATACAAAGGCTATCAGGTG
	RT-CLL12R	AGTGTTTGCCGGATGCAGTC
ACLL3	RT-CLL7F	AGGGCCCTTCTATTTCTAAAGG
	RT-CLL7R	CACGTGGCTAGATTCATATCG
ACLL4	RT-CLL6F (purif)	CAACGGGTATAGGAGCTTCAC
	RT-CLL6R (purif)	ACTTCTTTGTCCGGAAACGGG
ACLL5	RT-CLL4F	CCTAATGTCCAAGTCCAAGAG
	RT-CLL4R	CTTCCTCGTCCGGTAACGGC
ACLL6	RT-CLL1F	GGTGCATACCGGAGATCTTGG
	RT-CLL1R	CAGGATGTGATACAAGAAGACC
ACLL7	RT-CLL2F	GGTCCCGGTGTCATGAAAGGATAC
	RT-CLL2R	CAGTTGGAGATTTTGGTATAGAGTTGTCC
ACLL8	RT-CLL8F	GGGCCTTCTATCGCCAAAGG
	RT-CLL8R	CGTGCTACGTAAGCCATCGG
ACLL9	CLL5200F (purif)	CCTGTATCTCCTCCGTTGATTG
	CLL5200R (purif)	CTCTGTCAAGCCATAGCCCTG
APT1 (control)	APT1 F	GTTGCAGGTGTTGAAGCTAGAGGT
	APT1 R	TGGCACCAATAGCCAACGCAATAG
Actin1 (control)	AtActin3F	GCGACAATGGAACTGGAAT
	AtActin3R	GGATAGCATGTGGAAGTGCATACC
Poplar		
ACLL1	RT-POP1F	GAATGCGCCAAGAATTTGCCG
	RT-POP1R	AGGAGGGAGAGGCTTTGCAG
ACLL2	RT-POP16F	GATATGAGGTTCCACGGTCCC
	RT-POP16R	ACTTGAGACTGATAGTAACTTCC
ACLL3	RT-POP33F	CAGGGAAGCATGCTAACACAGG
	RT-POP33R	CAGTTTAGAAGTCAGGGAGCAC
ACLL4	RT-POP26F	CATCATCAACTATTGATTCAGAGG
	RT-POP26R	GGAAATGGTATTACAGCAGCATC
ACLL5	RT-POP27F	ATCGATTCAGAGGGATGGTTAAG
	RT-POP27R	CAGGAAACGGTATTACAGCAGC
ACLL10	RT-POP17F	CACACGTGGAAATAGTACAG
	RT-POP17R	CTCATCTCCTACATAGCCTTTC
ACLL11	RT-POP28F	GCCAACTGTCATGAAGGTTATG
	RT-POP28R	CTCTTCATCAGGATACGGAATC
ACLL12	RT-POP19F	CACAGGCTGAAATAATGCAGGG
	RT-POP19R	CATCTCCTACATAACCTTC
ACLL13	pop24RT-F	CCAGTGTGTTATGCAAGGTTAC
	pop24RT-R	TGCCTCTTCATCTGGCAACGG
c672 (control)	c672F	GACGGTATTTTAGCTATGGAATTG
	c672R	CTGATAACACAAGTTCCCTGC

2.6.2 Semi-quantitative RT-PCR

For the semi-quantitative RT-PCR described in Chapter 4 (Figure 4.1), gene-specific and intron-spanning primers were used in PCR reactions to amplify corresponding cDNA sequences as follows: general PCR conditions were 95°C for 3 min, followed by 30 cycles of (94°C for 30 sec, 57°C for 30s, 72°C 1min) followed by 72°C for 3mins, using Taq polymerase in a 25μl total reaction. PCR products were separated on 1% ethidium bromide agarose gels, and photographed under UV transilluminator using AlphaImager 1220. Actin 1 (At2g37620) was used as control.

2.6.3 GUS histochemical assay

The GUS histochemical assay solution was prepared by mixing an aqueous solution of 100 mM $NaPO_4$, pH 7.0, 0.5% X-GLUC (bromochloroindoyl-b-glucuronide) with an aqueous solution of 2 mM $K_3Fe(CN)_6$ and 2 mM $K_4Fe(CN)_6$ in 0.1% TritonX. Young Arabidopsis leaf blades were wounded with scissors, cut from the plant after 1h and placed in an 1.5mL plastic tube with cold GUS solution. The tubes were vacuum infiltrated for 15 minutes. The samples were incubated at 37°C for 2h or until a blue color could be seen. The reaction was stopped by removal of the assay buffer and the addition of 95% ethanol. Samples were cleared by incubation in 95% ethanol overnight. Stained Arabidopsis leaves were visualized using a Leica dissecting microscope and Spot32 camera and software, at the UBC BioImaging Facility.

2.7 Sub-cellular localization of ArathACLL4 and PoptrACLL5

Agrobacterium strains carrying GFP::ArathACLL4 and the GFP::PoptrACLL5 were used to transform tobacco leaf discs for generating a transgenic tobacco plant, as described in section 2.3.3. Transgenic plantlets were screened for fluorescence indicating *GFP* expression under an epifluorescence microscope (Zeiss Axioplan 2) and plants with both high and low levels of GFP expression were selected for analysis by confocal microscopy (Zeiss Meta Confocal). Plants expressing *GFP::ArathACLL4* were used as a positive control and plantlets with no visible GFP fluorescence and wild type tobacco plants were used as negative controls. Chloroplast auto-fluorescence was excited with a 488-nm argon laser and was detected after passage through a long-pass 650-nm emission filter. GFP fluorescence was excited with a 488-nm laser and was detected after passage through a band pass 505-530-nm emission filter. Images were reconstructed using the ImageJ software suite (http://rsb.info.nih.gov/ij/index.html).

2.8 Identification and characterization of an *ACLL5* insertion mutant

2.8.1 Genetic methods

Seeds for an ArathACLL5 (At1g62940) transposon insertion line (stock code N123936; synonymous SM_3.37225) were obtained from the The European Arabidopsis Stock Centre (NASC) Arabidopsis Biological Resource Center. Homozygous insertion lines were identified by PCR-based screening for both the presence of the transposon insertion and the absence of an intact endogenous gene. Primers are listed on Table 2.4. Primer combinations used were as follows: CLL4F and EcoR1 reverse for detection of endogenous gene, and CLL4F and dspn1 for detection of the transposon insertion. PCR

analysis confirmed the homozygosity of the insertion insert in all plants displaying male sterile phenotype. The PCR fragments generated by CLL4F and dspn1 were sequenced to determine the exact location of the insertion in the *ACLL5* gene. Real-time PCR (section 2.6.1) and RT-PCR (section 2.6.2) were used to determine mRNA levels in the mutants. Genetic crosses of wild-type pollen to a homozygous *acll5* mutant plant were performed to obtain F2 generation plants. The pattern of insertion segregation *ACLL5* transposon insertion in the F2 generation was tested by chi-square statistical analysis on observed phenotypes and genotypes.

Table 2.4: Primers used for genotyping *ACLL5* transposon insertion lines

Gene	Primer	Sequence (5' - 3')
ArathACLL5	CLL4F	ATGGAGAGTCAAAAGCAAGAAGATAATG
	EcoRIreverse	CATTGTCGGTATCTCCGCATTTGTC
transposon	dspm1	CTTATTTCAGTAAGAGTGTGGGGTTTTGG

2.8.2 Phenotypic analysis of the *acll5-1* mutant

For scanning electron microscopy observations, using a Hitachi S4700 SEM, inflorescences of wild type and homozygous mutant lines were fixed overnight in 2% glutaraldehyde, washed and post-fixed in 1% osmium tetraoxide in 0.05 M PIPES buffer, and dehydrated using a series of graded ethanol solutions (30% to 100%). Dried samples were gold sputter coated (Nanotech SEMPrep II Sputter Coater). To obtain cross sections of developing anthers, inflorescences of wild type and mutant lines were fixed in FAA (4% paraformaldehyde, 15% acetic acid, and 50% ethanol) overnight and directly dehydrated without post-fixation. Samples were then transferred to a propylene oxide solution and slowly infiltrated with Spurr's epoxy resin (Canemco). For bright-field

microscopy, 1μm 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. All procedures described were performed in the UBC BioImaging Facility (http://www.emlab.ubc.ca).

2.8.3 *In situ* hybridization (Experiment performed by S.McKim, UBC)

Arabidopsis Col-0 inflorescences were embedded in Paraplast (Sigma), sectioned at 8μm thickness and mounted onto pre-charged slides. For antisense *AtCLL5* probe synthesis, a 1629bp DNA template corresponding to the entire *AtCLL5* cDNA was amplified by PCR from flower cDNA using the forward primer CLL4F (Table 2.4) and 5'-GATAATACGACTCACTATAGGCTACTTCTTGTTGATGCTGAGGATC-3' reverse primer which incorporates a T7 polymerase binding site. Digoxigenin (DIG) -labeled probes were transcribed off the template using T7 polymerase (Roche). Probes were shortened to 200bp fragments by limited carbonate hydrolysis, quantified and hybridized to slides. Tissue fixation, embedding, probe design, hybridization and signal detection are described in Hooker *et al.* (2002).

CHAPTER 3 - GENOME-WIDE PHYLOGENETIC ANALYSIS AND COMPARATIVE GENOMICS OF THE PLANT-SPECIFIC ACYL:COENZYMEA LIGASE-LIKE (*ACLL*) GENE FAMILY IN ARABIDOPSIS AND POPLAR

3.1 Introduction

The enzyme 4-coumarate:CoA ligases (4CLs) play important roles in phenylpropanoid metabolism by generating CoA esters of hydroxycinnamic acids. These cinnamyl CoA esters are used as intermediates in the biosynthesis of a large array of phenolic secondary natural products, including monolignols and flavonoids (Hahlbrock and Scheel, 1989). The first *4CL* gene cloned was derived from parsley (Douglas *et al.*, 1987), and it has subsequently been shown that 4CL enzymes are encoded by multi-gene families in all vascular plants examined to date (Cukovic *et al.*, 2001; Hamberger *et al.*, in press). Analysis of enzymatic properties of recombinant enzymes has revealed that 4CL isoenzymes have differential activity towards different hydroxycinnamyl substrates (Allina *et al.*, 1998; Ehlting *et al.*, 1999; Hamberger and Hahlbrock, 2004; Hu *et al.*, 1998; Lee and Douglas, 1996; Lindermayr *et al.*, 2002; Stuible and Kombrink, 2001). The analysis of the *4CL* gene family in the fully sequenced Arabidopsis (Ehlting *et al.*, 1999; Hamberger and Hahlbrock, 2004) and poplar (Tuskan *et al.*, 2006) genomes showed that 4CL is encoded by four and five genes respectively. Differential *4CL* gene expression patterns in Arabidopsis and poplar, coupled with 4CL isoenzyme substrate utilization preferences, suggest that *4CL* genes and enzymes have undergone subfunctionalization and neofunctionalization for biosynthesis of monolignols and flavonoids (Ehlting *et al.*, 1999; Hamberger and Hahlbrock, 2004; Hu *et al.*, 1998).

4CL enzymes are members of the adenylate-forming enzyme superfamily, which share a common reaction involving formation of an adenylate intermediate, and includes enzymes involved in fatty acid chain elongation (Shockey *et al.*, 2003). Following the generation of sequence data from plant genomes, a number of genes encoding adenylate-forming enzymes of unknown specific function related to true *4CLs* (*4CL-like, ACLL* genes) have been annotated, and these may function as unknown enzymes in plant metabolism and secondary metabolism. For example, initial Arabidopsis genome sequence data revealed the presence of Arabidopsis *4CL-like* genes (Cukovic *et al.*, 2001), and later, eight members of a larger set of Arabidopsis adenylate-forming enzymes annotated in the completed Arabidopsis genome were classified as *4CL-like* genes because of their close phylogenetic relationship to true *4CLs* (Shockey *et al.*, 2003; this study). It has been proposed, however, that enzymes encoded by *4CL-like* genes may not have activity towards the known 4CL substrates, but instead, activate acyl molecules derived from fatty acid metabolism (Shockey *et al.*, 2003).

Jasmonic acid (JA) is an important plant signaling molecule, generated from the membrane lipid linolenic acid via the octadecanoid pathway (Liechti and Farmer, 2002). JA and its volatile methyl ester, methyjasmonate (MeJA), are known to be important plant growth regulators and stress signaling molecules mediating responses to various developmental and environmental cues, such as wounding and herbivory (Farmer *et al.*, 2003; Li *et al.*, 2005; Sasaki-Sekimoto *et al.*, 2005). The role of JA in regulation of gene expression has been well documented, and many of the enzymatic steps in its biosynthesis have been characterized (Schaller *et al.*, 2004; Schilmiller *et al.*, 2007). The

latter steps of the octadecanoid pathway occur in the peroxisome, in which the activated CoA ester of the plastid-derived precursor 12-oxo-phytodienoic acid (OPDA) is generated, followed by three rounds of acyl chain shortening by beta-oxidation, (Li *et al.*, 2005). Therefore enzymes involved in this part of the pathway are predicted to be targeted to this organelle, likely via a C-terminal peroxisomal target signal (PTS1) or N-terminal signal (PTS2) as deduced from sequence analysis of plant peroxisome proteins (Reumann, 2004), and previously reported for *in vivo* import in *Trypanosoma brucei* (Sommer *et al.*, 1992). Proposed substrates for certain Arabidopsis ACLLs derive from the octadecanoid pathway, including OPDA and 3-oxo-2(2'[Z]-pentenyl)-cyclopentane-1-octanoic acid (OPC8) (Koo *et al.*, 2006; Schneider *et al.*, 2005). However, the biological functions of most the *4CL-like* genes are still unknown.

There is an increasing amount of publicly available gene expression data in Arabidopsis, such as expression data generated by various microarray experiments. These data are searchable using bioinformatic tools such as Expression Angler (Toufighi *et al.*, 2005) and PRIMe (http://prime.psc.riken.jp). Therefore, networks of co-expressed genes can be visualized by mining existing gene expression data. For enzyme-encoding genes, such co-expression analysis can provide clues regarding possible metabolic pathways to which enzymes of unknown function may belong, based on their correlated co-expression with genes encoding other enzymes (Ehlting *et al.*, 2006). In addition, the completion of the rice (Yuan *et al.*, 2003) and poplar (Tuskan *et al.*, 2006) genomes, together with the reference genome of Arabidopsis, opens the door to the application of comparative

genomic approaches to understanding the evolution and potential functions of conserved genes of unknown specific function such as those in the *ACLL* gene family.

In an initial analysis, the complete set of Arabidopsis *ACLL* genes, formerly called *4CLL* genes, was identified based on their similarity to genes encoding *bona fide* 4CL enzymes (Ehlting *et al.*, 2005). In this chapter, I identified all *ACLL* genes in the fully sequenced poplar and rice genomes. In addition, I obtained full-length *ACLL* sequences from a maize genome database (sequences provided by Dr. Brad Barbazuk), retrieved nucleotide sequences from publicly available plant genome databases, and searched eukaryotic and prokaryotic genome databases for *ACLL* genes in diverse taxa. Phylogenetic reconstructions based on amino acid sequence alignments showed that *ACLL* genes belong to a land plant-specific clade of adenylate-forming enzymes more closely related to true *4CLs* than any other adenylate-forming enzyme. Furthermore, each fully sequenced plant genome has representatives in each of five well-defined *ACLL* clades, four of which contain proteins predicted to be localized in the peroxisome. This suggests that ACLL enzymes perform important, conserved roles in plant metabolism. I profiled the developmental and stress-induced expression of Arabidopsis and poplar homologues representing all five clades, and similarities in expression patterns across these taxa allowed me to identify putative orthologues, and suggested subfunctionalization of *ACLL* genes in these two lineages. In addition, using Arabidopsis co-expression analysis, I was able to predict the function of poplar homologues in one ACLL clade related to JA biosynthesis, a hypothesis that was further tested by monitoring stress-induced gene expression and subcellular localization.

3.2 Results

3.2.1 Phylogenetic analysis of 100 ACLLs from plants and microorganisms

The adenylate-forming enzyme superfamily of genes includes members from all organisms, including prokaryotes and eukaryotes (Conti *et al.*, 1996), and are distinguished by the presence of conserved structural elements that define this superfamily (Conti *et al.*, 1997). The phenylpropanoid enzyme 4CL is one member of this family that has been extensively studied due to its important role in the phenylpropanoid pathway. As a first step towards a genome-wide survey of *ACLL* genes most closely related to 4CLs, I identified 100 predicted 4CL-like (ACLL) proteins from genomic databases using an *in silico* similarity search based on the amino-acid sequences of Arabidopsis 4CL proteins (Ehlting *et al.*, 1999; Hamberger and Hahlbrock, 2004), as described in Materials and Methods. In this analysis, I focused on the three complete genome sequences available for angiosperms (Arabidopsis, poplar, and rice), the genomes of maize, *Physcomitrella patens*, *Chlamydomonas reinhardtii*, and the genomes of selected other microorganisms (fungi and bacteria) for which complete or substantial genome sequence data were available.

In order to determine if a plant-specific clade of ACLL enzymes could be circumscribed, I carried out phylogenetic analysis of all 100 aligned ACLL translated nucleotide sequences using PhyML 4 (Guindon and Gascuel, 2003) and generated the maximum likelihood tree shown in Figure 3.1. This analysis revealed two general groups of ACLL proteins. One large group contained representatives from all organisms analyzed, including bacteria, fungi, *Chlamydomonas*, *Physcomitrella*, and angiosperms, which are

relatively distantly related to true 4CLs. Many representatives in this group are probably adenylate-forming enzymes with metabolic functions in primary metabolism or related functions common to all or many prokaryotic and eukaryotic cells. As an example of possible functions of such enzymes, one clade in this group contains the Arabidopsis ACN1 protein, an acetate:CoA ligase which functions as an entry point to the glyoxylate cycle during seed germination (Turner *et al.*, 2005). I did not carry out any further analyses of plant genes or enzymes in this large group.

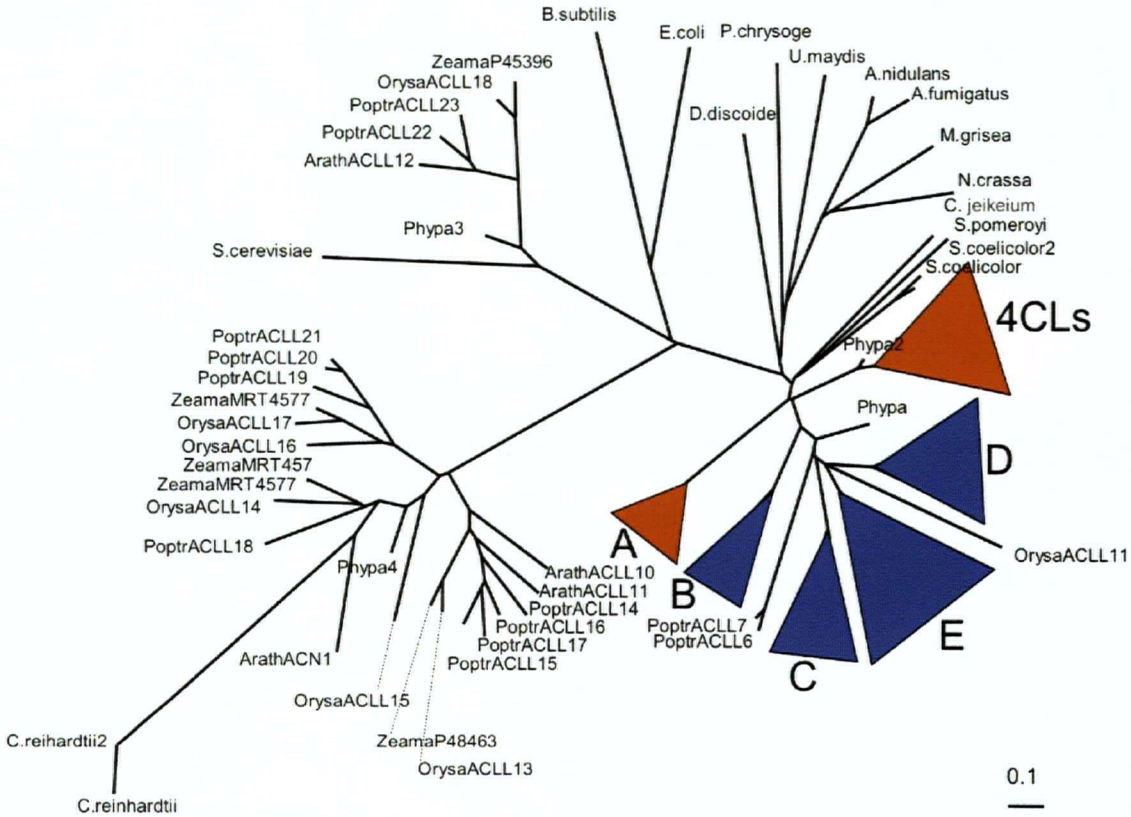


Figure 3.1: Phylogenetic tree of 100 ACLLs corresponding to translated nucleotide sequences from full-length cDNAs and ESTs from various organisms. Solid triangles represent the clades of plant-specific ACLLs most closely related to true 4CL enzymes.

The second group of ACLL proteins, containing previously annotated Arabidopsis 4CL-like (4CLL) proteins (Ehlting *et al.*, 2005; Shockey *et al.*, 2003), was striking in that it was land plant-specific. All angiosperm genomes, as well as the *Physcomitrella* genome, encoded proteins contained in this group, while no representatives from other eukaryote species, including *Chlamydomonas*, were found. Based on their deduced phylogenetic relationships to each other, the ACLLs in this group could be further divided into five clades (Figure 3.1; clades A to E), which are phylogenetically closely related *bona fide* 4CLs. In each of the clades there is at least one representative of each of the four angiosperm plant species analyzed (*Arabidopsis*, poplar, rice, and maize; data not shown), demonstrating that these proteins are evolutionarily conserved in the angiosperm lineage and that common ancestors in each clade were present before the divergence of monocots and eudicots.

Analyses of the *Physcomitrella* EST dataset revealed an ACLL protein monophyletic to true 4CLs, suggesting that the 4CL clade likely originated early during the evolution of land plants, consistent with the postulated role of phenylpropanoids in the adaptation to the land environment (Douglas, 1996). Interestingly, a second *Physcomitrella* ACLL protein in this group appears basal to clades C, D and E, suggesting that the proteins from these clades originated from the same common ancestor, also early in land plant evolution. Completion of the *Physcomitrella* genome sequence should reveal whether additional *Physcomitrella* ACLLs exist.

3.2.2 Most ACLLs contain the PTS1 (Peroxisomal Target Signal 1)

Almost all proteins in clades B, C, D and E contain the PTS1 peroxisomal target sequence in their C-termini, which suggests they are targeted to this organelle. To my knowledge, ArathACLL9 from clade E and ArathACLL4 from clade D are the only enzymes for which this localization has been experimentally demonstrated, using a GFP-tagging approach (Koo *et al.*, 2006; Schneider *et al.*, 2005). Interestingly, all fungal adenylate-forming enzymes identified also have peroxisomal target signals. None of the ACLLs in clade A, the ACLL clade most closely related to *bona fide* 4CLs, contained the PTS1 sequence, suggesting that loss of this sequence may have played a role in the acquisition of 4CL and clade A functions.

3.2.3 Species-specific *ACLL* gene family evolution

In order to gain insights into species-specific retention and expansion of the plant-specific *ACLL* genes for each clade, I next analyzed each clade in more detail, focusing on the complete *ACLL* gene families from the three angiosperm genomes for which whole genome sequence information is available: Arabidopsis, poplar, and rice. As shown in Figure 3.2, all three species contained ACLL proteins in each of clades A-E. My annotation of the complete set of *ACLL* genes in poplar, and their locations on poplar linkage groups, is given in Table 3.1.

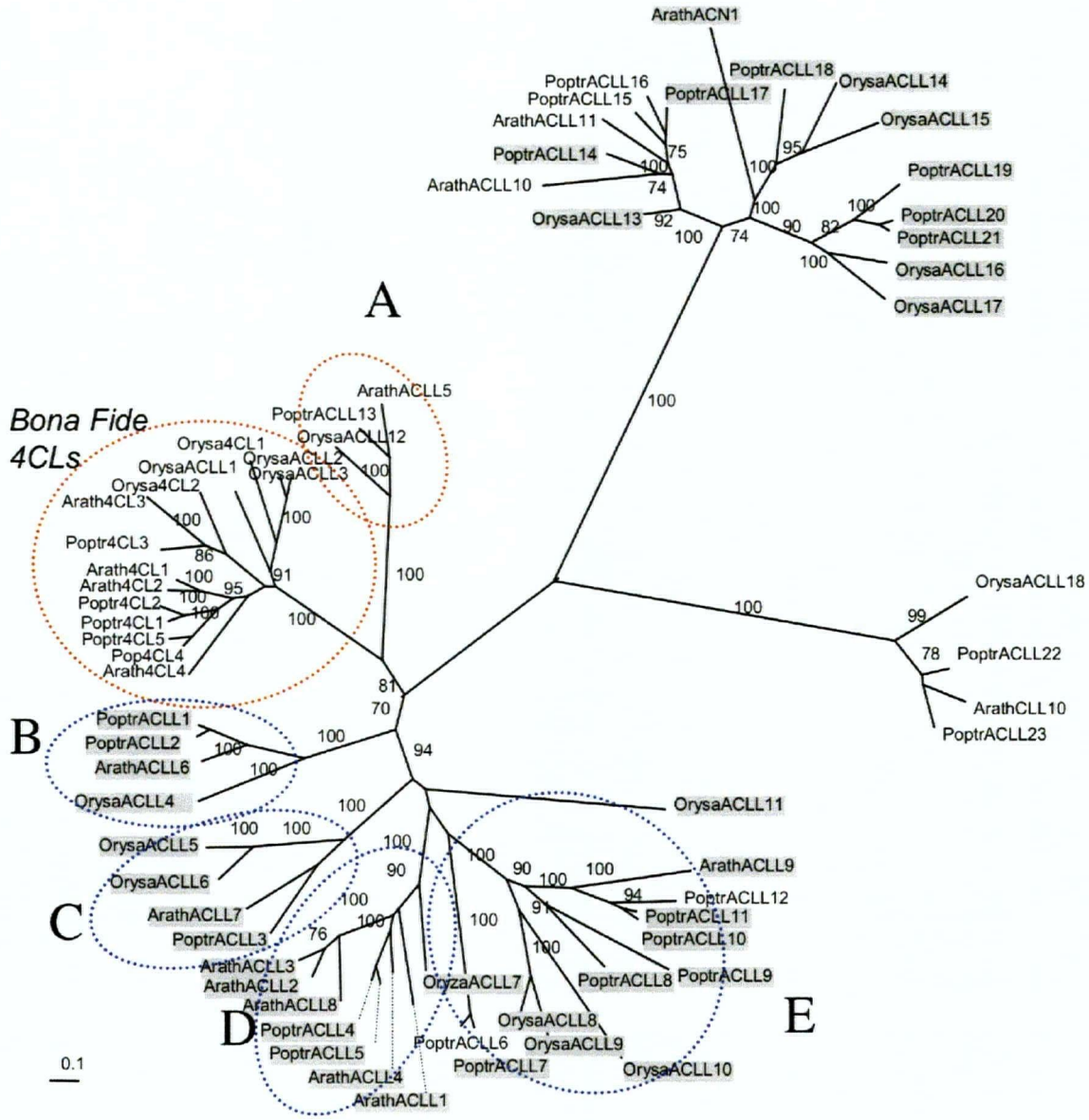


Figure 3.2: Phylogenetic relationship of plant-specific ACLLs, including translated nucleotide sequences from Arabidopsis, poplar and rice. Bootstrap values are shown on the branches. Values below 70% were removed from the tree. Clades are circled and contain at least one representative of each plant species. Proteins in shaded boxes contain the PTS1 (Peroxisomal Target Signal 1).

Table 3.1: Annotation of *Populus trichocarpa* and *Oryza sativa ACLL* genes.

Gene name	Clade	Gene model	Location	Coordinates
Poplar*				
PoptrACLL1	B	eugene3.01230068	scaffold 123	568900-572342
PoptrACLL2	B	estEXT_fgenesh1_pg_v1.C_LG_IV0024	LG IV	8408652-8412261
PoptrACLL3	C	fgenesh4_pg.C_LG_III000781	LG III	9834289-9841997
PoptrACLL4	D	eugene3.00020113	LG II	738720-743055
PoptrACLL5	D	fgenesh4_pm.C_LG_V000686	LG VIII	17267514-17272723
PoptrACLL6	E	fgenesh4_pm.C_LG_X000932	LG X	19581039-19583937
PoptrACLL7	E	fgenesh4_pm.C_LG_VIII000094	LG VIII	1492142-1494543
PoptrACLL8	E	eugene3.00640074	scaffold 64	517028-522599
PoptrACLL9	E	fgenesh4_pm.C_LG_X000174	LG X	6952628-6055525
PoptrACLL10	E	estEXT_fgenesh4_pm.C_LG_XV0272	LG XV	6923655-6933433
PoptrACLL11	E	eugene3.00120875	LG XII	11162718-11168082
PoptrACLL12	E	grail3.0015024001	LG XII	11177018-11180781
PoptrACLL13	A	eugene3.00010460	LG I	3983954-3987069
Rice				
OrysaACLL1	4CL	Os08g34790		
OrysaACLL2	4CL	Os02g08100		
OrysaACLL3	4CL	Os06g44620		
OrysaACLL4	B	Os03g05780		
OrysaACLL5	C	Os10g42800		
OrysaACLL6	C	Os08g04770		
OrysaACLL7	D	Os03g04000		
OrysaACLL8	E	Os01g67530		
OrysaACLL9	E	Os01g67540		
OrysaACLL10	E	Os07g17970		
OrysaACLL11	N/D**	Os07g44560		
OrysaACLL12	A	Os04g24530		

*Gene model and locations from JGI *Populus trichocarpa* web browser v.1.1
** not defined by phylogenetic results

While the number of *ACLL* genes within each genome was similar (13 in poplar, 12 in rice, and 9 in Arabidopsis), the number of genes in each clade varied between species, and certain clades were greatly enriched with genes from a particular species. For example, clade D is an Arabidopsis rich clade, with 5 Arabidopsis representatives, two from poplar, and only one rice member. On the other hand, clade E is poplar rich with seven poplar *ACLL* genes, three from rice and one from Arabidopsis. Clade A, unique in containing ACLL proteins lacking the PTS1 targeting signal, is the only clade that contained a single representative from each species. Thus, while the origin of the ACLLs clades clearly predates the divergence of monocots and dicot lineages, the variable numbers of genes in most clades reveals species-specific genome evolution. For example, four of the five Arabidopsis genes in clade D are found in tandem in chromosome 1 (Figure 3.3), which suggests that the duplication of the original gene that gave rise to clade D was a result of tandem gene duplication events and selection for retention of copies in the Arabidopsis genome. Two of the poplar ACLL genes in clade E (*ACLL11* and *ACLL12*) appear to have arisen by tandem duplication on linkage group XII (Table 3.1). However, other members of this and other clades that have poplar gene models anchored to linkage groups are physically unlinked and on different linkage groups. This suggests that tandem gene duplication was not the only factor responsible for the diversification of the poplar *ACLL* gene family, in keeping with the apparent greater role of tandem gene duplications in Arabidopsis genome evolution relative to poplar (Tuskan *et al.*, 2006). Many of the poplar *ACLL* genes may rather have been retained after the salicoid whole genome duplication in the *Populus* lineage, in which chromosome doubling and subsequent rearrangement is thought to have increased the *Populus*

chromosome number from n=10 to the current n=19 (Tuskan *et al.*, 2006). For example, in the poplar rich clade E, *ACLL10*, *ACLL11* and *ACLL12*, which are located on duplicated homologous linkage groups XII and XV, and *ACLL6* and *ACLL7*, which are located on duplicated homologous linkage groups XIII and X (Table 3.1; Tuskan *et al.*, 2006) are likely to have arisen in this manner.

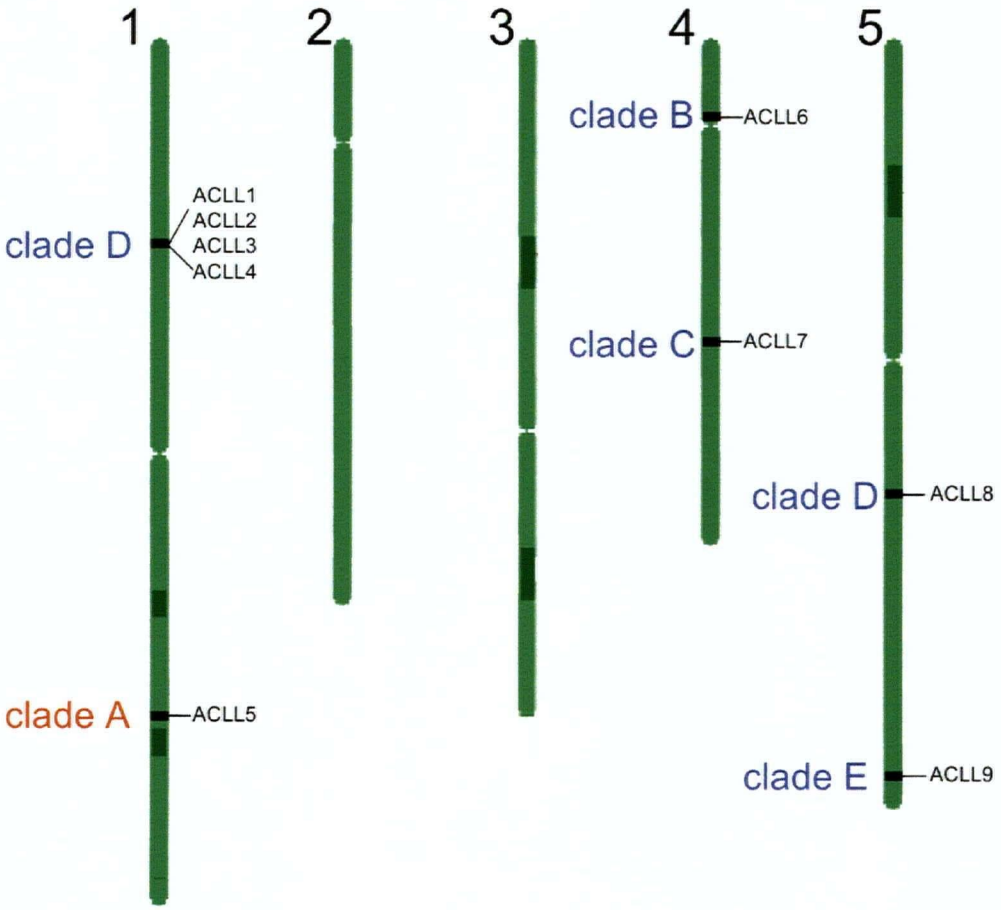


Figure 3.3: Schematic representation of the *Arabidopsis thaliana* genome, showing the location of ACLL genes with the respective clades. Clade D is the only clade containing more that one ACLL in Arabidopsis, containing 4 genes originated by tandem duplication in chromosome 1 and one gene in chromosome 5.

Also noteworthy amongst the poplar clade E ACLLs is the loss of C-terminal PST1 peroxisomal targeting sequences in two members (ACLL6 and ACLL12), suggesting that functional diversification, or neofunctionalization, may have taken place at the level of enzyme localization in the poplar lineage after gene duplication. Taken together, these data show two scenarios in the evolution of the *ACLL* genes: conservation of *ACLL* gene number, and likely function, in all three angiosperm lineages for some clades (A, B, and C, with 1-2 members from each lineage), and family expansion with possible diversification of function taking place in a lineage-specific manner for other clades (D and E).

3.2.4 Comparative analysis of Arabidopsis and poplar 4CL and ACLL proteins

In order to assess the amino acid sequence diversity of ACLL proteins relative to that of well-characterized *4CL* gene family members, the levels of amino-acid sequence conservation among poplar and Arabidopsis 4CL enzymes were compared to the levels ACLL protein sequence conservation in each clade. Similar levels of interspecies identity may indicate retention of function among ACLL enzymes between lineages, as observed for the enzymes in the *bona fide* 4CL clade. Table 3.2 shows the identity values of Arabidopsis and poplar 4CLs and ACLLs in relation to each other for each clade. Identity was calculated for both overall sequence (OS) and putative substrate binding domain region (SBD) and are shown as OS%/SBD%. Identity values between Arabidopsis and poplar 4CLs, which use the same or similar substrates, are over 65%/70%, suggesting that ACLLs with similar or higher levels of conservation in sequence may also have conserved functions. The data show that OS%/SBD% identity values were 75%/81%

between ArathACLL5 and PoptrACLL13 of clade A, the highest level of conservation seen between poplar and Arabidopsis proteins in this study. In clade B, with one Arabidopsis and two poplar genes, values were 74%/75% and 74%/73%, also demonstrating high conservation of sequence. This result may indicate that ACLLs in clades A and B have conserved functions in both poplar and Arabidopsis.

In clade C, with one ACLL copy in Arabidopsis and poplar, the sequence conservation level was of 58%/65%, suggesting that genes in this clade have diverged and could carry out species-specific functions. In clade D, enriched with Arabidopsis sequences, identity values between poplar and Arabidopsis sequences were low, comparable to that of clade C, with the exception of ArathACLL4, which reached values of 72%/75% when compared to the two poplar proteins in this clade, PotrACLL4 and PoptrACLL5. This result may suggest that ArathACLL4 function has been conserved in both species, while other Arabidopsis ACLLs in the same clade may have diverged in function. For clade E, which contains 7 poplar genes for one Arabidopsis gene, the three poplar homologues most similar to ArathACLL9 were analyzed, and the results showed that amino-acid identity values were generally low. The highest values were between ArathACLL9 and PoptrACLL11 62%/73%, dropping to 59%/66% for PoptrACLL10 and 55%/64% for PoptrACLL12. With the exception of PoptrACLL11, results suggest that the poplar ACLLs in clade E may have diverged to fulfill other biochemical and/or biological roles in a species-specific manner.

Table 3.2: Amino acid identity comparison of full-length amino acid sequences of 4CLs and ACLLs in the different clades. Results of pairwise similarity are shown as full-length sequence% / predicted substrate binding domain%.

4CLs	Arath4CL1	Arath4CL2	Arath4CL3	Arath4CL4	Poptr4CL1	Poptr4CL2	Poptr4CL3	Poptr4CL4	Poptr4CL5
Arath4CL1		78%/87%	56%/67%	62%/72%	66%/77%	65%/76%	58%/70%	67%/73%	68%/74%
Arath4CL2			60%/66%	63%/71%	69%/75%	68%/75%	60%/69%	67%/74%	69%/74%
Arath4CL3				55%/61%	62%/66%	62%/67%	69%/75%	59%/66%	69%/66%
Arath4CL4					59%/68%	60%/69%	52%/62%	60%/64%	62%/66%
Poptr4CL1						85%/85%	63%/70%	72%/76%	74%/79%
Poptr4CL2							62%/67%	72%/77%	74%/79%
Poptr4CL3								60%/70%	60%/70%
Poptr4CL4									89%/90%
Poptr4CL5									

Clade A	ArathACLL5	PoptrACLL13
ArathACLL5		75%/81%
PoptrACLL13		

Clade B	ArathACLL6	PoptrACLL1	PoptrACLL2
ArathACLL6		74%/75%	74%/73%
PoptrACLL1			91%/88%
PoptrACLL2			

Clade C	ArathACLL7	PoptrACLL3
ArathACLL7		58%/65%
PoptrACLL3		

Clade D	ArathACLL1	ArathACLL2	ArathACLL3	ArathACLL4	ArathACLL8	PoptrACLL4	PoptrACLL5
ArathACLL1		58%/50%	59%/58%	65%/63%	53%/56%	63%/66%	61%/62%
ArathACLL2			83%/75%	64%/65%	70%/73%	60%/60%	61%/57%
ArathACLL3				67%/69%	71%/72%	65%/63%	64%/61%
ArathACLL4					60%/67%	72%/75%	72%/74%
ArathACLL8						61%/63%	60%/60%
PoptrACLL4							88%/90%
PoptrACLL5							

Clade E	ArathACLL9	PoptrACLL10	PoptrACLL11	PoptrACLL12
ArathACLL9		59%/66%	62%/73%	55%/64%
PoptrACLL10			87%/85%	72%/67%
PoptrACLL11				72%/74%
PoptrACLL12				

3.2.5 Comparative expression analysis of Arabidopsis and poplar genes

In order to gain clues as to the possible functions of ACLL proteins, I examined the gene expression patterns of all Arabidopsis *ACLL* genes, as well as representative poplar genes in clades A-E by quantitative real-time reverse transcription-PCR. Data on expression in different organs are shown in Figure 3.4.

These results revealed that Arabidopsis and poplar *ACLL* homologues tended to have similar developmental expression patterns in clades in which there are single Arabidopsis and poplar ACLL representatives. A striking example is clade A, in which *ArathACLL5* expression was strongly flower-preferred, and *PoptrACLL13*, while also showing expression in phloem and bark, showed a similar pattern of flower-preferred expression. Interestingly, *PoptrACLL13* expression is specific to male flowers, and a putative *ArathACLL5* orthologue in tobacco shows an anther preferred expression pattern (Varbanova *et al.*, 2003). Together, these data suggest a role for ACLL enzymes in clade A in a biochemical pathway important in anther and/or pollen development. Another example is the predominant expression of both Arabidopsis and poplar representatives of clade C in leaves, with less predominant expression in stem/xylem/phloem and flowers. Clade B contains two poplar genes and a single Arabidopsis member. Genes in this clade were expressed in all organs, but both poplar *PoptrACLL2* and *ArathACLL6* showed highest expression in mature leaves, and lower expression in other organs and tissues. Interestingly, the *PoptrACLL1* expression profile differed from that of *PoptrACLL2*, with highest expression in flowers, bark, and young leaves, suggesting subfunctionalization of expression patterns, as predicted as one possible outcome of genes retained after duplication events (Duarte *et al.*, 2006).

More complex expression patterns were observed in clades where substantial expansion of gene family members in either Arabidopsis or poplar has occurred. In clade D, the duplicated poplar genes *PoptrACLL4* and *PoptrACLL5* appeared to have very similar expression patterns across a range of organs and tissues, with low expression only in

xylem and male flowers (Figure 3.4). However, the transcribed portions of the two poplar genes were nearly identical, making it impossible to design gene-specific PCR primers, and the products amplified using primers for each gene were contaminated with products from the other gene (data not shown). In contrast, for the five representatives of the Arabidopsis members of this clade, I saw distinct and complementary expression patterns throughout almost all organs tested in Arabidopsis, which suggests subfunctionalization in expression of these duplicated genes in clade D. The only Arabidopsis gene from clade E, *ArathACLL9*, was most highly expressed in seedlings, followed by flowers. The expression patterns of the three poplar homologues most closely related to *ArathACLL9* (out of the seven poplar genes present in this group) were largely complementary to each other, covering expression in leaves, roots and male flowers, and did not parallel that of *ArathACLL9*.

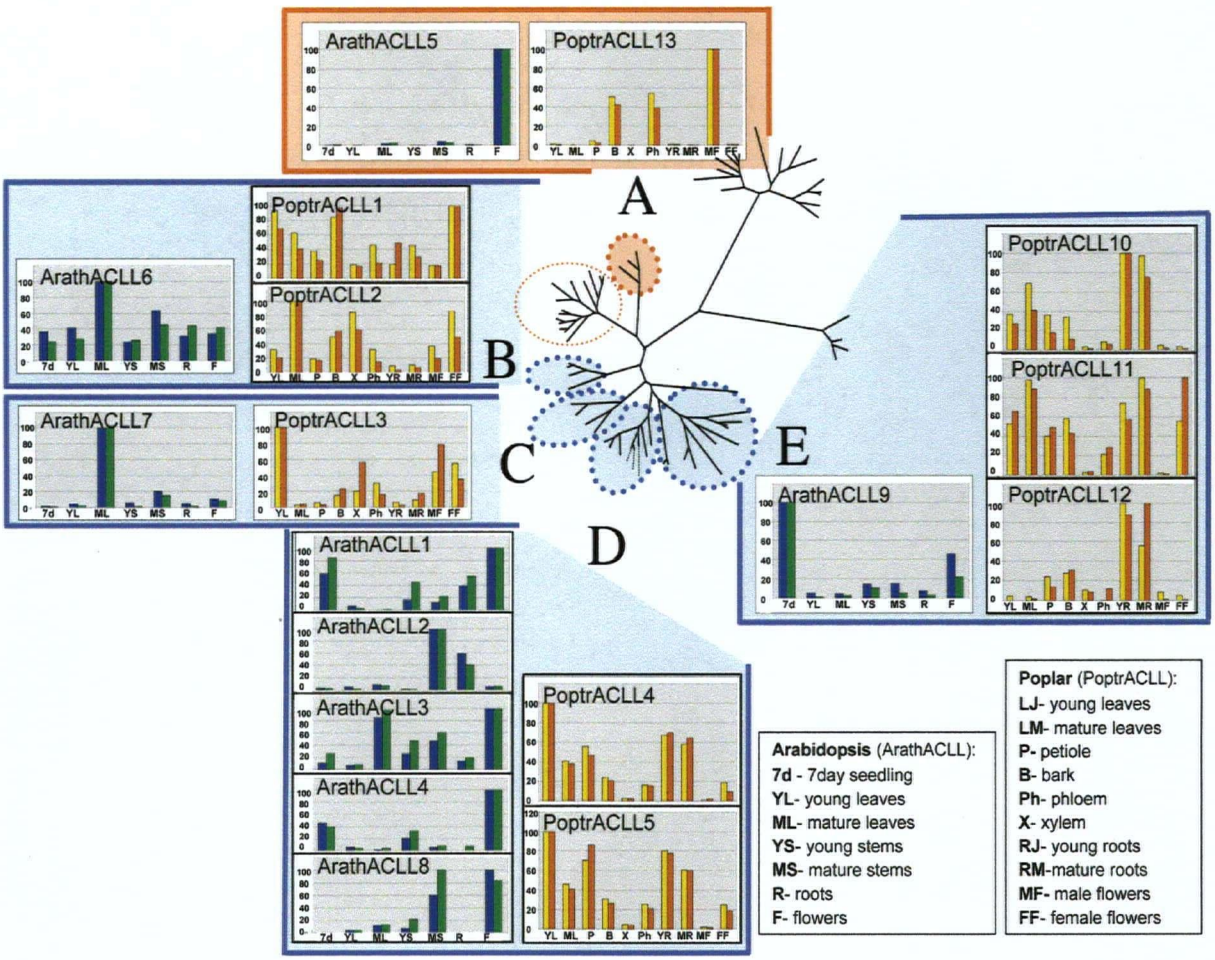


Figure 3.4: Tissue expression profile of *ACLLs* in Arabidopsis and poplar. Expression was determined by real-time PCR relative to the tissue with the highest level of expression, set at 100%. Calibrator genes used were *APT1* for Arabidopsis and *c672* for poplar. Two technical replicates per tissue were tested.

3.2.6 Identification of stress responsive *ACLL* genes

One important clue in the quest towards identifying biological roles for the ACLLs came from the presence of the peroxisomal target signal (PTS1) in most of the Arabidopsis, rice, and poplar ACLLs in clades B, C, D and E. Given that the ACLLs in this study are part of a plant-specific group (Figure 3.1), I considered possible plant-specific peroxisomal functions. For example, plant peroxisomes, in addition to ubiquitous functions in primary metabolism, perform special roles in synthesizing plant hormones such as auxin and jasmonates (Nyathi and Baker, 2006). Jasmonates, in particular, play important roles in response to stress (Farmer *et al.*, 2003). Therefore, *ACLL* genes encoding proteins targeted to the peroxisome were subjected to further expression analyses in order to gain insights into possible environmental influence on their expression.

I generated transgenic Arabidopsis plants expressing chimeric constructs of selected Arabidopsis *ACLL* promoters fused to *GUS* reporter genes, and then analyzed GUS activity histochemically. Promoter regions were defined as the genomic sequences directly upstream of the ATG start codon, between 1.5 and 2-kb in length. At least five independently transformed lines were generated per gene, and at least 5 individuals of each transgenic Arabidopsis line were examined. Transgenic plants were subject to mechanical wound treatments as described in Material and Methods, and representative results consistently observed are shown in Figure 3.5A. Out of all plants tested, *GUS* expression was stronger at the wound sites of transgenic plants containing promoter constructs from clade D genes (*ACLL2*, *ACLL3* and *ACLL4*), indicating activation of

these *ACLL* promoters upon wounding. Promoters of *ACLL* genes in clades B, C and E, and of *ACLL1* and *ACLL8* in clade D were not wound activated by this assay.

Wound responsiveness of genes encoding peroxisomally targeted ACLLs was further tested by measuring gene expression using quantitative real-time reverse transcription PCR at various times after mechanical wounding in both Arabidopsis and poplar (Figure 3.5B). Arabidopsis *4CL2*, known to be wound inducible (Ehlting *et al.*, 1999), was used as a positive control for Arabidopsis treatments, and was up-regulated by over 5 fold at 4h after wounding, as expected (data not shown).

In clade B, *ArathACLL6* and *PoptrACLL2* showed no wound responsiveness, whereas *PoptrACLL1* expression was slightly up-regulated (1.6-fold) after 6h wounding. These data suggest that *ArathACLL6* and *PoptrACLL2* do not function in wound-related biochemical pathways, but that *PoptrACLL1* may have gained this function after duplication of the gene in the *Populus* lineage. In clade C, *ArathACLL7* expression was down-regulated to less than half the level of the unwounded control, and a similar result was obtained for the single poplar homologue in this group, *PoptrACLL3*. Arabidopsis and poplar genes in clade D were particularly responsive to wounding, with two out of five Arabidopsis *ACLL* genes showing wound induction, up to 14-fold 1h after wounding, and the poplar homologues *PoptrACLL4/5* were induced by up to 5-fold 2h after wounding. *ArathACLL4*, which has been shown to be an OPDA-CoA ligase (Koo *et al.*, 2006) and showed strong wound induction of the *ArathACLL4*::*GUS* reporter gene, showed no wound induction by real-time PCR. A possible reason for this may be the time

points chosen for harvesting the tissue after wounding, which could have missed the 'window" of transient up-regulated expression of this gene. *ArathACLL1*, for which no developmental expression was detected in leaves by reverse transcription PCR (Figure 3.4), showed only weakly detectable expression on the basis of multiple promoter-GUS lines (Figure 3.5A), and no expression above background levels was detectable by reverse transcription PCR after wounding (data not shown).

Finally, in clade E, *ArathACLL9* expression was down-regulated after 1h and expression stayed below the original levels even at 24h. Of the 3 closest poplar homologues, *PoptrACLL12* showed a similar expression pattern, whereas *PoptrACLL10* and *PoptrACLL11* showed little or no change in expression in response to wounding. The enzyme encoded by *ArathACLL9* has been shown to have activity *in vitro* with octodecanoid precursors in JA biosynthesis, suggesting that it may be involved in its biosynthesis in the peroxisome (Schneider *et al.*, 2005). However, the lack of wound-inducible expression suggests that *ArathACLL9* and its closest poplar homologues are not involved in stress-induced octadecanoid metabolism.

Overall, the results from promoter-GUS expression and real-time PCR showed consistent wound induced up-regulation for both the poplar gene members and certain Arabidopsis members of clade D, whereas genes in other clades showed little or no wound responsiveness. This suggests, as shown for *ArathACLL4* (*OPDA::CoA ligase*; Koo *et al.*, 2006) that members of clade D are likely to have functions in stress response pathways localized in the peroxisome.

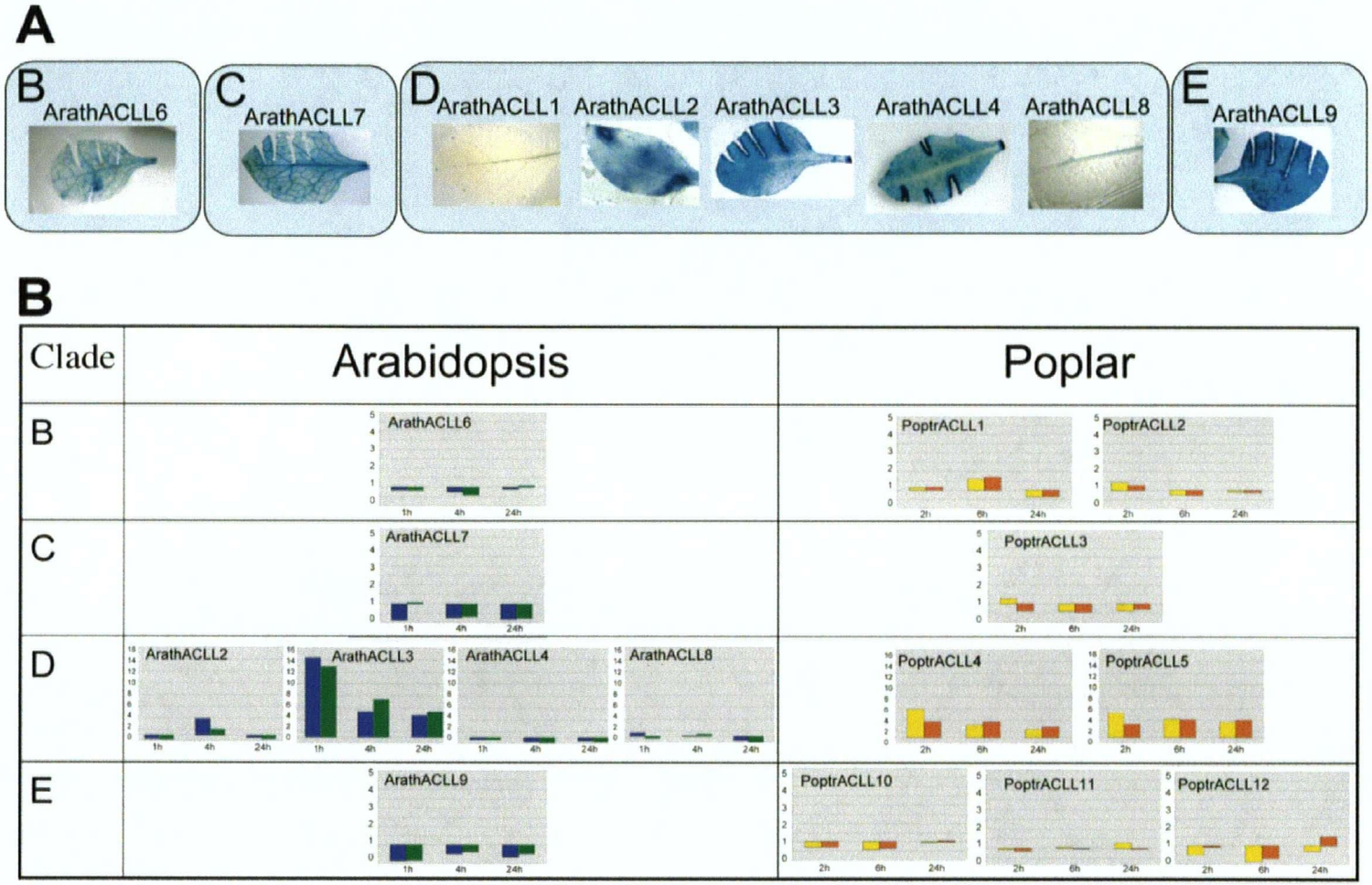


Figure 3.5: Effect of wound stress on peroxisomal *ACLLs*. (**A**) Histochemical GUS staining in transgenic Arabidopsis plants expressing beta-glucoronidase (*GUS*) gene driven by the corresponding *ACLL* promoter. (**B**) Real-time PCR data of time course of wound response in Arabidopsis and poplar. Y axis represents fold change relative to unwounded control.

3.2.7 *In silico* co-expression analysis of Arabidopsis *ACLL* genes

With the increasing amount of information becoming available for global expression data in Arabidopsis, public databases have been successfully used to identify co-expressed genes that could be participating in the same biological process and/or biochemical pathway (Ehlting *et al.*, 2006; Persson *et al.*, 2005). Therefore, in the effort to gain insights into the possible functions of peroxisomal ACLLs, I performed an *in silico* co-expression analysis using public *Arabidopsis* microarray datasets. Using the PRIMe Correlated Gene Search tool, I queried all microarray experiments in all datasets available from the RIKEN Institute (http://prime.psc.riken.jp). The top 100 most highly co-regulated genes in the dataset, showing highest Pearson co-expression coefficient values, are listed in Appendix 1. When possible, a graphic network of interactions was generated using the Pajek program (V. Batagelj *et al.*, 2003) to facilitate interpretation of results (Figure 3.6).

Out of all eight peroxisomal *ACLLs* analyzed, only co-expression data for *ArathACLL3* and *ArathACLL4*, both in clade D, provided extractible information that could be associated with a biological function in plants. *ArathACLL3* demonstrated extremely high coefficient values, of above 0.9, with all 100 co-expressed genes. Around 25% of the genes in this list were associated with lipid metabolism, and about 10% were correlated with seed germination. *ArathACLL3* was directly connected to a gibberellin oxidase in the network of co-expressed genes (Figure 3.6). These results suggest that ArathACLL3 could have a function in lipid metabolism related to seed development or germination. However, due to the extremely high values of co-expression between all genes on this

list, it is quite difficult to distinguish a biochemical pathway that could require this CoA ligase.

For *ArathACLL4*, the 100 most highly co-expressed genes had coefficients of co-expression varying from 0.64 and 0.86. Among these genes, I identified a network of co-regulated genes that participate in the JA pathway, including those encoding lipoxygenases (LOX), allene oxide synthase (AOS), allene oxide cyclase (AOC) and OPDA reductase (OPR3). In addition, other stress related genes were co-expressed with *ArathACLL4* such as the transcription factor WRKY40, which has been shown to be up-regulated after infection with *Pseudomonas syringae* or treatment with salicylic acid (Dong *et al.*, 2003), and *RLPK3*, shown to be activated by oxidative stress and pathogen attack (Czernic *et al.*, 1999). This result suggested that *ArathACLL4* participates in defense response. More specifically, ArathACLL4 could be involved in the JA biosynthetic pathway. This hypothesis was independently confirmed by Koo *et al.* 2006, who demonstrated *in vitro* biochemical activity of ArathACLL4 with OPDA and OPC8 as substrates, both components of the JA pathway, and localization of this enzyme in the peroxisome.

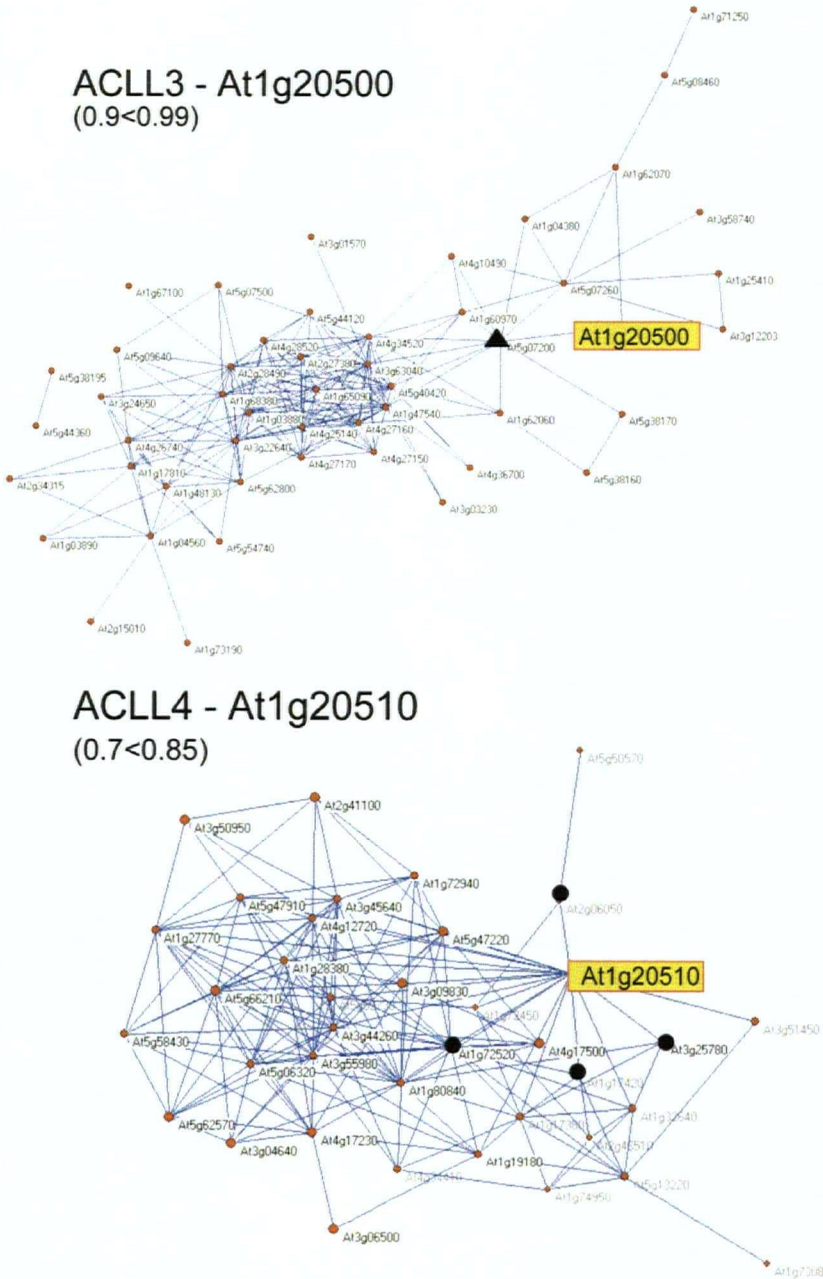


Figure 3.6: Pajek co-expression networks generated from PRIMe Correlated Gene Search tool data (http://prime.psc.riken.jp). (**A**) Co-expression network of ArathACLL3. **0.99** At5g07200 gibberellin 20-oxidase (▲) was the most highly co-expressed gene with ACLL3. Tissue and Development dataset (237 data). (**B**) Co-expression network of ArathACLL4. Co-expressed JA synthesis genes (●): **0.851** At3g25780 allene oxide cyclase (AOC), **0.847** At2g06050 12-oxophytodienoate reductase, (OPR3), **0.831** At1g72520 lipoxygenase, **0.830** At1g17420 lipoxygenase are among the most highly co-expressed with ACLL4. Stress treatment dataset (298 data). Pearson coefficients are highlighted in bold.

Interestingly, no gene was co-expressed with *ArathACLL7* in clade C. Also, while more than 45 genes were co-expressed with *ArathACLL2*, *ArathACLL8* and *ArathACLL9* individually, at high co-expression coefficients (starting above 0.6), no network of co-expressed genes could be generated for these genes. This result could indicate that co-expressed genes in these sets are more highly co-expressed among each other than with the corresponding *ACLL*s, in which case the co-expression would be only circumstantial, and without biological meaning. Genes co-expressed with *ArathACLL1* and *ArathACLL6* generated networks with scattered connections, which did not provide any clues regarding biological function. However, it is worth noting that *ArathACLL1* was most highly co-expressed (Pearson coefficient = 0.8) with the gene encoding 3-hydroxy-3-methylglutaryl-CoA reductase 2 (HMGR2), which catalyzes the synthesis of mevalonate, a rate-limiting step in the mevalonic acid pathway of isoprenoid biosynthesis (Enjuto *et al.*, 1995).

In conclusion, co-expression analyses applied to individual *Arabidopsis ACLL* genes provided limited functional information with the exception of *ArathACLL4*, now known to encode an OPDA/OPC8-CoA ligase as predicted based on co-expression analysis. However, the data presented here on *ACLL1* and *ACLL3* may be useful for generating hypotheses once more is known about these ACLLs.

3.2.8 Poplar genes activated by additional stress treatments

In this study I demonstrated that among the poplar *ACLL* genes in clade D, *PoptrACLL4* and *PoptrACLL5*, encode highly similar proteins most closely related to *ArathACLL4*,

which encodes an ODPA/OPC8-CoA ligase (Koo *et al.*, 2006). Therefore it was possible to speculate that both poplar genes, which share 90% nucleic acid identity, and are up-regulated following wounding, could likewise encode ODPA/OPC8-CoA ligases.

As one test of this hypothesis, and to further test the stress responsiveness of the poplar *ACLL* genes analyzed in Figure 3.5, I measured the expression of poplar *ACLL* genes by quantitative real-time reverse transcription PCR, using RNA isolated after treatment of poplar trees with a battery of additional stressors: herbivory by the forest tent caterpillar (*Malacosoma disstria*), simulated herbivory (SH; wound + insect regurgitant) and exposure to MeJA, a volatile derivative of JA. These data are summarized in Figure 3.7.

The results of this analysis showed differences in the responses of poplar *ACLL* genes to these stresses. In clade B, *PoptrACLL2* showed no stress responsiveness, whereas *PoptrACLL1* expression was strongly up-regulated by SH and herbivory by 6h after treatment onset, consistent with wound activation of this gene (Figure 3.5B). In a separate microarray expression profiling experiment, *ArathACLL6*, the Arabidopsis homologue in clade B, expression was not activated by diamondback moth herbivory (J. Ehlting and J. Bohlmann, personal communication). These data suggest that *ArathACLL6* and *PoptrACLL2* do not function in wound or herbivory related pathways, but that *PoptrACLL1* may have gained this function after duplication of the gene in *Populus*. Interestingly, *PoptrACLL3* (in clade C) was strongly up regulated by SH after 2h, but not by other treatments tested. For the three poplar homologues in clade E that were tested, these stress treatments either had no effect or resulted in down regulation of gene

expression. *PoptrACLL12* expression was the most down regulated, with levels dropping below half or original values after 1h and staying below the original levels even at 24h. No consistent change in expression could be observed for *PoptrACLL10* and *PoptrACLL11*.

The enzyme encoded by the Arabidopsis homologue in clade E, *ArathACL9,* has been shown to have activity with precursors in JA biosynthesis, suggesting that it may be involved in JA biosynthesis in the peroxisome (Schneider *et al.*, 2005). However, the lack of wound (Figure 3.5) and herbivory (J. Ehlting and J. Bohlmann, personal communication) activation of *ArathACLL9*, and lack of wound, SH, herbivory, and MeJA activation of the most closely related poplar *ACLL* genes in clade E, does not support a role for these enzymes in the stress induced synthesis of JA.

Finally, in clade D, which contains poplar and Arabidopsis genes responsive to wound stress, *PoptrACLL4* and *PoptrACLL5* were remarkably up-regulated by herbivory, SH, and MeJA, with the latter treatment leading to a 10-20 fold increase in expression. In the separate microarray expression profiling experiment mentioned above, expression of Arabidopsis homologues *ArathACLL4* and *ArathACLL8* were activated by diamondback moth herbivory (J. Ehlting and J. Bohlmann, personal communication). These data are consistent with a role for certain Arabidopsis and poplar ACLL clade D enzymes in biochemical pathway(s) related to defense against wounding and/or herbivory, with *PotrACLL4* and *PoptrACLL5* being likely orthologs of the *ArathACLL4/ODPA/OPC8-*

CoA ligase gene, based on their phylogenetic relationship to *ArathACLL4* and strong wound, herbivory, and MeJA induced expression.

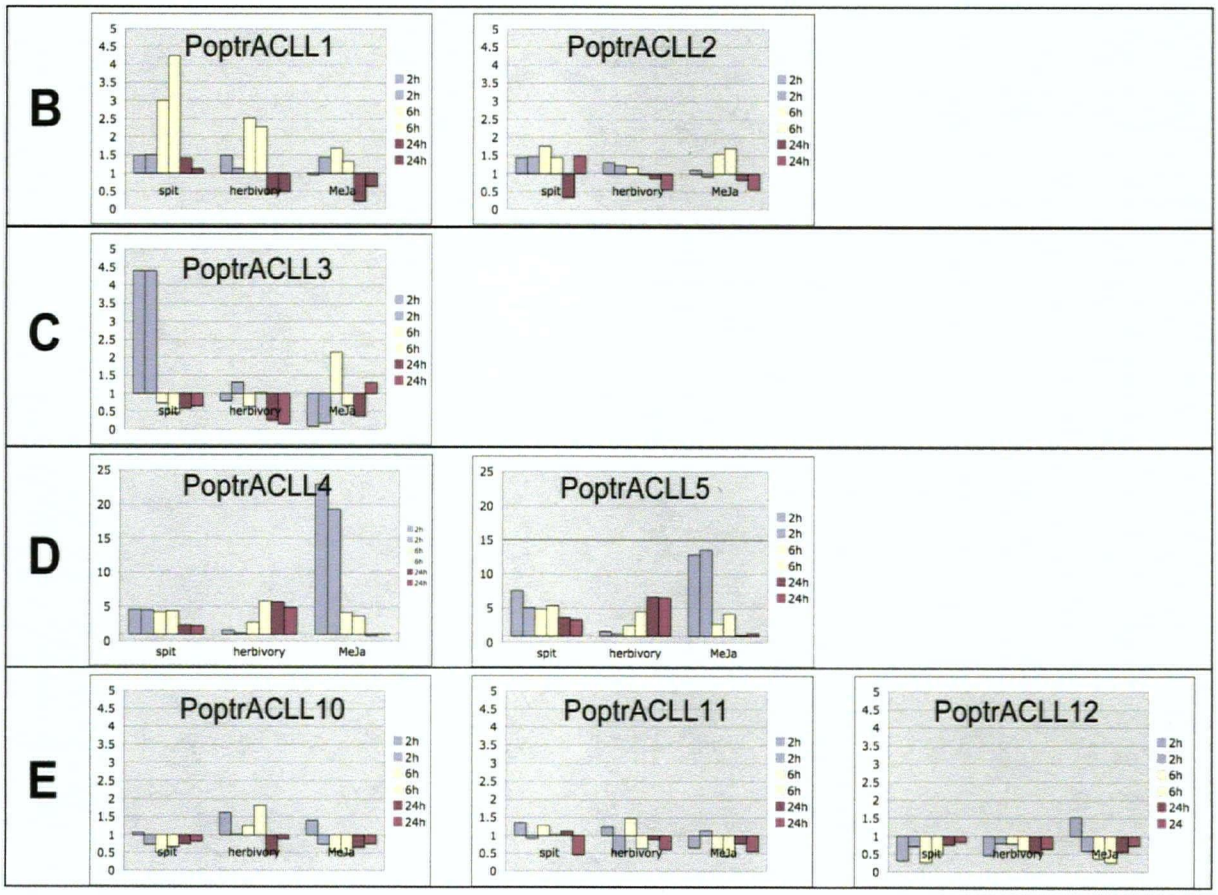


Figure 3.7: Real-time PCR data showed expression of poplar *ACLLs* after simulated herbivory (SH or "spit"=mechanical wound + insect regurgitant), herbivory by the forest tent caterpillar (*Malacosoma disstria*), and exposure to MeJA. Results are given in two replicas for each treatment (same color bars). Y axis represent fold change in gene expression compared to expression at zero time point.

3.2.9 Sub-cellular localization of PoptrACLL5

In order to determine if the poplar homologues of *ArathACLL4* in clade D actually encode peroxisomally localized enzymes, consistent with their postulated roles as OPDA:CoA ligases, I tested the subcellular localization of one of the two highly similar homologues PoptrACLL5, using ArathACLL4, experimentally shown to be localized to peroxisomes (Koo *et al.*, 2006) as a positive control. I generated N-terminally tagged GFP-PoptrACLL5 and GFP-ArathACLL4 protein fusions. Constructs were expressed under the control of the 35S promoter in transgenic tobacco plants generated by tissue culture (in collaboration with K. Turner, BC Institute of Technology). Eight transgenic plants derived from independent calli were obtained from *Agrobacterium*-infected tobacco leaf disks, and GFP signal localization was analyzed using confocal microscopy. Figure 3.8A shows the results of this analysis. The clearest GFP signal was found in guard cells of the leaf epidermis in the transgenic lines. In the negative controls of transgenic lines not expressing GFP and wild type untransformed tobacco plants, only chlorophyll-derived autofluorescence of chloroplasts was observed, indicating that the fluorescence observed attributed to GFP is not naturally occurring in tobacco cells. However, in GFP:ArathACLL4 transgenic lines, punctate peroxisome-like GFP fluorescent sub-cellular structures were observed, similar to those described by Koo *et al.* (2006) and consistent with the deduced peroxisomal localization of ArathACLL4 (Figure 3.8A; Koo *et al.*, 2006). GFP-PoptrACLL5 expressing tobacco lines (Figure 3.8A) showed GFP fluorescence patterns in the guard cells that were indistinguishable from those in the GFP:ArathACLL4 positive control. At higher magnification (Figure 3.8B), the round fluorescent structures in a GFP-PoptrACLL5 expressing tobacco line clearly

resembled the peroxisomes described in other studies (Koo *et al.*, 2006; Schneider *et al.*, 2005), and were around 1μm in diameter, consistent with the reported size of this organelle (Nyathi and Baker, 2006).

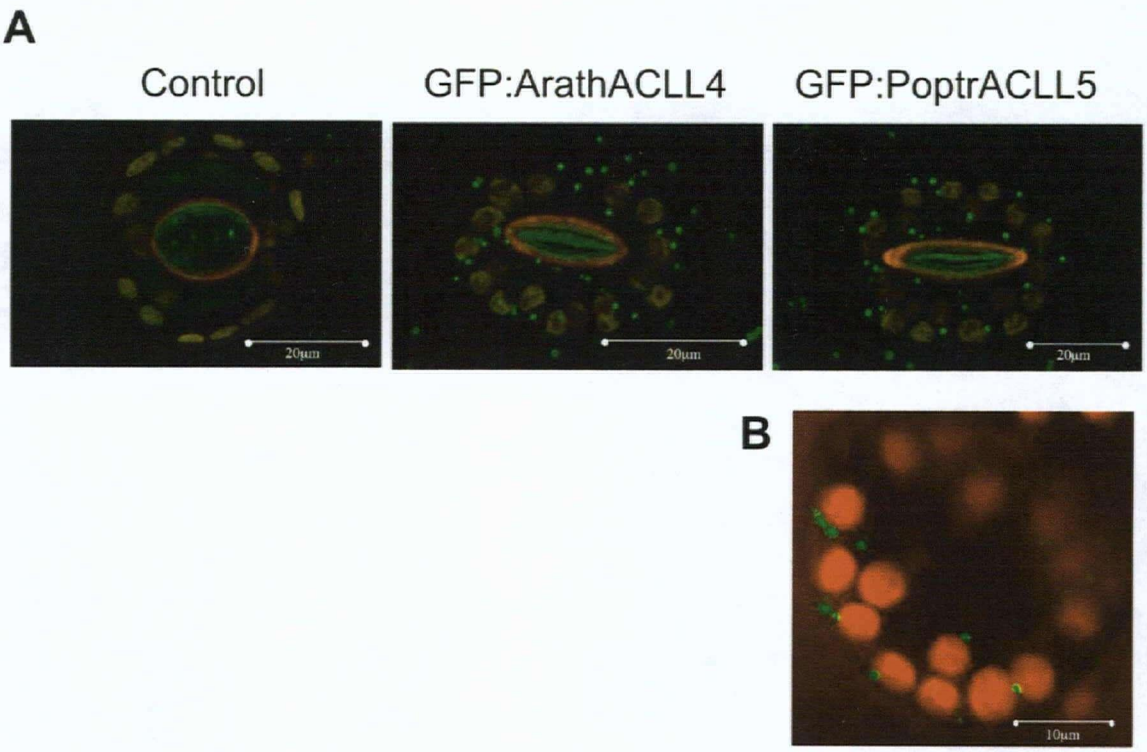


Figure 3.8: Confocal microscopy image showing sub-cellular localization of ArathACLL4 (OPDA/OPC8-CoA ligase) and the poplar homologue PoptrACLL5 in guard cells of transgenic tobacco lines. Yellow signal derives from chlorophyll autofluorescence (**A**) GFP signal (green) localized to similar structures consistent with a peroxisomal localization for both Arabidopsis and poplar clade D enzymes. (**B**) Detailed view of GFP:PoptrACLL5 showing close proximity of peroxisomes (green) with chloroplasts (red).

3.3 Discussion

3.3.1 *4CL* and *ACLL* gene evolution

Before this study, little was known about the group of adenylate-forming enzymes most closely related to plant 4CL enzymes (i.e, ACLL enzymes). Important initial contributions to the phylogeny of ACLLs (Cukovic *et al.*, 2001; Shockey *et al.*, 2003) demonstrating their close relationship to true 4CLs and other enzymes in fatty acid metabolism support my phylogenetic results. Also, recent discoveries regarding biochemical functions of two such enzymes (Koo *et al.*, 2006; Schneider *et al.*, 2005), lend further support for stress induction of gene expression and co-expression analyses presented here.

In this part of my thesis, using genome sequence information from land plants and various microorganisms, I show that *ACLL* genes, like true *4CL* genes, are a land plant-specific gene super family (Figure 3.1). Furthermore, genome sequence information from Arabidopsis, poplar, and rice shows that the five clades of *ACLL* genes are conserved in monocot and dicot lineages, with at least one representative in each of the 3 fully sequenced angiosperm genomes (Figure 3.2), demonstrating the early origin of such genes during angiosperm evolution. The two diverging branches, separating 4CLs and clade A *ACLLs* from the peroxisomal *ACLLs*, evident in Figure 3.2, suggests an early common ancestor of both *4CLs* and clade A *ACLLs*, and a common ancestor for peroxisomal ACLLs. Therefore, the phylogenetic reconstruction allowed me to infer that there were at least two *ACLL* ancestral genes in land plants, which gave rise to the five *ACLL* clades present today.

Given the ancient origins of the enzymes encoded by genes in these clades it is quite likely that ACLL enzymes from different clades perform similar but distinct functions in current species. In addition, given that the adenylate-forming enzymes most closely related to the ACLLs that are not unique to plants are targeted to the peroxisome, it is reasonable to speculate that the ancestral enzymes of both the cytosolic and peroxisomal 4CL and ACLL clades were recruited from peroxisomal enzymes. Since the major role of peroxisomes is in lipid metabolism, and CoA ligases are widely used in β-oxidation of these molecules, it is tempting to speculate that ACLLs and 4CLs were recruited from fatty acid metabolism early in land plant evolution, to perform their current functions. It may be relatively easy for genes to lose the peroxisomal targeting signal in their encoded proteins. An evidence for this is the fact that two poplar ACLL proteins (PoptrACLL6 and PoptrACLL12), with close homologues in the peroxisomal clade E (Figure 3.2), have lost their C-terminal peroxisomal targeting sequences and are presumably localized in the cytosol. Additional genome sequence information from basal land plant species will help to more accurately infer the evolutionary history of *ACLL* genes.

In this context, it is interesting that two *Physcomitrella* ACLL sequences were identified, which could represent genes ancestral to the current suite of 4CL and ACLL genes in angiosperms. Completion of the *Physcomitrella* genome sequence should reveal whether additional *Physcomitrella* ACLLs exist. It is possible that analysis of a loss of function mutation in *Physcomitrella* genes will provide hints regarding the biochemical functions of ACLLs in the other plant species, in addition to indicating the possible biochemical

function of this apparently ancestral protein. Such work could shed light both on the origin of the ACLL family of proteins, and shed light on how enzymes are recruited to novel biochemical pathways.

3.3.2 *ACLL* gene family structure and expression patterns

Despite the larger poplar genome, in which many genes are duplicated relative to Arabidopsis as a result of the salicoid whole genome duplication event (Tuskan *et al.*, 2006), my results demonstrate that there is no relationship between species and the number of genes in a given clade. Overall, my results indicate that certain *ACLL* gene families, as defined by clades A-E, have undergone differential expansion in individual species over the course of lineage-specific genome evolution. Presumably, duplicated genes were retained due to selective pressures for elaboration of biochemical pathways requiring ACLL activity, which may vary according to the diverse life histories of Arabidopsis, poplar, and rice. One example of species-specific gene family diversification is the *4CL* gene family itself in Arabidopsis and poplar. While both poplar and Arabidopsis have only one class I *4CL*, involved in flavonoid and soluble phenolic biosynthesis, class II genes evolved in a lineage-specific manner. The three Arabidopsis class II *4CLs* seem to have originated by a combination of segmental duplication involving the chromosomal region where *Arath4CL4* resides, followed by tandem duplications giving origin to *Arath4CL1* and *Arath4CL2* (Hamberger and Hahlbrock, 2004). In poplar, the four class II *4CL* loci are unlinked and scattered over four different chromosomes, indicating a different mechanism of gene duplication (Hamberger *et al.*, submitted). In Arabidopsis, the three class II genes have different expression patterns and

even encode enzymes with specialized properties, as mentioned in Chapter 1. Similarly to *4CLs*, specific CoA ligation requirements could be driving *ACLL* gene family diversification in a lineage-specific manner.

Interestingly, the conservation of sequence identity between Arabidopsis and poplar *ACLL* genes within individual clades was not uniform, suggesting more rapid evolution of genes in certain clades. My analysis revealed that for clades A, B and D, Arabidopsis and poplar species share ACLL homologues that have been strongly conserved since divergence of the poplar and Arabidopsis lineages. This indicates that these enzymes may perform key roles in plant metabolism, conserved in these two dicot lineages. For example, in clade A poplar and Arabidopsis ACLL amino acid sequence identity (PoptrACLL13 and ArathACLL5) is strongly conserved. Amino acid sequence conservation, especially in the putative ACLL substrate binding domains inferred from adenylate-forming enzyme protein structure (Stuible and Kombrink, 2001), is a good indication of possible conservation of the substrate utilization. Thus, the highly conserved homologous genes belonging to clades such as clade A have a high likelihood of being orthologues, i.e., having the same biological/biochemical function in different organisms. On the other hand, the relatively divergent poplar and Arabidopsis sequences and the expansion of, for example, clade E sequences in poplar, suggest species-specific retention of duplicated genes and their neofunctionalization. This could reflect evolution of poplar-specific metabolic pathways involving these ACLL enzymes. In clade C, the single copy poplar and Arabidopsis proteins appear to have less constraint on sequence divergence, perhaps due to the nature of the substrates used, or partial redundancy of

clade C enzyme function with other ACLLs, allowing potential acquisition of new species-specific functions.

Another indicator of functional conservation across species is the conservation of gene expression patterns. Genes derived from a common ancestor that perform the same function in related organisms might retain similar expression patterns, especially if gene duplication has not occurred, leading to subfunctionalization of gene expression. A striking example of conservation of gene expression patterns was observed for clade A *ACLL* genes. While the expression pattern of the rice representative is unknown, both the Arabidopsis and poplar *ACLL* genes in this clade have flower-preferred expression patterns, according to our gene expression data. This is supported by Arabidopsis microarray data from Douglas and Ehlting (2005), and microarray expression data mined from public databases (data not shown), which indicates that *ArathACLL5* has a strongly anther-preferred expression pattern. Similarly *PoptrACLL13* is preferentially expressed in male flowers, with no detectable expression in female flowers. The relatively high level of sequence conservation of these genes, and their shared expression patterns suggest functional conservation of clade A enzymes in Arabidopsis and poplar, and that they play metabolic roles associated with anther development.

Clade B is interesting since there are two *ACLL* copies in poplar (*PoptrACLL1/2*), encoding enzymes with same identity values when compared to the Arabidopsis protein homologue (ArathACLL6). My results showed that one of the poplar homologues (*PoptrACLL2*) had an expression pattern most similar to *ArathACLL6* (expression

throughout all organs and tissue types, but most highly expressed in mature leaves; Figure 3.4). Combined with their high level of amino acid sequence conservation, this conservation of developmental expression suggests that these poplar and Arabidopsis genes could be functional othologues. However, the two poplar homologues in these clades appear to have undergone subfuntionalization and neofunctionalization, suggesting an expanded function of poplar enzymes in this clade. Expression analysis showed that *PoptrACLL1* and *PoptrACLL2* have highly complementary patterns for expression in all organ and tissue types analyzed (i.e. in organs and tissues where *PoptrACLL1* is low, *PoptrACLL2* is high, and vice versa; Figure 3.4). This appears to be a classical example of subfunctionalization, where duplicated genes acquire specialized expression patterns, which when combined are equal to the expression pattern before duplication (Duarte *et al.*, 2006). Analysis of stress-induced expression of clade B genes suggests as well neofunctionalization of one member of the duplicated poplar gene pair. While neither *ArathACLL6* nor *PoptrACLL2* is stress inducible (Figures 3.5 and 3.7; J. Ehlting and J. Bohlmann, personal communication), *PoptrACLL1* is clearly induced by wounding, simulated herbivory and herbivory (Figures 3.5 and 3.7). This suggests that, whatever its biochemical function, the enzyme encoded by the duplicated *PoptrACLL1* gene has been recruited to serve in a defense-related metabolic pathway, in addition to its developmental function. This new function is apparently novel in the poplar lineage.

In clade C there is a single copy gene for Arabidopsis and poplar. Interestingly, although identity values for these two genes are the lowest among the ACLL clades (58%/65%), expression patterns seem to be conserved with highest expression in leaves, stem and

flowers. One hypothesis would be that, despite their similar expression patterns, there has been functional divergence of the poplar and Arabidopsis (and, possibly, rice) ACLL enzymes in this clade, possibly in the type of substrate accepted or preferred. Alternatively, enzymes in this clade may still perform conserved functions despite the low identity. Further evidence for functional divergence of the poplar gene comes from its strong induction by simulated herbivory, although expression of the Arabidopsis gene has not been tested under similar conditions.

Clade D is interesting in that four of the five Arabidopsis *ACLL* genes, with exception of *ArathACLL8*, originated via tandem duplications on chromosome 1 (Figure 3.3). One member of this group, *ArathACLL4*, has been shown to encode an OPDA:CoA ligase, required for JA biosynthesis in the peroxisome (Koo *et al.*, 2006). This function is consistent with the wound inducible expression of the *ArathACLL4*::*GUS* fusion gene (Figure 3.5A), and its expression pattern assessed in public microarray databases (data not shown). However, it is striking that these duplicated genes share largely non-overlapping developmental expression patterns (Figure 3.4), as well as variable responses to wounding, herbivory, and MeJA treatment (Figure 3.5). Thus, diversification of functions of the tandemly duplicated members of the Arabidopsis genes in this clade could have occurred, and may be related to their differential expression patterns. Since not all Arabidopsis *ACLL* genes in clade D were inducible by wounding, it may be that they perform functions other than in JA biosynthesis, or participate in developmentally regulated JA biosynthesis. This is supported by the fact that identity levels among Arabidopsis ACLLs in this clade are above 60%. As mentioned previously, my findings

from co-expression data are not consistent with a role for *ACLL3* in JA biosynthesis, even though this gene is strongly wound-induced (Figure 3.5). However, many of the genes highly co-expressed with *ACLL3* are related to seed development, which is a developmental process also known to be regulated by JA. Given the known biochemical function of ArathACLL4 as an OPDA/OPC8-CoA ligase, ACLLs in this clade are good candidates for participating in the octadecanoid pathway. Additional experiments, such as biochemical characterization of heterologously expressed enzymes, will be important to address this question.

My data indicate that the apparent subfunctionalization and possible neofunctionalizaton of duplicated clade D Arabidopsis *ACLL* genes has not occurred in poplar and rice, in which clade D genes have not expanded in number. Poplar contains two highly similar genes *PoptrACLL4* and *PoptrACLL5*, located on different linkage groups and a single rice gene (Figure 3.2). The close phylogenetic relationship between the *ArathACLL4/OPDA CoA ligase* gene and *PoptrACLL4* and *PoptrACLL5* (Figure 3.2), their high amino acid similarities (72%/75% amino acid identity), coupled with the wound, herbivory, and MeJA inducible expression of the poplar genes (Figures 3.5 and 3.7), strongly suggest that the poplar ACLL enzymes encoded by *PoptrACLL4* and *PoptrACLL5* in this clade are peroxisomally localized OPDA:CoA ligases, involved in acyl chain shortening step required for JA biosynthesis, indeed performing the same functions as *ArathACLL4*. Given the lack of diversification of poplar and rice genes in this clade relative to gene family expansion in Arabidopsis, it is possible that diverse

octadecanoid signaling pathways are more prevalent in Arabidopsis than in these two species.

Clade E represents a contrasting case to clade D, in that the numbers of poplar and rice genes have expanded to 7 and 5, respectively, relative to a single Arabidopsis gene, and two of the poplar genes have undergone apparent neofunctionalization by loss of peroxisomal targeting sequences (Figure 3.2). The amino acid identity values between enzymes in this clade are quite low (Table 3.2) and there was no discernable similarity in developmental expression patterns between the Arabidopsis *ACLL* (*ArathACLL9*) and its three most closely related poplar homologues (Figure 3.4). While analysis of ArathACLL9 enzyme activity showed *in vitro* activity with octadecanoid precursors in the JA biosynthetic pathway, leading to the suggestion that it may be involved in JA biosynthesis (Schneider *et al.*, 2005), its lack of wound or herbivory inducible expression Figures 3.5 and 3.7), coupled with the lack of co-expressed genes associated with JA biosynthesis (this study; Koo *et al.*, 2006), suggests that it could be involved in some other aspect of fatty acid metabolism (Koo *et al.*, 2006). Taken together, the data suggest that clade E contains genes encoding enzymes that are not strongly conserved between species, and may perform specialized functions specific to certain lineages.

3.3.3 Summary and combining data to make functional inferences

The first clue available for inferring any kind of function for ACLLs came from the conservation of sequence motifs that define the adenylate-forming enzyme superfamily in addition to the sequence similarity with true 4CLs (Cukovic *et al.*, 2001; Shockey *et al.*,

2003) suggesting that ACLLs are CoA ligases. My phylogenetic analyses showed that ACLLs are a plant-specific group, thus are unlikely to be performing ubiquitous functions in general metabolism. Additional clues for functions came from comparative genomic analyses presented here, focusing on poplar and Arabidopsis, but also including ACLLs from other plant species (discussed in Chapter 4 for clade A). The presence of the peroxisomal target sequence PTS1, indicating a putative localization in that organelle and therefore a function in plant–specific processes of that organelle, provided another significant functional clue. Combining promoter activity assays together with analysis of changes in gene expression in response to stress treatments revealed a set of ACLLs, mostly in clade D, which have apparent functions in biochemical responses to wounding and insect herbivory. Using a gene expression data mining approach, I identified the JA biochemical pathway as one in which the ArathACLL4 enzyme likely participates as an OPDA:CoA ligase. This result was independently confirmed by conclusive biochemical and genetic data by Koo *et al.* (2006).

Other stress responsive *ACLL* genes in clade D did not show co-expression with enzymes in the JA biosynthetic pathway in the public microarray data. However, co-expression analysis may be less robust in identifying genes in common pathways or processes when the genes have broad developmental expression patterns and are not highly inducible. Therefore, it could be that there is functional redundancy among the Arabidopsis *ACLL* genes in this group, as has been suggested by Koo *et al.* (2006). They found that JA accumulation is only partially compromised in an *ArathACLL4* loss of function mutant, suggesting that other related genes encode enzymes with the same function. The role of

other clade D ACLL, if they are OPDA:CoA ligases, could be primarily in constitutive, developmental JA biosynthesis, but this activity could be sufficient to provide sufficient flux through the pathway upon stress activation of JA biosynthesis such that stress-activated JA accumulation still occurs. It has been shown that OPDA is constitutively present in untreated wild-type leaves (Stenzel *et al.*, 2003), so it is possible that esterified OPDA is stored in plants for rapid response to wound and herbivory attacks. It is important to note that *ArathACLL4* is constitutively expressed in flowers, and that other *ACLL*s in clade D have highly complementary constitutive expression levels in all tissues analyzed. JA is also a signal molecule for various developmental processes, including root growth (Staswick *et al.*, 1992), flower and seed development (Feys *et al.*, 1994; Li *et al.*, 2004) and apical meristem development (Cenzano *et al.*, 2003). So, if one or more *ACLL*s in clade D also encode enzymes in the JA biosynthetic pathway in the tissues where they are expressed, it is possibile that they may play a role in developmental metabolism regulated by JA. With this in mind, I checked public expression data for clade D *ACLL* expression after methyl jasmonate (MeJA) treatment, using the eFP browser (http://bbc.botany.utoronto.ca/efp/cgi-bin/efpWeb.cgi). In the available experiments, done with 7day seedlings treated with 10μM MeJA and harvested between 30mins and 3hours after exposure, *ACLL1* was not inducible by MeJA or any other hormone tested. However, *ACLL2*, *ACLL8* and, as expected *ACLL4*, showed up regulation, consistent with a role of additional clade D *ACLL*s in JA related pathways. *ACLL3* does not map to a probe set on the Affymetrix ATH1 GeneChip dataset used by the eFP browser, therefore no information could be obtained for this gene using this tool.

The comparative genomics and expression profiling approaches allowed me to predict with some confidence that the two similar poplar clade D genes, *PotrACLL4* and *PoptrACLL5*, encode enzymes that carry out the same function as ArathACLL4 as an OPDA:CoA ligase. The poplar homologues in clade D were highly up-regulated by a variety of stress treatments, in particular after induction with MeJA, a derivative signaling molecule of JA. It has been shown that JA has the property of regulating its own synthesis via a positive feedback loop, which is dependent on JA synthesis and JA signaling (Bonaventure *et al.*, 2007; Jensen *et al.*, 2002). Therefore, up-regulation upon contact with MeJA is a predicted response for genes involved in the JA pathway, in addition to genes involved in downstream events. Other evidence supporting the biochemical functions of PoptrACLL4 and PoptrACLL5 as OPDA:CoA ligases include their high amino acid similarity to ArathACLL4 (Table 3.2) and the experimentally demonstrated sub-cellular localization of PoptrACLL5 in the peroxisome (Figure 3.8). Biochemical characterization of recombinant poplar enzymes and/or the generation and phenotypic characterization of RNAi-mediated loss of *PoptrACLL4/PoptrACLL5* function poplar plants would be necessary to test this hypothesis. Unlike Arabidopsis, in which there appear to be partially redundant genes encoding OPDA:CoA ligase in clade D in addition to *ArathACLL4*, no extensive redundancy appears to be present in poplar. Thus, simultaneous knockdown of *PoptrACLL4* and *PoptrACLL5* expression would be predicted to severely impact JA biosynthesis and accumulation, allowing definitive tests to be carried out regarding the role of this signaling molecule in defense and development.

CHAPTER 4 – THE *ARABIDOPSIS THALIANA* FLOWER-SPECIFIC ACYL-COENZYMEA LIGASE GENE *ACLL5* IS CONSERVED IN ANGIOSPERMS AND IS REQUIRED FOR MALE FERTILITY

4.1 Introduction

Anther development, culminating in the formation of mature male gametophytes (microspores, or pollen grains) is a complex process that is central to angiosperm life history (Ma, 2005). Anther development and microsporogenesis have been subjects of intense study and are well documented and characterized in many plants, including models such as tobacco and *Arabidopsis thaliana* (Goldberg *et al.*, 1993; Ma, 2005; Sanders *et al.*, 1999; Scott *et al.*, 2004).

Stages of anther development and microsporogensis are precisely timed and tightly controlled, and are characterized by specific events ranging from initial cell differentiation from the floral meristem to pollen formation, maturation and release during anther dehiscence (Goldberg *et al.*, 1993; Ma, 2005; Sanders *et al.*, 1999; Scott *et al.*, 2004). In tobacco and *Arabidopsis*, anther development has been divided into stages based on anatomical, morphological, cellular and molecular events (Goldberg *et al.*, 1993; Ma, 2005; Sanders *et al.*, 1999). Molecular genetic studies, particularly in *Arabidopsis*, have shed light on many events in anther development and microsporogenesis (Ma, 2005). However, many biochemical and cellular processes specific to anther development and their regulation are still unknown.

One event of fundamental importance during pollen maturation is the deposition of the pollen wall, necessary for pollen protection, dispersal and pollen-stigma recognition. 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, a maternal cell layer that surrounds the inner side of the anther locules, is responsible for the production and secretion of the exine, generally known as a mixture of proteins, lipids and aromatic molecules that comprises the outermost layer of the pollen wall (Ma, 2005; Scott *et al.*, 2004). After synthesis and deposition of the pollen wall, the tapetum cells are degraded via programmed cell death (PCD), and pollen grains continue to develop and mature.

Although the exact composition of the exine and other components of the pollen wall, as well as the regulation of its synthesis and deposition, are not completely understood, it is known that functional tapetum cells are essential for the development of viable pollen grains, presumably due to their crucial role in biosynthesis and secretion of the exine (Vizcay-Barrena and Wilson, 2006; Zhang *et al.*, 2006). The precise chemical composition of the exine has been long debated. The major component of the exine is termed sporopollenin, a complex biopolymer characterized by its extreme stability and resistance to degradation. As a result, there are a limited number of techniques available for exine chemical analysis (Blokker *et al.*, 2006), but the major components of sporopollenin are long-chain fatty acids and poorly characterized phenolic molecules coupled by ether linkages (Scott *et al.*, 2004). Genetic approaches that identify enzymes

and biochemical pathways required for exine production promise to aid in the elucidatation of its composition and functions (Ma, 2005).

Several male sterile mutants that have been isolated and characterized in *Arabidopsis* (Ma, 2005; Sanders *et al.*, 1999; Taylor *et al.*, 1998) shed some light on the cell biology and biochemistry of pollen maturation. Obvious examples include mutants defective in meiosis that result in abnormal pollen grains, such as *pollenless3* (Sanders *et al* 1999). Male sterile mutants have also identified genes required for normal tapetum development, demonstrating the intimate relationship between tapetum function and male fertility. The mutant *dysfunctional tapetum1* (*dyt1*) is an example of a postmeiotic male sterile mutant. The *DYT1* gene has been cloned, and exhibits strong tapetum preferred expression. *DYT1* encodes a bHLH transcription factor believed to be required for the proper expression of tapetum genes (Zhang *et al.*, 2006), since loss of *DYT1* function results in reduced expression of tapetum-preferential genes and abnormal pollen development. The *male sterile1* (*ms1*) mutant, which is defective in tapetum programmed cell death and does not produce pollen grains in an otherwise normal appearing mature anther, provides a clear example of the requirement for a functional tapetum for pollen grain development. (Vizcay-Barrena and Wilson, 2006). The *male sterile2* (*ms2*) mutant, defective in sporopollenin deposition, develops microspores that collapse shortly after release from tetrads, showing no signs of pollen wall formation. The MS2 protein accumulates specifically in the tapetum and is suggested to be a long chain fatty acid reductase, possibly involved in biosynthesis of a long-chain aliphatic sporopollenin polymer (Aarts *et al.*, 1997). In the *defective in exine formation* (*dex1*) mutant, like *ms2*, abnormal microspores develop after their release from tetrads. Although sporopollenin is produced,

it does not anchor to the microspores, which eventually collapse. DEX1 is a novel protein of unknown function but may function at the plasmalemma. Accumulation of the DEX1 protein is not restricted to floral buds, but it is found in many other organs in the plant (Paxson-Sowders *et al.*, 2001). Other genes required for exine production and male fertility defined by mutations include *YRE/WAX2/FLP1*, encoding an enzyme that may be involved in wax biosynthesis (Ariizumi *et al.*, 2003; Chen *et al.*, 2003; Kurata *et al.*, 2003), and *NEF1*, which encodes a possible transporter protein (Ariizumi *et al.*, 2004).

While these forward genetic analyses have identified certain biochemical and regulatory events involved in anther and pollen development in Arabidopsis, they provide a far from complete picture of these events. In particular, the biochemical and cellular events involved in tapetum function and exine formation, as well as the biochemistry and functions of the exine and sporopollenin remain poorly defined, despite their importance for microspores and male fertility.

Novel classes of Arabidopsis genes encoding enzymes related to, yet functionally distinct from phenylpropanoid genes have been defined (phenylpropanoid-like genes) (Costa *et al.*, 2003; Ehlting *et al.*, 2005; Raes *et al.*, 2003), many of which are conserved in the fully sequenced genomes of poplar and rice (Hamberger *et al.*, submitted; Tuskan *et al.*, 2006). An example of such a gene superfamily is the group of genes encoding enzymes related to the phenylpropanoid enzyme 4CL (Costa *et al.*, 2003; Cukovic *et al.*, 2001; Ehlting *et al.*, 2005; Shockey *et al.*, 2003). I have characterized this family of genes, which I now designate as *Acyl-CoA ligase-like* (*ACLL*) genes, in genomes of Arabidopsis,

poplar, rice, and other plants (de Azevedo Souza *et al.*, in preparation; see Chapter 3). These studies show that *ACLL* genes, together with *4CL* genes, are a land-plant-specific clade of adenylate-forming enzymes. The *ACLL* clade that contains the Arabidopsis *ACLL5* gene (At1g62940) is more closely related to true *4CL* genes than other *ACLLs* are. Additional analysis using sequence information from the poplar and rice genomes revealed that *ACLL5* is a highly conserved single copy gene in Arabidopsis, poplar and rice, which suggests they originated from a common ancestor present before the divergence of monocot and eudicot lineages (de Azevedo Souza *et al.*, in preparation; see Chapter 3). Furthermore, *ACLL5* and its poplar orthologue are expressed in a strongly flower-preferred manner (de Azevedo Souza *et al.*, in preparation; see Chapter 3; (Douglas and Ehlting, 2005), suggesting a function in flowers.

Here I describe *acll5*, a novel male-sterile mutant of the *ACLL5* gene. Characterization of this mutant suggests that *ACLL5* has a tapetum-specific function, and encodes an enzyme that may be involved the production of an aromatic constituent(s) of the exine which is required for post-meiotic pollen development and male fertility.

4.2 Results

4.2.1 Identification of an *ACLL5* insertion mutant

The flower-preferred expression pattern of *ACLL5*, including evidence that is expressed in the male organ, as well as the conservation of single copy *ACLL5* homologues in the fully sequenced rice and poplar genomes (Chapter 3), suggest that this gene may play an important role in male reproductive organ development in angiosperms. In order to determine the biological function of ACLL5, I identified a potential *acll5* transposon insertion loss of function mutant. Seeds for the line were obtained from NASC (stock code N123936; standard name SM_3.37225). A segregating population derived from the original insertion line was genotyped by PCR as described in Materials and Methods. Within this population, *acll5* homozygous plants were identified. The *ACLL5* insertion segregated as a single Mendelian locus, which I designated *acll5-1,* to my knowledge the first described *ACLL5* mutant allele. The location of the transposon insertion, in the second exon of *ACLL5*, was verified by sequencing the PCR amplification product generated from line SM_3.37225 genomic DNA using a gene-specific primer and a primer specific to the transposon-tag (Figure 4.1A).

I tested *ACLL5* expression in the *acll5* mutant plants, using semi-quantitative and quantitative RT-PCR, with template cDNA derived from both wild type and mutant flowers. Figure 4.1B shows that no *ACLL5* mRNA could be detected in the mutant by semi-quantitative RT-PCR, and quantitative RT-PCR also failed to detect any mRNA (data not shown), indicating that *acll5-1* is a null allele of *ACLL5*.

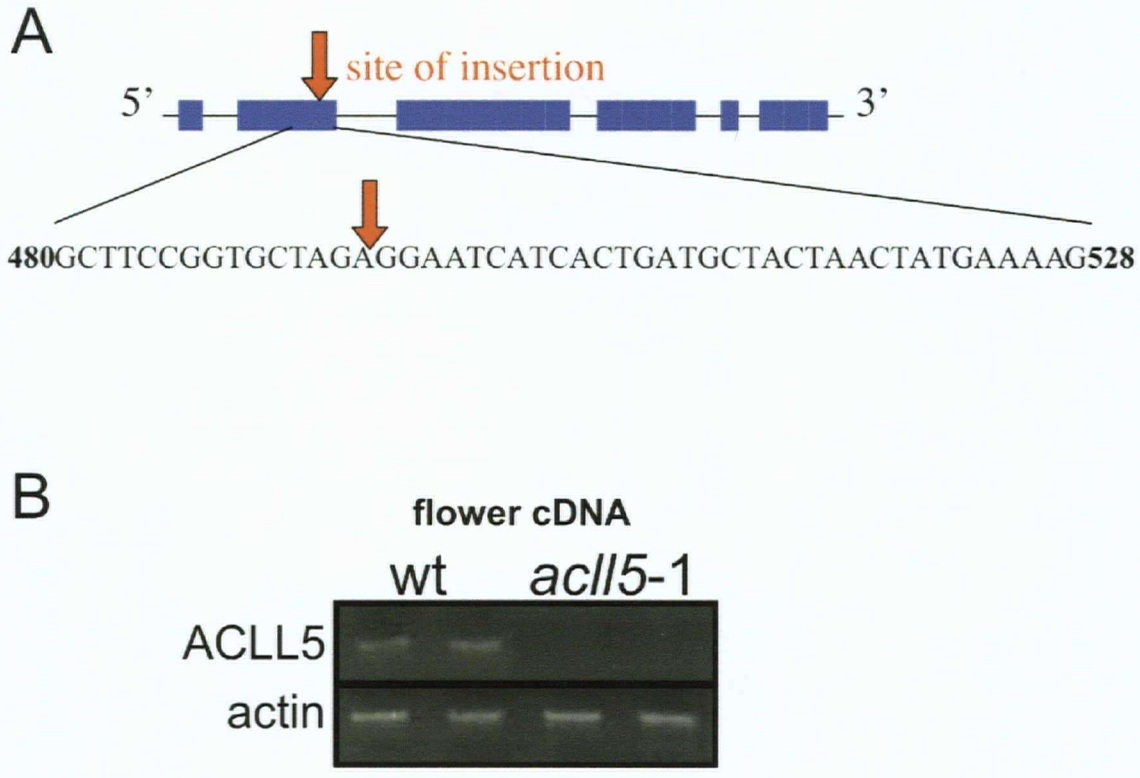


Figure 4.1: (**A**) Schematic representation of the *ACLL5* (At1g62940) gene, showing the location of the transposon insertion in line SM_3.37225 (*acll5-1*). (**B**) *ACLL5* expression in wild-type (wt) and *acll5*-1 homozygote lines. RT-PCR analysis (30 cycles) was carried out on duplicate samples from cDNA prepared from wt or *acll5-1* flowers, using *ACLL5* and actin-specific primers.

4.2.2 *acll5* mutation is correlated with male sterility and absence of pollen grains

Initial phenotypic analysis of the homozygous *acll5* plants suggested that the plants were self-sterile, since siliques failed to mature and produce seeds. Careful examination of the mutant flowers revealed anthers with a darker appearance than wild-type anthers, and that were devoid of pollen grains (Figure 4.2). At a time when wild-type flowers had self pollinated and produced siliques with developing seeds, development of *acll5* mutant flowers culminated in undeveloped siliques and absence of seed production (Figure 4.2). There were no other obvious morphological differences between *acll5* mutant and wild-type plants when grown to maturity (Figure 4.2).

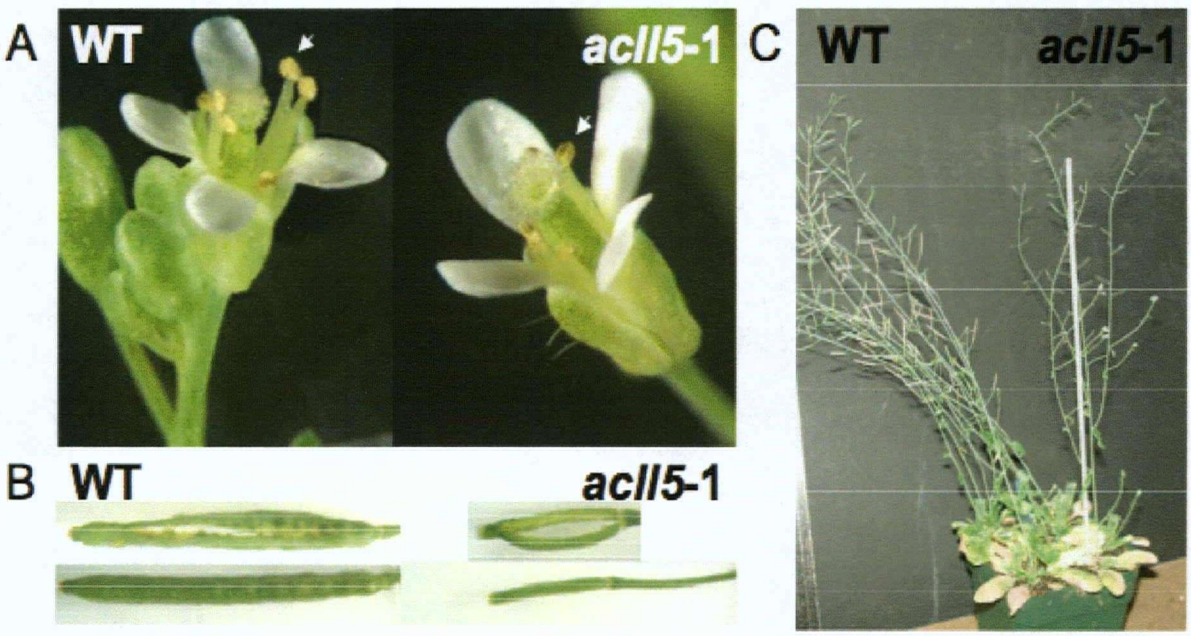


Figure 4.2: Phenotypic characterization *acll5*-1 homozygous plants. (**A**) Flowers from wild-type (WT) and *acll5*-1 plants. Arrows indicate mature anthers, which are dehiscing and releasing mature pollen grains onto the stigmatic surface in WT, but which appear dark and without obvious pollen in *acll5*-1. (**B**) Mature siliques from WT and *acll5*-1 plants. No seed formation could be observed in the mutant. (**C**) Mature WT and *acll5*-1 plants.

To verify if the failure of self-pollination in *acll5* mutant plants was due to a defect in anther development and/or pollen production, I used SEM to view these processes at higher resolution in several wild-type and mutant flowers. Figure 4.3 shows that, although *acll5* mutant anthers underwent dehiscence, no pollen grains could be observed in the anthers, and abundant pollen was observed in dehiscing wild-type anthers examined in parallel (Figure 4.3).

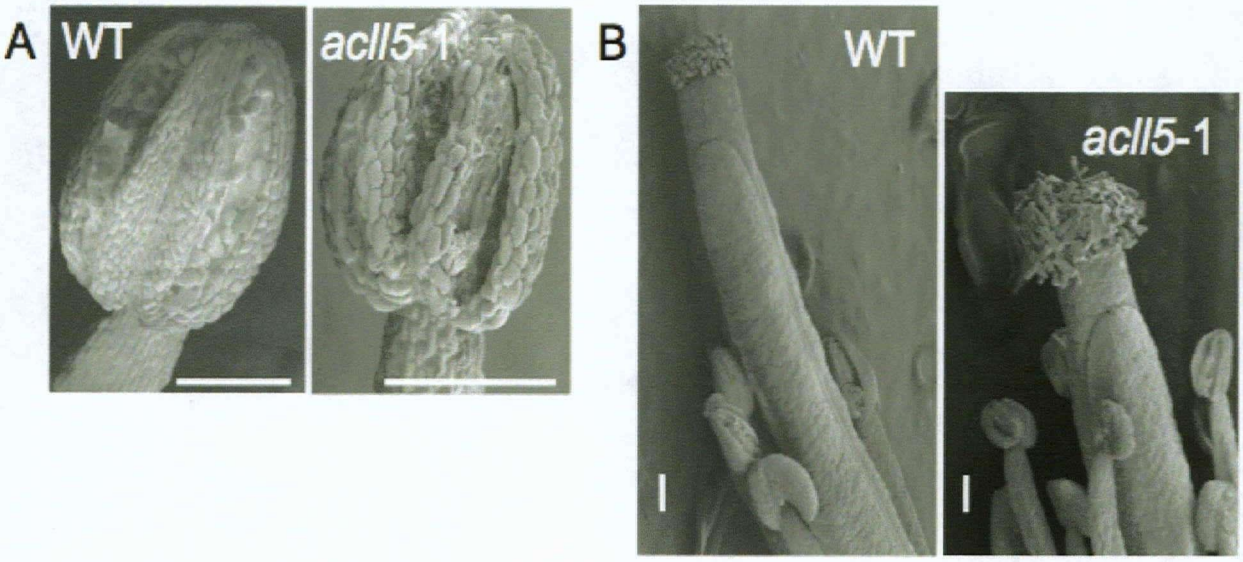


Figure 4.3: SEM of wild type (WT) and homozygous *acll5*-1 anthers. (**A**) Anther dehiscence was observed to occur normally in the mutant, however no pollen grains could be found. (**B**) There was no carpel elongation for silique development in *acll5*-1 plants. Scale bar = 100um.

Furthermore, wild-type carpels were clearly elongated whereas carpel elongation was defective in the mutant. These data imply that loss of *ACLL5* function in the *acll5* mutant may lead to a defect in pollen production and development, and consequent male sterility and absence of self-pollination. Unfertilized ovules were present in the mutant siliques,

suggesting that the mutant was female fertile (Figure 4.2B). To test this hypothesis I used *acll5* mutant plants as pollen recipients in crosses using pollen from wild-type plants. Siliques developed normally from such crosses, which produced F1 seeds that were able to germinate normally with no apparent loss in fecundity relative to self-pollinated wild-type plants.

4.2.3 Genetic characterization of the *acll5* mutant

In addition to showing female fertility of the *acll5* mutant, the crosses mentioned above were used to investigate the genetic basis for the observed male sterility phenotype. F1 heterozygote plants were allowed to self-pollinate and the resulting F2 population was then analyzed for co-segregation of the male sterile phenotype with *acll5-1*. The results showed that the mutant phenotype was inherited in a Mendelian fashion in all 184 plants analyzed, with one quarter of the F2 progeny displaying male sterility ($x^2 = 0.437$; $p>0.4$; $n=184$). This demonstrates that the mutant phenotype is caused by a single locus. Next, in order to establish if the single-locus male sterile phenotype is caused by the *acll5-1* mutant allele, I determined the genotype of 183 F2 plants for which there were phenotypic data. *ACLL5* alleles segregated in a Mendelian ratio of 1:2:1 ($x^2 = 2.08$; $p > 0.1$), and inheritance of the male sterile phenotype co-segregated with *acll5-1* (51/183 *acll5-1* homozygotes male sterile, 0/183 *acll5-1* heterozygotes and *ACLL5* homozygotes male sterile).

4.2.4 Phenotypic analysis of anther development in the *acll5-1* mutant

All together, my data suggested that the transposon insertion in *ACLL5* generated a loss-of-function allele, resulting in male sterility. Therefore, in order to investigate further which point anther development and pollen production were impaired in the mutant, I analyzed in detail the development of mutant and wild-type anthers. Anther sections taken from flowers at different developmental stages were stained with toluidine blue, visualized in bright field microscopy, and the anthers categorized into developmental stages (Sanders *et al.*, 1999). Representative sections from wild type and *acll5* mutant flowers are shown in Figure 4.4.

Early stages of anther development in *acll5* appeared normal. At stage 5, in both mutant and wild-type anthers, four defined locules were established and visible pollen mother cells had appeared. Subsequently, the pollen mother cells undergo meiosis and tetrads are formed, connected by a callose wall (Sanders *et al.*, 1999), which is degraded by stage 8, releasing microspores. Figure 4.4 shows that, in both wild-type and mutant anthers, individual microspores could be seen at stage 8, indicating that the callose wall had degenerated, releasing the microspores from the tetrads in a normal manner in the mutant.

At stage 9, as described (Sanders *et al.*, 1999), the wild-type microspores became vacuolated, and the exine wall started to become visible, as evidenced by toluidine blue staining of the walls of developing pollen grains. However, normal development seemed to be arrested in the *acll5* mutant at this stage. Although the vacuoles were evident in some mutant microspores, they appeared malformed and distorted in shape, and exine

walls were not clearly evident as in wild-type (Figure 4.4). In the place of clearly developing microspores seen in the locules of wild-type anthers, stage 9 mutant locules were filled with these misshapen structures.

At stage 10, wild-type microspores continued to enlarge and develop, and the tapetum layer showed initial signs of degeneration. In contrast, in the mutant anthers at an age equivalent to stage 10, I observed degradation of both microspores and the tapetum wall (Figure 4.4). Thus, in the mutant few if any microspores were observed at this stage, and the tapetum layer degenerated earlier than expected in normal development. At stage 11, the tapetum cell layer was greatly degraded in wild-type anthers, and clear darkly staining pollen grains were seen. In contrast, although fully mature anthers appeared normal in the mutant, the locules were empty, devoid of any pollen grains (Figure 4.4). These data allowed me to pinpoint the developmental stage at which the *acll5* male sterile phenotype became apparent.

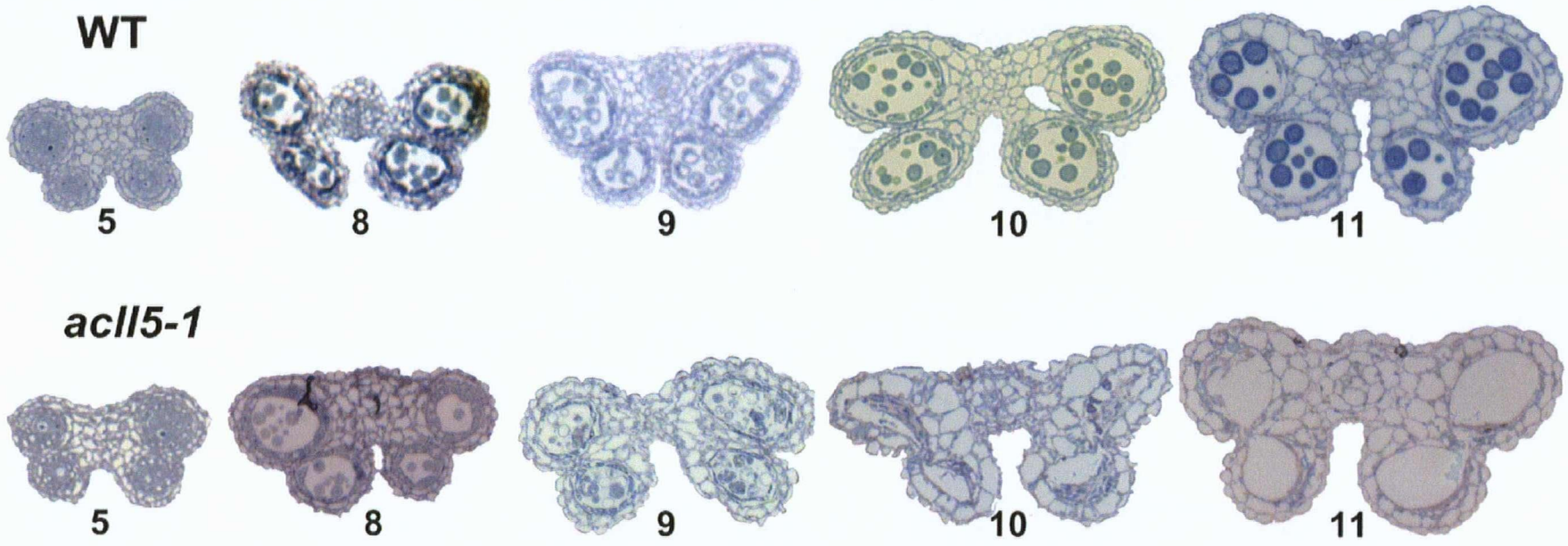


Figure 4.4: Anther cross-sections (1μm) of wild type and homozygous *acll5*-1 mutant. Developmental stages are according to Sanders *et al.* 1999. Anther development appears normal in the mutant up until stage 8. In stage 9 initial degradation of microspores is apparent in the mutant. The tapetum cell layer has normal appearance when compared to wild type. Microspore degradation is complete in *acll5*-1 and results in a normal looking anther devoid of any pollen grains.

4.2.5 *In situ* hybridization analysis of *ACLL5* expression in developing anthers

Our results demonstrated that the mutation does not impair the early stages of microspore production in the developing anthers. Instead, microspores are present but fail to complete maturation into fully developed pollen grains and degenerate well before maturation, together with premature degradation of the tapetum cell layer. The mutant phenotype is first visible at stage 9 of anther development. To gain insights into the spatio-temporal pattern of *ACLL5* expression in the anthers, and to determine if *ACLL5* expression could be correlated to the timing and location of the *acll5* phenotype observed, *in situ* hybridization experiments were performed using an *ACLL5*-derived probe hybridized to developing wild-type flowers. The experimental procedure was carried out by our collaborator Sarah McKim, University of British Columbia, while I interpreted the data.

Figure 4.5 shows that *ACLL5* was strongly and specifically expressed in the tapetum cell layer of developing anthers. Strong expression was first evident at stage 7, right before the separation of the microspores from the tetrads. Gene expression was dramatically reduced in stage 8 anthers, observed in the same flower (Figure 4.5A), and it decreased to background levels in later developmental stages, observed in different flowers (Figure 4.5B). These results demonstrate that *ACLL5* has a transient and tapetum preferred expression pattern and is most highly expressed in the stages immediately preceding the appearance of the visible phenotype. One interesting observation is the presence of anthers of slightly different developmental stages in the same flower. The development of the anther from the outer short stamen occurs immediately after the anther from the inner

long stamen. This timing is likely an outcome of the initiation from the flower primordia occurring at consecutive time points for both types of stamen (Kunst *et al.*, 1989), and explains the presence of stage 7 and stage 8 anthers on the same flower.

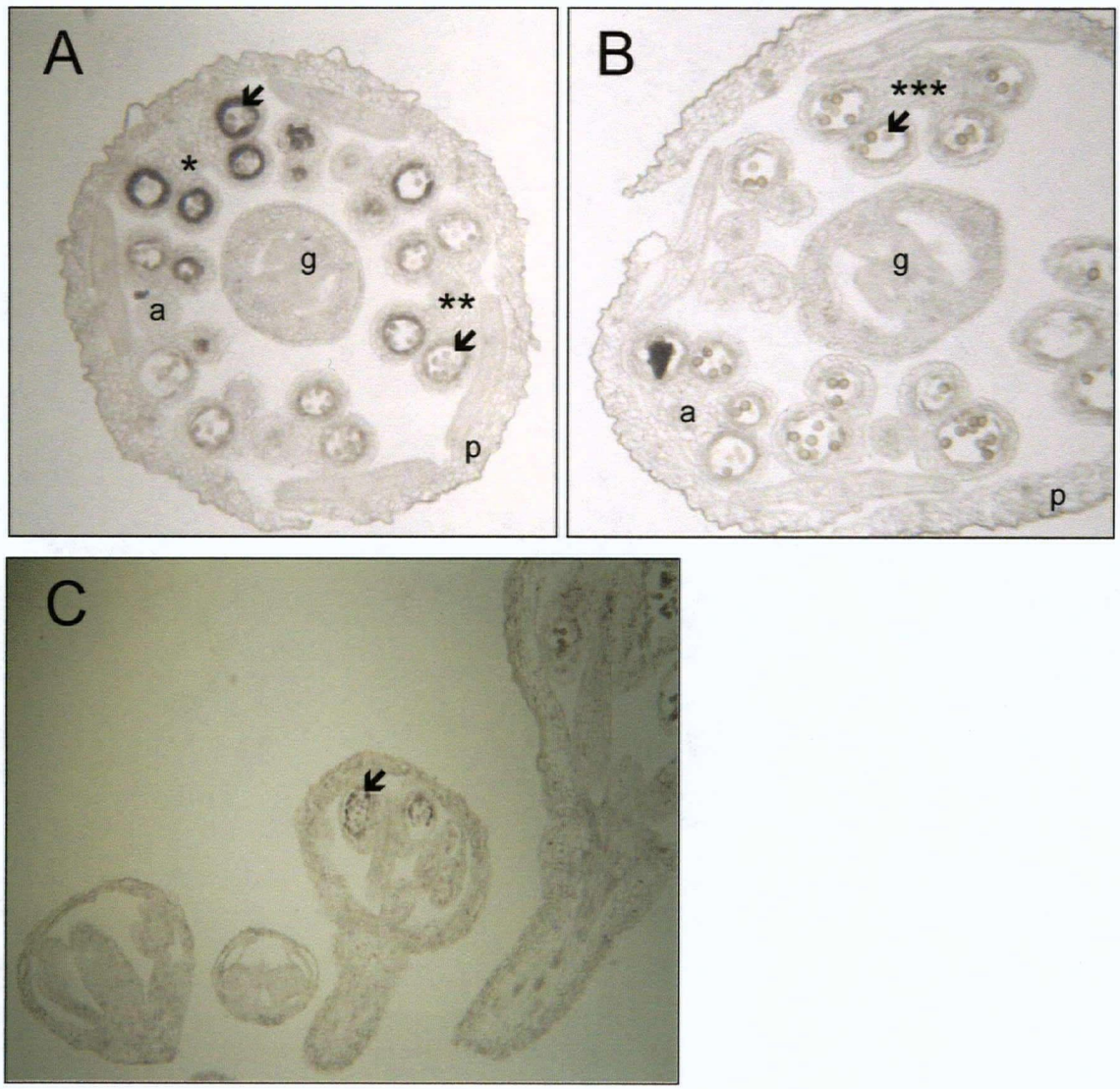


Figure 4.5: *In situ* hybridization analysis in developing wild-type flowers showing *ACLL5* expression specific to the tapetum cell layer (➹). (**A**) Cross-section through an immature flower with anthers at slightly different stages of development, hybridized to an anti-sense *ACLL5* probe. Tapetum-specific *ACLL5* expression was highest in stage 7 anthers (*) just before tetrad separation, and was reduced in stage 8 anthers (**) when individual microspores can be observed. (**B**) Cross-section through an older flower, hybridized with the same probe. Little *ACLL5* expression was observed in stage 10 anthers (***). a, anther; g, gynoecium; p, petal. (**C**) Longitudinal section of inflorescence showing developing flowers, and highest signal in stage 7 anthers.

4.2.6 Co-expression analysis of *ACLL5* in *Arabidopsis*

Co-expression analysis using global expression data available in public databases has been successfully used to identify genes participating in the same biological process and/or biochemical pathway (Ehlting *et al.*, 2006; Persson *et al.*, 2005). If *ACLL5* is transcriptionaly regulated together with other genes encoding enzymes in the same hypothetical biochemical pathway in tapetum cells prior to microspore release, one should expect to identify this group of genes by their co-expression with *ACLL5* in the various datasets. Therefore, in the effort to gain insights into biochemical pathways that ACLL5 could be participating in, I performed an *in silico* co-expression analysis using public *Arabidopsis* microarray datasets, as described in Materials and Methods.

Using the Correlated Gene Search tool, I queried 237 microarray experiments in the Tissue and Development dataset (http://prime.psc.riken.jp), using a cutoff Pearson co-expression coefficient of 0.80. This analysis identified 56 co-expressed genes, most of unknown function (complete list in Appendix I). Not surprisingly, expression of all genes in this group was very specific to floral tissues, as could be seen by their individual expression patterns available at the eFP browser (http://bbc.botany.utoronto.ca). Among those genes most highly co-expressed with *ACLL5*, I identified genes annotated as related to lipid metabolism and others with similarity to genes involved in monolignol and flavonoid metabolism, which could be expected to encode enzymes in a pathway involving ACLL5 and a CoA ester intermediate in the biosynthesis of fatty acid or phenolic constituents of the exine (Figure 4.6). Interestingly, *MS2,* involved in fatty acid metabolism and with possible fatty acyl-coenzyme A reductase activity (Aarts *et al.*,

1997) is the only one of these genes, of fatty-acid or phenylpropanoid-like annotation, of known biological function that was co-regulated with *ACLL5* in our search.

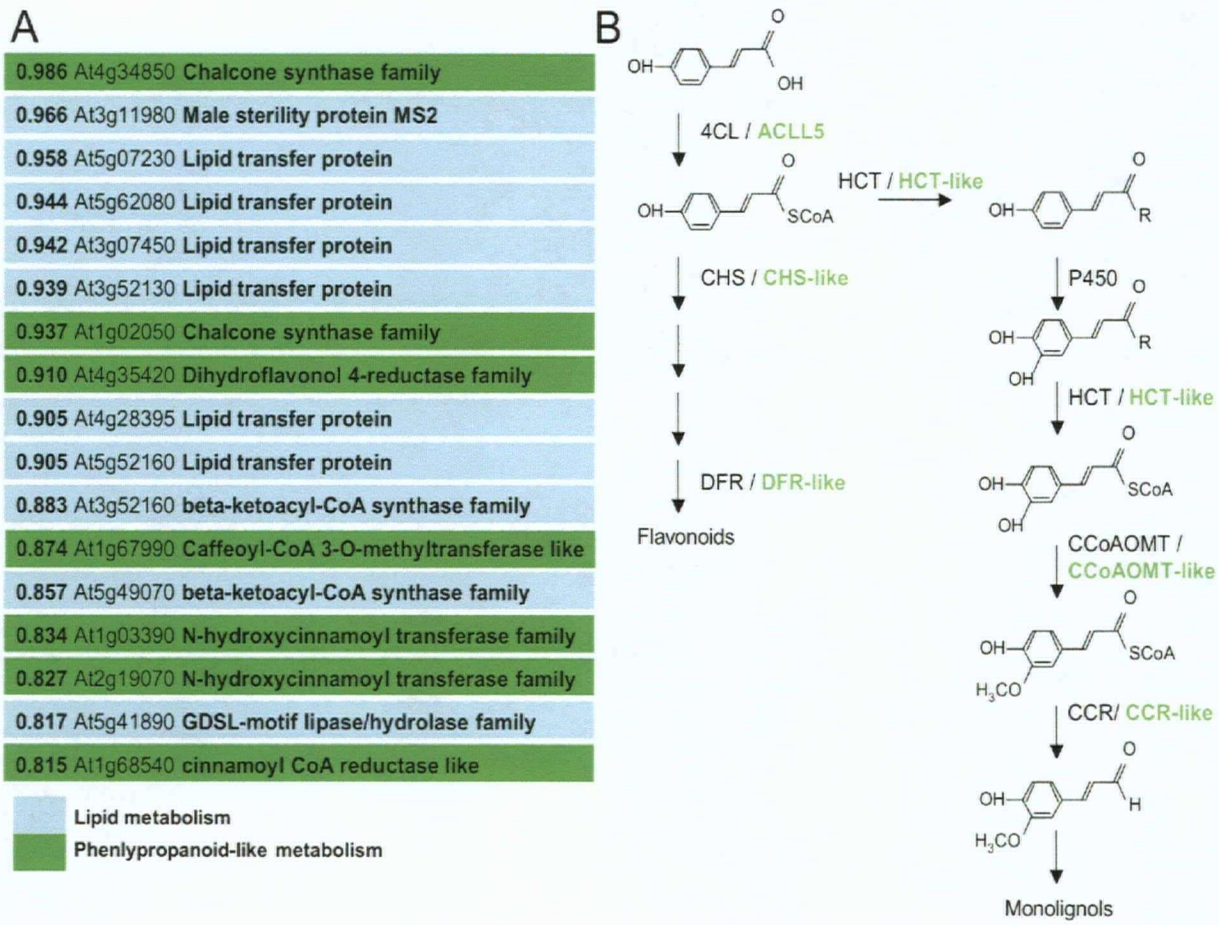


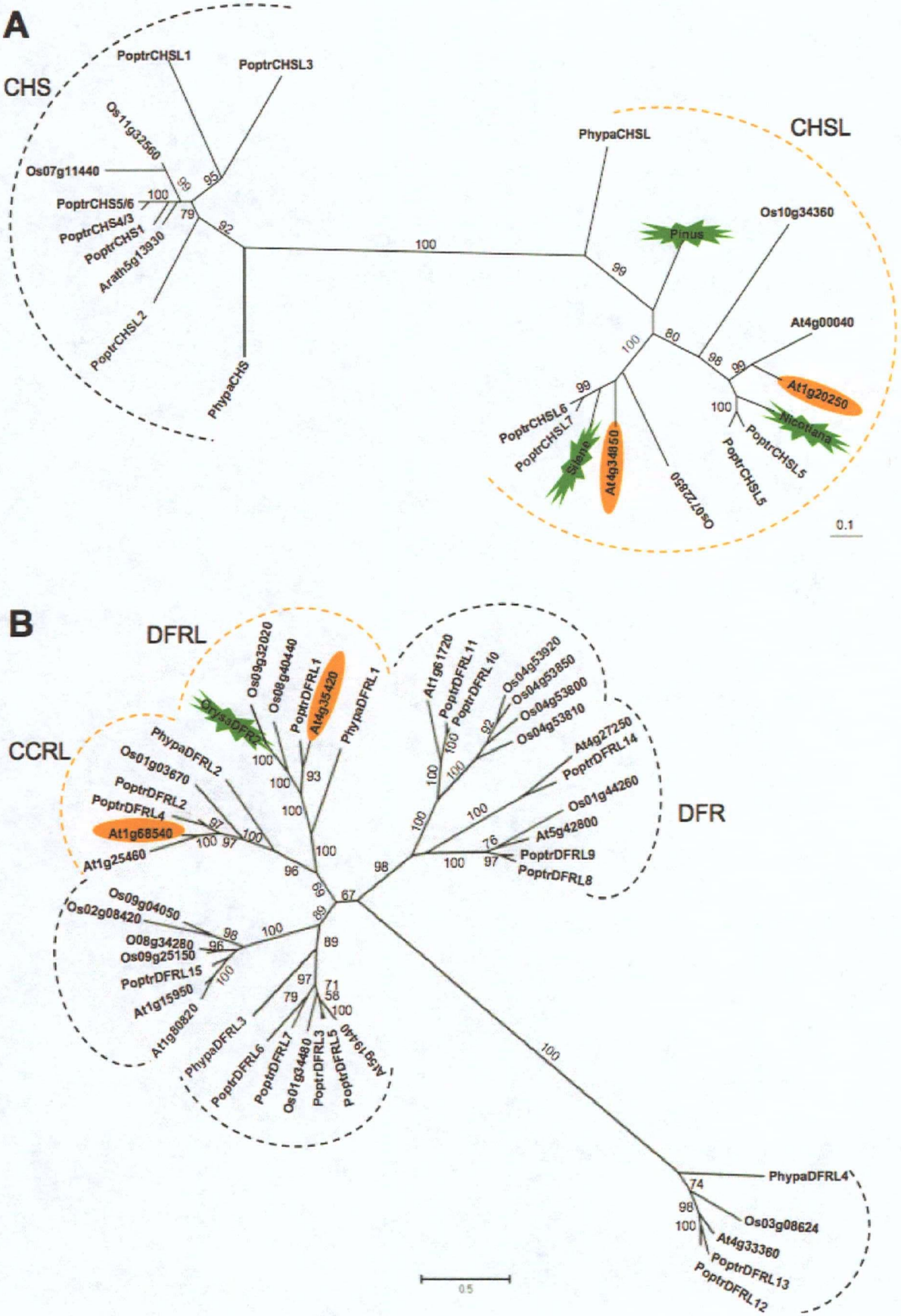
Figure 4.6: (**A**) Selected genes with high co-expression coefficients with *ACLL5*/At1g62940 involved in lipid and phenylpropanoid-like metabolism (http://prime.psc.riken.jp/ Correlated Gene Search: Tissue and development v.1, 237 data, threshold 0.80). (**B**) Putative duplicated phenylpropanoid pathway.

A number of co-expressed genes annotated as encoding phenylpropanoid-like enzymes were found co-expressed with *ACLL5*. As shown in Figure 4.6A, co-expressed phenylpropanoid-like genes included those encoding two chalcone synthase (CHS)-like enzymes and a dihydroflavonol reductase (DFR)-like enzyme, both related to enzymes involved in flavonoid biosynthesis (Figure 4.6B). CHS uses 4-coumaryl-CoA as one of its substrates. This list also included two genes encoding hydroxycinnamyl CoA shikimate/quinate hydroxycinnamyltransferase (HCT)-like enzymes, a gene encoding a caffeoyl-CoA *O*-methyltransferase (CCOMT)-like enzyme and a gene encoding a cinnamyl-CoA reductase (CCR)-like enzyme. Interestingly, this set of genes encodes sets of enzymes mimicking the pathways leading to the production of monolignols, known to act in sequence leading from hydroxycinnamyl-CoA to the corresponding alcohol (Figure 4.6B), and to flavonoids, in which CHS catalyzes condensation of 4-coumaryl-CoA with malonyl CoA at the entry point of flavonoid metabolism. None of these genes have yet been biochemically or biologically characterized but their co-expression with *ACLL5* suggests the existence of one or more anther expressed phenylpropanoid-like pathways, analogous to the well-characterized monolignol and flavonoid biosynthetic pathways, both of which could involve ACLL5 (Figure 4.6).

Phenylpropanoid-like genes such as the *CCR-like* and *CCOMT-like* genes co-expressed with *ACLL5* have been described in several reports (Costa *et al.*, 2003; Ehlting *et al.*, 2005; Raes *et al.*, 2003), and occur in clades distinct from those known to be involved in phenylpropanoid and monolignol biosynthesis. However, less phylogenetic information is available for *CHS-like* and *DFR-like* genes. In order to obtain further information about

the *CHS-like* and *DFR-like* genes co-expressed with *ACLL5*, I carried out phylogenetic analyses of homologues identified by *in silico* searches of sequence information in Arabidopsis and other plant species. In addition to Arabidopsis, poplar, rice and *Physcomitrella CHS*, *CHS-like*, *DFR*, and *DFR-like* genes, *CHS-like* and *DFR-like* genes from tobacco (Varbanova *et al.*, 2003), pine (Walden *et al.*, 1999), rice (Yau *et al.*, 2005) and *Silene latifolia* (Ageez *et al.*, 2005) were identified from the literature on the basis of high expression during uninucleate microspore development (Varbanova *et al.*, 2003; Walden *et al.*, 1999; Yau *et al.*, 2005) and during bursts of tapetum cell activity (Ageez *et al.*, 2005).

These analyses, shown in Figure 4.7A and 4.7B, indicated that the *CHS-like* and *DFR-like* genes co-expressed with *ACLL5* are in distinct clades from those containing *bona fide* Arabidopsis *CHS* and *DFR* genes. These clades also contain *CHS-like* and *DFR-like* homologues from other species expressed in tapetum cells and/or in concert with microsporogenesis. In addition, further phylogenetic analysis showed that a tobacco *4CL-like* gene expressed in the tapetum during microsporogenesis (Atanassov *et al.*, 1998) is in the same clade as *ACLL5* and its poplar and rice homologues (Figure 4.7C). Also, expression of the single rice homologue gene in this clade, based on massively parallel signature sequencing (MPSS) data (http://mpss.udel.edu/rice/), is strongly preferred in the immature panicles (inflorescence). These data support a role for the Arabidopsis *ACLL5* co-expressed *CHS-like* and *DFR-like* genes in anther and microspore development, possibly functioning together with *ACLL5* in one or more exine-specific biochemical pathways conserved in angiosperms.


A
CHS
CHSL
PoptrCHSL1
PoptrCHSL3
Os11g32560
Os07g11440
PoptrCHS5/6
PoptrCHS4/3
PoptrCHS1
Arath5g13930
PoptrCHSL2
PhypaCHS
PhypaCHSL
Pinus
Os10g34360
At4g00040
At1g20250
Nicotiana
PoptrCHSL5
PoptrCHSL6
PoptrCHSL7
Silene
At4g34850
Os07g22850
0.1
B
DFRL
CCRL
DFR
Os09g32020
Os08g40440
PoptrDFRL1
At4g35420
PhypaDFRL1
OrysaDFR2
PhypaDFRL2
Os01g03670
PoptrDFRL2
PoptrDFRL4
At1g68540
At1g25460
At1g61720
PoptrDFRL11
PoptrDFRL10
Os04g53920
Os04g53850
Os04g53800
Os04g53810
At4g27250
PoptrDFRL14
Os01g44260
At5g42800
PoptrDFRL9
PoptrDFRL8
Os09g04050
Os02g08420
O08g34280
Os09g25150
PoptrDFRL15
At1g15950
At1g80820
PhypaDFRL3
PoptrDFRL6
PoptrDFRL7
Os01g34480
PoptrDFRL3
PoptrDFRL5
At5g19440
PhypaDFRL4
Os03g08624
At4g33360
PoptrDFRL13
PoptrDFRL12
0.5

C

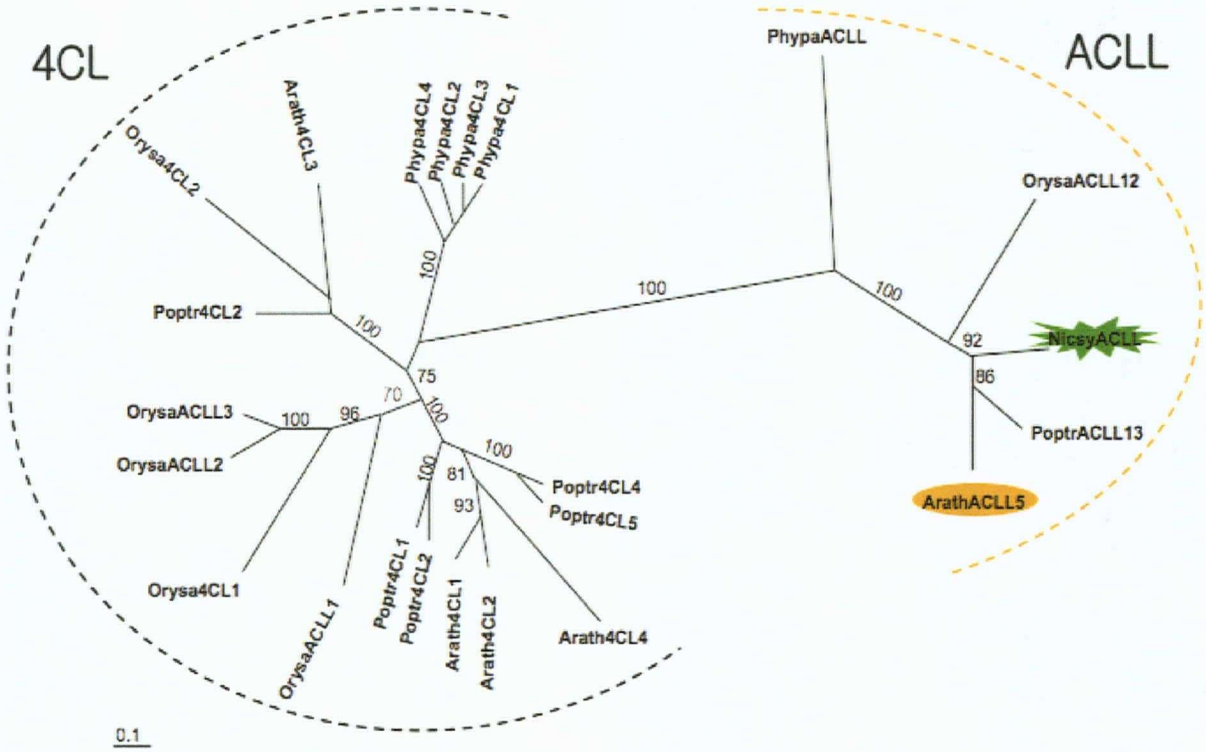


Figure 4.7: Phylogenetic analyses of protein sequences of Arabidopsis, poplar, rice, Physcomitrella and other species genes expressed during microspore development (star). Maximum-likelihood (ML) tree was built using 100 or 500 bootstrap replicates in PhyML 2.4.4. Bootstrap values above 70 are shown on branches. *ACLL5* and co-expressed genes are highlighted (oval). (**A**) Chalcone synthase (CHS) and chalcone synthase-likes (CHSL). (**B**) Dehydroflavonol reductase (DFR) and dehydroflavonol reductase like families. (**C**) 4-coumarate-CoA ligase (4CL) and acyl-CoA ligase 5 and homologues (ACLL).

4.3 Discussion

All together, our data support the hypothesis that *ACLL5*, a gene transiently and preferentially expressed in the tapetum, is required for normal development and maturation of pollen grains, and is involved in pollen wall formation. Furthermore, our phylogenetic analyses provide evidence for *ACLL5* homologues from poplar, rice, and tobacco that may have a similar function in those species. In addition, given that ACLL5 is an enzyme closely related to *bona fide* 4CLs, and that *in silico* co-expression results revealed a number of genes co-expressed with *ACLL5* that encode phenylpropanoid and lipid metabolism related enzymes, it is reasonable to speculate that ACLL5 has a phenylpropanoid-like or fatty acid substrate, providing precursors for biosynthesis of an essential sporopollenin polymer.

The exact composition of the pollen wall is not well defined yet, and may even vary greatly between species. However, it is generally accepted that tapetum cell function is conserved among plant species and is required for exine production and deposition (Scott *et al.*, 2004). Sporopollenin, a heterogeneous polymer found in the pollen exine layer, is composed of long chain fatty acids and phenolic compounds. The phenolic monomers are coupled by ether linkages, which are characteristic of polyphenolic polymers such as lignin (Scott *et al.*, 2004). It has been shown that enzymes involved in the phenylpropanoid pathway such as PAL and CHS are required for male fertility in some plants (Matsuda *et al.*, 1996; van der Meer *et al.*, 1992), reinforcing an essential role for phenylpropanoids in exine composition. However, the Arabidopsis mutant *tt4* (*transparent testa4*), a loss of function lesion in the single copy Arabidopsis *CHS* gene,

encoding the first enzyme committed to flavonoid biosynthesis, exhibits normal pollen development (Ylstra *et al.*, 1996). This suggests that flavonoids themselves are not required for pollen viability and sporopollenin biosynthesis, and that other phenolic compounds may play such roles (Boavida *et al.*, 2005).

The tapetum contribution to exine synthesis and deposition starts while the microspores are still attached in tetrads and continues through the vacuolated stages until the first pollen mitosis is almost completed (Boavida *et al.*, 2005). The spatio-temporal pattern of *ACLL5* gene expression, revealed by *in situ* hybridization, is consistent with the timing of this function of the tapetum, and *ACLL5* expression in the tapetum during stages 7 and 8 of anther development, characterized by tetrad formation and microspore release, supports the hypothesis that ACLL5 is necessary for production of phenylpropanoid or fatty acid related molecules in the early steps of exine biosynthesis. The spatio-temporal pattern of *ACLL5* expression is also consistent with the timing of the defect in pollen development during anther maturation in the *acll5* mutant. Loss of the *ACLL5* function in tapetum cells in stage 7 anthers, when *ACLL5* is most highly expressed in tapetum cells (Figure 4.5) is consistent with a defect in biosynthesis of a critical secreted sporopollenin component(s), leading to defective microspores which, when released from tetrads in stage 8 anthers, fail to develop normal exine and are aborted in development at stage 9, as observed in the *acll5* mutant (Figure 4.4).

The ACLL5 substrate and the nature of the biochemical pathway that uses the CoA esters, that are the presumed product of the enzyme, are unknown. Genes closely related

to *ACLL5* are found in poplar, rice, and tobacco (Chapter 3; Figure 4.7C), suggesting that the metabolic pathway in which it participates is conserved amongst angiosperms. *In silico* analysis of Arabidopsis genes strongly co-expressed with *ACLL5*, and tapetum-expressed genes in other species known from the literature, provide clues for possible enzymes that could function with ACLL5 in one or more common pathways. Prominent among the Arabidopsis co-expressed genes are those encoding phenylpropanoid-like enzymes (Figure 4.6). These include two *CHS-like* genes and a *DFR-like* gene. Interestingly, *CHS-like* and *DFR-like* genes with anther and/or tapetum-preferred expression patterns have been described in several other species. Phylogenetic analysis of *CHS* and *CHS-like* genes in Arabidopsis, poplar, rice, and other species (Figure 4.7A) shows that the Arabidopsis co-expressed *CHS-like* genes At4g34850 and At1g02050 occur in a clade distinct from true *CHS*. This clade contains representatives from poplar and rice, as well as tapetum-expressed genes from pine, *Silene*, and tobacco, suggesting that ACLL5 could function upstream of a CHS-like polyketide synthase enzyme in the biosynthesis of a structural polyketide component of sporopollenin, distinct from flavonoids. Similarly, the co-expressed *DFR-like* gene At4g35420 occurs in a clade distinct from true *DFR* genes (Figure 4.7B), which also contains poplar and rice representatives. Interestingly, one of the rice *DFR-like* genes in this clade is expressed in the tapetum of developing anthers (Yau *et al.*, 2005). The co-expressed *DFR-like* gene At4g35420 could be a reductase involved in modification of a *CHS-like* derived polyketide constituent of sporopollenin, or could be reductase that acts directly on the CoA ester product of the ACLL5-catalyzed reaction, analogous to CCR in monolignol biosynthesis (Lauvergeat *et al.*, 2001).

Also of interest among the phenylpropanoid-like co-expressed genes are a *CCR-like* gene and a *CCOMT-like* gene, (*CCRL6* and *CCOMT2*; Ehlting *et al.*, 2005), as well as two genes encoding N-hydroxycinnamyl transferase family proteins (HCT-likes). The *CCRL6* gene occurs in a clade of genes encoding reductases distinct from *DFR*, the *ACLL5* co-expressed *DRF-like* gene At4g35420, and *bona fide CCR* genes, and this clade also contains rice and poplar members (Figure 4.7B). As illustrated in Figure 4.6, one interpretation is that *ACLL5* and this set of co-expressed phenylpropanoid-like genes encode enzymes in a pathway required for biosynthesis of a phenolic sporopollenin constituent, which is analogous to the well-characterized sequence of enzymatic steps leading to monolignol biosynthesis.

Thus, co-expression analysis has revealed potentially novel phenylpropanoid-like biochemical pathways in which ACLL5 could play a key role by providing CoA ester substrates, and one or more of these pathways could be involved in biosynthesis of crucial sporopollenin constituents. This hypothesis could be further tested by determining the spatio-temporal expression patterns of the co-expressed genes during anther development, to establish how well they coincide with *ACLL5* expression, and to obtain loss of function alleles of the genes to determine if, like *ACLL5*, they are required for male fertility and pollen development.

Phenylpropanoid genes have been shown to be transcriptionally regulated by MYB transcription factors, which control many aspects of phenylpropanoid metabolism in plants (Douglas, 1996; Hauffe *et al.*, 1991; Rogers and Campbell, 2004). The gene

encoding the transcription factor MYB99 is strongly co-expressed with *ACLL5* (Appendix 1), and could thus play a role in regulating the pathways of secondary metabolism related to ACLL5 function. To look for evidence of co-regulation of *ACLL5* and co-expressed phenylpropanoid-like genes, I performed an *in silico* search of the PLACE 25.0.1 (http://bbc.botany.utoronto.ca/ntools/cgi-bin/BAR_Promomer.cgi) database to look for consensus matches of regulatory elements in the promoter regions of these genes. This search identified a plant MYB binding site (MYBPLANT), that has been reported to activate genes members of the phenylpropanoid metabolism in *Antirrhinum majus* (snapdragon) flowers (Sablowski *et al.*, 1994). Additional consensus *cis* elements related to flower development, pollen development and MYB binding were also present in the promoter regions of *ACLL5* and co-expressed phenylpropanoid-like genes (Table 4.1).

Table 4.1: Consensus *cis* element matches in *ACLL5* and co-expressed phenylpropanoid-like gene promoter regions.

Element	Sequence	Description
23BPUASNSCYCB1	ACAAA	MYB binding core required for M-phase expression. Related to cell cycle.
AGAMOUSATCONSENSUS	[AGT]CC[AT][AT][AT]	Indispensible for AGAMOUS function in flower development
AGATCONSENSUS	[AT]CC[AT][AT][AT]	Indispensible for AGAMOUS function in flower development
AGL1ATCONSENSUS	[AGT]CC[AT][AT][AT]	Sequence for AtAGL1, MADS-Box domain gene expressed in transition to flowering
AGL3ATCONSENSUS	[AT]C[CT]A[AT][AT]	Sequence for AtAGL3, MADS-Box domain gene expressed in above-ground vegetative organs
CIACADIANLELHC	A[ACGT][ACGT][ACGT][AT]	Region necessary for circadian region in tomato
MYBPLANT	ACC[AT]A[AC]	Plant MYB binding site, sequence related to box P in promoters of phenylpropanoid genes
PALBOXPPC	[AC][AC]C[AC]A[AC]	Box P. One of three cis-acting element boxes of phanylalanine amonia lyase (PAL)

Given the high lipid content of sporopollenin in the exine, and the fact that fatty acid-related enzymes are also co-expressed with *ACLL5*, an alternative to an ACLL5 function in a phenylpropanoid-like metabolic pathway is the possibility that this enzyme plays a crucial role in fatty acid metabolism and uses a fatty acid derived substrate. For example, ACLL5 could participate in the same pathway as *MS2*, which is strongly co-expressed

with *ACLL5* and apparently encodes a long chain fatty acid reductase required for sporopollenin deposition. It could also play a role in another lipid-related pathway, such as those that yield in the production of the lipid rich pollen coat formed after sporopollenin deposition. The pollen coat fills the gaps between the exine structures, and confers important functions such as protection from dehydration and pollen-stigma recognitions (Boavida *et al.*, 2005). As opposed to the exine, the pollen coat is easily extractible and has been extensively analyzed, revealing that its major components are non-polar esters of medium, long-chain and very long chain fatty acids, as well as lipases and other proteins attached to the pollen surface. However, deposition of the pollen coat must be timed to occur after the deposition of the exine. The temporal pattern of *ACLL5* expression, as seen in our *in situ* experiments (Figure 4.5), concomitant to callose wall dissolution at stage 7 of anther development, is not as well correlated with the timing of pollen coat deposition as with the timing of exine deposition. Therefore *ACLL5* is more likely to participate in the biosynthesis of either phenolic or fatty acid constituents of sporopollenin in the exine, prior to the deposition of the pollen coat.

Pollen development and pollen wall biosynthesis are very complex processes, involving more genes that any other single developmental process in plants (Scott *et al.*, 2004), many of which are expressed in similar spatio-temporal patterns during anther development. In this light, it should be kept in mind that genes that show high co-expression coefficients with *ACLL5* in microarray experiments are not necessarily co-regulated or part of the same biochemical pathway. Still, it is interesting to observe the various classes of co-expressed genes that might, collectively, orchestrate processes in

pollen development and pollen wall formation. One class of proteins with several representatives encoded by genes that are co-expressed with *ACLL5* is Lipid Transfer Proteins (LTP). It has been suggested that LTPs can bind to fatty acids and acyl-CoA esters facilitating the secretion and deposition of lipophilic molecules onto cell walls (Arondel *et al.*, 2000). Analogously, LTPs could be participating in a similar process transporting molecules from the tapetum cells to the pollen wall. The same could be speculated for the ABC transporter co-expressed with *ACLL5*, which belongs to the WBC subfamily of ABC transporters (Sanchez-Fernandez *et al.*, 2001). Members of the WBC subfamily have been shown to transport lipids to the cell walls (Pighin *et al.*, 2004) and could function in the transport of related molecules to the pollen wall. Other co-expressed genes with *ACLL5* include those encoding for glycosyl hydrolases. Such class of enzymes would be necessary for callose degradation for microspore release from the tetrads during stage 7 of anther development. In addition, two uncharacterized genes enoding cytochrome P450 enzymes were co-expressed with *ACLL5*, which could be involved in hyrodroxylation of phenolic constituents of sporopollenin, or modification of other sporopollenin constituents.

Another possible role for *ACLL5* that would be consistent with the phenotype observed in the *acll5* mutant is participation in a vital tapetum-specific metabolic pathway unrelated to exine formation. The loss of ACLL5 enzyme function could result in improper tapetum function and degeneration of the tapetum cells, which would result indirectly in improper pollen wall formation resulting in abortion of the pollen grains. The degradation of the tapetum is a normally occurring and tightly regulated physiological process. In normal

development, the tapetum starts to degenerate only after stage 10 (Sanders *et al.*, 1999). Although in the *acll5* mutant the tapetum cells appeared to degenerate earlier than in wild-type, I could not determine if the early degeneration of the tapetum is the result of accelerated programmed cell death or a non-specific process. Additional detailed ultrastructural analysis of the tapetum cells in the mutant to verify early signs of PCD would be helpful to address this question. However, I did not observe any obvious aberrations in the tapetum appearance before the degradation of the microspores, suggesting that tapetum function itself is not strongly affected in the *acll5* mutant. In addition, our *in situ* hybridization results showed that *ACLL5* expression in the tapetum is very transient, being restricted largely to stages 7 and 8 of anther development, and returning to background levels well before initiation of tapetum degeneration. Therefore, the early degeneration of the tapetum in the *acll5* mutant is more likely a consequence, and not a cause, of microspore malformation and subsequent degradation, and the *ACLL5* expression pattern is more consistent with a role for *ACLL5* in the production of sporopollenin compounds in the exine of the pollen wall, rather than in functioning of the tapetum itself.

What still remains unclear is the exact cause of degradation of the microspores in the *acll5* mutant. From an evolutionary perspective, it would be an extreme disadvantage for the plant to release its genetic material in a defective "package". Healthy pollen grains that will survive the obstacles between anther release, stigma recognition and germination are crucial for the survival of the species. It is likely that there are "check points" to verify the fidelity of the pollen developmental process, with a mechanism to eliminate

defective microspores, and that this is engaged in the *acll5* mutant. An alternative hypothesis would be that without the physical strength provided by a normal pollen wall coating in the *acll5* mutant, the pollen grains simply collapse due to physical pressures.

As described in the introduction, detailed chemical analysis of the pollen wall, and elucidation of sporopollenin chemical structure is a daunting task given the biochemical nature of the exine. Therefore, combined genetic and bioinformatics approaches, such as those taken on this study, are important to generate hypotheses regarding the structure and biological function of the pollen wall, and the nature of the biochemical networks required for its development and deposition. This study opens the door to further testing of the hypothesis that *ACLL5* is involved a pathway required for the biosynthesis of uncharacterized phenolic or lipid-based constituents of the pollen wall, for example by chemical analysis of sporopollenin in the *acll5* mutant, and investigation of potential male sterile phenotypes of co-expressed genes.

Finally, this work demonstrates the usefulness of comparative genomics in understanding the role of a particular gene in a given biological system. Figure 4.7C and previous results (Chapter 3) show that single-copy *ACLL5* homologs are present in poplar and rice and maize, and that a closely related tapetum-specfic gene is present in tobacco, supporting a conserved biological function for this enzyme in angiosperms. Furthermore, by taking advantage of the dioecious nature the poplar species, I showed that the poplar *ACLL5* homolog is preferentially expressed in male flowers (Chapter 3), consistent with a role for this enzyme in anther development. The Arabidopsis *ACLL5* gene and its poplar

homologue share 80% identity at the amino acid level, which is comparable to the level of identity between the *4CL* representatives of both species. The availability of both poplar and Arabidopsis *ACLL5* homologues allows future testing of the hypothesis that they are orthologous genes with the same biochemical function, for example by complementation of the *acll5* phenotype with the poplar *ACLL5* homologue. Experiments are underway to test this hypothesis.

CHAPTER 5- CONCLUSIONS AND FUTURE DIRECTIONS

5.1 A timeline of discoveries

The study of the *ACLL* gene family presented in this thesis was a discovery-based project in which the starting null hypothesis was ACLL enzymes have 4CL activity given their close relationship to *bona fide* 4CL enzymes. This hypothesis was rejected in the early phases of this project by preliminary *in vitro* biochemistry studies done in the Douglas lab with *ACLL6* expressed heterologously in *E. coli* (unpublished data), and additional biochemical experiments done by other labs interested in *ACLL* genes (E. Kombrink, personal communication). Therefore, not knowing which kind of substrates ACLLs could be active with, information regarding *ACLL* gene function had to be generated from little available data. Given initial evidence that all *ACLL* genes are expressed, based in EST support, a number of experiments and analyses were carried out in order to obtain clues to the functions of *ACLL* genes in Arabidopsis.

In parallel with the studies of my thesis, more information about Arabidopsis genes became publicly available. First, after performing ACLL amino acid alignments for phylogenetic analyses discussed in Chapter 3, I realized that most ACLL proteins contain a conserved C terminus peroxisomal target signal (PTS1), which was subsequently confirmed by the publication of a database of putative Arabidopsis peroxisomal proteins (Reumann *et al.*, 2004). My attention then focused on possible ACLL functions in the peroxisome. Analysis of the expression of *ACLL* promoter-*GUS* fusion constructs in transgenic Arabidopsis revealed that *ACLL* expression was not limited to any particular

tissue or cell-type (data not shown), which made it difficult to generate specific hypotheses about putative functions based on developmental expression patterns. However, results obtained from analysis of gene fusion expression revealed response to wounding for some clade D *ACLL* genes.

With my access to information from the initial stages of the poplar genome sequence assembly in the early 2005, I was able to identify potential poplar *ACLL* homologues and a comparative genome approach between Arabidopsis and poplar became possible (Chapter 3). I focused on comparison of poplar and *ArathACLL* developmental expression patterns and their response to stress treatments, which confirmed that some clade D Arabidopsis and poplar homologues have increased expression upon wound and herbivory treatments. In 2005 it was reported that the protein encoded by the clade E Arabidopsis *ACLL9* gene converts the octadecanoid pathway intermediate OPDA to the corresponding CoA ester *in vitro* (Schneider *et al.*, 2005), suggesting a role for this gene in the JA pathway. However, a putative role of ArathaCLL9 in defense-related JA biosynthesis was not corroborated by our gene expression data, which did not show increase in gene expression of *ACLL9* after mechanical wounding or stress activated expression of its closest poplar homologues. At a similar time, I used newly available tools for identifying co-expressed genes based on data from public microarray experiments, such as Expression Angler (Toufighi *et al.*, 2005) and Prime Correlated Gene Search (http://prime.psc.riken.jp). This lead me to suggest a function of ArathACLL4 in the JA pathway, which was independently confirmed in experiments performed by Koo *et al.* (2006)

The comparative genomics approach revealed highly conserved *ACLL* genes with similar expression patterns, with a striking example being those *ArathACLL5* and *PoptrACLL13* in clade A. Expression of each gene was flower-preferred, suggesting a common and conserved function in that organ in both organisms. In 2006 I was able to isolate a homozygous line of an *acll5* loss of function transposon insertion mutant from a segregating population. This mutant had a male sterile phenotype, consistent with anther-localized of gene expression as revealed by mining of microarray expression data. Further analysis of mutant phenotype, collaborative *in situ* hybridization experiments, as well as co-expression analysis allowed me to generate specific hypotheses regarding the roles of ACLL5 and homologues from other species in a tapetum-localized biochemical pathway required for pollen exine biosynthesis (Chapter 4).

5.2 Future Directions

5.2.1 Mutant analysis

As seen for *ArathACLL5*, functionally analyzed in Chapter 4, mutants affecting proper gene expression can be powerful tools to assess gene function. Arabidopsis knock-out lines for a large fraction of Arabidopsis genes are available and can be identified by simple *in silico* searches (Salk Institute, http://signal.salk.edu/cgi bin/tdnaexpress).

I obtained knock-out lines for all nine *ACLLs* and I have been able to isolate homozygous mutants lines for four of them, including *ACLL5*. Lines for *ACLL1*, *ACLL4* and *ACLL9* were isolated but no visible phenotype could be observed under laboratory growing

conditions (data not shown). I did not elaborate on these lines in this thesis as the same knock-out mutants for *ACLL4* and *ACLL9* showing no phenotype had been described elsewhere (Koo *et al.*, 2006; Schneider *et al.*, 2005). Single knock-out mutants are not always informative since genes might have redundant functions, particularly genes that are part of gene families. *ACLL1* and *ACLL4* are in the same clade (D), and are located in tandem making nearly impossible the generation of double mutants by cross-pollination and genetic recombination. As discussed in Chapter 3, based on the known function of *ACLL4*, it is possible that other genes in clade D have similar functions, encoding ODPA/OPC8-CoA ligases. If this hypothesis is correct, a mutant with all clade D genes knocked out should be unable to make JA, and show a phenotype related to JA deficiency. Such a mutant would be valuable for the study of plant defense and developmental mechanisms that depend on JA signaling.

An alternative to insertion knock-outs for reverse genetic analysis relies on a natural mechanism of gene silencing in response to virus attack (Waterhouse and Helliwell, 2003). Most plant viruses have single-stranded RNA genomes, which are released into the host plant cell upon infection. Double stranded RNA (dsRNA) is formed by replication of viral RNA. The presence of dsRNA triggers a plant defense response, resulting in cleavage of specific RNA by an enzyme termed Dicer. Dicer-generated ~22nt RNA segments are then associated with an endonuclease forming a complex that will cut any RNA that hybridizes to this complex. Therefore, in nature, this process culminates in the degradation of homologous viral RNA molecule. Double stranded RNA can be artificially generated in the plant cells by expressing a construct leading to the

transcription of a mRNA that complements itself to form a hairpin structure. This hairpin is recognized as dsRNA and the plant response described above is activated, leading to the degradation of the corresponding endogenous mRNA. This RNAi gene silencing approach can generate plants with mRNA levels ranging from near wild type to undetectable (Waterhouse and Helliwell 2003). One strategy for silencing expression of all genes in clade D could take advantage of their similarity in sequence by creation of an RNAi silencing construct targeted to a region of sequence conserved in all clade D genes, potentially leading to silencing all five Arabidopsis *ACLL* genes in this clade, thus eliminating the entire suite of putatively similar enzymatic function. Alternatively, genes in clade D can be heterologously expressed for the production of ACLL enzymes for testing for OPDA/OPC8 activity *in vitro*.

As seen in Chapter 3, I identified a network of genes that are co-expressed with *ArathACLL4* (Figure 3.6). In addition to co-expressed genes that encode enzymes that may be in the same biochemical pathway, this kind of analysis can also identify potential regulatory genes. Central within the *ArathACLL4* network is a transcription factor encoded by At3g44260. This CCR-NOT transcription complex protein could potentially be involved in regulation of the octadecanoid pathway biosynthetic genes. A knock-out mutant for this gene could be used to test changes in expression of *ACLL4*, the remaining 4 gene members of clade D, and the co-expressed octadecanoid pathway genes. If this hypothesis was confirmed and At3g44260 is a central regulator of the octadecanoid pathway, then a single mutant would be sufficient to generate plants deficient in JA biosynthesis via the octadecanoid pathway. This mutant could also potentially shed light

on the functions of other clade D *ACLL* genes, or alternatively, be useful for understanding alternative pathways of JA synthesis or alternative defense and regulatory mechanisms in Arabidopsis.

Similarly, in Chapter 4, I discussed the possibility of the transcription factor gene *MYB99* being a regulator of *ACLL5* and other genes co-expressed in the tapetum that encode enzymes involved in exine biosynthesis. It would be interesting test *MYB99* loss of function mutants for loss of expression of potential target genes such as *ACLL5* and co-expressed genes, and also to test the *myb99* mutant phenotype with regards to male sterility and impairment of pollen development. This approach would help test the hypothesis that the *ACLL5* co-expressed genes are in fact co-regulated, and provide support for roles in a common biochemical pathway required for exine biosynthesis during pollen development. In addition, phenotypic analysis of null mutants of co-expressed genes, and especially the effect of such mutations on pollen development and male fertility would confirm their importance in this biological process and would be consistent with functions in the same biochemical pathway as *ACLL5*. Experiments are underway in the Douglas Lab to address these questions.

5.2.2 *acll5* mutant complementation

One well accepted method for proving gene function is by genetic complementation of a loss of function mutation. As described in Chapter 4, the loss of function mutant *acll5-1* yields a male-sterile phenotype. Although co-segregation analysis shows that the mutation is tightly linked to the phenotype, complementation of the *acll5* phenotype with

a wild-type *ACLL5* gene would provide undisputable proof that the mutation in *ACLL5* is responsible for the male sterile mutant phenotype.

In order to accomplish this, I have built a construct composed of *ACLL5* genomic region driven by the 2Kb *ACLL5* native promoter and cloned in the pGreen 0029 T-DNA vector (Hellens *et al.*, 2000). In preliminary work, plants heterozygous for the *acll5* mutation were transformed with an *Agrobacterium* carrying this construct. However, no transformants were obtained, and this experiment will have to be repeated using a different binary vector such as pCambia (http://www.cambia.org). If the male sterile phenotype is indeed due to disruption of the *ACLL5* gene, T1 transformant plants derived from heterozygous background are predicted to have wild-type (male fertile) phenotypes, and this trait would be inherited by T2 progeny of such lines.

Based on the results of the comparative genomics approach, it would be interesting test the ability of *ACLL5* homologues from other plant species to complement the Arabidopsis *acll5*-1 mutation. For example *PoprtACLL13*, which is preferentially expressed in the male flowers (Chapters 3 and Chapter 4), and the single rice homologue in clade A could be tested. This heterologous complementation approach is a valuable tool to confirm gene function in organisms for which there are less information and/or resources available than in Arabidopsis. For example, it can take several years for a poplar tree to flower, but by testing complementation of the Arabidopsis mutant it would take only months to test whether *PoptrACLL13* has a function similar or identical to *ArathACLL5* in pollen development.

5.2.3 Mutant studies in poplar and other plant species

In poplar, reverse genetic approaches employing RNAi induced gene silencing are possible. Despite the necessity of generating transgenic plants by the labor intensive and time consuming process of plant regeneration in tissue culture following co-cultivation of leaf discs with *Agrobacterium*, it is possible to obtain transgenic poplar plants with reduced levels of gene expression via this technology (Meyer *et al.*, 2004). Given the knowledge that *ArathACLL4* encodes an OPDA:CoA ligase that functions in JA biosynthesis (Koo *et al.*, 2006), and that *PoptrACLL4* and *PoptrACLL5* may have the same function based on expression and phylogenetic analyses (Chapter 3), an RNAi strategy could be used to generate transgenic poplar plants with reduced or null levels of *PoptrACLL4/5* expression. Since, in contrast to Arabidopsis, these two highly similar genes appear to be the only genes encoding OPDA:CoA ligase in poplar, such transgenic plants would be predicted to have reduced or undetectable levels of JA. Such plants would be valuable for the study of the role of JA in plant defense against herbivory and other stresses, of the postulated roles for alternative signaling molecules for plant defense *in vivo*, such as upstream intermediates in the octadecanoid pathway that are generated in the peroxisome and have been postulated to play roles in defense signaling (Stintzi *et al.*, 2001). A similar approach using the Arabidopsis *opr3* mutant, defective in the isoform of OPDA reductase required for JA biosynthesis, has been successfully used to determine the role of OPDA in wound-induced signal transduction. Genes previously known to be JA-dependant were up-regulated in the *opr3* mutant (Stintzi *et al.*, 2001).

Using a similar strategy, RNAi silencing of *PoptrACLL13* in transgenic poplar would be predicted to lead to defects in pollen development and male fertility in that plant. While it normally takes five or more years for poplar to flower, it has recently been shown that over-expression of either of the two poplar *FT* genes leads to rapidly accelerated flowering, sometimes observed even in tissue culture (Bohlenius *et al.*, 2006; Hsu *et al.*, 2006), making it feasible to express RNAi constructs in early flowering poplar plants. If this approach were successful in generating male sterile poplar trees, it could be used as a tool to generate pollen-less trees for biotechnology purposes (C. de Azevedo Souza and C.J. Douglas, US Provisional Patent "A method for generating male sterile plants"). This technology could be particularly useful in trees such as poplar that are wind pollinated, since they release large amounts of pollen. Male sterile transgenic trees that fail to form pollen would be desirable to prevent gene dispersal from transgenic or exotic poplar trees via cross-pollination to wild relatives.

5.2.4 Biochemical characterization Poplar clade D ACLLs

In Chapter 3 I suggested a function in JA biosynthesis and plant defense for clade D poplar *ACLL* homologues, which are the closest relatives to *ArathACLL4*, which encodes an OPDA/OPC8-CoA ligase required for JA biosynthesis (Koo *et al.*, 2006). The hypothesis that the poplar homologues *PoptrACLL4* and *PoptrACLL5* also encode OPDA/OPC8-CoA ligases is supported by their strong up-regulated expression after wounding, herbivory, and MeJA treatments, and the experimentally demonstrated peroxisomal localization of PoptrACLL5. However, characterization of the PoptrACLL5

enzyme by expression of the recombinant protein in *E. coli* was unsuccessful (data not shown).

The question of the enzymatic properties of PoptrACLL4/5, and particularly its ability to use OPDA/OPC8 as a substrate is still very interesting in light of the known activity of enzyme encoded by the *ArathACLL4* homologue. The use a eukaryotic host for heterologous expression such as yeast instead of *E. coli* might be a useful alternative to address this question.

5.2.5 4CL/ACLL structural information and identification of substrates

With the exception of ArathACLL4 and ArathACLL9, no published experimental information is available regarding ACLL substrates. There is currently no protein structural information available for 4CL, not to mention ACLL enzymes, which could aid in making predictions as substrates and structural features relevant to substrate predictions. Two known crystal structures of adenylate-forming enzymes are those of the firefly *Photinus pyralis* luciferase (EC 1.13.12.7) (Conti *et al.*, 1996) and the bacterium *Brevibacillus brevis* gramicidin *S*-synthetase 1 (PheA; CAA33603) (Conti *et al.*, 1997). While these enzymes share limited sequence identity to 4CL (less than 20% identity), information from these structures has allowed prediction of the nature of the 4CL substrate binding pocket, and has been used to predict amino acid residues that determine 4CL substrate binding (Stuible and Kombrink, 2001). Based on information obtained from the crystal structure of the phenylalanine-activating domain of PheA, and amino-acid sequence comparisons between PheA and 4CL, 10 amino acid residues were

identified that could form the 4CL substrate binding pocket (Stuible and Kombrink, 2001). The authors took advantage of the fact that a single member of the Arabidopsis 4CL family (4CL2) is unable to accept ferulate as a substrate, allowing them to pinpoint amino acid residues absent, or not conserved, in the 4CL2 putative substrate binding pocket, and therefore are candidates for causing the lack of activity of 4CL2 towards ferulate.

In later studies, the same group used homology modeling to predict 4CL2 tertiary structure by alignment to the known PheA structure (Schneider *et al.*, 2003). Although the structure of enzyme luciferase, which is more closely related to 4CL and PheA, is also known, PheA was chosen for this study due to the similarity of substrate structures. This allowed a more accurate prediction the orientation of the 4CL substrates in the putative binding pocket. Using the 3D model, 12 amino acids were identified, in the substrate binding pocket, that were predicted to be close enough to the substrate to form electron interactions. A site directed mutagenesis approach of targeted amino acids was used successfully to allow the design of ferulic acid, sinapic acid, and cinnamic acid-activating At4CL2 variants (Schneider *et al.*, 2003). The previous knowledge of 4CL substrates was therefore indispensable for testing the hypothesis of amino acids responsible for substrate specificity in 4CL2.

In the future, one might be able to use this same approach to make predictions regarding the substrate specificities of ACLL enzymes. A crystal structure of ArathACLL4 would be particularly advantageous given that the substrate is known. This data would show

which amino acids are important for substrate recognition. Therefore, homology modeling of additional members of clade D ACLLs, including poplar representatives, could provide further insights into similarities of putative substrate binding pockets and indicating if OPDA/OPC8 is a suitable substrate for these enzymes. Further knowledge of crystal structures of closely related enzymes such as 4CLs will also provide insights into how similar the binding pockets of ACLLs are to 4CLs. It would be, for example, interesting to compare a 4CL structure with that of ArathACLL5 in clade A, which could have a phenolic substrate based on co-expression analysis, as discussed in Chapter 4.

A large scale screening approach for identification of ACLL substrates *in vitro* has been used successfully for ArathACLL9 (Schneider *et al.*, 2005). The method consists of using the property of the adenylate-forming enzyme luciferase, in the presence of the substrate luciferin and ATP, for generation of light involving ATP hydrolysis and formation of an AMP-bound substrate intermediate. Since ACLL enzymes, as adenylate-forming enzymes, also require ATP, luciferase activity is used in a visual assay for ATP depletion. Thus, loss of luciferase activity when co-incubated with a recombinant ACLL enzyme and a potential substrate indicates potential ACLL activity against the substrate, if the activity is high enough to deplete the ATP concentration. In theory all ACLLs can be screened using this method. Although this approach can be a powerful tool for identification of substrates, one limitation of this method lies in the necessity of having a large enough library of potential substrates to find potential substrates for which an ACLL has high enough activity to deplete ATP in the assay. Since ACLLs may have very specific substrate preferences, this could limit the chances of successfully finding the

correct substrate using this assay. Another variable that could confound the assay is possible necessity of adjusting enzyme assay conditions. However, as more is learned about ACLLs, the multitude of possible substrates will be narrowed to fewer more likely candidates, making this a potential powerful approach to screen candidate substrates.

5.2.6 Continuous mining of data

With the ongoing efforts to functionally characterize all genes in Arabidopsis, there is an enormous amount of information continually generated about single genes, biochemical pathways and biological processes. I found that co-expression studies are an especially useful tool for generating hypotheses regarding biological and biochemical roles for genes of unknown function. The reverse is also true. Given a set of genes with known functions in a common process or pathway, this approach is useful to identify genes encoding enzymes or other proteins that function in uncharacterized parts of the pathway or process. The more expression information there is available, the more robust the data pointing to these relationships should become. Therefore, in the near future data on the networks of genes co-expressed with *ACLLs* in clades that are still poorly characterized might become easier to interpret, and may allow us to make educated guesses regarding ACLL functions. For example, in Chapter 3 I showed that *ArathACLL3* is part of a large set of co-expressed genes, but no obvious functional relationships to *ArathACLL3* could be derived from that the data. However, as more expression profiling experiments under different environmental, developmental, and genotype-specific conditions are performed, this network of relationships may become more clear, revealing potential biochemical partners of ArathACLL3.

Libraries of insertional mutants, as discussed in the section 5.2.1 of this chapter, are also constantly being enlarged and more lines of DNA insertions in selected ACLLs might become available. For example, if an additional mutant allele for ArathACLL5 becomes available, it would be desirable to verify whether this mutant has the same male sterile phenotype as *acll5-1*. This would be an additional indication that the *acll5-1* mutation is really the cause of the observed phenotype. This type of evidence would eliminate the necessity of mutant rescue by genetic complementation as suggested in section 5.2.2.

Given the above, a constant mining of publicly available data is important to support functional analysis of genes of unknown function such as *ACLL* genes. The large amount of knowledge and large numbers of tools available are possibly the best advantages when working with a model organism such as Arabidopsis.

More information becomes continually available in the published literature that provides direct or indirect insights into a biological question of interest. For example, shortly before completing this chapter, the biological and biochemical function of the Arabidopsis *CYP703A2* was published (Morant *et al.*, 2007). CYP703A2 is a single copy plant-specific P450 enzyme that was shown to be specifically involved in pollen development. The expression pattern of *CYP703A2* is the same as that of *ArathACLL5* (expressed during pollen formation) and belongs to our list of genes co-expressed with *ArathACLL5* (Chapter4). Mutants lacking expression of *CYP703A2* have reduced male fertility and impaired pollen wall development with the absence of exine. Biochemical

characterization of *CYP703A2* heterologously expressed in yeast identified lauric acid and in-chain hydroxy lauric acid as the substrate and product, respectively. It is known that sporopollenin, major component of the pollen exine, is composed of fatty acids and phenolic units, so it should be expected that genes encoding these compounds are co-expressed in the same tissue, during the same developmental stage. An interesting aspect of the *CYP703A2* mutant phenotype is the absence of detectible phenylpropanoids in the sporopollenin, which could indicate that these components can only be attached to the pollen wall if polymerized with fatty acids. The substrate of ArathACLL5 remains a mystery. My data strongly indicate a role in sporopollenin production and I favor a function in the synthesis of the phenolic components due to the close relationship to true 4CLs and *ArathACLL5* co-expression with genes encoding phenylpropanoid-like enzymes. However, I cannot rule out a possible role in the synthesis of the fatty acid components of sporopollenin. Therefore, I suggest that in-chain hydroxy lauric acid is a candidate ArathACLL5 substrate that should be tested.

5.3 Final remarks

This work demonstrates the usefulness of comparative genomics in understanding the roles of particular genes in given biological systems. I have used information available from the model plant *Arabidopsis thaliana* as a tool for gene discovery and to generate functional hypothesis regarding homologous genes in *Populus* (poplar). This shows the importance of having model organisms with large repositories of information and tools available, such as public global expression data using microarrays, together with less developed model systems such as *Populus* (poplar) (Jansson and Douglas, 2007).

Information obtained from targeted expression studies in poplar, such as herbivory and MeJA induced up-regulation of gene expression and, in particular, the male flower expression facilitated by the diecious nature of the poplar species, demonstrated that comparative genomics is a two way road. Whereas one single organism might be more feasible to be explored collectively, the knowledge of more than one "model" species will allow us to progress to a holistic view of gene functions in all plant species.

The data presented in the results chapters (Chapter 3 and 4) open new routes to the study of ACLL function in plants, as it allows for various insights into the comparative genomics of gene family evolution, identifies a crucial function for *ArathACLL5* in pollen development, a key process in the perpetuation of life, and suggests new genes in this process for further studies. I hope this newly paved road becomes well traveled.

REFERENCES

Aarts, M.G.M., 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 Journal, 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.

Allina, S.M., Pri-Hadash, A., Theilmann, D.A., Ellis, B.E. and Douglas, C.J. (1998) 4-Coumarate:coenzyme A ligase in hybrid poplar. Properties of native enzymes, cDNA cloning, and analysis of recombinant enzymes. Plant Physiol, 116, 743-754.

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.

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.

Arondel, V.V., Vergnolle, C., Cantrel, C. and Kader, J. (2000) Lipid transfer proteins are encoded by a small multigene family in *Arabidopsis thaliana*. Plant Science, 157, 1-12.

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.

Becker-Andre, M., Schulze-Lefert, P. and Hahlbrock, K. (1991) Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate: CoA ligase genes in potato. J Biol Chem, 266, 8551-8559.

Beggs, C.J., Stolzer-Jehle, A. and Wellmann, E. (1985) Isoflavonoid formation as an indicator of UV stress in bean (*Phaseolus vulgaris* L.) Leaves : The significance of photorepair in assessing potential damage by increased solar UV-B Radiation. Plant Physiol, 79, 630-634.

Blanc, G., Hokamp, K. and Wolfe, K.H. (2003) A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. Genome Res, 13, 137-144.

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.

Boavida, L.C., Becker, J.D. and Feijo, J.A. (2005) The making of gametes in higher plants. Int J Dev Biol, 49, 595-614.

Bohlenius, H., Huang, T., Charbonnel-Campaa, L., Brunner, A.M., Jansson, S., Strauss, S.H. and Nilsson, O. (2006) CO/FT Regulatory Module Controls Timing of Flowering and Seasonal Growth Cessation in Trees. Science, 312, 1040-1043.

Bonaventure, G., Gfeller, A., Proebsting, W.M., Hortensteiner, S., Chetelat, A., Martinoia, E. and Farmer, E.E. (2007) A gain-of-function allele of TPC1 activates oxylipin biogenesis after leaf wounding in Arabidopsis. Plant J, 49, 889-898.

Cenzano, A., Vigliocco, A., Kraus, T. and Abdala, G. (2003) Exogenously applied jasmonic acid induces changes in apical meristem morphology of potato stolons. Ann Bot (Lond), 91, 915-919.

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.

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.

Conti, E., Franks, N.P. and Brick, P. (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. Structure, 4, 287-298.

Conti, E., Stachelhaus, T., Marahiel, M.A. and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. Embo Journal, 16, 4174-4183.

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.

Croteau, R., Kutchan T.M. and Lewis, N. (2000) Natural Products (Secondary metabolites) in Buchannan, B.B., Gruissem, W. and Jones, R. L. (Eds.) Biochemistry and Molecular Biology of Plants. American Society of Plant Physiologists, Rockville. 1250-1318.

Cukovic, D., Ehlting, J., VanZiffle, J.A. and Douglas, C.J. (2001) Structure and evolution of 4-coumarate:coenzyme A ligase (4CL) gene families. Biol Chem, 382, 645-654.

Czernic, P., Visser, B., Sun, W., Savoure, A., Deslandes, L., Marco, Y., Van Montagu, M. and Verbruggen, N. (1999) Characterization of an *Arabidopsis thaliana* receptor-like protein kinase gene activated by oxidative stress and pathogen attack. The Plant Journal, 18, 321-327.

de Azevedo Souza, C. and Douglas, C.J. A method for generating male sterile plants. Provisional Patent USPTO Priority Date Nov 3-2006 (UBC UILO reference 07-082).

Deluca, M. (1976) Firefly luciferase. Adv Enzymol Relat Areas Mol Biol, 44, 37-68.

Dixon, R.A. and Paiva, N.L. (1995) Stress-induced phenylpropanoid metabolism. Plant Cell, 7, 1085-1097.

Dong, J., Chen, C. and Chen, Z. (2003) Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response. Plant Mol Biol, 51, 21-37.

Douglas, C., Hoffmann, H., Schulz, W. and Hahlbrock, K. (1987) Structure and elicitor or u.v.-light-stimulated expression of two 4-coumarate:CoA ligase genes in parsley. Embo J, 6, 1189-1195.

Douglas, C.J. (1996) Phenylpropanoid metabolism and lignin biosynthesis: from weeds to trees. Trends in Plant Science, 1, 171-178.

Douglas, C.J. and Ehlting, J. (2005) *Arabidopsis thaliana* full genome longmer microarrays: a powerful gene discovery tool for agriculture and forestry. Transgenic Res, 14, 551-561.

Duarte, J.M., Cui, L., Wall, P.K., Zhang, Q., Zhang, X., Leebens-Mack, J., Ma, H., Altman, N. and dePamphilis, C.W. (2006) Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. Mol Biol Evol, 23, 469-478.

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. Plant J, 19, 9-20.

Ehlting, J., Mattheus, N., Aeschliman, D.S., Li, E., Hamberger, B., Cullis, I.F., Zhuang, J., Kaneda, M., Mansfield, S.D., Samuels, L., Ritland, K., Ellis, B.E., Bohlmann, J. and Douglas, C.J. (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.

Ehlting, J., Provart, N.J. and Werck-Reichhart, D. (2006) Functional annotation of the Arabidopsis P450 superfamily based on large-scale co-expression analysis. Biochem Soc Trans, 34, 1192-1198.

Ehlting, J., Shin, J.J. and Douglas, C.J. (2001) Identification of 4-coumarate:coenzyme A ligase (4CL) substrate recognition domains. Plant J, 27, 455-465.

Enjuto, M., Lumbreras, V., Marin, C. and Boronat, A. (1995) Expression of the Arabidopsis HMG2 Gene, Encoding 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase, Is Restricted to Meristematic and Floral Tissues. Plant Cell, 7, 517-527.

Facchini, P.J. and St-Pierre, B. (2005) Synthesis and trafficking of alkaloid biosynthetic enzymes. Current Opinion in Plant Biology, 8, 657-666.

Farmer, E.E., Almeras, E. and Krishnamurthy, V. (2003) Jasmonates and related oxylipins in plant responses to pathogenesis and herbivory. Curr Opin Plant Biol, 6, 372-378.

Feys, B., Benedetti, C.E., Penfold, C.N. and Turner, J.G. (1994) Arabidopsis mutants selected for resistance to the phytotoxin coronatine are male sterile, insensitive to methyl jasmonate, and resistant to a bacterial pathogen. Plant Cell, 6, 751-759.

Goldberg, R.B., Beals, T.P. and Sanders, P.M. (1993) Anther development: basic principles and practical applications. Plant Cell, 5, 1217-1229.

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., Lamb, C.J., Purwin, C., Ebel, J., Fautz, E. and Schafer, E. (1981) Rapid response of suspension-cultured parsley cells to the elicitor from Phytophthora megasperma var. sojae: Induction of the Enzymes of General Phenylpropanoid Metabolism. Plant Physiol, 67, 768-773.

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. Proc Natl Acad Sci U S A, 101, 2209-2214.

Hamberger, B., Ellis M., Friedmann M., de Azevedo Souza, C., Barbazuk, 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, in press.

Harding, S.A., Leshkevich, J., Chiang, V.L. and Tsai, C.J. (2002) Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme a ligases with spatially distinct metabolic roles in quaking aspen. Plant Physiol, 128, 428-438.

Hauffe, K.D., Paszkowski, U., Schulze-Lefert, P., Hahlbrock, K., Dangl, J.L. and Douglas, C.J. (1991) A parsley 4CL-1 promoter fragment specifies complex expression patterns in transgenic tobacco. Plant Cell, 3, 435-443.

Hellens, R.P., Edwards, E.A., Leyland, N.R., Bean, S. and Mullineaux, P.M. (2000) pGreen: a versatile and flexible binary Ti vector for *Agrobacterium*-mediated plant transformation. Plant Mol Biol, 42, 819-832.

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.

Hooker, T.S., Millar, A.A., and Kunst, L. (2002). Significance of the expression of the CER6 condensing enzyme for cuticular wax production in Arabidopsis. Plant Physiol, 129, 1568–1580.

Hsu, C.-Y., Liu, Y., Luthe, D.S. and Yuceer, C. (2006) Poplar FT2 shortens the juvenile phase and promotes seasonal flowering. Plant Cell, 18, 1846-1861.

Hu, W.J., Kawaoka, A., Tsai, C.J., Lung, J., Osakabe, K., Ebinuma, H. and Chiang, V.L. (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). Proc Natl Acad Sci U S A, 95, 5407-5412.

Humphreys, J.M. and Chapple, C. (2002) Rewriting the lignin roadmap. Current Opinion in Plant Biology, 5, 224-229.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature, 436, 793-800.

Itoh, T., Tanaka, T., Barrero, R.A., Yamasaki, C., Fujii, Y., Hilton, P.B., Antonio, B.A., Aono, H., Apweiler, R., Bruskiewich, R., Bureau, T., Burr, F., Costa de Oliveira, A., Fuks, G., Habara, T., Haberer, G., Han, B., Harada, E., Hiraki, A.T., Hirochika, H., Hoen, D., Hokari, H., Hosokawa, S., Hsing, Y.I., Ikawa, H., Ikeo, K., Imanishi, T., Ito, Y., Jaiswal, P., Kanno, M., Kawahara, Y., Kawamura, T., Kawashima, H., Khurana, J.P., Kikuchi, S., Komatsu, S., Koyanagi, K.O., Kubooka, H., Lieberherr, D., Lin, Y.C., Lonsdale, D., Matsumoto, T., Matsuya, A., McCombie, W.R., Messing, J., Miyao, A., Mulder, N., Nagamura, Y., Nam, J., Namiki, N., Numa, H., Nurimoto, S., O'Donovan, C., Ohyanagi, H., Okido, T., Oota, S., Osato, N., Palmer, L.E., Quetier, F., Raghuvanshi, S., Saichi, N., Sakai, H., Sakai, Y.,

Sakata, K., Sakurai, T., Sato, F., Sato, Y., Schoof, H., Seki, M., Shibata, M., Shimizu, Y., Shinozaki, K., Shinso, Y., Singh, N.K., Smith-White, B., Takeda, J., Tanino, M., Tatusova, T., Thongjuea, S., Todokoro, F., Tsugane, M., Tyagi, A.K., Vanavichit, A., Wang, A., Wing, R.A., Yamaguchi, K., Yamamoto, M., Yamamoto, N., Yu, Y., Zhang, H., Zhao, Q., Higo, K., Burr, B., Gojobori, T. and Sasaki, T. (2007) Curated genome annotation of *Oryza sativa* ssp. japonica and comparative genome analysis with *Arabidopsis thaliana*. Genome Res, 17, 175-183.

Jansson, S. and Douglas, C.J. (2007) *Populus*: a model system for Plant Biology. Annu Rev Plant Biol, 58, 435-458.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. Embo J, 6, 3901-3907.

Jensen, A.B., Raventos, D. and Mundy, J. (2002) Fusion genetic analysis of jasmonate-signalling mutants in Arabidopsis. Plant J, 29, 595-606.

Ke, J., Behal, R.H., Back, S.L., Nikolau, B.J., Wurtele, E.S. and Oliver, D.J. (2000) The role of pyruvate dehydrogenase and acetyl-coenzyme A synthetase in fatty acid synthesis in developing Arabidopsis seeds. Plant Physiol, 123, 497-508.

Koo, A.J., Chung, H.S., Kobayashi, Y. and Howe, G.A. (2006) Identification of a peroxisomal acyl-activating enzyme involved in the biosynthesis of jasmonic acid in Arabidopsis. J Biol Chem, 281, 33511-33520.

Koopmann, E., Logemann, E. and Hahlbrock, K. (1999) Regulation and functional expression of cinnamate 4-hydroxylase from parsley. Plant Physiol., 119, 49-56.

Kunst, L., Klenz, J.E., Martinez-Zapater, J. and Haughn, G.W. (1989) AP2 gene determines the identity of perianth organs in flowers of *Arabidopsis thaliana*. Plant Cell, 1, 1195-1208.

Kurata, T., Kawabata-Awai, C., Sakuradani, E., Shimizu, S., Okada, K. and Wada, T. (2003) The YORE-YORE gene regulates multiple aspects of epidermal cell differentiation in Arabidopsis. Plant J, 36, 55-66.

Lauvergeat, V., Lacomme, C., Lacombe, E., Lasserre, E., Roby, D. and Grima-Pettenati, J. (2001) Two cinnamoyl-CoA reductase (CCR) genes from *Arabidopsis thaliana* are differentially expressed during development and in response to infection with pathogenic bacteria. Phytochemistry, 57, 1187-1195.

Lee, D. and Douglas, C.J. (1996) Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family. cDNA structure, gene inheritance and expression, and properties of recombinant proteins. Plant Physiol, 112, 193-205.

Li, C., Schilmiller, A.L., Liu, G., Lee, G.I., Jayanty, S., Sageman, C., Vrebalov, J., Giovannoni, J.J., Yagi, K., Kobayashi, Y. and Howe, G.A. (2005) Role of beta-oxidation in jasmonate biosynthesis and systemic wound signaling in tomato. Plant Cell, 17, 971-986.

Li, L., Zhao, Y., McCaig, B.C., Wingerd, B.A., Wang, J., Whalon, M.E., Pichersky, E. and Howe, G.A. (2004) The tomato homolog of CORONATINE-INSENSITIVE1 is required for the maternal control of seed maturation, jasmonate-signaled defense responses, and glandular trichome development. Plant Cell, 16, 126-143.

Liechti, R. and Farmer, E.E. (2002) The jasmonate pathway. Science, 296, 1649-1650.

Lindermayr, C., Mollers, B., Fliegmann, J., Uhlmann, A., Lottspeich, F., Meimberg, H. and Ebel, J. (2002) Divergent members of a soybean (*Glycine max* L.) 4-coumarate:coenzyme A ligase gene family. Eur J Biochem, 269, 1304-1315.

Lozoya, E., Hoffmann, H., Douglas, C., Schulz, W., Scheel, D. and Hahlbrock, K. (1988) Primary structures and catalytic properties of isoenzymes encoded by the two 4-coumarate: CoA ligase genes in parsley. Eur J Biochem, 176, 661-667.

Ma, H. (2005) Molecular genetic analyses of microsporogenesis and microgametogenesis in flowering plants. Annu Rev Plant Biol, 56, 393-434.

Martin, D., Tholl, D., Gershenzon, J. and Bohlmann, J. (2002) Methyl jasmonate induces traumatic resin ducts, terpenoid resin biosynthesis, and terpenoid accumulation in developing xylem of norway spruce stems. Plant Physiol., 129, 1003-1018.

Matsuda, N., Tsuchiya, T., Kishitani, S., Tanaka, Y. and Toriyama, K. (1996) Partial male sterility in transgenic tobacco carrying antisense and sense PAL cDNA under the control of a tapetum-specific promoter. Plant and Cell Physiology, 37, 215-222.

Meyer, S., Nowak, K., Sharma, V.K., Schulze, J., Mendel, R.R. and Hansch, R. (2004) Vectors for RNAi technology in poplar. Plant Biology, 6, 100-103.

Moore, R.C. and Purugganan, M.D. (2005) The evolutionary dynamics of plant duplicate genes. Current Opinion in Plant Biology, 8, 122-128.

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, tpc.106.045948.

Nielsen, R. (2006) Comparative genomics: difference of expression. Nature, 440, 161.

Nishiyama, T., Fujita, T., Shin, I.T., Seki, M., Nishide, H., Uchiyama, I., Kamiya, A.,

Carninci, P., Hayashizaki, Y., Shinozaki, K., Kohara, Y. and Hasebe, M. (2003) Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. Proc Natl Acad Sci U S A, 100, 8007-8012.

Noel, J.P., Austin, M.B. and Bomati, E.K. (2005) Structure-function relationships in plant phenylpropanoid biosynthesis. Curr Opin Plant Biol, 8, 249-253.

Nyathi, Y. and Baker, A. (2006) Plant peroxisomes as a source of signalling molecules. Biochim Biophys Acta, 1763, 1478-1495.

Paterson, A.H. (2006) Leafing through the genomes of our major crop plants: strategies for capturing unique information. Nat Rev Genet, 7, 174-184.

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 Physiology, 127, 1739-1749.

Persson, S., Wei, H., Milne, J., Page, G.P. and Somerville, C.R. (2005) Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proc Natl Acad Sci U S A, 102, 8633-8638.

Pighin, J.A., Zheng, H., Balakshin, L.J., Goodman, I.P., Western, T.L., Jetter, R., Kunst, L. and Samuels, A.L. (2004) Plant cuticular lipid export requires an ABC transporter. Science, 306, 702-704.

Raes, J., Rohde, A., Christensen, J.H., Van de Peer, Y. and Boerjan, W. (2003) Genome-wide characterization of the lignification toolbox in Arabidopsis. Plant Physiol., 133, 1051-1071.

Ralph, S., Oddy, C., Cooper, D., Yueh, H., Jancsik, S., Kolosova, N., Philippe, R.N., Aeschliman, D., White, R., Huber, D., Ritland, C.E., Benoit, F., Rigby, T., Nantel, A., Butterfield, Y.S.N., Kirkpatrick, R., Chun, E., Liu, J., Palmquist, D., Wynhoven, B., Stott, J., Yang, G., Barber, S., Holt, R.A., Siddiqui, A., Jones, S.J.M., Marra, M.A., Ellis, B.E., Douglas, C.J., Ritland, K. & Bohlmann, J. (2006) Genomics of hybrid poplar (*Populus trichocarpa x deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): Normalized and full-length cDNA libraries, expressed sequence tags (ESTs), and a cDNA microarray for the study of insect-induced defenses in poplar. Molecular Ecology, *15*,1275-1297.

Raven, P.H., Evert, R.F. and Curtis, H. (1996) Biology of Plants 5th ed edn. New York, NY: Worth Publishers.

Reumann, S. (2004) Specification of the peroxisome targeting signals type 1 and type 2 of plant peroxisomes by bioinformatics analyses. Plant Physiol, 135, 783-800.

Reumann, S., Ma, C., Lemke, S. and Babujee, L. (2004) AraPerox. A database of putative Arabidopsis proteins from plant peroxisomes. Plant Physiol, 136, 2587-2608.

Ritter, H. and Schulz, G.E. (2004) Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase. Plant Cell, 16, 3426-3436.

Rogers, L.A. and Campbell, M.M. (2004) The genetic control of lignin deposition during plant growth and development. New Phytologist, 164, 17-30.

Sablowski, R.W., Moyano, E., Culianez-Macia, F.A., Schuch, W., Martin, C. and Bevan, M. (1994) A flower-specific Myb protein activates transcription of phenylpropanoid biosynthetic genes. Embo J, 13, 128-137.

Sanchez-Fernandez, R., Davies, T.G., Coleman, J.O. and Rea, P.A. (2001) The *Arabidopsis thaliana* ABC protein superfamily, a complete inventory. J Biol Chem, 276, 30231-30244.

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.

Sasaki-Sekimoto, Y., Taki, N., Obayashi, T., Aono, M., Matsumoto, F., Sakurai, N., Suzuki, H., Hirai, M.Y., Noji, M., Saito, K., Masuda, T., Takamiya, K., Shibata, D. and Ohta, H. (2005) Coordinated activation of metabolic pathways for antioxidants and defence compounds by jasmonates and their roles in stress tolerance in Arabidopsis. Plant J, 44, 653-668.

Schaller, F., Schaller, A. and Stintzi, A. (2004) Biosynthesis and metabolism of jasmonates. Journal of Plant Growth Regulation, 23, 179-199.

Schilmiller, A.L., Koo, A.J. and Howe, G.A. (2007) Functional diversification of acyl-coenzyme a oxidases in jasmonic acid biosynthesis and action. Plant Physiol, 143, 812-824.

Schneider, K., Hovel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E. and Stuible, H.P. (2003) The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase. Proc Natl Acad Sci U S A, 100, 8601-8606.

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. J Biol Chem, 280, 13962-13972.

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

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. Plant Physiol, 132, 1065-1076.

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 Physiol, 129, 1710-1722.

Soltani, B.M., Ehlting, J. and Douglas, C.J. (2006) Genetic analysis and epigenetic silencing of At4CL1 and At4CL2 expression in transgenic Arabidopsis. Biotechnol J, 1, 1124-1136.

Sommer, J.M., Cheng, Q.L., Keller, G.A. and Wang, C.C. (1992) In vivo import of firefly luciferase into the glycosomes of *Trypanosoma brucei* and mutational analysis of the C-terminal targeting signal. Mol Biol Cell, 3, 749-759.

Staswick, P.E., Su, W. and Howell, S.H. (1992) Methyl jasmonate inhibition of root growth and induction of a leaf protein are decreased in an *Arabidopsis thaliana* mutant. Proc Natl Acad Sci U S A, 89, 6837-6840.

Staswick, P.E., Tiryaki, I. and Rowe, M.L. (2002) Jasmonate response locus JAR1 and several related Arabidopsis genes encode enzymes of the firefly luciferase superfamily that show activity on jasmonic, salicylic, and indole-3-acetic acids in an assay for adenylation. Plant Cell, 14, 1405-1415.

Stenzel, I., Hause, B., Miersch, O., Kurz, T., Maucher, H., Weichert, H., Ziegler, J., Feussner, I. and Wasternack, C. (2003) Jasmonate biosynthesis and the allene oxide cyclase family of *Arabidopsis thaliana*. Plant Mol Biol, 51, 895-911.

Steppuhn, A., Gase, K., Krock, B., Halitschke, R. and Baldwin, I.T. (2004) Nicotine's defensive function in nature. PLoS Biology, 2, e217.

Stintzi, A., Weber, H., Reymond, P., Browse, J. and Farmer, E.E. (2001) Plant defense in the absence of jasmonic acid: the role of cyclopentenones. Proc Natl Acad Sci U S A, 98, 12837-12842.

Stuible, H.P. and Kombrink, E. (2001) Identification of the substrate specificity-conferring amino acid residues of 4-coumarate:coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. J Biol Chem, 276, 26893-26897.

Taylor, P.E., Glover, J.A., Lavithis, M., Craig, S., Singh, M.B., Knox, R.B., Dennis, E.S. and Chaudhury, A.M. (1998) Genetic control of male fertility in *Arabidopsis thaliana*: structural analyses of postmeiotic developmental mutants. Planta, 205, 492-505.

The Arabidopsis Genome Initiative, A. (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature, 408, 796-815.

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.

Turner, G., Gershenzon, J., Nielson, E.E., Froehlich, J.E. and Croteau, R. (1999) Limonene synthase, the enzyme responsible for monoterpene biosynthesis in peppermint, is localized to leucoplasts of oil gland secretory cells. Plant Physiol., 120, 879-886.

Turner, J.E., Greville, K., Murphy, E.C. and Hooks, M.A. (2005) Characterization of Arabidopsis fluoroacetate-resistant mutants reveals the principal mechanism of acetate activation for entry into the glyoxylate cycle. J. Biol. Chem., 280, 2780-2787.

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R.R., Bhalerao, R.P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G.L., Cooper, D., Coutinho, P.M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J.C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D.R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C.J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y. and Rokhsar, D. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science, 313, 1596-1604.

V. Batagelj, ., A.M. and Springer, B. (2003) Pajek - Analysis and visualization of large networks in Jünger, M., Mutzel, P., (Eds.) Graph Drawing Software. Springer, Berlin. 77-103.

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

Varbanova, M.P., Atanassov, A.I. and Atanassov, I.I. (2003) Anther-specific coumarate CoA ligase-like gene from *Nicotiana sylvestris* expressed during uninucleate microspore development. Plant Science, 164, 525-530.

Vizcay-Barrena, G. and Wilson, Z.A. (2006) Altered tapetal PCD and pollen wall development in the Arabidopsis *ms1* mutant. J Exp Bot, 57, 2709-2717.

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.

Waterhouse, P.M. and Helliwell, C.A. (2003) Exploring plant genomes by RNA-induced gene silencing. Nat Rev Genet, 4, 29-38.

Yau, C.P., Zhuang, C.X., Zee, S.Y. and Yip, W.K. (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, 65-74.

Ylstra, B., Muskens, M. and Van Tunen, A.J. (1996) Flavonols are not essential for fertilization in *Arabidopsis thaliana*. Plant Mol Biol, 32, 1155-1158.

Yuan, Q., Ouyang, S., Liu, J., Suh, B., Cheung, F., Sultana, R., Lee, D., Quackenbush, J. and Buell, C.R. (2003) The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. Nucl. Acids Res., 31, 229-233.

Zhang, W., Sun, Y., Timofejeva, L., Chen, C., Grossniklaus, U. and Ma, H. (2006) Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. Development, 133, 3085-3095.

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. Plant Physiol, 113, 65-74.

APPENDIX 1

Genes co-expressed with ACLLs in Arabidopsis public microarray experiment datasets (http://prime.psc.riken.jp/?action=coexpression_index)

CLADE A

ArathACLL5 (At1g62940) all data v3 (1388)

0.908	At4g14080	glycosyl hydrolase family 17 protein / anther-specific protein (A6) identical to probable glucan endo-1,3-beta-glucosidase A6...
0.905	At4g20420	tapetum-specific protein-related similar to SaTAP 35 [Sinapis alba] GI:408108
0.891	At5g07230	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein identical to tapetum-specific protein A9...
0.879	At3g07450	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein similar to cysteine-rich 5B protein - Lycopersicon...
0.878	At4g34850	chalcone and stilbene synthase family protein similar to chalcone synthase homolog PrChS1, Pinus radiata, gb:U90341; similar to...
0.876	At3g42960	alcohol dehydrogenase (ATA1) identical to alcohol dehydrogenase (ATA1) GI:2501781 from [Arabidopsis thaliana]
0.873	At5g62080	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein similar to tapetum-specific protein a9 precursor...
0.868	At3g52130	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein similar to cysteine-rich 5B protein - Lycopersicon...
0.863	At1g01280	cytochrome P450 family protein similar to cytochrome P450 GB:BAA92894 GI:7339658 from [Petunia hybrida]
0.862	At5g16920	expressed protein
0.861	At3g11980	male sterility protein 2 (MS2) identical to male sterility protein 2 (MS2) SP:Q08891 (Arabidopsis thaliana)
0.854	At1g69500	cytochrome P450 family protein similar to Cytochrome P450 86A2 (SP:O23066) [Arabidopsis thaliana]contains Pfam profile:...
0.851	At1g61070	plant defensin-fusion protein, putative (PDF2.4) plant defensin protein family member, personal communication, Bart Thomma...
0.84	At3g13220	ABC transporter family protein contains Pfam profile: PF00005 ABC transporter; similar to white protein GB:Q27256 [Anopheles...
0.821	At3g23770	glycosyl hydrolase family 17 protein similar to A6 anther-specific protein SP:Q06915 [Arabidopsis thaliana]
0.788	At2g42940	DNA-binding family protein contains a AT hook motif (DNA binding motifs with a preference for A/T rich regions), Pfam:PF02178
0.781	At1g02813	expressed protein contains Pfam profile PF04398: Protein of unknown function, DUF538
0.773	At4g29980	expressed protein
0.771	At2g16910	basic helix-loop-helix (bHLH) family protein
0.77	At1g02050	chalcone and stilbene synthase family protein Similar to rice chalcone synthase homolog, gp|U90341|2507617 and anther specific...
0.77	At1g33430	galactosyltransferase family protein contains Pfam profile: PF01762 galactosyltransferase
0.763	At1g20150	subtilase family protein similar to subtilisin-type protease precursor GI:14150446 from [Glycine max]
0.742	At5g52160	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam protease inhibitor/seed storage/LTP...
0.739	At3g57620	glyoxal oxidase-related contains similarity to glyoxal oxidase precursor [Phanerochaete chrysosporium] gi|1050302|gb|AAA87594
0.727	At4g28395	lipid transfer protein, putative identical to anther-specific gene ATA7 [gi:2746339]; contains Pfam protease inhibitor/seed...
0.717	At4g12920	aspartyl protease family protein low similarity to CND41, chloroplast nucleoid DNA binding protein [Nicotiana tabacum]...
0.713	At1g75790	multi-copper oxidase type I family protein contains Pfam profile: PF00394 Multicopper oxidase
0.707	At1g06170	basic helix-loop-helix (bHLH) family protein contains Pfam profile:PF00010 helix-loop-helix DNA-binding domain
0.702	At3g52160	beta-ketoacyl-CoA synthase family protein beta-ketoacyl-CoA synthase - Simmondsia chinensis,PID:g1045614
0.685	At1g03390	transferase family protein similar to anthranilate N-hydroxycinnamoyl/benzoyltransferase from Dianthus caryophyllus...
0.68	At5g13380	auxin-responsive GH3 family protein similar to auxin-responsive GH3 product [Glycine max] GI:18591; contains Pfam profile...
0.678	At1g67990	caffeoyl-CoA 3-O-methyltransferase, putative similar to GI:2960356 [Populus balsamifera subsp. trichocarpa], GI:684942...
0.669	At5g24820	aspartyl protease family protein low similarity to CND41, chloroplast nucleoid DNA binding protein [Nicotiana tabacum]...
0.655	At5g48210	expressed protein
0.652	At1g71160	beta-ketoacyl-CoA synthase family protein similar to fatty acid elongase 3-ketoacyl-CoA synthase 1 GB:AAC99312, very-long-chain...
0.651	At1g30020	expressed protein contains Pfam profile PF04398: Protein of unknown function, DUF538
0.65	At1g13140	cytochrome P450 family protein similar to Cytochrome P450 86A2 (SP:O23066) [Arabidopsis thaliana]; contains Pfam PF|00067...
0.646	At4g35420	dihydroflavonol 4-reductase family / dihydrokaempferol 4-reductase family similar to dihydroflavonol 4-reductase (Rosa hybrid...
0.637	At4g14815	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam protease inhibitor/seed storage/LTP...
0.629	At5g60500	undecaprenyl pyrophosphate synthetase family protein / UPP synthetase family protein contains putative undecaprenyl diphosphate...
0.625	At5g62320	myb family transcription factor (MYB99) contains PFAM profile: myb DNA binding domain PF00249
0.615	At1g66850	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein similar to GP|3062791 Lipid transfer protein...
0.615	At2g19070	transferase family protein similar to anthranilate N-hydroxycinnamoyl/benzoyltransferase from Dianthus caryophyllus...
0.61	At1g74540	cytochrome P450, putative similar to cytochrome P450 GB:O48922 [Glycine max]; contains Pfam profile: PF00067 cytochrome P450
0.61	At3g51590	lipid transfer protein, putative similar to lipid transfer protein E2 precursor, Brassica napus, PIR:T07984 [GI:899224];...
0.61	At5g49070	beta-ketoacyl-CoA synthase family protein similar to very-long-chain fatty acid condensing enzyme CUT1 [GI:5001734],...
0.608	At1g22015	galactosyltransferase family protein contains Pfam profile: PF01762 galactosyltransferase
0.6	At4g29250	transferase family protein low similarity to CER2 Arabidopsis thaliana GI:1213594, anthocyanin 5-aromatic acyltransferase...

ArathACLL5 (At1g62940) tissue and develoment (237 data)

0.986	At4g34850	chalcone and stilbene synthase family protein similar to chalcone synthase homolog PrChS1, Pinus radiata, gb:U90341; similar to...
0.98	At3g42960	alcohol dehydrogenase (ATA1) identical to alcohol dehydrogenase (ATA1) GI:2501781 from [Arabidopsis thaliana]
0.975	At1g01280	cytochrome P450 family protein similar to cytochrome P450 GB:BAA92894 GI:7339658 from [Petunia hybrida]
0.975	At4g14080	glycosyl hydrolase family 17 protein / anther-specific protein (A6) identical to probable glucan endo-1,3-beta-glucosidase A6...
0.974	At4g20420	tapetum-specific protein-related similar to SaTAP 35 [Sinapis alba] GI:408108
0.966	At3g11980	male sterility protein 2 (MS2) identical to male sterility protein 2 (MS2) SP:Q08891 (Arabidopsis thaliana)
0.966	At5g16920	expressed protein
0.964	At3g57620	glyoxal oxidase-related contains similarity to glyoxal oxidase precursor [Phanerochaete chrysosporium] gi|1050302|gb|AAA87594
0.959	At3g23770	glycosyl hydrolase family 17 protein similar to A6 anther-specific protein SP:Q06915 [Arabidopsis thaliana]
0.958	At5g07230	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein identical to tapetum-specific protein A9...
0.953	At3g13220	ABC transporter family protein contains Pfam profile: PF00005 ABC transporter; similar to white protein GB:Q27256 [Anopheles...
0.951	At1g61070	plant defensin-fusion protein, putative (PDF2.4) plant defensin protein family member, personal communication, Bart Thomma...
0.95	At1g02813	expressed protein contains Pfam profile PF04398: Protein of unknown function, DUF538
0.946	At1g20150	subtilase family protein similar to subtilisin-type protease precursor GI:14150446 from [Glycine max]
0.944	At5g62080	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein similar to tapetum-specific protein a9 precursor...
0.942	At3g07450	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein similar to cysteine-rich 5B protein - Lycopersicon...
0.939	At3g52130	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein similar to cysteine-rich 5B protein - Lycopersicon...
0.937	At1g02050	chalcone and stilbene synthase family protein Similar to rice chalcone synthase homolog, gp|U90341|2507617 and anther specific...
0.934	At2g42940	DNA-binding family protein contains a AT hook motif (DNA binding motifs with a preference for A/T rich regions), Pfam:PF02178
0.928	At5g62320	myb family transcription factor (MYB99) contains PFAM profile: myb DNA binding domain PF00249
0.924	At1g69500	cytochrome P450 family protein similar to Cytochrome P450 86A2 (SP:O23066) [Arabidopsis thaliana]contains Pfam profile:...
0.92	At4g29980	expressed protein
0.914	At4g29250	transferase family protein low similarity to CER2 Arabidopsis thaliana GI:1213594, anthocyanin 5-aromatic acyltransferase...
0.91	At3g06100	major intrinsic family protein / MIP family protein contains Pfam profile: PF00230 major intrinsic protein; contains...
0.91	At4g35420	dihydroflavonol 4-reductase family / dihydrokaempferol 4-reductase family similar to dihydroflavonol 4-reductase (Rosa hybrid...
0.909	At2g16910	basic helix-loop-helix (bHLH) family protein
0.905	At4g28395	lipid transfer protein, putative identical to anther-specific gene ATA7 [gi:2746339]; contains Pfam protease inhibitor/seed...
0.905	At5g52160	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam protease inhibitor/seed storage/LTP...
0.89	At5g13380	auxin-responsive GH3 family protein similar to auxin-responsive GH3 product [Glycine max] GI:18591; contains Pfam profile...
0.883	At3g52160	beta-ketoacyl-CoA synthase family protein beta-ketoacyl-CoA synthase - Simmondsia chinensis,PID:g1045614
0.882	At3g50580	proline-rich family protein contains proline-rich extensin domains, INTERPRO:IPR002965
0.874	At1g67990	caffeoyl-CoA 3-O-methyltransferase, putative similar to GI:2960356 [Populus balsamifera subsp. trichocarpa], GI:684942...
0.867	At1g06170	basic helix-loop-helix (bHLH) family protein contains Pfam profile:PF00010 helix-loop-helix DNA-binding domain
0.866	At1g74140	hypothetical protein
0.861	At5g24820	aspartyl protease family protein low similarity to CND41, chloroplast nucleoid DNA binding protein [Nicotiana tabacum]...
0.857	At5g49070	beta-ketoacyl-CoA synthase family protein similar to very-long-chain fatty acid condensing enzyme CUT1 [GI:5001734],...
0.85	At4g12920	aspartyl protease family protein low similarity to CND41, chloroplast nucleoid DNA binding protein [Nicotiana tabacum]...
0.849	At1g30020	expressed protein contains Pfam profile PF04398: Protein of unknown function, DUF538
0.844	At5g16960	NADP-dependent oxidoreductase, putative similar to probable NADP-dependent oxidoreductase (zeta-crystallin homolog) P1...
0.842	At5g60090	protein kinase family protein contains protein kinase domain, Pfam:PF00069
0.836	At5g48210	expressed protein
0.834	At1g03390	transferase family protein similar to anthranilate N-hydroxycinnamoyl/benzoyltransferase from Dianthus caryophyllus...
0.829	At1g75790	multi-copper oxidase type I family protein contains Pfam profile: PF00394 Multicopper oxidase
0.827	At1g33430	galactosyltransferase family protein contains Pfam profile: PF01762 galactosyltransferase
0.827	At2g19070	transferase family protein similar to anthranilate N-hydroxycinnamoyl/benzoyltransferase from Dianthus caryophyllus...
0.826	At4g34210	E3 ubiquitin ligase SCF complex subunit SKP1/ASK1 (At11), putative E3 ubiquitin ligase; similar to Skp1 homolog Skp1a...
0.818	At5g43340	inorganic phosphate transporter identical to inorganic phosphate transporter [Arabidopsis thaliana] GI:3869190
0.817	At5g41890	GDSL-motif lipase/hydrolase family protein similar to family II lipase EXL3 (GI:15054386), EXL1 (GI:15054382), EXL2...
0.815	At1g68540	oxidoreductase family protein similar to cinnamoyl CoA reductase [Eucalyptus gunnii, gi:2058311], cinnamyl-alcohol...
0.812	At2g31210	basic helix-loop-helix (bHLH) family protein contains Pfam profile: PF00010 helix-loop-helix DNA-binding domain; PMID: 12679534
0.812	At4g22080	pectate lyase family protein similar to pectate lyase 2 GP:6606534 from [Musa acuminata]
0.812	At5g40940	hypothetical protein
0.81	At5g61110	hypothetical protein
0.805	At4g30040	aspartyl protease family contains Pfam domain, PF00026: eukaryotic aspartyl protease
0.803	At3g58290	meprin and TRAF homology domain-containing protein / MATH domain-containing protein similar to ubiquitin-specific protease 12...
0.802	At5g17200	glycoside hydrolase family 28 protein / polygalacturonase (pectinase) family protein similar to polygalacturonase [Lycopersicon...
0.799	At1g71160	beta-ketoacyl-CoA synthase family protein similar to fatty acid elongase 3-ketoacyl-CoA synthase 1 GB:AAC99312, very-long-chain...
0.796	At5g43110	pumilio/Puf RNA-binding domain-containing protein contains similarity to RNA-binding protein
0.794	At1g13140	cytochrome P450 family protein similar to Cytochrome P450 86A2 (SP:O23066) [Arabidopsis thaliana]; contains Pfam PF00067...
0.79	At5g60500	undecaprenyl pyrophosphate synthetase family protein / UPP synthetase family protein contains putative undecaprenyl diphosphate...
0.787	At4g14815	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam protease inhibitor/seed storage/LTP...
0.785	At1g22015	galactosyltransferase family protein contains Pfam profile: PF01762 galactosyltransferase
0.785	At1g36150	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein low similarity to glucoamylase S1/S2 [Precursor]...
0.782	At1g64030	serpin family protein / serine protease inhibitor family protein similar to phloem serpin-1 [Cucurbita maxima] GI:9937311,...
0.78	At5g65205	short-chain dehydrogenase/reductase (SDR) family protein contains INTERPRO family IPR002198 short chain dehydrogenase/reductase...
0.771	At1g22090	expressed protein contains Pfam profile PF04776: Protein of unknown function (DUF626)
0.771	At3g51590	lipid transfer protein, putative similar to lipid transfer protein E2 precursor, Brassica napus, PIR:T07984 [GI:899224];...
0.769	At1g74540	cytochrome P450, putative similar to cytochrome P450 GB:O48922 [Glycine max]; contains Pfam profile: PF00067 cytochrome P450
0.767	At1g79780	integral membrane protein, putative contains 1 transmembrane domain; contains plant integral membrane protein domain,...
0.762	At1g08065	carbonic anhydrase family protein similar to storage protein (dioscorin) [Dioscorea cayenensis] GI:433463; contains Pfam...
0.758	At5g14980	esterase/lipase/thioesterase family protein low similarity to monoglyceride lipase from [Homo sapiens] GI:14594904, [Mus...
0.757	At3g63100	glycine-rich protein
0.756	At2g03170	E3 ubiquitin ligase SCF complex subunit SKP1/ASK1 (At14), putative E3 ubiquitin ligase; similar to Skp1 homolog Skp1b...
0.75	At4g27330	sporocyteless (SPL) identical to sporocyteless SPL (MADS-box related protein) [Arabidopsis thaliana] gi|5566240|gb|AAD45344
0.747	At3g15870	fatty acid desaturase family protein similar to delta 9 acyl-lipid desaturase (ADS1) GI:2970034 from [Arabidopsis thaliana]
0.746	At3g23840	transferase family protein low similarity to hypersensitivity-related gene [Nicotiana tabacum] GI:1171577,...
0.746	At3g28470	myb family transcription factor (MYB35) similar to Atmyb103 GB:AAD40692 from [Arabidopsis thaliana]; contains PFAM profile: myb...
0.745	At1g56360	calcineurin-like phosphoesterase family protein contains Pfam profile: PF00149 calcineurin-like phosphoesterase
0.745	At5g17340	expressed protein weak similarity to M3.4 protein [Brassica napus] GI:4574746
0.744	At5g60080	protein kinase family protein contains protein kinase domain, Pfam:PF00069
0.742	At1g23810	paired amphipathic helix repeat-containing protein low similarity to transcriptional repressor SIN3B [Mus musculus] GI:2921547;...
0.742	At4g28580	magnesium transporter CorA-like family protein (MRS2-6) weak similarity to SP|Q01926 RNA splicing protein MRS2, mitochondrial...
0.741	At1g68875	expressed protein
0.733	At4g36350	calcineurin-like phosphoesterase family protein contains Pfam profile: PF00149 calcineurin-like phosphoesterase
0.733	At5g41090	no apical meristem (NAM) family protein contains Pfam PF02365: No apical meristem (NAM) domain; similar to unknown protein...
0.732	At1g44222	hypothetical protein
0.726	At1g07340	hexose transporter, putative similar to hexose transporter [Lycopersicon esculentum] GI:5734440; contains Pfam profile PF00083:...
0.724	At5g17830	hypothetical protein contains Pfam domain, PF04515: Protein of unknown function, DUF580
0.719	At4g33870	peroxidase, putative similar to peroxidase [Spinacia oleracea] gi|1781334|emb|CAA71494
0.716	At1g75030	pathogenesis-related thaumatin family protein identical to thaumatin-like protein [Arabidopsis thaliana] GI:2435406; contains...
0.713	At2g03740	late embryogenesis abundant domain-containing protein / LEA domain-containing protein similar to cold-regulated gene cor15b...
0.711	At1g23700	protein kinase family protein contains protein kinase domain, Pfam:PF00069
0.707	At1g28375	expressed protein
0.706	At1g75940	glycosyl hydrolase family 1 protein / anther-specific protein ATA27 contains Pfam PF00232 : Glycosyl hydrolase family 1 domain;...
0.705	At5g53510	oligopeptide transporter OPT family protein similar to SP|P40900 Sexual differentiation process protein isp4...
0.704	At1g61630	equilibrative nucleoside transporter, putative (ENT7) identical to putative equilibrative nucleoside transporter ENT7...
0.702	At4g24890	calcineurin-like phosphoesterase family protein contains Pfam profile: PF00149'calcineurin-like phosphoesterase
0.701	At1g73050	(R)-mandelonitrile lyase, putative / (R)-oxynitrilase, putative similar to mandelonitrile lyase from Prunus serotina...
0.7	At1g48940	plastocyanin-like domain-containing protein

CLADE B

ArathACLL6 (At4g05160) tissue and develoment (237 data)

0.877	At1g03950	SNF7 family protein contains Pfam domain, PF03357: SNF7 family
0.873	At3g48990	AMP-dependent synthetase and ligase family protein similar to peroxisomal-coenzyme A synthetase (FAT2) [gi:586339] from...
0.871	At2g23430	kip-related protein 1 (KRP1) / cyclin-dependent kinase inhibitor 1 (ICK1) identical to cyclin-dependent kinase inhibitor (ICK1)...
0.861	At1g21000	zinc-binding family protein similar to zinc-binding protein [Pisum sativum] GI:16117799; contains Pfam profile PF04640 :...
0.86	At3g51580	expressed protein
0.86	At4g24220	expressed protein protein induced upon wounding - Arabidopsis thaliana, PID:e257749
0.859	At1g53560	expressed protein
0.858	At4g17170	Rab2-like GTP-binding protein (RAB2) identical to Rab2-like protein (At-RAB2) GI:1765896 from [Arabidopsis thaliana]
0.857	At1g53570	mitogen-activated protein kinase kinase kinase (MAPKKK), putative (MAP3Ka) identical to MEK kinase (MAP3Ka)[Arabidopsis...
0.855	At1g72800	nuM1-related contains similarity with nuM1 GI:1279563 from [Medicago sativa]
0.855	At4g20260	DREPP plasma membrane polypeptide family protein contains Pfam profile: PF05558 DREPP plasma membrane polypeptide
0.854	At4g36760	aminopeptidase P similar to Xaa-Pro aminopeptidase 2 [Lycopersicon esculentum] GI:15384991; contains Pfam profile PF00557:...
0.853	At1g70810	C2 domain-containing protein similar to zinc finger and C2 domain protein GI:9957238 from [Arabidopsis thaliana]
0.852	At3g03890	expressed protein
0.848	At5g07220	BAG domain-containing protein contains Pfam:PF02179 BAG domain
0.848	At5g56180	actin-related protein, putative (ARP8) strong similarity to actin-related protein 8A (ARP8) [Arabidopsis thaliana] GI:21427473;...
0.847	At5g53330	expressed protein
0.846	At1g26920	expressed protein Location of EST 228A16T7A, gb|N65686
0.844	At3g11930	universal stress protein (USP) family protein similar to ER6 protein GB:AAD46412 GI:5669654 from [Lycopersicon esculentum];...
0.842	At3g01170	expressed protein
0.841	At4g26060	expressed protein
0.841	At5g63800	glycosyl hydrolase family 35 protein similar to beta-galactosidase GI:7939621 from [Lycopersicon esculentum]; contains Pfam...
0.839	At1g02170	latex-abundant family protein (AMC1) / caspase family protein contains similarity to latex-abundant protein [Hevea...
0.839	At1g80180	expressed protein
0.838	At1g04440	casein kinase, putative similar to casein kinase I [Arabidopsis thaliana] gi|1103318|emb|CAA55395; contains protein kinase...
0.838	At4g39140	expressed protein
0.838	At5g04080	expressed protein
0.836	At5g56150	ubiquitin-conjugating enzyme, putative strong similarity to ubiquitin-conjugating enzyme UBC2 [Mesembryanthemum crystallinum]...
0.833	At5g51070	ATP-dependent Clp protease ATP-binding subunit (ClpD), (ERD1) SAG15/ERD1; identical to ERD1 protein GI:497629, SP:P42762 from...
0.832	At1g18470	zinc finger (C3HC4-type RING finger) family protein contains Pfam profile: PF00097 zinc finger, C3HC4 type
0.832	At2g46030	ubiquitin-conjugating enzyme 6 (UBC6) E2; identical to gi|431267, SP:P42750, PIR:S52661; contains a ubiquitin-conjugating...
0.832	At3g11660	harpin-induced family protein / HIN1 family protein / harpin-responsive family protein similar to harpin-induced protein hin1...
0.832	At5g18490	expressed protein
0.832	At5g33290	exostosin family protein contains Pfam profile: PF03016 Exostosin family
0.831	At1g14000	protein kinase family protein / ankyrin repeat family protein contains Pfam profiles: PF00069 protein kinase domain, PF00023...
0.831	At5g54940	eukaryotic translation initiation factor SUI1, putative similar to SP|P32911 Protein translation factor SUI1 {Saccharomyces...
0.83	At1g15860	expressed protein
0.829	At1g08320	bZIP family transcription factor contains Pfam profile: PF00170 bZIP transcription factor
0.829	At5g17650	glycine/proline-rich protein glycine/proline-rich protein GPRP - Arabidopsis thaliana, EMBL:X84315
0.828	At1g20440	dehydrin (COR47) identical to dehydrin COR47 (Cold-induced COR47 protein) [Arabidopsis thaliana] SWISS-PROT:P31168
0.826	At1g58180	carbonic anhydrase family protein / carbonate dehydratase family protein similar to SP|P46512 Carbonic anhydrase 1 (EC 4.2.1.1)...
0.826	At3g05970	long-chain-fatty-acid--CoA ligase / long-chain acyl-CoA synthetase (LACS6) strong similarity to AMP-binding protein (MF39P)...
0.826	At5g47560	sodium/dicarboxylate cotransporter, putative similar to SWISS-PROT:Q13183 renal sodium/dicarboxylate cotransporter [Human]{Homo...
0.825	At5g39590	expressed protein
0.823	At3g51730	saposin B domain-containing protein contains Pfam profiles: PF00026 eukaryotic aspartyl protease, PF03489 surfactant protein B,...
0.822	At1g30000	glycoside hydrolase family 47 protein similar to GI:5579331 from [Homo sapiens]; contains Pfam profile PF01532: Glycosyl...
0.822	At5g11680	expressed protein predicted proteins, Arabidopsis thaliana
0.821	At1g47960	invertase/pectin methylesterase inhibitor family protein low similarity to SP|P83326 Pectinesterase inhibitor (Pectin...
0.821	At2g25450	2-oxoglutarate-dependent dioxygenase, putative similar to 2A6 (GI:599622) and tomato ethylene synthesis regulatory protein E8...
0.819	At4g29160	SNF7 family protein contains Pfam domain, PF03357: SNF7 family
0.818	At1g27290	expressed protein
0.818	At2g23450	protein kinase family protein contains protein kinase domain, Pfam:PF00069
0.818	At3g57090	expressed protein
0.818	At5g55850	nitrate-responsive NOI protein, putative similar to nitrate-induced NOI protein [Zea mays] GI:2642213
0.817	At1g49670	ARP protein (REF) identical to ARP protein GB:CAA89858 GI:886434 from [Arabidopsis thaliana]; contains Pfam profile PF00107:...
0.816	At1g32700	zinc-binding family protein similar to zinc-binding protein [Pisum sativum] GI:16117799; contains Pfam profile PF04640 :...
0.816	At2g30140	UDP-glucoronosyl/UDP-glucosyl transferase family protein contains Pfam profile: PF00201 UDP-glucoronosyl and UDP-glucosyl...
0.816	At5g18630	lipase class 3 family protein low similarity to Triacylglycerol Acylhydrolase (E.C.3.1.1.3) [Rhizomucor miehei] GI:230348;...
0.814	At1g10150	expressed protein similar to ESTs gb|T20511, gb|T45308, gb|H36493, and gb|AA651176
0.814	At2g02360	F-box family protein / SKP1 interacting partner 3-related contains similarity to SKP1 interacting partner 3 GI:10716951 from...
0.814	At3g13720	prenylated rab acceptor (PRA1) family protein contains Pfam profile PF03208: Prenylated rab acceptor (PRA1)
0.813	At1g49470	expressed protein contains Pfam profile PF04819: Family of unknown function (DUF716) (Plant viral-response family)
0.812	At1g27000	bZIP family transcription factor
0.812	At2g26670	heme oxygenase 1 (HO1) (HY1) identical to plastid heme oxygenase (HY1) [Arabidopsis thaliana] GI:4877362, heme oxygenase 1...
0.812	At4g24990	ubiquitin family protein contains INTERPRO:IPR000626 ubiquitin domain
0.811	At2g16710	hesB-like domain-containing protein similar to IscA (putative iron-sulfur cluster assembly protein) [Azotobacter vinelandii]...
0.811	At4g17830	peptidase M20/M25/M40 family protein similar to acetylornithine deacetylase (Acetylornithinase, AO; N-acetylornithinase, NAO)...
0.81	At5g60360	cysteine proteinase, putative / AALP protein (AALP) identical to AALP protein GI:7230640 from [Arabidopsis thaliana]; similar...
0.807	At1g04970	lipid-binding serum glycoprotein family protein low similarity to SP|P17213 Bactericidal permeability-increasing protein...
0.807	At3g11780	MD-2-related lipid recognition domain-containing protein / ML domain-containing protein weak similarity to...
0.806	At2g39710	aspartyl protease family protein contains profile Pfam PF00026: Eukaryotic aspartyl protease; contains Prosite PS00141:...
0.805	At1g49300	Ras-related GTP-binding protein, putative contains Pfam profile: PF00071 Ras family
0.805	At1g76070	expressed protein
0.805	At4g08930	thioredoxin-related contains weak similarity to Swiss-Prot:Q39239 thioredoxin H-type 4 (TRX-H-4). [Mouse-ear cress]
0.805	At4g30270	MERI-5 protein (MERI-5) (MERI5B) / endo-xyloglucan transferase / xyloglucan endo-1,4-beta-D-glucanase (SEN4) identical to...
0.805	At4g36400	FAD linked oxidase family protein low similarity to SP|Q12627 from Kluyveromyces lactis and SP|P32891 from Saccharomyces...
0.805	At5g24460	expressed protein
0.805	At5g40690	expressed protein
0.804	At3g54140	proton-dependent oligopeptide transport (POT) family protein contains Pfam profile: PF00854 POT family
0.804	At4g32760	VHS domain-containing protein / GAT domain-containing protein weak similarity to hepatocyte growth factor-regulated tyrosine...
0.803	At1g72510	expressed protein
0.803	At2g27310	F-box family protein contains Pfam PF00646: F-box domain;; similar to SKP1 interacting partner 2 (SKIP2) TIGR_Ath1:At5g67250
0.803	At5g40670	PQ-loop repeat family protein / transmembrane family protein similar to SP|O60931 Cystinosin {Homo sapiens}; contains Pfam...
0.803	At5g51640	leaf senescence protein-related (YLS7) annotation temporarily based on supporting cDNA gi|13122291|dbj|AB047810.1|; identical...
0.802	At4g16520	autophagy 8f (APG8f) identical to autophagy 8f [Arabidopsis thaliana] GI:19912161; contains Pfam profile PF02991: Microtubule...
0.801	At1g04960	expressed protein
0.801	At2g30550	lipase class 3 family protein similar to DEFECTIVE IN ANTHER DEHISCENCE1 [Arabidopsis thaliana] GI:16215706; contains Pfam...
0.8	At1g12140	flavin-containing monooxygenase family protein / FMO family protein similar to flavin-containing monooxygenase [Cavia...
0.8	At1g13990	expressed protein
0.8	At1g80310	expressed protein
0.799	At2g38480	integral membrane protein, putative contains 4 transmembrane domains; contains plant integral membrane protein domain,...
0.798	At3g23280	zinc finger (C3HC4-type RING finger) family protein / ankyrin repeat family protein contains Pfam profile: PF00097 zinc finger,...
0.798	At5g45410	expressed protein similar to unknown protein (pir||T05524)
0.797	At3g14050	RelA/SpoT protein, putative (RSH2) nearly identical to RelA/SpoT homolog RSH2 [Arabidopsis thaliana] GI:7141306; contains Pfam...
0.797	At4g32870	expressed protein hypothetical protein F17H15.20 Arabidopsis thaliana chromosome II BAC F17H15, PID:g3643606
0.797	At5g02600	heavy-metal-associated domain-containing protein low similarity to gi:3168840 copper homeostasis factor; contains Pfam...
0.797	At5g45550	mob1/phocein family protein contains Pfam profile: PF03637 Mob1/phocein family
0.796	At1g32410	vacuolar protein sorting 55 family protein / VPS55 family protein contains Pfam domain PF04133: Vacuolar protein sorting 55
0.796	At2g26230	uricase / urate oxidase / nodulin 35, putative identical to uricase SP:O04420 from [Arabidopsis thaliana]
0.796	At3g02910	expressed protein contains Pfam domain PF03674: Uncharacterised protein family (UPF0131)

CLADE C

ArathACLL7 (At4g19010) no coexpressed gene

CLADE D

ArathACLL1 (At1g20480) Tissue and development (data 237)

0.801	At2g17370	3-hydroxy-3-methylglutaryl-CoA reductase 2 / HMG-CoA reductase 2 (HMGR2) identical to SP|P43256...
0.781	At4g31340	myosin heavy chain-related contains weak similarity to Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain, type...
0.766	At3g05020	acyl carrier protein 1, chloroplast (ACP-1) identical to SP|P11829 Acyl carrier protein 1, chloroplast precursor (ACP)...
0.764	At5g16510	reversibly glycosylated polypeptide, putative similar to reversibly glycosylatable polypeptide (RGP1) [Pisum sativum]...
0.762	At5g15530	biotin carboxyl carrier protein 2 (BCCP2) identical to biotin carboxyl carrier protein isoform 2 [Arabidopsis thaliana]...
0.762	At5g50390	pentatricopeptide (PPR) repeat-containing protein contains INTERPRO:IPR002885 PPR repeats
0.758	At1g22170	phosphoglycerate/bisphosphoglycerate mutase family protein similar to SP|P31217 Phosphoglycerate mutase 1 (EC 5.4.2.1)...
0.753	At4g11820	hydroxymethylglutaryl-CoA synthase / HMG-CoA synthase / 3-hydroxy-3-methylglutaryl coenzyme A synthase identical to...
0.75	At1g78050	phosphoglycerate/bisphosphoglycerate mutase family protein similar to SP|P31217 Phosphoglycerate mutase 1 (EC 5.4.2.1)...
0.746	At3g20920	translocation protein-related contains weak similarity to Drosophila translocation protein 1 (GI:558181) [Drosophila melanogaster]
0.741	At2g43040	calmodulin-binding protein similar to pollen-specific calmodulin-binding protein MPCBP GI:10086260 from [Zea mays]; contains...
0.727	At2g35160	SET domain-containing protein (SUVH5) identical to SUVH5 [Arabidopsis thaliana] GI:13517751; contains Pfam profiles PF00856:...
0.725	At1g22800	expressed protein similar to Biotin synthesis protein bioC. {Serratia marcescens} (SP:P36571); ESTs gb|Z34075, gb|Z34835 and...
0.723	At1g76550	pyrophosphate--fructose-6-phosphate 1-phosphotransferase alpha subunit, putative / pyrophosphate-dependent:...
0.721	At5g04620	aminotransferase class I and II family protein similar to 8-amino-7-oxononanoate synthase, Bacillus sphaericus, PIR:JQ0512...
0.718	At2g29050	rhomboid family protein contains PFAM domain PF01694, Rhomboid family
0.715	At1g70770	expressed protein
0.713	At5g42780	zinc finger homeobox family protein / ZF-HD homeobox family protein similar to unknown protein (pir||T05568)
0.712	At1g18180	expressed protein
0.711	At2g20840	secretory carrier membrane protein (SCAMP) family protein contains Pfam domain, PF04144: SCAMP family
0.71	At5g66310	kinesin motor family protein contains Pfam domain, PF00225: Kinesin motor domain
0.709	At4g23490	fringe-related protein + weak similarity to Fringe [Schistocerca gregaria](GI:6573138);Fringe encodes an extracellular protein...
0.709	At5g62500	microtubule-associated EB1 family protein similar to EBF3-S (Microtubule-associated protein) [Homo sapiens] GI:12751131;...
0.708	At1g09870	histidine acid phosphatase family protein contains Pfam profile PF00328: Histidine acid phosphatase; similar to multiple...
0.707	At3g54250	mevalonate diphosphate decarboxylase, putative similar to mevalonate diphosphate decarboxylase [Arabidopsis thaliana]...
0.705	At2g46000	expressed protein
0.705	At3g08910	DNAJ heat shock protein, putative similar to SP|P25685 DnaJ homolog subfamily B member 1 (Heat shock 40 kDa protein 1) {Homo...
0.702	At4g12700	expressed protein
0.7	At1g67680	expressed protein
0.7	At3g09570	expressed protein
0.699	At2g29390	sterol 4-alpha-methyl-oxidase 1 (SMO1) nearly identical to sterol 4-alpha-methyl-oxidase GI:16973469 from [Arabidopsis...
0.699	At5g06830	expressed protein contains Pfam profile: PF05600 protein of unknown function (DUF773)
0.699	At5g42280	DC1 domain-containing protein contains Pfam profile PF03107: DC1 domain
0.698	At2g22900	galactosyl transferase GMA12/MNN10 family protein very low similarity to alpha-1,2-galactosyltransferase, Schizosaccharomyces...
0.696	At1g29310	protein transport protein sec61, putative similar to PfSec61 [Plasmodium falciparum] GI:3057044; contains Pfam profile PF00344:...
0.693	At3g22845	emp24/gp25L/p24 protein-related contains weak similarity to transmembrane protein (GI:1212965) [Homo sapiens]
0.69	At3g56640	exocyst complex subunit Sec15-like family protein contains Pfam profile PF04091: Exocyst complex subunit Sec15-like
0.688	At1g54830	CCAAT-box binding transcription factor Hap5a, putative similar to heme activated protein GI:6289057 from (Arabidopsis thaliana)...
0.686	At1g19970	ER lumen protein retaining receptor family protein similar to SP|P33946 ER lumen protein retaining receptor 1 (KDEL receptor 1)...
0.686	At1g79750	malate oxidoreductase, putative similar to malate oxidoreductase (NADP-dependent malic enzyme) GB:P34105 (Populus balsamifera...
0.686	At3g02230	reversibly glycosylated polypeptide-1 (RGP1) identical to reversibly glycosylated polypeptide-1 (AtRGP) [Arabidopsis thaliana]...
0.685	At1g76270	expressed protein contains Pfam PF03138: Plant protein family. The function of this family of plant proteins is unknown;...
0.685	At2g16760	expressed protein
0.684	At5g18550	zinc finger (CCCH-type) family protein contains Pfam domain, PF00642: Zinc finger C-x8-C-x5-C-x3-H type (and similar)
0.684	At5g59740	UDP-galactose/UDP-glucose transporter-related weak similarity to UDP-galactose/UDP-glucose transporter [Arabidopsis thaliana]...
0.683	At2g01140	fructose-bisphosphate aldolase, putative similar to plastidic aldolase NPALDP1 from Nicotiana paniculata [GI:4827251]; contains...
0.682	At1g75110	expressed protein
0.681	At5g17770	NADH-cytochrome b5 reductase identical to NADH-cytochrome b5 reductase [Arabidopsis thaliana] GI:4240116
0.681	At5g22940	exostosin family protein contains Pfam profile: PF03016 exostosin family
0.68	At4g38480	transducin family protein / WD-40 repeat family protein contains contains Pfam PF00400: WD domain, G-beta repeat (7 copies, 3...
0.679	At5g66460	(1-4)-beta-mannan endohydrolase, putative similar to (1-4)-beta-mannan endohydrolase [Coffea arabica] GI:10178872; contains...
0.677	At3g50960	expressed protein
0.676	At3g48410	hydrolase, alpha/beta fold family protein low similarity to 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase [Rhodococcus...
0.675	At1g74030	enolase, putative similar to Swiss-Prot:P15007 enolase (EC 4.2.1.11) (2-phosphoglycerate dehydratase)(2-phospho-D- glycerate...
0.674	At3g25110	acyl-[acyl carrier protein] thioesterase / acyl-ACP thioesterase / oleoyl-[acyl-carrier protein] hydrolase / S-acyl fatty acid...
0.674	At4g14695	expressed protein contains Pfam domain, PF03650: Uncharacterized protein family (UPF0041)
0.673	At1g06450	CCR4-NOT transcription complex protein, putative similar to SWISS-PROT:Q9UFF9 CCR4-NOT transcription complex, subunit 8...
0.673	At2g19600	K+ efflux antiporter, putative (KEA4) similar to glutathione-regulated potassium-efflux system protein KEFB, Escherichia coli,...
0.673	At3g45090	2-phosphoglycerate kinase-related contains weak similarity to 2-phosphoglycerate kinase (GI:467751) [Methanothermus fervidus]
0.673	At5g49460	ATP-citrate synthase, putative / ATP-citrate (pro-S-)-lyase, putative / citrate cleavage enzyme, putative strong similarity to...
0.672	At3g20560	thioredoxin family protein contains Pfam profile PF00085: Thioredoxin
0.672	At3g57650	acyl-CoA:1-acylglycerol-3-phosphate acyltransferase, putative similar to acyl-CoA:1-acylglycerol-3-phosphate acyltransferase...
0.671	At4g13710	pectate lyase family protein
0.669	At1g11680	obtusifoliol 14-demethylase (CYP51) identical to obtusifoliol 14-demethylase (GI:14624983) [Arabidopsis thaliana]
0.669	At1g54630	acyl carrier protein 3, chloroplast (ACP-3) nearly identical to SP|P25702 Acyl carrier protein 3, chloroplast precursor (ACP)...
0.669	At4g35560	expressed protein
0.668	At1g14970	expressed protein contains Pfam PF03138: Plant protein family. The function of this family of plant proteins is unknown;...
0.668	At5g01340	mitochondrial substrate carrier family protein contains Pfam profile: PF00153 mitochondrial carrier protein
0.667	At3g14000	expressed protein
0.667	At3g48440	zinc finger (CCCH-type) family protein contains Pfam domain, PF00642: Zinc finger C-x8-C-x5-C-x3-H type (and similar)
0.663	At2g03120	signal peptide peptidase family protein contains Pfam domain PF04258: Membrane protein of unknown function (DUF435)
0.662	At1g25510	aspartyl protease family protein contains Pfam domain, PF00026: eukaryotic aspartyl protease
0.661	At3g54930	serine/threonine protein phosphatase 2A (PP2A) regulatory subunit B', putative similar to SWISS-PROT:Q28653 serine/threonine...
0.659	At2g14835	zinc finger (C3HC4-type RING finger) family protein contains Pfam profile: PF00097 zinc finger, C3HC4 type (RING finger)
0.657	At1g05360	expressed protein Similar to Arabidopsis hypothetical protein PID:e326839 (gb|Z97337) contains transmembrane domains
0.656	At1g79360	transporter-related low similarity to SP|O76082 Organic cation/carnitine transporter 2 (Solute carrier family 22, member 5)...
0.656	At5g47520	Ras-related GTP-binding protein, putative similar to GTP-binding protein RAB11J GI:1370160 from [Lotus japonicus]
0.656	At5g48230	acetyl-CoA C-acyltransferase, putative / 3-ketoacyl-CoA thiolase, putative strong similarity to Acetoacetyl-coenzyme A thiolase...
0.656	At5g58190	expressed protein contains Pfam profile PF04146: YTS21-B-like family
0.654	At1g76510	ARID/BRIGHT DNA-binding domain-containing protein contains Pfam profile PF01388: ARID/BRIGHT DNA binding domain
0.654	At2g21520	SEC14 cytosolic factor, putative / phosphoglyceride transfer protein, putative contains Pfam PF00650 : CRAL/TRIO domain;...
0.654	At3g54320	ovule development protein, putative similar to ovule development protein aintegumenta (GI:1209099) [Arabidopsis thaliana]
0.654	At5g23530	expressed protein contains similarity to PrMC3 [Pinus radiata] GI:5487873
0.653	At5g10550	DNA-binding bromodomain-containing protein low similarity to kinase [Gallus gallus] GI:1370092; contains Pfam profile PF00439:...
0.652	At1g62020	coatomer protein complex, subunit alpha, putative contains Pfam PF00400: WD domain, G-beta repeat; similar to Coatomer alpha...
0.651	At3g16950	dihydrolipoamide dehydrogenase 1, plastidic / lipoamide dehydrogenase 1 (PTLPD1) identical to plastidic lipoamide dehydrogenase...
0.651	At4g22250	zinc finger (C3HC4-type RING finger) family protein contains Pfam profile: PF00097 zinc finger, C3HC4 type (RING finger)
0.65	At5g42630	myb family transcription factor (KAN4) contains Pfam profile: PF00249 myb-like DNA-binding domain; identical to cDNA GARP-like...
0.649	At1g23890	NHL repeat-containing protein contains Pfam profile PF01436: NHL repeat
0.649	At2g40620	bZIP transcription factor family protein identical to b-Zip DNA binding protein GI:2246376 from [Arabidopsis thaliana];...
0.648	At1g60810	ATP citrate-lyase -related similar to ATP citrate-lyase GI:949989 from [Rattus norvegicus]
0.648	At5g11230	phosphate translocator-related low similarity to phosphoenolpyruvate/phosphate translocator precursor [Mesembryanthemum...
0.647	At1g21070	transporter-related low similarity to GDP-Mannose transporter [Arabidopsis thaliana] GI:15487237; contains Pfam profile...
0.647	At4g39860	expressed protein
0.646	At1g10670	expressed protein
0.645	At5g22740	glycosyl transferase family 2 protein similar to beta-(1-3)-glucosyl transferase GB:AAC62210 GI:3687658 from [Bradyrhizobium...
0.644	At3g04980	DNAJ heat shock N-terminal domain-containing protein contains Pfam profile PF00226 DnaJ domain
0.644	At3g46200	MutT/nudix family protein similar to head organizer protein P17F11 GI:17976973 from [Xenopus laevis]; contains a NUDIX...
0.644	At5g43060	cysteine proteinase, putative / thiol protease, putative similar to cysteine proteinase RD21A precursor (thiol protease)...
0.643	At2g38700	mevalonate diphosphate decarboxylase (MVD1) identical to mevalonate diphosphate decarboxylase [Arabidopsis thaliana]...

ArathACLL2 (At1g20490) Stress treatments v.1 (298 data)

0.732	At5g58700	phosphoinositide-specific phospholipase C family protein contains Pfam profile: PF00388 phosphatidylinositol-specific...
0.721	At1g08920	sugar transporter, putative similar to ERD6 protein {Arabidopsis thaliana} GI:3123712, sugar-porter family proteins 1 and 2...
0.718	At5g58690	phosphoinositide-specific phospholipase C family protein contains Pfam profile: PF00388 phosphatidylinositol-specific...
0.695	At1g58270	meprin and TRAF homology domain-containing protein / MATH domain-containing protein similar to ubiquitin-specific protease 12...
0.694	At3g29575	expressed protein
0.691	At1g07430	protein phosphatase 2C, putative / PP2C, putative similar to GB:CAB90633 from [Fagus sylvatica]
0.688	At1g80110	expressed protein contains similarity to SKP1 interacting partner 3 [Arabidopsis thaliana] GI:10716951
0.685	At1g69260	expressed protein
0.684	At1g01650	protease-associated (PA) domain-containing protein contains protease associated (PA) domain, Pfam:PF02225
0.68	At5g65280	lanthionine synthetase C-like family protein contains Pfam domain, PF05147: Lanthionine synthetase C-like protein
0.665	At1g67300	hexose transporter, putative similar to hexose transporters from Solanum tuberosum [GI:8347246], Nicotiana tabacum...
0.665	At3g48510	expressed protein
0.662	At3g02480	ABA-responsive protein-related similar to ABA-inducible protein [Fagus sylvatica] GI:3901016, cold-induced protein kin1...
0.658	At5g04250	OTU-like cysteine protease family protein contains Pfam profile PF02338: OTU-like cysteine protease
0.657	At5g06760	late embryogenesis abundant group 1 domain-containing protein / LEA group 1 domain-containing protein low similarity to...
0.657	At5g15960	stress-responsive protein (KIN1) / stress-induced protein (KIN1) identical to SP|P18612 Stress-induced KIN1 protein...
0.655	At2g47780	rubber elongation factor (REF) protein-related similar to Small rubber particle protein (SRPP) (22 kDa rubber particle protein)...
0.653	At3g03170	expressed protein
0.651	At1g05100	protein kinase family protein contains protein kinase domain, Pfam:PF00069
0.648	At1g62570	flavin-containing monooxygenase family protein / FMO family protein low similarity to flavin-containing monooxygenase FMO3...
0.646	At2g39050	hydroxyproline-rich glycoprotein family protein contains QXW lectin repeat domain, Pfam:PF00652
0.646	At5g50720	ABA-responsive protein (HVA22e) identical to AtHVA22e [Arabidopsis thaliana] GI:11225589
0.642	At5g57050	protein phosphatase 2C ABI2 / PP2C ABI2 / abscisic acid-insensitive 2 (ABI2) identical to SP|O04719 Protein phosphatase 2C ABI2...
0.641	At2g15970	cold-acclimation protein, putative (FL3-5A3) similar to cold acclimation WCOR413-like protein gamma form [Hordeum vulgare]...
0.638	At2g34850	NAD-dependent epimerase/dehydratase family protein similar to UDP-galactose 4-epimerase from Cyamopsis tetragonoloba...
0.638	At5g59220	protein phosphatase 2C, putative / PP2C, putative ABA induced protein phosphatase 2C, Fagus sylvatica, EMBL:FSY277743
0.637	At1g56600	galactinol synthase, putative similar to galactinol synthase, isoform GolS-1 GI:5608497 from [Ajuga reptans]
0.637	At3g50970	dehydrin xero2 (XERO2) / low-temperature-induced protein LTI30 (LTI30) identical to dehydrin Xero 2 (Low-temperature-induced...
0.63	At3g62700	glutathione-conjugate transporter, putative similar to glutathione-conjugate transporter AtMRP4 GI:2959767 from [Arabidopsis...
0.63	At5g20900	expressed protein
0.629	At2g42540	cold-responsive protein / cold-regulated protein (cor15a) identical to cold-regulated protein cor15a [Arabidopsis thaliana]...
0.628	At2g47770	benzodiazepine receptor-related contains weak similarity to Peripheral-type benzodiazepine receptor (PBR) (PKBS) (Mitochondrial...
0.627	At4g33905	peroxisomal membrane protein 22 kDa, putative similar to 22 kDa peroxisomal membrane protein PMP22 [Mus musculus]...
0.625	At1g66830	leucine-rich repeat transmembrane protein kinase, putative contains Pfam profiles: PF00069: Eukaryotic protein kinase domain,...
0.624	At1g52890	no apical meristem (NAM) family protein contains Pfam PF02365: No apical meristem (NAM) domain; similar to NAM (no apical...
0.618	At1g52690	late embryogenesis abundant protein, putative / LEA protein, putative similar to SP|P13934 Late embryogenesis abundant protein...
0.618	At1g54830	CCAAT-box binding transcription factor Hap5a, putative similar to heme activated protein GI:6289057 from (Arabidopsis thaliana)...
0.618	At2g33380	calcium-binding RD20 protein (RD20) induced by abscisic acid during dehydration PMID:10965948; putative transmembrane channel...
0.618	At5g65990	amino acid transporter family protein similar to proton/amino acid transporter 1 [Mus musculus] GI:21908024; contains Pfam...
0.616	At5g13750	transporter-related
0.615	At1g17550	protein phosphatase 2C-related / PP2C-related similar to protein phosphatase 2C GI:3242077 from (Arabidopsis thaliana)
0.612	At3g28007	nodulin MtN3 family protein contains Pfam PF03083 MtN3/saliva family; similar to LIM7 GI:431154 (induced in meiotic prophase in...
0.612	At3g55610	delta 1-pyrroline-5-carboxylate synthetase B / P5CS B (P5CS2) identical to SP|P54888
0.611	At1g58360	amino acid permease I (AAP1) identical to amino acid permease I GI:22641 from [Arabidopsis thaliana]
0.611	At2g19810	zinc finger (CCCH-type) family protein contains Pfam domain, PF00642: Zinc finger C-x8-C-x5-C-x3-H type (and similar)
0.611	At2g41190	amino acid transporter family protein low similarity to vesicular GABA transporter [Rattus norvegicus] GI:2587061; belongs to...
0.611	At4g16760	acyl-CoA oxidase (ACX1) identical to acyl-CoA oxidase [Arabidopsis thaliana] GI:3044214
0.61	At4g19390	expressed protein
0.61	At4g26080	protein phosphatase 2C ABI1 / PP2C ABI1 / abscisic acid-insensitive 1 (ABI1) nearly identical to SP|P49597 Protein phosphatase...
0.61	At4g27410	no apical meristem (NAM) family protein (RD26) contains Pfam PF02365: No apical meristem (NAM) domain; Arabidopsis thaliana nap...
0.609	At3g09910	Ras-related GTP-binding protein, putative similar to GTP-binding protein GI:2723477 from [Arabidopsis thaliana] ;contains Pfam...
0.609	At5g52310	low-temperature-responsive protein 78 (LTI78) / desiccation-responsive protein 29A (RD29A)
0.608	At3g11410	protein phosphatase 2C, putative / PP2C, putative identical to protein phosphatase 2C (PP2C) GB:P49598 [Arabidopsis thaliana];...
0.607	At1g16850	expressed protein
0.607	At4g10960	UDP-glucose 4-epimerase, putative / UDP-galactose 4-epimerase, putative / Galactowaldenase, putative similar to UDP-galactose...
0.606	At3g27870	haloacid dehalogenase-like hydrolase family protein similar to Potential phospholipid-transporting ATPase (EC 3.6.3.1) from...
0.605	At3g25870	expressed protein
0.604	At4g27840	expressed protein
0.603	At2g47600	magnesium/proton exchanger (MHX1) identical to magnesium/proton exchanger AtMHX [Arabidopsis thaliana] gi|6492237|gb|AAF14229;...

ArathACLL3 (At1g20500)		tissue and development (237 data)
0.99	At5g07200	gibberellin 20-oxidase identical to GI:1109699
0.988	At1g62070	expressed protein
0.986	At5g50750	reversibly glycosylated polypeptide, putative strong similarity to reversibly glycosylated polypeptide-1 (AtRGP) [Arabidopsis...
0.985	At4g10490	oxidoreductase, 2OG-Fe(II) oxygenase family protein similar to naringenin,2-oxoglutarate 3-dioxygenase [Dianthus...
0.984	At1g04380	2-oxoglutarate-dependent dioxygenase, putative Strong similarity to Arabidopsis 2A6 (gb|X83096), tomato ethylene synthesis...
0.984	At5g07260	homeobox protein-related contains weak similarity to Homeobox protein FWA (Swiss-Prot:Q9FVI6) [Arabidopsis thaliana]
0.983	At5g38170	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam profile: PF00234 protease...
0.981	At1g60970	clathrin adaptor complex small chain family protein contains Pfam profile: PF01217 clathrin adaptor complex small chain
0.981	At1g62060	expressed protein
0.98	At3g58740	citrate synthase, glyoxysomal, putative strong similarity to SP|P49299 Citrate synthase, glyoxysomal precursor {Cucurbita...
0.979	At5g07210	two-component responsive regulator family protein / response regulator family protein contains Pfam profile: PF00072 response...
0.979	At5g38160	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam profile: PF00234 protease...
0.978	At5g08460	GDSL-motif lipase/hydrolase family protein similar to family II lipase EXL3 (GI:15054386), EXL1 (GI:15054382), EXL2...
0.976	At4g34520	fatty acid elongase 1 (FAE1) identical to fatty acid elongase 1 [GI:881615]
0.974	At1g25410	adenylate isopentenyltransferase 6 / adenylate dimethylallyltransferase / cytokinin synthase (IPT6) identical to adenylate...
0.974	At3g63040	expressed protein predicted protein, C.elegans
0.974	At5g38180	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam profile: PF00234 protease...
0.974	At5g40420	glycine-rich protein / oleosin
0.972	At1g47540	trypsin inhibitor, putative similar to SP|P26780 Trypsin inhibitor 2 precursor (MTI-2) {Sinapis alba}
0.972	At1g71250	GDSL-motif lipase/hydrolase family protein similar to family II lipases EXL3 GI:15054386, EXL1 GI:15054382, EXL2 GI:15054384...
0.972	At4g27160	2S seed storage protein 3 / 2S albumin storage protein / NWMU2-2S albumin 3 identical to SP|P15459
0.971	At4g36700	cupin family protein low similarity to preproMP27-MP32 from Cucurbita cv. Kurokawa Amakuri [GI:691752]; contains Pfam profile...
0.971	At5g44120	12S seed storage protein (CRA1) nearly identical to SP|P15455 [Plant Mol Biol 11:805-820 (1988)]; contains Pfam profile PF00190...
0.969	At1g28590	lipase, putative similar to lipase [Arabidopsis thaliana] GI:1145627; contains InterPro Entry IPR001087 Lipolytic enzyme,...
0.969	At2g44470	glycosyl hydrolase family 1 protein contains Pfam PF00232 : Glycosyl hydrolase family 1 domain; TIGRFAM TIGR01233:...
0.969	At4g28520	12S seed storage protein, putative / cruciferin, putative strong similarity to SP|P33525 Cruciferin CRU1 precursor (11S...
0.968	At1g78390	9-cis-epoxycarotenoid dioxygenase, putative / neoxanthin cleavage enzyme, putative / carotenoid cleavage dioxygenase, putative...
0.968	At3g03230	esterase/lipase/thioesterase family protein contains Interpro entry IPR000379
0.967	At2g27380	proline-rich family protein contains proline-rich extensin domains, INTERPRO:IPR002965
0.967	At3g12203	serine carboxypeptidase S10 family protein contains Pfam profile: PF00450 serine carboxypeptidase; similar to serine...
0.967	At4g25140	glycine-rich protein / oleosin
0.967	At4g27150	2S seed storage protein 2 / 2S albumin storage protein / NWMU2-2S albumin 2 identical to SP|P15458
0.965	At1g65090	expressed protein
0.965	At4g37360	cytochrome P450 family protein cytochrome P450 monooxygenase, Arabidopsis thaliana, PID:d1029478
0.964	At1g03880	12S seed storage protein (CRB) identical to 12S seed storage protein, gi|808937 [SP|P15456] [Plant Mol Biol 11:805-820 (1988)];...
0.963	At2g23580	hydrolase, alpha/beta fold family protein similar to ethylene-induced esterase [Citrus sinensis] GI:14279437, polyneuridine...
0.963	At5g47670	CCAAT-box binding transcription factor family protein / leafy cotyledon 1-related (L1L) supporting cDNA...
0.962	At4g27170	2S seed storage protein 4 / 2S albumin storage protein / NWMU2-2S albumin 4 identical to SP|P15460
0.961	At3g01570	glycine-rich protein / oleosin similar to oleosin GB:AAB58402 [Sesamum indicum]
0.961	At5g07500	zinc finger (CCCH-type) family protein contains Pfam domain, PF00642: Zinc finger C-x8-C-x5-C-x3-H type (and similar)
0.961	At5g22810	GDSL-motif lipase, putative similar to EXL3 (GP:15054386) [Arabidopsis thaliana]
0.96	At2g45420	LOB domain protein 18 / lateral organ boundaries domain protein 18 (LBD18) identical to LOB DOMAIN 18 [Arabidopsis thaliana]...
0.959	At1g68380	expressed protein contains Pfam profile PF03267: Arabidopsis protein of unknown function, DUF266
0.958	At2g28490	cupin family protein similar to preproMP27-MP32 [Cucurbita cv. Kurokawa Amakuri] GI:691752, allergen Gly m Bd 28K [Glycine max]...
0.958	At5g38195	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam protease inhibitor/seed storage/LTP...
0.958	At5g49190	sucrose synthase / sucrose-UDP glucosyltransferase (SUS2) nearly identical to SP|Q00917 Sucrose synthase (EC 2.4.1.13)...
0.956	At2g42860	expressed protein
0.955	At2g23220	cytochrome P450, putative
0.955	At3g04200	germin-like protein, putative contains Pfam profile: PF01072 germin family; similar to germin type2 GB:CAA63023 [SP|P92996]...
0.955	At5g44360	FAD-binding domain-containing protein similar to SP|P30986 reticuline oxidase precursor (Berberine-bridge-forming enzyme)...
0.954	At2g23260	UDP-glucoronosyl/UDP-glucosyl transferase family protein contains Pfam profile: PF00201 UDP-glucoronosyl and UDP-glucosyl...
0.953	At5g51490	pectinesterase family protein contains Pfam profile: PF01095 pectinesterase
0.951	At5g48100	laccase family protein / diphenol oxidase family protein similar to laccase [Pinus taeda][GI:13661197]
0.95	At1g21970	CCAAT-box binding transcription factor (LEC1) similar to CAAT-box DNA binding protein subunit B (NF-YB) (SP:P25209) (GI:22380)...
0.95	At3g22640	cupin family protein contains similarity to vicilin-like protein precursor [Juglans regia] GI:6580762, vicilin precursor...
0.95	At5g03810	GDSL-motif lipase/hydrolase family protein similar to family II lipase EXL3 (GI:15054386), EXL1 (GI:15054382), EXL2...
0.947	At1g67100	LOB domain protein 40 / lateral organ boundaries domain protein 40 (LBD40) identical to SP|Q9ZW96 LOB domain protein 40...
0.947	At5g62800	seven in absentia (SINA) family protein similar to SIAH1 protein [Brassica napus var. napus] GI:7657876; contains Pfam profile...
0.946	At3g28360	ABC transporter family protein similar to P-glycoprotein homologue GI:2292907 from [Hordeum vulgare subsp. vulgare]
0.945	At1g28030	oxidoreductase, 2OG-Fe(II) oxygenase family protein similar to GS-AOP loci [GI:16118889, GI:16118887, GI:16118891,...
0.945	At5g09640	sinapoylglucose:choline sinapoyltransferase (SNG2) GC donor splice site at exon 11 and 13; TA donor splice site at exon 10;...
0.944	At4g32490	plastocyanin-like domain-containing protein
0.943	At3g24250	glycine-rich protein
0.942	At1g15150	MATE efflux family protein similar to ripening regulated protein DDTFR18 [Lycopersicon esculentum] GI:12231296; contains Pfam...
0.942	At3g04280	two-component responsive regulator family protein / response regulator family protein contains Pfam profile: PF00072 response...
0.941	At5g39130	germin-like protein, putative identical to germin-like protein subfamily 1 member 16 (SP|Q9FIC8)
0.94	At1g05280	fringe-related protein Similar to hypothetical protein PID|e327464 (gb|Z97338) various hypothetical proteins from Arabidopsis...
0.94	At2g46960	cytochrome P450 family protein similar to cytochrome P450 72A1 (SP:Q05047) [Catharanthus roseus]; contains Pfam profile:...
0.94	At4g27140	2S seed storage protein 1 / 2S albumin storage protein / NWMU1-2S albumin 1 identical to SP|P15457
0.94	At5g50770	short-chain dehydrogenase/reductase (SDR) family protein similar to sterol-binding dehydrogenase steroleosin GI:15824408 from...
0.939	At1g18100	mother of FT and TF1 protein (MFT) identical to SP|Q9XFK7 MOTHER of FT and TF1 protein {Arabidopsis thaliana}; contains Pfam...
0.939	At2g28650	exocyst subunit EXO70 family protein contains Pfam domain PF03081: Exo70 exocyst complex subunit
0.938	At3g44460	basic leucine zipper transcription factor (BZIP67) identical to basic leucine zipper transcription factor GI:18656053 from...
0.937	At4g26740	embryo-specific protein 1 (ATS1) identical to embryo-specific protein 1 [Arabidopsis thaliana] GI:3335169
0.935	At3g24650	abscisic acid-insensitive protein 3 (ABI3) identical to abscisic acid-insensitive protein 3 GI:16146 SP:Q01593 from...
0.934	At1g28650	lipase, putative strong similarity to lipase [Arabidopsis thaliana] GI:1145627
0.934	At5g57260	cytochrome P450 71B10 identical to cytochrome P450 71B10 (SP:Q9LVD2) [Arabidopsis thaliana]
0.933	At1g17810	major intrinsic family protein / MIP family protein contains Pfam profile: MIP PF00230
0.932	At3g02590	delta 7-sterol-C5-desaturase, putative similar to delta7 sterol C-5 desaturase GI:5031219 from [Arabidopsis thaliana]
0.932	At3g13540	myb family transcription factor contains Pfam profile: PF00249 myb-like DNA-binding domain
0.932	At4g00220	LOB domain protein 30 / lateral organ boundaries domain protein 30 (LBD30) identical to LOB DOMAIN 30 [Arabidopsis thaliana]...
0.932	At5g54740	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein similar to 2S seed storage proteins from Arabidosis...
0.931	At1g27080	proton-dependent oligopeptide transport (POT) family protein similar to nitrate transporter NRT1-5 [Glycine max] GI:11933414;...
0.931	At3g27785	myb family transcription factor (MYB118) contains PFAM profile: PF00249 myb-like DNA binding domain
0.927	At1g16980	alpha, alpha-trehalose-phosphate synthase, UDP-forming, putative / trehalose-6-phosphate synthase, putative /...
0.927	At1g48130	peroxiredoxin (PER1) / rehydrin, putative identical to peroxiredoxin (Rehydrin homolog) [Arabidopsis thaliana]...
0.927	At1g48910	flavin-containing monooxygenase family protein / FMO family protein similar to flavin monoxygenase-like protein floozy [Petunia...
0.927	At1g80330	gibberellin 3-beta-dioxygenase, putative / gibberellin 3 beta-hydroxylase, putative similar to gibberellin 3 beta-hydroxylase...
0.927	At3g54940	cysteine proteinase, putative contains similarity to cysteine proteinase GI:479060 from [Glycine max]
0.927	At5g59170	proline-rich family protein contains proline-rich extensin domains, INTERPRO:IPR002965
0.925	At5g07190	embryo-specific protein 3, putative similar to embryo-specific protein 3 GI:3335171 from [Arabidopsis thaliana]
0.924	At5g03800	exostosin family protein / pentatricopeptide (PPR) repeat-containing protein contains Pfam profiles: PF03016 exostosin family,...
0.924	At5g12460	fringe-related protein similarity to predicted proteins + similar to hypothetical protein GB:AAC23643 [Arabidopsis thaliana] +...
0.924	At5g57920	plastocyanin-like domain-containing protein
0.922	At5g51210	glycine-rich protein / oleosin
0.922	At5g55370	long-chain-alcohol O-fatty-acyltransferase family protein / wax synthase family protein contains similarity to wax synthase...
0.921	At1g71120	GDSL-motif lipase/hydrolase family protein similar to family II lipases EXL3 GI:15054386 from [Arabidopsis thaliana]; contains...
0.918	At2g15325	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam protease inhibitor/seed storage/LTP...
0.916	At1g04560	AWPM-19-like membrane family protein contains Pfam PF05512: AWPM-19-like family; similar to late embryogenesis abundant...
0.915	At2g23550	hydrolase, alpha/beta fold family protein similar to ethylene-induced esterase [Citrus sinensis] GI:14279437, polyneuridine...

ArathACLL4 (At1g20510) stress treatments (298 data)

0.861	At1g17380	expressed protein
0.851	At3g25780	allene oxide cyclase, putative / early-responsive to dehydration protein, putative / ERD protein, putative similar to allene...
0.847	At2g06050	12-oxophytodienoate reductase (OPR3) / delayed dehiscence1 (DDE1) nearly identical to DELAYED DEHISCENCE1 [GI:7688991] and to...
0.831	At1g72520	lipoxygenase, putative similar to lipoxygenase gi:1495804 [Solanum tuberosum], gi:1654140 [Lycopersicon esculentum],...
0.83	At1g17420	lipoxygenase, putative similar to lipoxygenase gi:1495804 [Solanum tuberosum], gi:1654140 [Lycopersicon esculentum]
0.827	At1g73080	leucine-rich repeat transmembrane protein kinase, putative similar to receptor protein kinase GI:1389566 from [Arabidopsis...
0.798	At1g17750	leucine-rich repeat transmembrane protein kinase, putative similar to receptor-like protein kinase INRPK1 GI:1684913 from...
0.793	At3g09830	protein kinase, putative similar to protein kinase [Lophopyrum elongatum] gi|13022177|gb|AAK11674
0.791	At1g74950	expressed protein
0.79	At3g51450	strictosidine synthase family protein similar to hemomucin [Drosophila melanogaster][GI:1280434], strictosidine synthase...
0.788	At4g34410	AP2 domain-containing transcription factor, putative ethylene-responsive element binding protein homolog, Stylosanthes hamata,...
0.787	At5g13220	expressed protein
0.784	At3g44860	S-adenosyl-L-methionine:carboxyl methyltransferase family protein similar to defense-related protein cjs1 [Brassica...
0.777	At1g28480	glutaredoxin family protein contains INTERPRO Domain IPR002109, Glutaredoxin (thioltransferase)
0.776	At1g32640	basic helix-loop-helix (bHLH) protein (RAP-1) identical to bHLH protein GB:CAA67885 GI:1465368 from [Arabidopsis thaliana]
0.768	At5g12340	expressed protein ; expression supported by MPSS
0.766	At2g44840	ethylene-responsive element-binding protein, putative
0.754	At1g19180	expressed protein
0.751	At1g30135	expressed protein
0.751	At1g76040	calcium-dependent protein kinase, putative / CDPK, putative similar to calcium-dependent protein kinase GB:AAC25423 GI:3283996...
0.744	At3g08720	serine/threonine protein kinase (PK19) identical to serine/threonine-protein kinase AtPK19 (Ribosomal-protein S6 kinase...
0.741	At1g28380	expressed protein
0.735	At1g06620	2-oxoglutarate-dependent dioxygenase, putative similar to 2A6 (GI:599622) and tomato ethylene synthesis regulatory protein E8...
0.731	At4g24380	expressed protein contains Pfam profile: PF03959 domain of unknown function (DUF341)
0.731	At5g42650	allene oxide synthase (AOS) / hydroperoxide dehydrase / cytochrome P450 74A (CYP74A) identical to Allene oxide synthase,...
0.723	At2g13790	leucine-rich repeat family protein / protein kinase family protein
0.721	At1g44350	IAA-amino acid hydrolase 6, putative (ILL6) / IAA-Ala hydrolase, putative virtually identical to gr1-protein from [Arabidopsis...
0.721	At4g23180	receptor-like protein kinase 4, putative (RLK4) nearly identical to receptor-like protein kinase 4 [Arabidopsis thaliana]...
0.72	At3g01830	calmodulin-related protein, putative similar to regulator of gene silencing calmodulin-related protein GI:12963415 from...
0.72	At4g17230	scarecrow-like transcription factor 13 (SCL13)
0.719	At1g70700	expressed protein
0.718	At2g14290	F-box family protein contains F-box domain Pfam:PF00646
0.711	At5g61900	copine BONZAI1 (BON1) nearly identical to BONZAI1 [Arabidopsis thaliana] GI:15487382; contains Pfam profile PF00168: C2 domain
0.71	At3g23250	myb family transcription factor (MYB15) similar to myb-related transcription factor GB:CAA66952 from [Lycopersicon esculentum]
0.709	At3g13050	transporter-related low similarity to apical organic cation transporter [Sus scrofa] GI:2062135, SP|Q02563 Synaptic vesicle...
0.707	At2g22880	VQ motif-containing protein contains PF05678: VQ motif
0.707	At2g33580	protein kinase family protein / peptidoglycan-binding LysM domain-containing protein protein kinase [Arabidopsis thaliana]...
0.706	At1g74430	myb family transcription factor (MYB95) contains Pfam profile: PF00249 myb-like DNA-binding domain
0.706	At5g44070	phytochelatin synthase 1 (PCS1) identical to phytochelatin synthase [Arabidopsis thaliana] gi|18254401|gb|AAL66747; identical...
0.705	At3g11820	syntaxin 121 (SYP121) / syntaxin-related protein (SYR1) contains Pfam profiles: PF00804 syntaxin and PF05739: SNARE domain;...
0.702	At2g32140	disease resistance protein (TIR class), putative domain signature TIR exists, suggestive of a disease resistance protein.
0.7	At3g17690	cyclic nucleotide-binding transporter 2 / CNBT2 (CNGC19) identical to cyclic nucleotide-binding transporter 2 (CNBT2)...
0.698	At1g22810	AP2 domain-containing transcription factor, putative Contains similarity to transcription factor (TINY) isolog T02O04.22...
0.696	At1g72280	endoplasmic reticulum oxidoreductin 1 (ERO1) family protein contains Pfam domain, PF04137: Endoplasmic Reticulum Oxidoreductin...
0.696	At1g72450	expressed protein
0.694	At5g53050	hydrolase, alpha/beta fold family protein contains Pfam profile PF00561: hydrolase, alpha/beta fold family
0.689	At1g07000	exocyst subunit EXO70 family protein similar to leucine zipper protein GI:10177020 from [Arabidopsis thaliana] contains Pfam...
0.689	At2g46510	basic helix-loop-helix (bHLH) family protein contains Pfam profile: PF00010 helix-loop-helix DNA-binding domain
0.687	At2g27690	cytochrome P450, putative similar to Cytochrome P450 94A1 (P450-dependent fatty acid omega-hydroxylase) (SP:O81117) {Vicia...
0.686	At4g12720	MutT/nudix family protein similar to SP|P53370 Nucleoside diphosphate-linked moiety X motif 6 {Homo sapiens}; contains Pfam...
0.684	At4g30430	senescence-associated family protein similar to senescence-associated protein 5 [Hemerocallis hybrid cultivar]...
0.683	At4g23190	protein kinase family protein contains Pfam PF00069: Protein kinase domain
0.683	At4g34390	extra-large guanine nucleotide binding protein, putative / G-protein, putative similar to extra-large G-protein (XLG)...
0.683	At5g57890	anthranilate synthase beta subunit, putative strong similarity to anthranilate synthase beta chain GI:403434 [Arabidopsis...
0.681	At1g80840	WRKY family transcription factor similar to WRKY transcription factor GB:BAA87058 GI:6472585 from [Nicotiana tabacum]
0.681	At5g25930	leucine-rich repeat family protein / protein kinase family protein contains similarity to Swiss-Prot:P47735 receptor-like...
0.68	At4g10390	protein kinase family protein contains protein kinase domain, Pfam:PF00069
0.679	At2g27310	F-box family protein contains Pfam PF00646: F-box domain;; similar to SKP1 interacting partner 2 (SKIP2) TIGR_Ath1:At5g67250
0.679	At5g47220	ethylene-responsive element-binding factor 2 (ERF2) identical to SP|O80338 Ethylene responsive element binding factor 2...
0.678	At2g25460	expressed protein
0.676	At3g19970	expressed protein
0.676	At5g13190	expressed protein
0.675	At1g16370	transporter-related low similarity to organic cation transporter OCTN1 from [Homo sapiens] GI:2605501, [Mus musculus]...
0.675	At1g69840	band 7 family protein strong similarity to hypersensitive-induced response protein [Zea mays] GI:7716466; contains Pfam profile...
0.673	At2g34600	expressed protein
0.673	At5g14700	cinnamoyl-CoA reductase-related similar to cinnamoyl-CoA reductase from Pinus taeda [GI:17978649], Saccharum officinarum...
0.671	At1g20310	expressed protein
0.67	At1g27770	calcium-transporting ATPase 1, plasma membrane-type / Ca(2+)-ATPase isoform 1 (ACA1) / plastid envelope ATPase 1 (PEA1)...
0.67	At3g44400	disease resistance protein (TIR-NBS-LRR class), putative domain signature TIR-NBS-LRR exists, suggestive of a disease...
0.67	At4g39890	Ras-related GTP-binding family protein contains Pfam profile: PF00071 Ras family
0.67	At5g66640	LIM domain-containing protein-related contains low similarity to Pfam profile PF00412: LIM domain
0.669	At3g16860	phytochelatin synthetase-related contains Pfam PF04833: Phytochelatin synthetase-like conserved region
0.669	At5g05300	expressed protein similar to unknown protein (gb|AAF01528.1); expression supported by MPSS
0.668	At4g14680	sulfate adenylyltransferase 3 / ATP-sulfurylase 3 (APS3) identical to ATP sulfurylase (APS3) [Arabidopsis thaliana] GI:1575327
0.667	At1g19210	AP2 domain-containing transcription factor, putative similar to AP2 domain transcription factor GI:4567204 from [Arabidopsis...
0.665	At3g55950	protein kinase family protein contains protein kinase domain, Pfam:PF00069; similar to cytokinin-regulated kinase 1 [Nicotiana...
0.665	At4g23220	protein kinase family protein contains Pfam PF00069: Protein kinase domain
0.664	At1g26730	EXS family protein / ERD1/XPR1/SYG1 family protein similar to PHO1 protein [Arabidopsis thaliana] GI:20069032; contains Pfam...
0.664	At2g24850	aminotransferase, putative similar to nicotianamine aminotransferase from Hordeum vulgare [GI:6498122, GI:6469087]; contains...
0.664	At3g02840	immediate-early fungal elicitor family protein similar to immediate-early fungal elicitor protein CMPG1 (GI:14582200)...
0.664	At4g39030	enhanced disease susceptibility 5 (EDS5) / salicylic acid induction deficient 1 (SID1) identical to SP|Q945F0; contains Pfam...
0.663	At2g26530	expressed protein
0.662	At1g03370	C2 domain-containing protein / GRAM domain-containing protein contains Pfam profiles PF00168: C2 domain; contains PF02893: GRAM...
0.661	At5g05140	transcription elongation factor-related low similarity to transcription elongation factor TFIIS.h [Mus musculus] GI:3288547,...
0.659	At1g12610	DRE-binding protein, putative / CRT/DRE-binding factor, putative similar to DREB1A GI:3738224 from [Arabidopsis thaliana] and...
0.658	At1g53885	senescence-associated protein-related similar to senescence-associated protein SAG102 (GI:22331931) [Arabidopsis thaliana];
0.658	At2g22860	phytosulfokines 2 (PSK2) identical to phytosulfokines 2 (PSK2) from [Arabidopsis thaliana]
0.658	At3g11840	U-box domain-containing protein low similarity to immediate-early fungal elicitor protein CMPG1 [Petroselinum crispum]...
0.657	At4g21390	S-locus lectin protein kinase family protein contains Pfam profiles: PF00954 S-locus glycoprotein family, PF00069 protein...
0.654	At1g51780	IAA-amino acid hydrolase 5 / auxin conjugate hydrolase (ILL5) identical to auxin conjugate hydrolase ILL5 [Arabidopsis...
0.654	At4g24160	hydrolase, alpha/beta fold family protein contains Pfam profile PF00561: hydrolase, alpha/beta fold family
0.651	At1g71697	choline kinase, putative similar to GmCK2p choline kinase gi|1438881|gb|AAC49375
0.651	At5g52050	MATE efflux protein-related contains Pfam profile PF01554: Uncharacterized membrane protein family
0.648	At3g15210	ethylene-responsive element-binding factor 4 (ERF4) identical to ethylene responsive element binding factor 4 SP:O80340 from...
0.648	At4g14365	zinc finger (C3HC4-type RING finger) family protein / ankyrin repeat family protein contains Pfam profile: PF00097 zinc finger,...
0.647	At3g06500	beta-fructofuranosidase, putative / invertase, putative / saccharase, putative / beta-fructosidase, putative similar to neutral...
0.647	At5g19110	extracellular dermal glycoprotein-related / EDGP-related similar to extracellular dermal glycoprotein EDGP precursor [Daucus...
0.646	At3g21070	ATP-NAD kinase family protein contains Pfam domain, PF01513: ATP-NAD kinase
0.646	At5g01540	lectin protein kinase, putative similar to receptor lectin kinase 3 [Arabidopsis thaliana] gi|4100060|gb|AAD00733; contains...
0.646	At5g63450	cytochrome P450, putative

ArathACLL8 (At5g38120) Tissue and development (237data)

0.846	At1g70720	invertase/pectin methylesterase inhibitor family protein low similarity to pectinesterase from Lycopersicon esculentum...
0.811	At3g51410	expressed protein contains Pfam profile PF03087: Arabidopsis protein of unknown function
0.806	At2g17470	expressed protein contains Pfam profile PF01027: Uncharacterized protein family UPF0005
0.798	At4g14750	calmodulin-binding family protein contains Pfam profile PF00612: IQ calmodulin-binding motif
0.796	At1g59640	basic helix-loop-helix (bHLH) family protein
0.789	At1g15360	AP2 domain-containing transcription factor family protein Similar to SP|P16146 PPLZ02 protein {Lupinus polyphyllus}; contains...
0.781	At1g75880	family II extracellular lipase 1 (EXL1) EXL1 (PMID:11431566); similar to anter-specific proline-rich protein (APG) SP:P40602...
0.777	At2g34340	expressed protein contains Pfam profile PF04520: Protein of unknown function, DUF584
0.761	At1g75300	isoflavone reductase, putative identical to SP|P52577 Isoflavone reductase homolog P3 (EC 1.3.1.-) {Arabidopsis thaliana};...
0.741	At4g10240	zinc finger (B-box type) family protein zinc-finger protein R2931, Oryza sativa, PIR3:JE0116
0.73	At3g16170	acyl-activating enzyme 13 (AAE13) similar to malonyl CoA synthetase GB:AAF28840 from [Bradyrhizobium japonicum]; contains Pfam...
0.727	At1g32780	alcohol dehydrogenase, putative similar to alcohol dehydrogenase GB:CAA37333 GI:297178 from [Solanum tuberosum]; contains Pfam...
0.726	At4g16590	glucosyltransferase-related low similarity to beta-(1-3)-glucosyl transferase [Bradyrhizobium japonicum] GI:3687658
0.719	At2g44770	phagocytosis and cell motility protein ELMO1-related contains weak similarity to ELMO1 [Mus musculus] gi|16118551|gb|AAL14464
0.714	At5g20240	floral homeotic protein PISTILLATA (PI) contains Pfam profiles PF01486: K-box region and PF00319: SRF-type transcription factor..
0.712	At1g16750	expressed protein contains Pfam profile PF04784: Protein of unknown function, DUF547
0.707	At2g38110	phospholipid/glycerol acyltransferase family protein low similarity to SP|O87707 CicA protein {Caulobacter crescentus};...
0.706	At1g35310	Bet v I allergen family protein similar to Csf-2 [Cucumis sativus][GI:5762258][J Am Soc Hortic Sci 124, 136-139 (1999)] ;...
0.705	At1g78960	lupeol synthase, putative / 2,3-oxidosqualene-triterpenoid cyclase, putative similar to lupeol synthase GI:1762150 from...
0.703	At3g10570	cytochrome P450, putative similar to cytochrome P450 77A3 GB:O48928 [Glycine max]
0.7	At4g13790	auxin-responsive protein, putative similar to small auxin up RNA (SAUR-AC1) (SP:S70188) [Arabidopsis thaliana]
0.696	At5g50335	expressed protein
0.692	At3g44610	protein kinase family protein similar to viroid symptom modulation protein (protein kinase)[Lycopersicon esculentum]...
0.689	At2g42900	expressed protein
0.688	At3g50630	kip-related protein 2 (KRP2) / cyclin-dependent kinase inhibitor 2 (ICK2) / cdc2a-interacting protein identical to...
0.685	At1g61680	terpene synthase/cyclase family protein similar to 1,8-cineole synthase [GI:3309117][Salvia officinalis]; contains Pfam...
0.683	At4g32460	expressed protein contains Pfam profile PF04862: Protein of unknown function, DUF642
0.68	At1g66350	gibberellin regulatory protein (RGL1) similar to GB:CAA75492 from [Arabidopsis thaliana]; contains Pfam profile PF03514: GRAS...
0.68	At2g41830	cyclin-related contains Pfam profile PF02984: Cyclin, C-terminal domain
0.676	At4g13000	protein kinase family protein contains protein kinase domain, Pfam:PF00069
0.675	At1g11000	seven transmembrane MLO family protein / MLO-like protein 4 (MLO4) identical to membrane protein Mlo4 [Arabidopsis thaliana]...
0.675	At5g17540	transferase family protein similar to hypersensitivity-related gene product HSR201 - Nicotiana tabacum, EMBL:X95343; contains...
0.673	At5g40350	myb family transcription factor (MYB24) similar to Myb26 GI:1841475 from [Pisum sativum]
0.672	At4g22730	leucine-rich repeat transmembrane protein kinase, putative leucine rich repeat receptor-like kinase, Oryza sativa, PATCHX:E267533
0.669	At5g45960	GDSL-motif lipase/hydrolase family protein
0.668	At5g49130	MATE efflux family protein contains Pfam profile PF01554: MatE Uncharacterized membrane protein family
0.664	At1g78680	gamma-glutamyl hydrolase (GGH1) / gamma-Glu-X carboxypeptidase / conjugase identical to SP|O65355 Gamma-glutamyl hydrolase...
0.664	At5g16440	isopentenyl-diphosphate delta-isomerase I / isopentenyl diphosphate:dimethylallyl diphosphate isomerase I (IPP1) identical to...
0.656	At1g01600	cytochrome P450, putative similar to cytochrome P450 GI:10442763 from [Triticum aestivum]
0.655	At2g24210	myrcene/ocimene synthase (TPS10) nearly identical to GI:9957293; contains Pfam profile: PF01397 terpene synthase family
0.654	At1g63710	cytochrome P450, putative similar to cytochrome P450 GB:O23066 [Arabidopsis thaliana]
0.654	At2g22960	serine carboxypeptidase S10 family protein contains Pfam profile: PF00450 serine carboxypeptidase ;similar to...
0.654	At3g48460	GDSL-motif lipase/hydrolase family protein similar to lipase [Arabidopsis thaliana] GI:1145627; contains InterPro Entry...
0.653	At5g50710	hypothetical protein
0.652	At1g35180	expressed protein similar to hypothetical protein GB:AAF69173 GI:7767676 from [Arabidopsis thaliana]
0.646	At3g01980	short-chain dehydrogenase/reductase (SDR) family protein contains Pfam profiles: PF00106 short chain dehydrogenase, PF00678...
0.645	At3g01750	ankyrin repeat family protein contains ankyrin repeats, Pfam:PF00023
0.644	At1g23600	expressed protein contains Pfam profile PF02713: Domain of unknown function DUF220; expression supported by MPSS
0.644	At2g16260	glycine-rich RNA-binding protein, putative similar to Glycine-rich RNA-binding protein from {Daucus carota} SP|Q03878, {Sinapis...
0.644	At4g01080	expressed protein
0.643	At1g22460	expressed protein similar to axi 1 [Nicotiana tabacum] GI:559921; contains Pfam profile PF03138: Plant protein family
0.642	At2g47050	invertase/pectin methylesterase inhibitor family protein low similarity to pollen-specific pectin esterase [Brassica rapa...
0.641	At3g01140	myb family transcription factor (MYB106) similar to transforming protein (myb) homolog GB:S26605 from [Petunia x hybrida]
0.639	At3g23450	pseudogene, similar to unnamed protein product blastp match of 19% identity and 2.8e-15 P-value to...
0.639	At4g33390	hypothetical protein contains Pfam profile PF05701: Plant protein of unknown function (DUF827)
0.638	At3g28007	nodulin MtN3 family protein contains Pfam PF03083 MtN3/saliva family; similar to LIM7 GI:431154 (induced in meiotic prophase in...
0.634	At3g27810	myb family transcription factor (MYB3) (MYB21) contains Pfam profile: PF00249 myb-like DNA-binding domain ;identical to ATMYB3...
0.633	At3g51150	kinesin motor family protein contains Pfam domain, PF00225: Kinesin motor domain
0.63	At3g55310	short-chain dehydrogenase/reductase (SDR) family protein contains similarity to 3-oxoacyl-[acyl-carrier protein] reductase...
0.629	At1g10060	branched-chain amino acid aminotransferase 1 / branched-chain amino acid transaminase 1 (BCAT1) nearly identical to SP|Q93Y32...
0.629	At1g16705	p300/CBP acetyltransferase-related protein-related similar to p300/CBP acetyltransferase-related protein 2 [Arabidopsis...
0.625	At1g19650	SEC14 cytosolic factor, putative / phosphoglyceride transfer protein, putative similar to SP:P24859 from [Kluyveromyces...
0.625	At2g27920	serine carboxypeptidase S10 family protein similar to retinoid-inducible serine carboxypeptidase precursor (GI:15146429) [Mus...
0.622	At1g08510	acyl-[acyl carrier protein] thioesterase / acyl-ACP thioesterase / oleoyl-[acyl-carrier protein] hydrolase / S-acyl fatty acid...
0.622	At1g11410	S-locus protein kinase, putative similar to receptor-like protein kinase [Arabidopsis thaliana] gi|4008008|gb|AAC95352;...
0.622	At2g40475	expressed protein
0.622	At5g55720	pectate lyase family protein similar to pectate lyase 1 GP:6606532 from [Musa acuminata]
0.621	At3g18850	phospholipid/glycerol acyltransferase family protein contains Pfam profile: PF01553 Acyltransferase
0.619	At3g55700	UDP-glucoronosyl/UDP-glucosyl transferase family protein glucuronosyl transferase homolog, Lycopersicon esculentum, PIR:S39507...
0.617	At2g42540	cold-responsive protein / cold-regulated protein (cor15a) identical to cold-regulated protein cor15a [Arabidopsis thaliana]...
0.617	At2g42990	GDSL-motif lipase/hydrolase family protein similar to family II lipase EXL3 (GI:15054386), EXL1 (GI:15054382), EXL2...
0.616	At1g11850	expressed protein
0.615	At2g04570	GDSL-motif lipase/hydrolase family protein similar to family II lipase EXL3 (GI:15054386), EXL1 (GI:15054382), EXL2...
0.614	At2g42620	F-box family protein (ORE9) E3 ubiquitin ligase SCF complex F-box subunit; identical to F-box containing protein ORE9...
0.614	At4g27790	calcium-binding EF hand family protein contains INTERPRO:IPR002048 calcium-binding EF-hand domain
0.613	At3g14380	integral membrane family protein similar to unknown protein GB:AAD50013 from [Arabidopsis thaliana]; contains TIGRFAM TIGR01569...
0.612	At4g34940	armadillo/beta-catenin repeat family protein contains Pfam profile: PF00514 armadillo/beta-catenin-like repeat
0.611	At3g08990	yippee family protein similar to qdgl-1 [Coturnix coturnix] GI:10441650, Yippee protein [Drosophila melanogaster] GI:5713279;...
0.61	At3g29370	expressed protein
0.609	At4g15980	pectinesterase family protein contains Pfam profile: PF01095 pectinesterase
0.608	At1g61350	armadillo/beta-catenin repeat family protein armadillo/beta-catenin-like repeats, Pfam:PF00514
0.608	At4g27840	expressed protein
0.608	At5g15780	pollen Ole e 1 allergen and extensin family protein contains Pfam profile PF01190: Pollen proteins Ole e I family
0.604	At2g20870	cell wall protein precursor, putative identical to Putative cell wall protein precursor (Swiss-Prot:P47925) [Arabidopsis...
0.602	At5g49330	myb family transcription factor contains Pfam profile: PF00249 myb-like DNA binding domain; identical to cDNA putative...
0.601	At2g13570	CCAAT-box binding transcription factor, putative similar to CAAT-box DNA binding protein subunit B (NF-YB) (SP:P25209)...
0.601	At3g11210	GDSL-motif lipase/hydrolase family protein contains Pfam profile PF00657: Lipase/Acylhydrolase with GDSL-like motif
0.601	At5g48900	pectate lyase family protein similar to pectate lyase GP:14531296 from [Fragaria x ananassa]; non-consensus AG donor splice...
0.6	At5g23970	transferase family protein similar to acetyl CoA: benzylalcohol acetyltransferase; BEAT [Clarkia...

CLADE E

ArathACLL9 (At5g63380) hormone treatments (236 data)

0.827	At4g33530	potassium transporter family protein similar to K+ transporter HAK5 [Arabidopsis thaliana] GI:7108597; KUP/HAK/KT Transporter...
0.814	At5g40890	chloride channel protein (CLC-a) identical to GI:1742952 (gb|AAC05742.1)
0.812	At3g23820	NAD-dependent epimerase/dehydratase family protein similar to nucleotide sugar epimerase from Vibrio vulnificus GI:3093975...
0.811	At5g13710	sterol 24-C-methyltransferase, putative similar to SP:P25087 Sterol 24-C-methyltransferase, Delta(24)-sterol C-...
0.808	At2g04780	fasciclin-like arabinogalactan-protein (FLA7) identical to gi_13377782_gb_AAK20860
0.807	At1g45688	expressed protein
0.801	At1g79340	latex-abundant protein, putative (AMC7) / caspase family protein similar to latex-abundant protein [Hevea brasiliensis]...
0.801	At5g03760	glycosyl transferase family 2 protein similar to beta-(1-3)-glucosyl transferase GB:AAC62210 GI:3687658 from [Bradyrhizobium...
0.797	At1g01620	plasma membrane intrinsic protein 1C (PIP1C) / aquaporin PIP1.3 (PIP1.3) / transmembrane protein B (TMPB) identical to plasma...
0.795	At2g39900	LIM domain-containing protein similar to pollen specific LIM domain protein 1b [Nicotiana tabacum] GI:6467905, PGPS/D1 [Petunia...
0.795	At5g47770	farnesyl pyrophosphate synthetase 1, mitochondrial (FPS1) / FPP synthetase 1 / farnesyl diphosphate synthase 1 identical to...
0.79	At3g23810	adenosylhomocysteinase, putative / S-adenosyl-L-homocysteine hydrolase, putative / AdoHcyase, putative strong similarity to...
0.789	At4g22010	multi-copper oxidase type I family protein similar to pollen-specific BP10 protein [SP|Q00624][Brassica napus]; contains Pfam...
0.788	At1g24170	glycosyl transferase family 8 protein contains Pfam profile: PF01501 glycosyl transferase family 8
0.787	At1g48480	leucine-rich repeat transmembrane protein kinase, putative contains similarity to many predicted protein kinases
0.787	At5g49720	endo-1,4-beta-glucanase KORRIGAN (KOR) / cellulase (OR16pep) identical to endo-1,4-beta-D-glucanase KORRIGAN [Arabidopsis...
0.785	At4g29220	phosphofructokinase family protein similar to phosphofructokinase [Amycolatopsis methanolica] GI:17432243; contains Pfam...
0.785	At4g34640	farnesyl-diphosphate farnesyltransferase 1 / squalene synthase 1 (SQS1) identical to SP|P53799 Farnesyl-diphosphate...
0.784	At1g70370	BURP domain-containing protein / polygalacturonase, putative similar to polygalacturonase isoenzyme 1 beta subunit...
0.783	At3g01810	expressed protein
0.783	At5g15350	plastocyanin-like domain-containing protein contains plastocyanin-like domain Pfam:PF02298
0.782	At1g04430	dehydration-responsive protein-related similar to early-responsive to dehydration stress ERD3 protein [Arabidopsis thaliana]...
0.782	At5g20700	senescence-associated protein-related similar to senescence-associated protein SAG102 (GI:22331931) [Arabidopsis thaliana];
0.781	At5g46700	senescence-associated protein, putative similar to senescence-associated protein 5 [Hemerocallis hybrid cultivar]...
0.78	At1g66200	glutamine synthetase, putative similar to glutamine synthetase, cytosolic isozyme (Glutamate-- ammonia ligase, GS1) [Lotus...
0.78	At5g44020	acid phosphatase class B family protein similar to SP|P15490 STEM 28 kDa glycoprotein precursor (Vegetative storage protein A)...
0.78	At5g67420	LOB domain protein 37 / lateral organ boundaries domain protein 37 (LBD37) identical to LOB DOMAIN 37 [Arabidopsis thaliana]...
0.778	At1g27930	expressed protein contains Pfam profile PF04669: Protein of unknown function (DUF579)
0.778	At5g17820	peroxidase 57 (PER57) (P57) (PRXR10) identical to SP|Q43729 Peroxidase 57 precursor (EC 1.11.1.7) (Atperox P57) (PRXR10)...
0.777	At1g67330	expressed protein contains Pfam profile PF04669: Protein of unknown function (DUF579)
0.777	At3g23000	CBL-interacting protein kinase 7 (CIPK7) identical to CBL-interacting protein kinase 7 [Arabidopsis thaliana]...
0.777	At4g12730	fasciclin-like arabinogalactan-protein (FLA2) identical to gi_13377778_gb_AAK20858
0.776	At1g04040	acid phosphatase class B family protein similar to SP|P15490 STEM 28 kDa glycoprotein precursor (Vegetative storage protein A)...
0.776	At1g05210	expressed protein
0.776	At1g62660	beta-fructosidase (BFRUCT3) / beta-fructofuranosidase / invertase, vacuolar identical to beta-fructosidase GB:CAA67560...
0.775	At1g12500	phosphate translocator-related low similarity to glucose-6-phosphate/phosphate-translocator precursor [Zea mays] GI:2997589,...
0.775	At1g20010	tubulin beta-5 chain (TUB5) nearly identical to SP|P29513 Tubulin beta-5 chain {Arabidopsis thaliana}
0.774	At1g09780	2,3-biphosphoglycerate-independent phosphoglycerate mutase, putative / phosphoglyceromutase, putative strong similarity to...
0.774	At3g16460	jacalin lectin family protein contains Pfam profile: PF01419 jacalin-like lectin domain; similar to myrosinase binding protein...
0.773	At1g70410	carbonic anhydrase, putative / carbonate dehydratase, putative similar to SP|P42737 Carbonic anhydrase 2 (EC 4.2.1.1)...
0.773	At2g32380	expressed protein
0.773	At4g37450	arabinogalactan-protein (AGP18) identical to gi_11935088_gb_AAG41964
0.771	At1g02500	S-adenosylmethionine synthetase 1 (SAM1) identical to S-adenosylmethionine synthetase 1 (Methionine adenosyltransferase 1,...
0.771	At3g05890	hydrophobic protein (RCI2B) / low temperature and salt responsive protein (LTI6B) identical to SP|Q9ZNS6 Hydrophobic protein...
0.77	At1g45130	beta-galactosidase, putative / lactase, putative similar to beta-galactosidase [Lycopersicon esculentum] GI:7939619,...
0.77	At2g15970	cold-acclimation protein, putative (FL3-5A3) similar to cold acclimation WCOR413-like protein gamma form [Hordeum vulgare]...
0.77	At2g36880	S-adenosylmethionine synthetase, putative similar to S-adenosylmethionine synthetase 3 (Methionine adenosyltransferase 3,...
0.77	At4g11290	peroxidase, putative identical to peroxidase ATP19a [Arabidopsis thaliana] gi|1546692|emb|CAA67337
0.77	At4g23850	long-chain-fatty-acid--CoA ligase / long-chain acyl-CoA synthetase nearly identical to acyl-CoA synthetase (MF7P) from Brassica...
0.77	At4g25570	cytochrome B561 family protein contains Pfam domain, PF03188: Cytochrome b561
0.769	At1g23480	glycosyl transferase family 2 protein similar to cellulose synthase from Agrobacterium tumeficiens [gi:710492] and...
0.768	At1g03870	fasciclin-like arabinogalactan-protein (FLA9) identical to gi_13377784_gb_AAK20861
0.767	At1g10670	expressed protein
0.767	At5g67400	peroxidase 73 (PER73) (P73) (PRXR11) identical to SP|Q43873 Peroxidase 73 precursor (EC 1.11.1.7) (Atperox P73) (PRXR11)...
0.766	At1g50430	7-dehydrocholesterol reductase / 7-DHC reductase / sterol delta-7-reductase (ST7R) / dwarf5 protein (DWF5) identical to...
0.765	At2g36870	xyloglucan:xyloglucosyl transferase, putative / xyloglucan endotransglycosylase, putative / endo-xyloglucan transferase,...
0.765	At4g26010	peroxidase, putative peroxidase ATP13a - Arabidopsis thaliana, PID:e264765; identical to cDNA class III peroxidase ATP35,...
0.764	At5g17330	glutamate decarboxylase 1 (GAD 1) sp|Q42521
0.763	At5g49460	ATP-citrate synthase, putative / ATP-citrate (pro-S-)-lyase, putative / citrate cleavage enzyme, putative strong similarity to...
0.762	At1g04680	pectate lyase family protein similar to pectate lyase GP:14531296 from [Fragaria x ananassa]
0.762	At1g65960	glutamate decarboxylase 2 (GAD 2) similar to glutamate decarboxylase (gad) GI:294111 from [Petunia hybrida]
0.762	At3g13520	arabinogalactan-protein (AGP12) identical to gi|10880501|gb|AAG24280
0.762	At3g52470	harpin-induced family protein / HIN1 family protein / harpin-responsive family protein similar to harpin-induced protein hin1 (...
0.762	At4g19120	early-responsive to dehydration stress protein (ERD3) identical to ERD3 protein [Arabidopsis thaliana] GI:15320410; contains...
0.762	At5g15230	gibberellin-regulated protein 4 (GASA4) / gibberellin-responsive protein 4 identical to SP|P46690 Gibberellin-regulated protein...
0.761	At2g36830	major intrinsic family protein / MIP family protein contains Pfam profile: MIP PF00230
0.76	At1g75680	glycosyl hydrolase family 9 protein similar to endo-beta-1,4-glucanase GB:AAC12685 GI:3025470 from [Pinus radiata]
0.759	At1g55330	arabinogalactan-protein (AGP21)
0.759	At3g49670	leucine-rich repeat transmembrane protein kinase, putative CLAVATA1 receptor kinase, Arabidopsis thaliana, EMBL:ATU96879
0.759	At3g49940	LOB domain protein 38 / lateral organ boundaries domain protein 38 (LBD38) identical to SP|Q9SN23 LOB domain protein 38...
0.759	At4g08685	pollen Ole e 1 allergen and extensin family protein contains Pfam domain, PF01190: Pollen proteins Ole e 1 family
0.759	At5g64100	peroxidase, putative identical to peroxidase ATP3a [Arabidopsis thaliana] gi|1546698|emb|CAA67340
0.757	At2g44160	methylenetetrahydrofolate reductase 2 (MTHFR2) identical to SP|O80585 Methylenetetrahydrofolate reductase (EC 1.5.1.20)...
0.755	At5g55730	fasciclin-like arabinogalactan-protein (FLA1) identical to gi|13377776||AAK20857|13377775|gb|AF333970
0.754	At5g43830	expressed protein similar to auxin down-regulated protein ARG10 [Vigna radiata] GI:2970051, wali7 (aluminum-induced protein)...
0.753	At2g26730	leucine-rich repeat transmembrane protein kinase, putative
0.753	At5g56870	beta-galactosidase, putative / lactase, putative similar to beta-galactosidase precursor GI:3869280 from [Carica papaya]
0.752	At1g01430	expressed protein similar to hypothetical protein GB:CAB80917 GI:7267605 from [Arabidopsis thaliana]
0.751	At3g51670	SEC14 cytosolic factor family protein / phosphoglyceride transfer family protein similar to polyphosphoinositide binding...
0.751	At4g12420	multi-copper oxidase, putative (SKU5) identical to multi-copper oxidase-related protein (SKU5)(GI:18158154) [Arabidopsis...
0.751	At5g19250	expressed protein
0.75	At2g37180	plasma membrane intrinsic protein 2C (PIP2C) / aquaporin PIP2.3 (PIP2.3) / water-stress induced tonoplast intrinsic protein...
0.75	At3g16390	jacalin lectin family protein similar to myrosinase-binding protein homolog [Arabidopsis thaliana] GI:2997767, epithiospecifier...
0.75	At4g12390	invertase/pectin methylesterase inhibitor family protein low similarity to pectinesterase from Arabidopsis thaliana SP|Q42534,...
0.748	At1g28290	pollen Ole e 1 allergen and extensin family protein similar to arabinogalactan protein [Daucus carota] GI:11322245; contains...
0.748	At1g30120	pyruvate dehydrogenase E1 component beta subunit, chloroplast identical to pyruvate dehydrogenase E1 beta subunit [Arabidopsis...
0.748	At1g66280	glycosyl hydrolase family 1 protein contains Pfam PF00232 : Glycosyl hydrolase family 1 domain; TIGRFAM TIGR01233:...
0.748	At3g12710	methyladenine glycosylase family protein similar to SP|P05100 DNA-3-methyladenine glycosylase I (EC 3.2.2.20)...
0.748	At5g05960	protease inhibitor/seed storage/lipid transfer protein (LTP) family protein contains Pfam protease inhibitor/seed storage/LTP...
0.748	At5g60920	phytochelatin synthetase, putative / COBRA cell expansion protein COB, putative similar to phytochelatin synthetase...
0.747	At1g28130	auxin-responsive GH3 family protein similar to auxin-responsive GH3 product [Glycine max] GI:18591; contains Pfam profile...
0.747	At2g38390	peroxidase, putative similar to peroxidase isozyme [Armoracia rusticana] gi|217934|dbj|BAA14144; identical to cDNA class III...
0.747	At3g56240	copper homeostasis factor / copper chaperone (CCH) (ATX1) identical to gi:3168840 Pfam profile PF00403: Heavy-metal-associated...
0.746	At4g37800	xyloglucan:xyloglucosyl transferase, putative / xyloglucan endotransglycosylase, putative / endo-xyloglucan transferase,...
0.746	At5g18500	protein kinase family protein contains protein kinase domain, Pfam:PF00069
0.746	At5g43060	cysteine proteinase, putative / thiol protease, putative similar to cysteine proteinase RD21A precursor (thiol protease)...
0.745	At3g27170	chloride channel protein (CLC-b) identical to CLC-b chloride channel protein GB:CAA96058 from [Arabidopsis thaliana] (J. Biol....
0.745	At5g11890	expressed protein
0.745	At5g44130	fasciclin-like arabinogalactan-protein, putative similar to gi_13377784_gb_AAK20861
0.744	At1g53290	galactosyltransferase family protein contains Pfam profile: PF01762 galactosyltransferase ;contains similarity to Avr9 elicitor...