UNIVERSITÉ DE STRASBOURG

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

UNIVERSITÉ DE STRASBOURG et UNIVERSITY OF VICTORIA

École Doctorale des Sciences de la Vie et de la Santé

Institut de Biologie Moléculaire des Plantes (CNRS) et
Centre for Forest Biology (UVic)

THÈSE présentée par

Annette Veronika ALBER

soutenue le 21 Octobre 2016

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline : Biochimie et biologie moléculaire

Phenolic 3-hydroxylases in land plants: biochemical diversity and molecular evolution

---

THÈSE EN CO-TUTELLE dirigée par :

Dr WERCK-REICHHART Danièle	Directeur de recherche, CNRS, IBMP
Prof EHLTING Jürgen	Professeur, University of Victoria

RAPPORTEURS :

Prof HEHN Alain	Professeur, Université de Lorraine
Prof CONSTABEL C. Peter	Professeur, University of Victoria

AUTRES MEMBRES DU JURY :

Prof HUGUENEY Philippe	Professeur, Université de Strasbourg
Prof BORASTON Alisdair	Professeur, University of Victoria
Prof JETTER Reinhard	Professeur, University of British Columbia

Abstracts

Plants produce a rich variety of natural products to face environmental constraints. Enzymes of the cytochrome P450 CYP98 family are key actors in the production of phenolic bioactive compounds. They hydroxylate phenolic esters for lignin biosynthesis in angiosperms, but also produce various other bioactive phenolics. We characterized CYP98s from a moss, a lycopod, a fern, a conifer, a basal angiosperm, a monocot and from two eudicots. We found that substrate preference of the enzymes has changed during evolution of land plants with typical lignin-related activities only appearing in angiosperms, suggesting that ferns, similar to lycopods, produce lignin through an alternative route. A moss *CYP98* knock-out mutant revealed coumaroyl-threonate as CYP98 substrate *in vivo* and showed a severe phenotype. Multiple CYP98s per species exist only in the angiosperms, where we generally found one isoform presumably involved in the biosynthesis of monolignols, and additional isoforms, resulting from independent duplications, with a broad range of functions *in vitro*.

Les plantes produisent une grande variété de produits naturels pour faire face aux conditions environnementales. Les enzymes de la famille CYP98 des cytochromes P450 sont des enzymes clés dans la production des composés dérivés de la voie des phénylpropanoïdes. Ces enzymes sont impliquées dans l'hydroxylation des esters phénoliques pour la biosynthèse des monolignols chez les angiospermes, mais elles sont également impliquées dans la production de divers autres composés phénoliques solubles. Nous avons caractérisé des CYP98 représentatifs des mousses, Lycopodes, fougères, Gymnospermes, Angiospermes basales, Monocotylédones et Eudicotylédones et démontré que leur préférence de substrat a changé au cours de l'évolution. Un mutant knock-out de *CYP98* de mousse a révélé un phénotype sévère et que le *p*-coumaroyl-thréonate est substrat de l'enzyme *in vivo.* Une duplication des *CYP98s* ne peut être observée que dans le génome des Angiospermes, qui présentent généralement une isoforme potentiellement impliquée dans la biosynthèse de la lignine et autres isoformes, résultant de duplications indépendantes, dont le spectre de substrats est plus large *in vitro*.

Table of contents

List of tables

List of figures

List of abbreviations

4CL	4-coumarate:CoA ligase
ALDH	Aldehyde-dehydrogenase
ATR1	Arabidopsis P450 reductase 1
aTRAM	automated target restricted assembly method
BLAST	Basic Local Alignment Search Tool
C2T	Caffeoyl-2-threonate
C3'H	*p*-Coumaroylshikimate/quinate 3'-hydroxylase; CYP98
C4H	Cinnamate 4-hydroxylase; CYP73
C4T	Caffeoyl-4-threonate
CAD	Cinnamyl-alcohol dehydrogenase
CCOMT	Caffeoyl-CoA *O*-methyltransferase
CCR	Cinnamoyl-CoA reductase
CGA	Chlorogenic acid
CoA	Coenzyme A
COMT	Caffeic acid *O*-methyltransferase
CSE	Caffeoyl-shikimate-esterase
CYP	Cytochromes P450
DAD	Diode array detector
DC-INA	2,6-Dichloroisonicotinic acid
DC-SA	3,5-Dichlorosalicylic acid
DDC	Duplication, Degeneration, Complementation
DNA	Deoxyribonucleic acid
EAC	Escape from adaptive conflict
F5H	Coniferaldehyde / coniferyl alcohol 5-hydroxylase; CYP84
G unit	Guaiacyl or coniferyl alcohol unit. Monolignol
GC-MS	Gas chromatography-mass spectrometry
H unit	*p*-Hydroxyphenyl or *p*-coumaryl alcohol unit. Monolignol
HCC	Hydroxycinnamic conjugate
HCT	Hydroxycinnamoyl-CoA: shikimate/quinate hydroxycinnamoyltransferase
HPLC	High-performance liquid chromatography
IAD	Innovation, Amplification, Divergence
kDA	kilo Dalton
MeJA	Methyl jasmonate
MRM	Multiple reaction monitoring
MS	Mass-spectrometry

mya	Million years ago
NADPH	Nicotinamide adenine dinucleotide phosphate
Natural products	Specialized compounds; secondary metabolites
NMR	Nuclear magnetic resonance spectroscopy
p	para or “4“ position on the phenolic ring
P450s	Cytochrome P450 enzymes
PAL	Phenylalanine ammonia lyase
pC2T	*p*-Coumaroyl-2-threonate
pC4T	*p*-Coumaroyl-4-threonate
Phe	Phenylalanine
PPOs	Polyphenol oxidases
ref8	*reduced epidermal fluorescence 8* mutant (*Arabidopsis thaliana*)
RNA	Ribonucleic acid
RT	Retention time
S unit	Syringyl or sinapyl alcohol unit. Monolignol
SA	Salicylic acid
Sm	*Selaginella moellendorffii*
SRS	Substrate recognition site
TPS	Terpene synthase
Trp	Tryptophan
Tyr	Tyrosine
UPLC	Ultra-high-performance liquid chromatography
UV	Ultraviolet
WGD	Whole genome duplication
WT	Wild type

Acknowledgements

I am thankful for the Working on Walls Collaborative Research and Training Experience Program provided by the Natural Sciences and Engineering Research Council of Canada. Through this program and the financial support it has provided, I have been able to perform international, collaborative research with passionate scientists throughout the training network, and I have been able to participate in workshops and conferences internationally. I would also like to thank the Collège Doctoral Européen for financial support and for organizing seminars and lectures on topical European issues.

I am especially grateful to my PhD supervisors Danièle Werck and Jürgen Ehlting. Thank you for your willingness to set up this exceptional collaboration between Canada and France. I have thoroughly enjoyed being a member of your excellent research teams.

I would also like to thank the members of my supervisory committee, Danièle Werck, Jürgen Ehlting, C. Peter Constabel and Alisdair Boraston. Your guidance and feedback during the last five years have been invaluable.

I thank the members of my thesis jury for kindly accepting my thesis for review.

I express my very great appreciation to Hugues Renault. Your day-to-day help and guidance in the laboratory, your ceaseless support of my work, and your valuable suggestions have been very much appreciated. Thank you also for giving me a good start in Strasbourg when I first arrived!

I thank the former and current lab members and staff of the University of Victoria Centre for Forest Biology and the CNRS Institut de biologie moléculaire des plantes in Strasbourg. Many people provided technical and scientific assistance with my project. I also appreciated the smiling faces and many good scientific discussions. In particular, I would like to thank L.

Herrgott and N. Baumberger for the expression of the moss HCT; P. Ullmann for her help with enzymes and substrates; R. Lugan for his help with UPLC analysis; A. Alioua for his help with real-time PCR; A. Lesot for her help with microsome preparation; the team of French gardeners; B. Binges and S. Robbins for their help with my plants; Heather Down for her help with technical equipment; O. Corea and Daisie Huang for their help with poplar sequence data; F. Disdier for his computer help; B. Ehlting for guiding me into molecular biology almost ten years ago; J. Iglesias, T. Ilc, A. Coulter, V. Veljanovski, L. Tran and A. Wong, labmates who became close friends of mine; and J. Aldana, JE Bassard, A. Berna, F. Bernier, B. Boachon, K. Boateng, Y. Carrington, R. Chedgy, C. Gavira, JF Ginglinger, B. Grausem, D. Gray, J. Guo, J. Hannemann, B. Hawkins, T. Heitz, D. Heintz, A. Hemmerlin, A. James, L. Kriegshauser, C. Le, Z. Liu, D. Ma, R. Menard, N. Navrot, C. Parage, F. Philippon, E. Pineau, F. Pinot, N. Prior, H. Schaller, M. Vance, G. Verdier, P. von Aderkas, E. Widemann, K. Yoshida.

I would like to thank all collaborating laboratories for their support of my work. In particular, I want to thank the team of Ralf Reski, especially Gertrud Wiedemann, in Freiburg Germany, and the team of Tobias Köllner, especially Jan Günther at the Max Planck Institute for Chemical Ecology in Jena, Germany.

Lastly, I would like to thank my friends and family. To my dearest friends all around the world thank you for your great encouragement. To Oliver, my partner and best friend (and my favourite cook, musician and artist!), thank you for your unceasing support over the past five years. Four words from you, “you can do it,” always helped to re-energize me. Thank you also for being crazy enough to help me raise our daughter when my research took me to three different countries. Together, we have taught her that anything is possible with teamwork. To my parents, thank you for your support and child care. Oliver and I could not have pursued our academic goals simultaneously without your help!

Dedication

I dedicate this thesis to my family, who encouraged me from early on to travel and open my mind to other cultures. Their support in this international PhD was incredible, and together with my close friends they taught me that “feeling home” is not a local concept. Thank you.

“Travel is fatal to prejudice, bigotry, and narrow-mindedness,
and many of our people need it sorely on these accounts.
Broad, wholesome, charitable views of men and things cannot be acquired
by vegetating in one little corner of the earth all one's lifetime.”

- Mark Twain

1. General introduction

Part of this chapters is adapted from (Alber and Ehlting, 2012).

1.1. Land plant evolution, plant secondary metabolism and molecular evolution

Plants dominate almost all terrestrial environments. Being the most abundant primary producers they nourish nearly all life on earth's surface. Land plants evolved an incredible ecological, structural and chemical diversity with several hundreds of thousands of species. Yet they all share a common ancestor with green, red, and glaucophyte algae forming the Plantae or Archaeplastida (Adl et al., 2005). With the exception of few parasitic plant lines, photosynthesis is a common feature of all plants (Reyes-Prieto et al., 2007).

1.1.1. Evolutionary systematics of green plants (Viridiplantae).

Land plants (embryophytes) evolved from freshwater green algae (Kranz et al., 1995) and together they form the green plant lineage (Viridiplantae)(Figure 1.1). Based on gene sequence comparison and comparative morphology, extant land plants can be classified into four major groups: bryophytes, lycophytes, fern group and seed plants (Bremer et al., 1987; Kranz et al., 1995). The bryophytes embrace mosses, hornworts and liverworts, with the latter being considered as basal, and mosses forming a sister group to the tracheophytes (vascular plants), which include the lycophytes and euphyllophytes (Bremer et al., 1987; Mishler et al., 1994; Rensing et al., 2008). Lycophytes include the clubmosses, spikemosses and quillworts. In contrast, euphyllophytes have expanded to a huge diversity and include the majority of extant land plants, comprising the Polypodiopsida (fern group) and the spermatophytes (seed plants). The fern group (Polypodiopsida) include the Equisetidae (horsetails), Psilotidae (grape ferns, whisk ferns), Polypodiidae (ferns) and Marattiidae (Rothfels et al., 2015). The seed plants include all gymnosperms, such as conifers, and the largest extant group of land plants, the angiosperms (flowering plants).

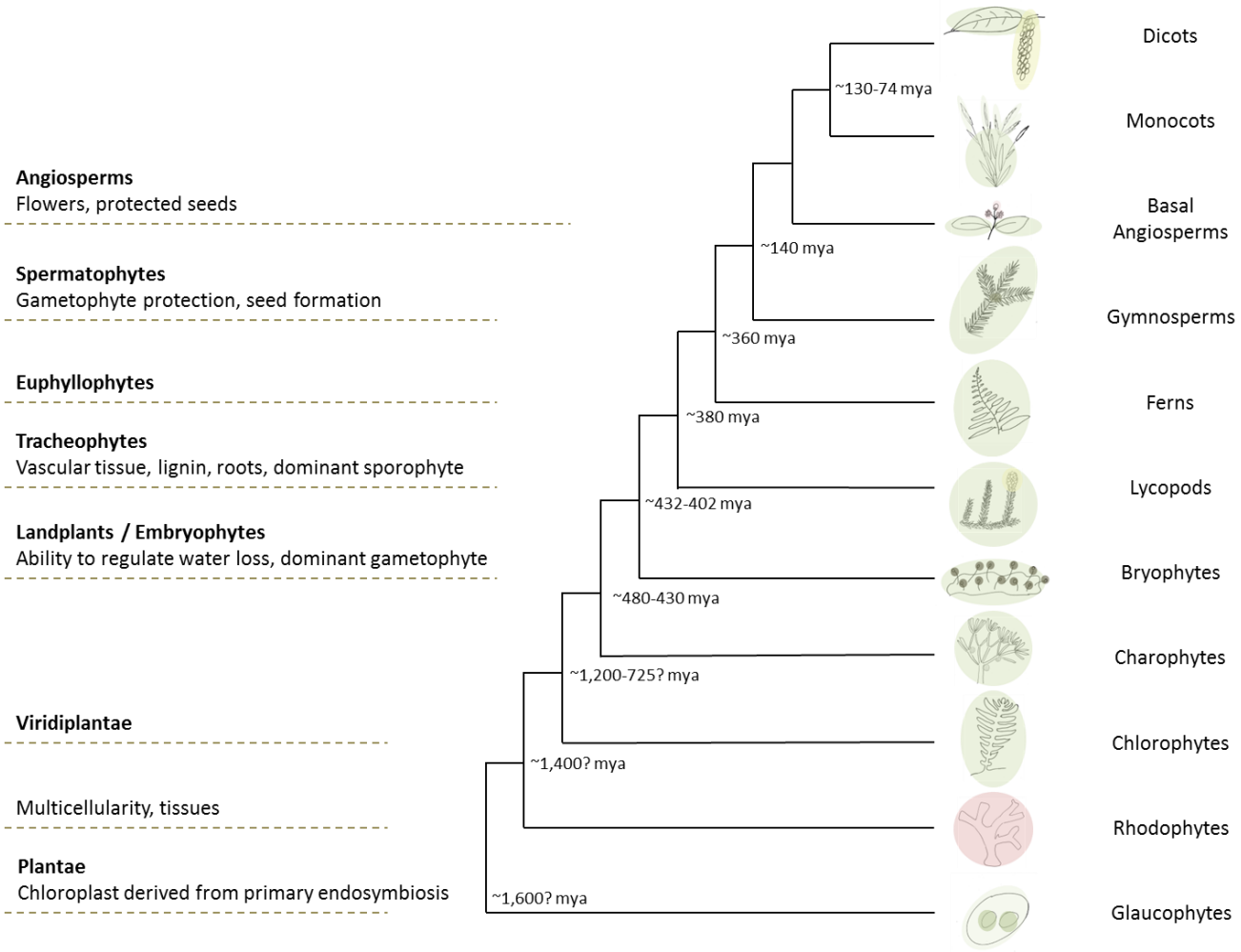


Figure 1.1 Cladogram showing plant evolution.

Time is indicated in million years ago (mya) at branching points. Dating is dependent on the method of investigation and sometimes not fully elucidated to date ("?") (Chaw et al., 2004; Palmer et al., 2004; Bowman et al., 2007; Delaux et al., 2012; Christin et al., 2013; Delwiche and Cooper, 2015; Field et al., 2015). Classes are indicated right to the cladogram, names and some acquired functions are displayed on the left.

1.1.2. Natural product metabolism is an adaptation to life on land

With their transition to land, plants encountered various new stresses. They had to cope with damaging UV-light, desiccation, rapid, wide and extreme temperature fluctuations, and the loss of structural support (Raven, 1984). One central adaption was likely the evolution of diverse

new specialized metabolite-based protection mechanisms, which were acquired and established early during land plant evolution (Sarkar et al., 2009; Delaux et al., 2012). Among them, and possibly most critical, was the development and expansion of phenylpropanoid metabolism (Douglas, 1996; Weng and Chapple, 2010).

Plants are sessile organisms, which adapt their metabolism to face environmental constraints. Instead of avoidance through motility, they construct physical barriers and produce specialized compounds to cope with hostile environments. These specialized compounds, also called plant natural products, have pivotal functions in plant development and chemical ecology (Dixon, 2001; Hartmann, 2007). Plant natural products fulfil distinct roles under a given set of conditions. Contrary to primary metabolites, their roles do not include vital involvement in development and growth. One major group of natural products are the phenylpropanoids which include lignin, flavonoids, and countless other soluble phenolic derivatives (Vogt, 2010). Lignin is a compound of importance as it provides structural support to allow for long distance water transport and erect growth of vascular plants (Weng and Chapple, 2010). Other specialised compounds are involved in interactions with other organisms or the abiotic environment. For example, some are defence-related compounds with antimicrobial properties, others are feeding deterrents, or they act as UV absorbing sunscreens, while other provide protection against abiotic stress (Dixon et al., 2002; Wink, 2003; Bartwal et al., 2013; Baetz and Martinoia, 2014). We as humans benefit from the bioactivity of several classes of these compounds and use many of them as pharmaceuticals (or their precursors), pesticides, cosmetic ingredients and as aromas, scents, or dyes (Wallace, 2004; Korkina, 2007; El-Seedi et al., 2012; Buchanan et al., 2015). Across the plant lineage, 200,000 natural products are thought to exist (Vogt, 2010) and at least 36,000 structures have been identified (Wink, 2003), of these 6,000 are phenylpropanoids, including coumarins, lignans and flavonoids. The immense diversity of plant natural products and their adaptive roles in chemical ecology and plant development makes them prime candidates to study basic molecular evolutionary models. Numerous molecular evolutionary models have been proposed, as outlined in more detail in the following paragraph. It is noteworthy that very few strong supportive examples have been identified. This is likely because most adaptive traits analysed in organismal evolution are

complex and caused by multiple genes, which makes it very difficult to connect the evolution of organismal traits with molecular evolutionary events of individual genes or gene families. Studying plant natural products provides an exception to this rule because a link between an adaptive trait, for example the ability to produce a particular natural product involved in defence, and a single enzyme responsible for this trait, may be made directly in some cases.

1.1.3. Molecular evolutionary models

The emergence, expansion, and diversification of plant natural products are driven by molecular evolutionary events affecting genes encoding metabolic enzymes and their regulators. Gene duplication and subsequent mechanisms such as neofunctionalization and metabolic diversification play important roles (Pichersky and Lewinsohn, 2011; Weng et al., 2012; Chen et al., 2013; Chae et al., 2014).

Gene ancestry in the context of species evolution can be reconstructed using molecular phylogenetic analyses. For this, sequences from presumably homologous loci are aligned to reconstruct evolutionary relationships (Nei and Kumar, 2000). The resulting phylogenetic trees can give evidence about gene duplications in the species' phylogenetic history. Genes that have presumably undergone multiple duplications and gene losses need careful consideration when interpreting phylogenetic analysis. While a species tree represents a pattern of lineages and their relationship over time, a gene tree is a model, summarizing how genes evolved through substitution, duplication, conversion and gene loss (Dittmar and Liberles, 2011). Genes encoding enzymes in natural product metabolism are good candidates to test evolutionary models. Due to their crucial role in survival and reproductive fitness of plants they have undergone strong, and various natural selection periods during their evolution (Weng, 2014).

Gene duplications arise through various mechanisms such as whole genome duplication, chromosomal rearrangements, unequal crossing over, transposition, or retroposition.

All types of gene duplications allow for divergence and modifications of duplicates. Whole genome duplications (WGDs) even allow for changes in complex gene networks. Gene duplications are an important evolutionary force and occurred comparably frequently in plants, particularly in ferns and angiosperms (Soltis et al., 2009; Dodsworth et al., 2015; Wolf et al., 2015) but also in gymnosperms (Li et al., 2015b). Following a WGD, structural chromosomal

rearrangements and gene losses occur rapidly. This diploidization process makes it difficult to determine the number and timing of WGDs. Angiosperm genomes today vary widely in size and chromosome distribution. Variation occurs even between close relatives, due to WGDs and subsequent events (Bowers et al., 2003; Tang et al., 2008). An example of a recent WGD within the angiosperms is the salicoid-specific WGD which happened 65 mya (Tuskan et al., 2006). Gene duplications in *Populus trichocarpa* arose from this single genome-wide event, concerning about 92% of the whole *P .trichocarpa* genome (Tuskan et al., 2006).

Gene duplicates will only be maintained if together they confer an adaptive advantage over having just one copy. Most gene duplicates therefore acquire deleterious mutations and they become pseudogenes rapidly, which are not retained in a population (Näsvall et al., 2012). Events following any gene duplication can also lead to functional variation between two duplicated genes. The accumulation of mutations can lead to the gain of a new function, ancestral functions can be separated and optimized, or changes in gene dosage may occur (Conant and Wolfe, 2008). If gene duplicates have been maintained in extant genomes, an adaptive advantage for having both or even multiple copies must be assumed. Diverse theoretical models exist, which describe events following gene duplications, but might not be exhaustive. Events happening in reality are often far more complex. The major models describing these retention mechanisms will be briefly addressed and are summarized in Figure 1.2.

Neofunctionalization

In the neofunctionalization model, two functionally redundant duplicates exist initially. Purifying pressure against mutations that change the original function acts on one duplicate, to conserve the original function. Functional redundancy causes relaxed selection pressure on the other duplicate, allowing for the random gain of a new function. Mutations are selected that gain and then optimize a novel function. If the novel gene function is beneficial, retention in population can occur (Conant and Wolfe, 2008). An example of neofunctionalization after gene duplication is the study of *CYP98A8* and *CYP98A9* of *Arabidopsis thaliana*, where duplicates of an enzyme involved in lignin biosynthesis evolved rapidly under relaxed selection to become involved in the biosynthesis of pollen coat constituents (Matsuno, et al., 2009).

Subfunctionalization

In the subfunctionalization model the ancestor is multifunctional. Functions are distributed onto the two copies after duplication (Conant and Wolfe, 2008). Two often described models of subfunctionalization are the “Duplication, Degeneration, Complementation (DDC)” model and the “Escape from adaptive conflict (EAC)” model. DDC describes a model in which mutations can occur neutrally in the gene duplicates, as long as the ancestral functions can be maintained by both duplicates together (Force et al., 1999). In the EAC model the ancestral gene is multifunctional and mutations are not neutral. Optimization of one function by mutation may be at the expense of the other function, and vice versa, leading to both functions being sub-optimal. Duplication is a possible way out of this adaptive conflict. Separate optimizations under positive selection of either function can take place after duplication (Hughes, 1994).

Dosage effects

In this model, the ancestral gene has a second, minor, function. Duplication(s) occur(s) and the duplicates provide an increase of this minor function through gene dosage effects. Duplicates can overcome low efficiency problems of a novel function.

One model of dosage effects is the “Innovation, Amplification, Divergence (IAD)” model. Duplications bring a novel, but weak function to a level where it may become adaptive. Beneficial mutations can then accumulate in the duplicates to improve the new function while the ancient function of the gene can be retained on another copy (Näsvall et al., 2012)

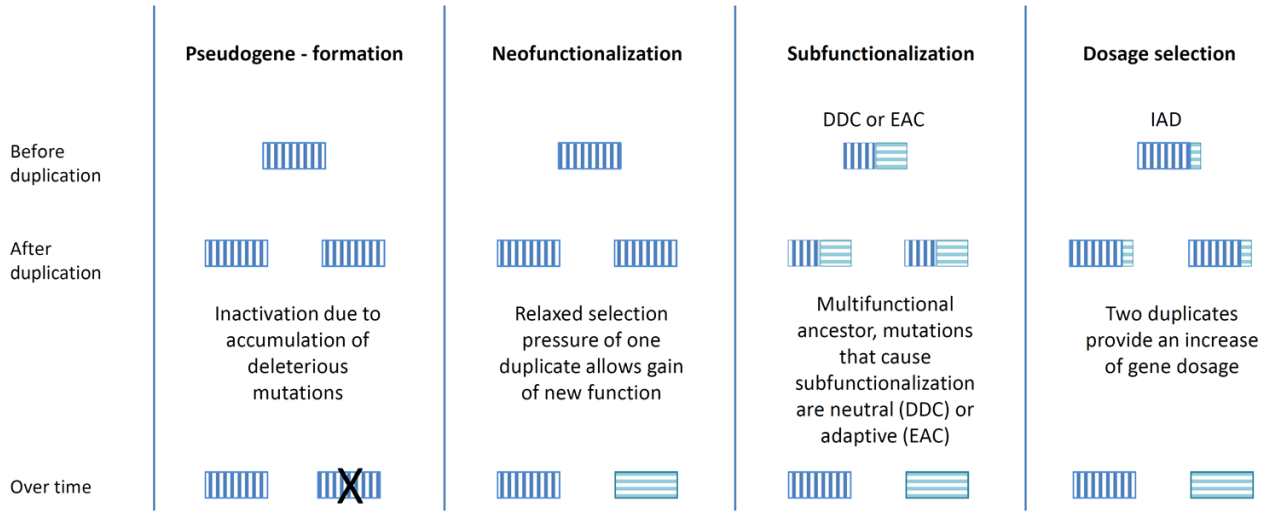


Figure 1.2 Molecular evolutionary models

Four theoretical molecular evolutionary models are shown. For each model, the gene and its function are shown before duplication in the top row. In the middle, the genes and functions are shown after gene duplication. The bottom row shows the fate of the genes and their functions over time.

Abbreviations: DDC Duplication, Degeneration, Complementation; EAC Escape from Adaptive Conflict; IAD Innovation, Amplification, Divergence.

1.2. Cytochromes P450

1.2.1. Definition and function

The focus here is on the functional diversity and molecular evolution of a particular enzyme family, the cytochrome P450 family CYP98, which is involved in the phenylpropanoid pathway. Cytochromes P450 are particularly useful for evolutionary and biochemical studies because they frequently catalyse rate limiting steps and define flow into the immensely diverse specialised compound pathways. Especially in plants, they compose a huge family in which diverse selection pressures are expected to act.

Cytochromes P450 (P450s) are heme containing enzymes that are found in all organisms, from bacteria to humans (Nelson, 1999). The classical P450 catalytic cycle is described in Figure 1.3. P450 stands for Pigment absorbing at 450 nm, the absorption maximum of a difference UV-Vis

spectrum between CO associated, reduced enzyme and reduced enzyme (Figure 1.4). P450 enzymes are classified as monooxygenases. The catalysed reactions are usually based on the activation of molecular oxygen with the insertion of one of its atoms into the substrate and the reduction of the other one to form water (Mansuy, 1998; Werck-Reichhart and Feyereisen, 2000). The typical reaction catalysed can be summarized as:

$RH + O_2 + NADPH + H^+ \rightarrow ROH + H_2O + NADP^+$ (Figure 1.3).

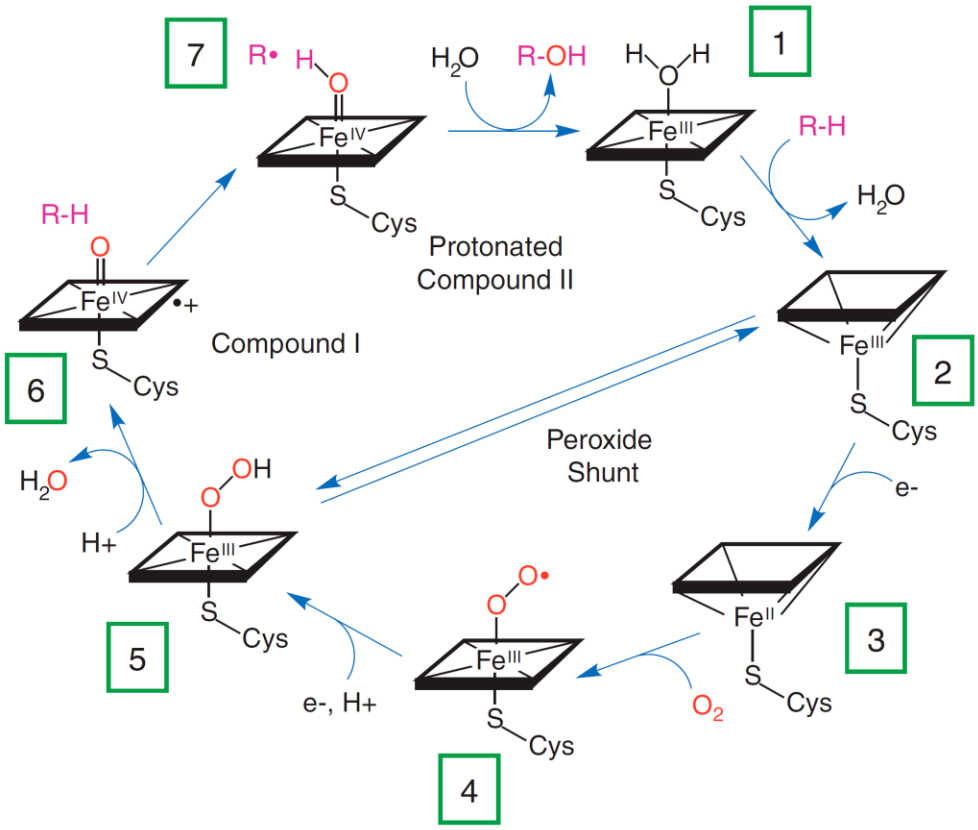


Figure 1.3 **The catalytic cycle of cytochromes P450** (Ener et al., 2010)

1) P450 in low-spin resting state, with bound H_2O molecule. **2)** Substrate is bound and the H_2O molecule released. **3)** Substrate binding causes change from low to high spin. The iron is reduced. **4)** An oxygen molecule binds to the active site of the P450. **5)** The iron is further reduced and the distal oxygen protonated. The O-O bond is cleaved, leading to a ferric hydroxoperoxo complex. One H_2O molecule is released and a highly reactive ferryl-oxo complex is formed. **6)** The ferryl-oxo complex abstracts hydrogen from the substrate. **7)** The substrate radical and the heme-bound hydroxyl combine. The hydroxylated product dissociates. Water binds to the heme and the P450 is in the resting state again.

P450s described in plants are membrane anchored and need to be coupled to an electron-donating protein to be active. They catalyse a wide variety of redox reactions in plant metabolism and are encoded by a superfamily typically encompassing more than two hundred members in higher plants (Mizutani and Ohta, 2010; Bak et al., 2011). P450 mediated reactions are essentially irreversible and are located at important branch points in many metabolic networks. Thus, they are the major "gatekeepers" that irreversibly channel carbon into distinct sub-branches of metabolic networks.

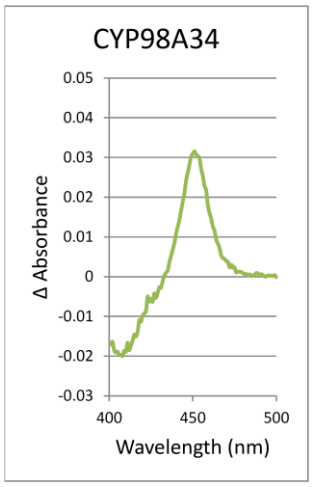


Figure 1.4 Differential spectrum of CYP98A34 from *Physcomitrella patens*.

Enzyme expressed in yeast microsomes is reduced by sodium dithionite and the sample split to two spectrophotometer cuvettes. One sample is associated with CO and the differential spectrum between the two samples read. A peak at 450 nm absorbance shows functional enzyme. The amount of functional enzyme can be calculated from the spectrum.

1.2.2. Nomenclature and classification

Systematic nomenclature of P450s is based on protein sequence identity and phylogeny (Nelson et al., 1996). Members of the same family usually share at least 40% amino acid identity, and subfamilies share at least 55% amino acid identity. P450 attribution to a family/subfamily is dictated by phylogenetic analysis. Within a (sub) family, individual genes are numbered in order of identification and submission to a nomenclature committee, regardless of the species they originate from (Figure 1.5). Plant P450 family numbers range from CYP71 to

CYP99 and from CYP710 to CYP772 (D.Nelson, Cytochrome P450 Nomenclature Files, http://drnelson.uthsc.edu/Nomenclature.html).

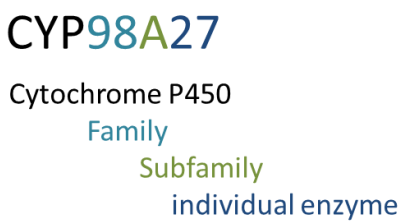


Figure 1.5 Cytochromes P450 nomenclature

1.2.3. Protein structure

P450 enzymes possess highly conserved domains. Membrane anchor and globular part of the protein are linked by a proline-rich hinge. A heme-binding cysteine is absolutely conserved and surrounded by a conserved region (Figure 1.6). The I-helix is involved in oxygen binding and activation. An amino acid triade (ERR), formed by glutamate and arginine of the K-helix and the arginine of a highly conserved PERF motif, locks the heme into position and ensures stabilization of the core structure (Hasemann et al., 1995; Werck-Reichhart and Feyereisen, 2000).

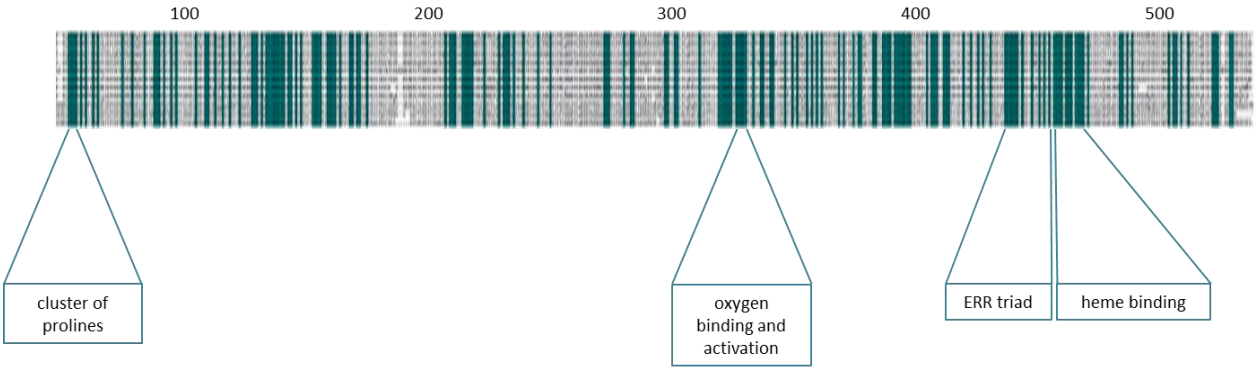


Figure 1.6 Common P450 structures.

Minimized Alignment of CYP98 protein sequences covered in this thesis. Conserved regions are shaded in green/blue. The cluster of prolines links the membrane anchor and globular part of the protein. The site of oxygen binding and activation is part of the I-helix. The ERR triad locks the heme into position and contributes to the stabilization of the core structure. The heme binding cysteine is absolutely conserved. It is located in a conserved region.

1.2.4. P450 Functional diversity in plants

A search of the term “Cytochrome P450” on the web interface of the National Center for Biotechnology (www.ncbi.nlm.nih.gov) results in more than 95,000 protein sequences (non-redundant sequences of RefSeq) thereof more than 16,000 plant P450s. The “P450 statistics, April 6, 2016” on the cytochrome P450 webpage state 13,978 plant P450s, among a total of 35,166 P450s (http://drnelson.uthsc.edu/P450.statsfile.html). About 1% of all genes of sequenced model plants such as *A. thaliana*, *P. trichocarpa* and *Oryza sativa* consist of P450s (Nelson, 2006). P450s in plants are involved in the biosynthesis and/or catabolism of various compounds such as structural polymers (lignin, cutin, sporopollenin, suberin), hormones and signalling molecules, lipids, UV protectants, antioxidants, pigments, odorants, flavours, defence compounds, phytoalexins and feeding deterrents (Schuler and Werck-Reichhart, 2003; Powles and Yu, 2010; Bak et al., 2011). They also play important roles in response to exposure to herbicides or pollutants (Werck-Reichhart et al., 2000; Schuler and Werck-Reichhart, 2003).

The expansion of the P450 gene family in plants can be largely attributed to the expansion of specialised compounds in plants. P450s occupy central positions in all secondary metabolic pathways. The focal P450 family of this thesis, CYP98, participates in the biosynthesis of phenylpropanoid derived secondary metabolites. This pathway and the CYP98 family will therefore be introduced in more detail in the following paragraph.

1.3. The phenylpropanoid pathway

Phenylpropanoids form a diverse class of plant natural products that, as the name implies, contain at their core an aromatic C6 phenyl group and a C3 propenoid sidegroup (Figure 1.7). They share a common origin from the aromatic amino acid phenylalanine (Phe) and the phenylpropanoid pathway begins with the deamination of Phe by the enzyme phenylalanine ammonia lyase (PAL) to form cinnamic acid, the precursor of all phenylpropanoids.

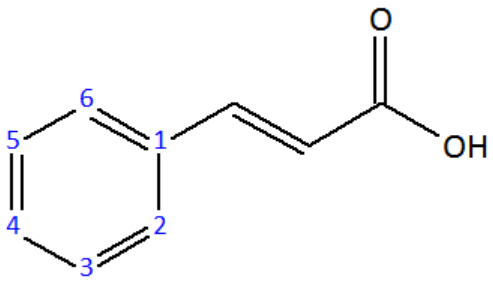


Figure 1.7 Structure of cinnamate, precursor of all phenylpropanoids

Subsequent hydroxylations and decorations of the aromatic ring and/or of the propenoid sidegroup form the tremendously diverse phenylpropanoid-based metabolites (Figure 1.8) (Alber and Ehlting, 2012). Simple phenylpropanoids may have evolved in plants originally to offer UV protection, as their absorbance maximum lies within the UV range. Some of these then offered additional bioactive functions, such as antimicrobial activity or astringency, that bore multiple benefits for the plant (Lowry et al., 1980). After the conquest of land and even with the first protective mechanisms established, plants were still physically small as they were lacking mechanisms of mechanical reinforcement (Bateman and Crane, 1998). Tracheophytes (vascular plants) acquired lignin in their cell walls and gained physical rigidity for erect growth and long distance water transport, which allowed a larger body size (Weng and Chapple, 2010).

The quantity and diversity of phenylpropanoids range dramatically between species. Some are present in most plants, while other may be found only in specific taxa (Clifford, 2000; Dixon, 2001; Petersen and Simmonds, 2003; Petersen et al., 2009). While the core phenylpropanoid pathway and the biosynthesis of monolignols are well characterized, knowledge about the biosynthesis of the majority of soluble compound classes is still fragmentary. Knowing the genes encoding their biosynthetic enzymes is a prerequisite for testing their roles in chemical ecology. Reverse genetic approaches for these enzymes help to identify their role in chemical ecology. Information about the genes also helps to elucidate the molecular evolutionary mechanisms that shaped their current diversity.

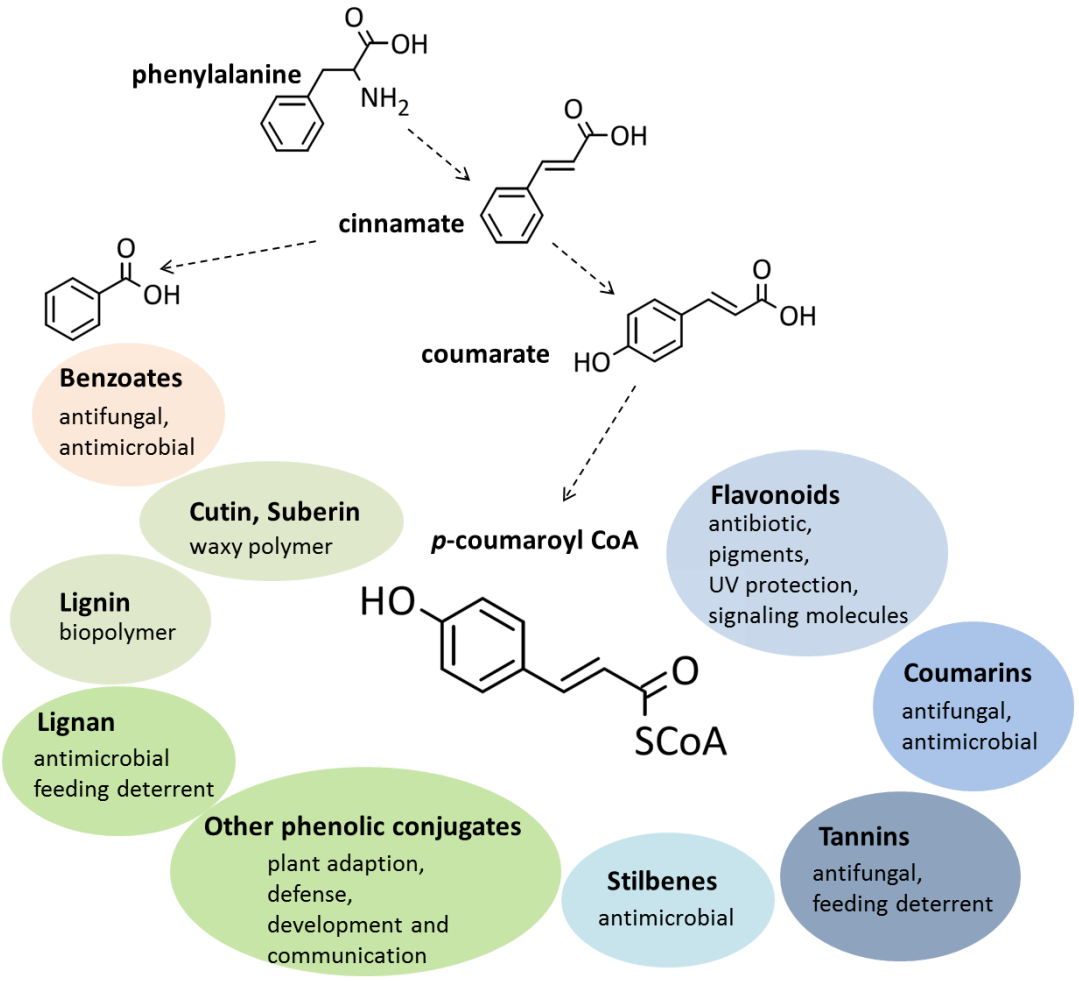


Figure 1.8 Diversification of phenylpropanoids

The structure of phenylalanine is shown at the entry point to the phenylpropanoid pathway. The intermediates cinnamate and coumarate are shown, leading to the major branching molecule of the pathway, *p*-coumaroyl Coenzyme A (CoA). Major phenylpropanoid pathway derived specialised metabolites and their presumed biological functions are shown in the colored circles.

1.3.1. Hydroxycinnamic conjugates

Among the phenylpropanoids, hydroxycinnamic conjugates (HCCs) are the focus here and include for example caffeate or ferulate conjugated with a huge variety of alcohols or amines. Many of these conjugates have antioxidant, antiviral, antibacterial and anti-inflammatory activities, which may imply primary roles in abiotic stress, pathogen and herbivore defence, but

little functional data exists about their actual ecological function (Petersen and Simmonds, 2003; Gülçin, 2006; Chao et al., 2009). Hydroxycinnamic conjugates are important for plants in acclimation to cold (Solecka and Kacperska, 2003). Caffeoyl-quinate, or chlorogenic acid (CGA) is known to be involved in defence against herbivores (Barbehenn et al., 2010). A large variety of hydroxycinnamoyl esters are known across the plant kingdom. For example, the genus *Populus*, which includes poplars, aspens, and cottonwoods, accumulates a rich diversity of HCCs that differ in their composition and abundance between different species and even within a single species. These HCCs include for example caffeate or ferulate esterified with i) quinate or shikimate ii) aromatic alcohols such as benzyl alcohol derivatives, phenylethanol, or monolignols, and iii) alkanols or alkenols including prenyl-alcohol and geraniol, and/or iv) glycerol derivatives (Figure 1.9) (Greenaway et al., 1988; Greenaway and Whatley, 1990a; English et al., 1991; Greenaway et al., 1991a; Greenaway et al., 1991b; English et al., 1992; Isidorov and Vinogorova, 2003; Dudonné and Poupard, 2011).

Beyond esters, also phenolamides or hydroxycinnamic acid amides occur in plants in a rich variety and constitute a considerable proportion of plant natural products (Martin-Tanguy, 1985; Bassard et al., 2010; Macoy et al., 2015a). Cinnamic acid, coumaric acid, caffeic acid, ferulic acid and sinapic acid can be conjugated with arylamines such as tyramine, tryptamine, octopamine and anthranilate, or polyamines such as spermidine and putrescine to form these phenolamides. Phenolamide deposition in the cell wall near pathogen infected or wound-healing regions is thought to have strengthening functions, decreasing the digestibility of the cell wall and creating a barrier for pathogens. For example, they are involved in defence against fungal penetration, building papillae deposited at the inner side of the cell wall. This arrests fungal penetration into host plant tissues. In addition, inhibitory effects on virus multiplication could be shown (Facchini et al., 2002; Edreva et al., 2007). The level of phenolamides increases rapidly in plants upon insect attack. This increase is due to the reconfiguration of the expression of genes involved in the production of phenolamides. It has been shown in *Nicotiana attenuata* that this induced defence reaction is mediated by a multi-hormonal signalling network (Gaquerel et al., 2014).

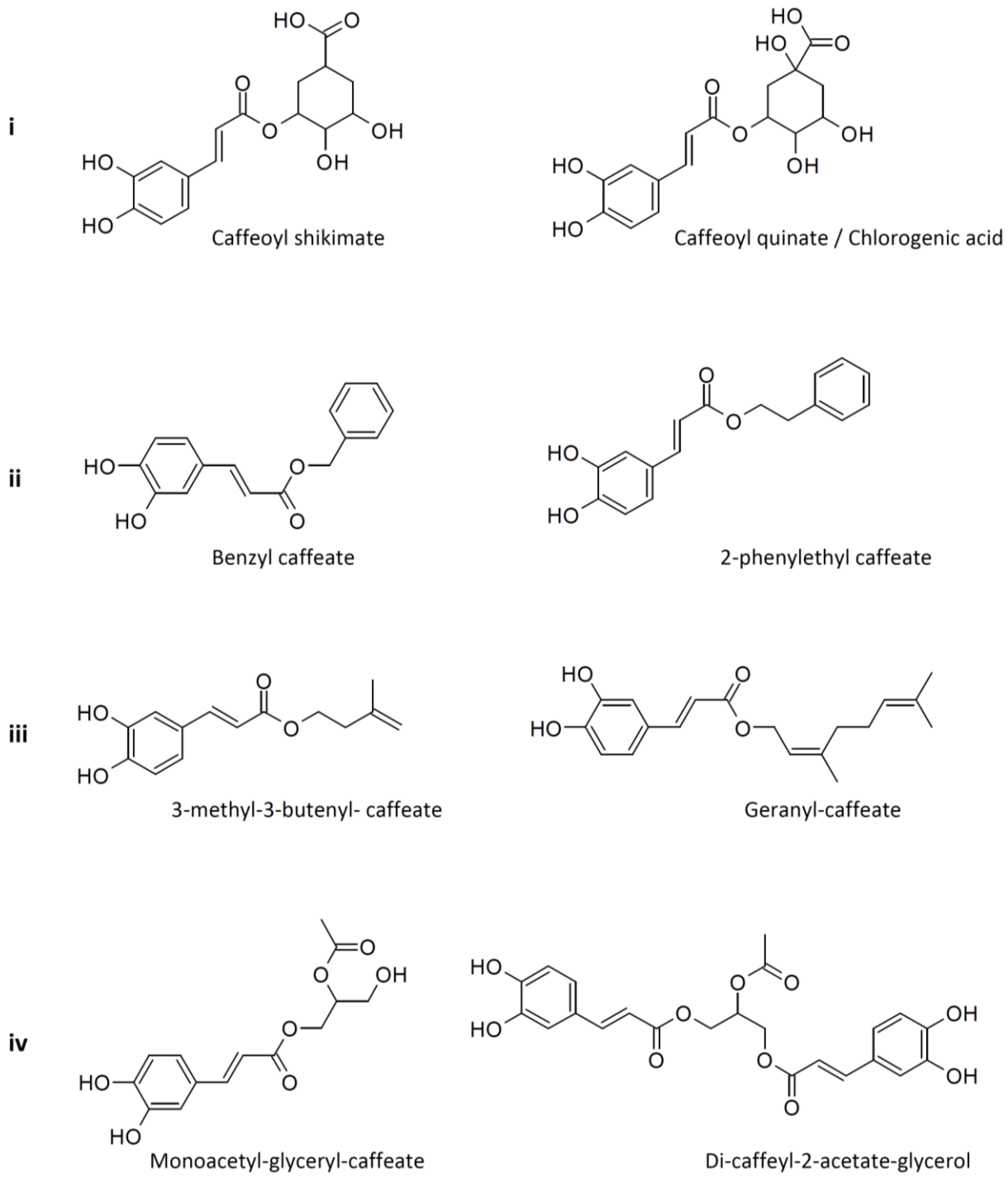


Figure 1.9 Examples of known hydroxycinnamoyl conjugates of *Populus* species.

Shown are **i**: caffeate esterified with shikimate or quinate. **ii**: Caffeate esterified with aromatic alcohols. **iii**: Caffeate esterified with alkenols. **iv**: Caffeate esterified with glycerol derivatives.

Avenanthramides, phenolamides containing hydroxy-anthranilates, were shown to be phytoalexins in oat (*Avena sativa*), involved in defence mechanisms of oat leaves against crown rust fungus (*Puccinia coronata*) (Mayama et al., 1981; Mayama et al., 1982). Elicitor treatment in oat leaves transcriptionally activates genes of the phenylpropanoid pathway and leads to an accumulation of phenylpropanoid enzymes potentially involved in their biosynthesis (including PAL and hydroxycinnamoyl-CoA:hydroxyanthranilate *N*-hydroxycinnamoyltransferase). It was therefore concluded that avenanthramides are made from phenylalanine and anthranilate, a precursor of tryptophan (Ishihara et al., 1999a; Ishihara et al., 1999b).

1.3.2. Lignin

Besides the multitude of functional and structural diversity of soluble HCCs, it is also clear that one hydroxycinnamoyl-ester, hydroxycinnamoyl-shikimate, functions as an intermediate in lignin biosynthesis, at least in angiosperms (Schoch et al., 2001; Humphreys and Chapple, 2002). Lignin monomers, or monolignols, are synthesized through the phenylpropanoid pathway. Lignin is quantitatively the most important final product of the pathway. The term lignin - introduced by de Candolle in 1819 - is derived from the Latin word lignum, meaning wood. Lignin is the second most abundant biopolymer on earth constituting 30% of non-fossil organic carbon (Boerjan et al., 2003). It is an aromatic heteropolymer that is incorporated into cell walls during secondary thickening, for example during wood formation. Integration of this hydrophobic polymer into the cellulose network causes the mechanical strength and hydrophobicity of secondary cell walls that allows long distance water transport and enables the erect growth of land plants (Sarkanen and Ludwig, 1971; Chabannes and Ruel, 2001; Jones et al., 2001). Thus, the ability of lignin biosynthesis contributed largely to the takeover of land by vascular plants. However, the origin of lignin or at least of phenylpropanoid biosynthesis predates vascular plant evolution. Lignin-like aromatic polymers have been identified in some green and even red algae. The red alga *Calliarthron cheilosporioides* makes H G and S lignin (Martone et al., 2009) (Figure 1.10). Homologs of genes needed to make *p*-coumaryl alcohol (Figure 1.10) units are already present in marine photosynthetic algea (Labeeuw et al., 2015).

In contrast, bryophytes do not produce lignin, but orthologs of most characterized monolignol biosynthetic genes are present in the bryophyte *P. patens*. (Gunnison and Alexander, 1975; Delwiche et al., 1989; Martone et al., 2009; Xu et al., 2009).

The complex racemic aromatic heteropolymers found in lignin are mainly derived from three hydroxycinnamyl alcohol monomers differing in their degree of methoxylation: *p*-coumaryl, coniferyl and sinapyl alcohols (Figure 1.10).

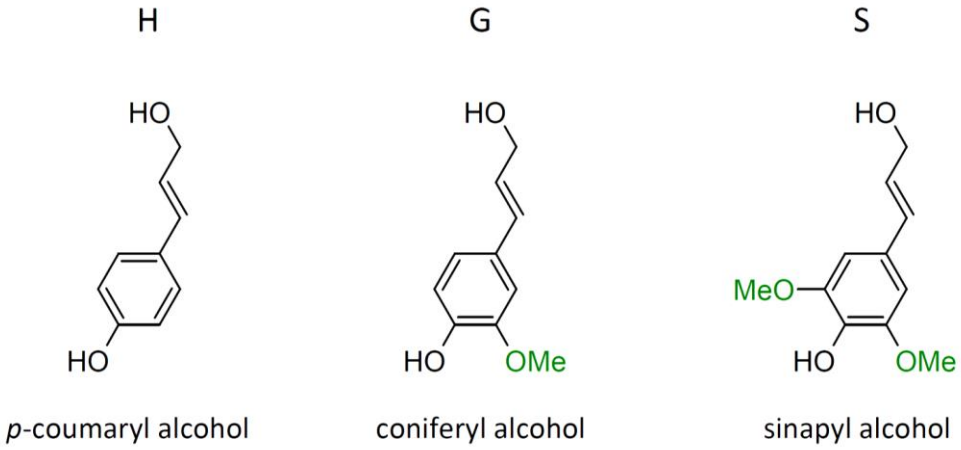


Figure 1.10 Hydroxycinnamyl alcohol monomers, which are the three major lignin building blocks.

Incorporated into the lignin polymer these monolignols produce *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) phenylpropane units respectively (Boerjan et al., 2003) (Figure 1.10;Figure 1.11). In general, the lignin found in ferns and gymnosperms consists mainly of G units, with a small proportion of H units, whereas the lignin of angiosperms mainly consists of G and S units, with only traces of H units. Lignins from grasses (monocots) incorporate G and S units at comparable levels, but they contain more H units than eudicots (Baucher and Monties, 1998). Species that possess only H lignin units were not described to date. Brown and green algae possess homologs of the 4CL, CCR and CAD genes, necessary to synthesize *p*-coumaryl alcohol (Labeeuw et al., 2015). However, they do not possess homologs of important phenylpropanoid entry point genes such as PAL and C4H. The biosynthesis of G and subsequently S lignin units requires hydroxylation at the third position and subsequently the fifth position of the phenolic ring. These 3' and 5' hydroxylations are important for the cross-linked structure and properties of the lignin polymer and are present in essentially all lignins analyzed.

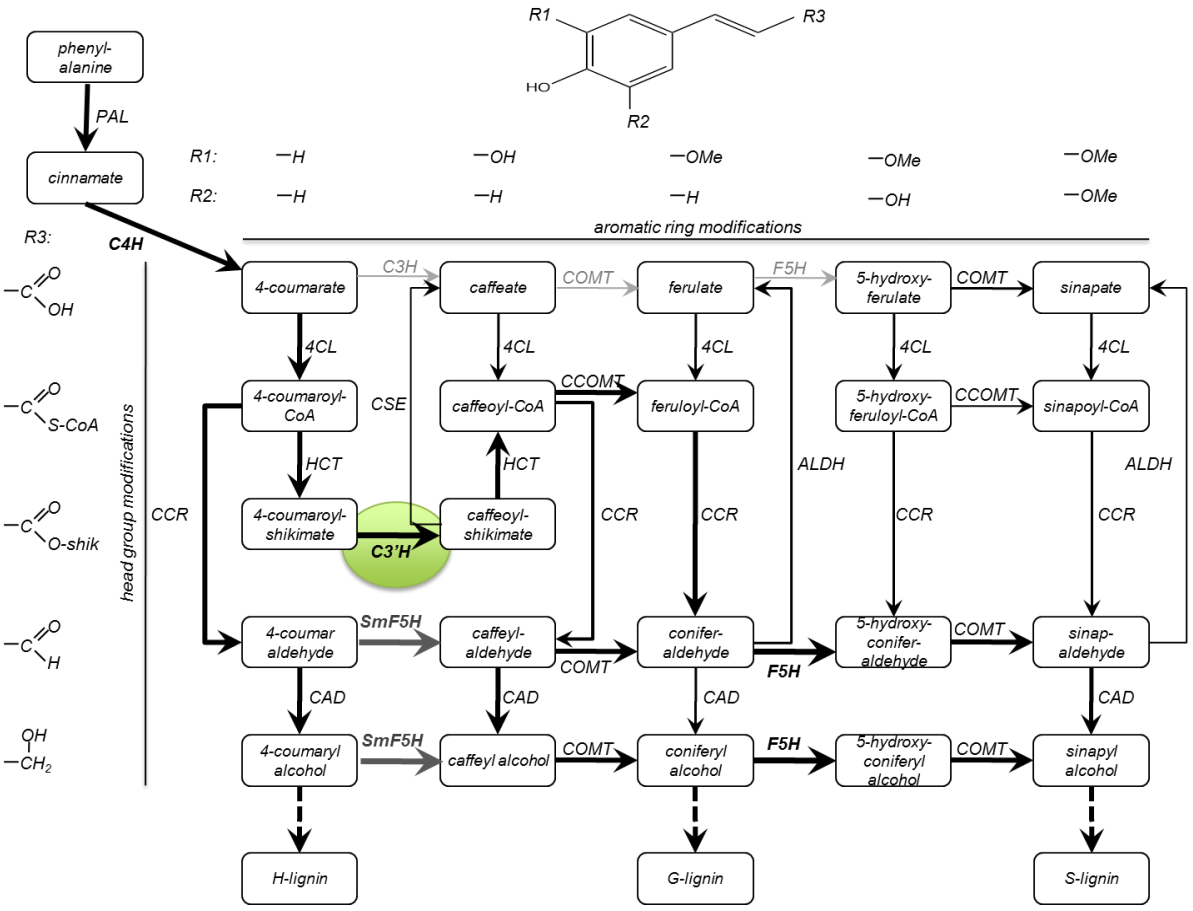


Figure 1.11 The phenylpropanoid grid.

Adapted from (Alber and Ehlting, 2012)**.** The basic structure of phenylpropanoids is shown on top. Head group modifications are shown top to bottom, and aromatic ring modifications are shown left to right. The individual residues (R1–R3 in the general structure) are shown at each level. Reactions that have been characterized to occur with kinetic properties rendering a physiological function likely are drawn in black (or dark grey if they occur only in the lycopod *Selaginella moellendorffii* [Sm]). Those occurring with low efficiency are shown in light grey. The currently accepted path through the grid to H-, G-, and S-lignin is highlighted by bold arrows. The enzymes catalysing each step are abbreviated as PAL phenylalanine ammonia lyase; C4H cinnamate 4-hydroxylase; CCR cinnamoyl-CoA reductase; 4CL 4-coumarate:CoA ligase; C3'H 4-coumaroylshikimate/quinate 3'-hydroxylase; CSE Caffeoyl-shikimate-esterase; HCT hydroxycinnamoyl-CoA: shikimate/quinate hydroxycinnamoyltransferase; CAD cinnamyl alcohol dehydrogenase; COMT caffeic acid O-methyltransferase; CCOMT caffeoyl-CoA O-methyltransferase; F5H coniferaldehyde / coniferyl alcohol 5-hydroxylase; ALDH aldehyde dehydrogenase.

Up to 20% of all fixed carbon might be channelled into this pathway and especially in woody tissues flux through the phenylpropanoid pathway into lignin is large. Lignin is the primary carbon sink derived from the shikimate pathway that produces the aromatic amino acids including Phe. However, as described above, Phe itself and the vast array of other metabolites derived from it also serve vital biological functions, e.g. as protein building blocks, defence compounds or signalling molecules (Tzin and Galili, 2010). Thus, flux through the pathway must be regulated tightly to ensure production of large amounts of precursors for lignin biosynthesis when and where needed, but also to ensure sufficient availability of precursors for less abundant products. Regulation of the pathway clearly occurs at the transcriptional level, as evidenced by the temporal and spatial variation of gene expression during development and in response to environmental cues. Most genes encoding phenylpropanoid enzymes are highly co-expressed in tissues and organs undergoing lignification, and many are induced by biotic and abiotic stresses. Most lignin biosynthetic genes share a common expression pattern when compared across hundreds of developmental samples (Ehlting et al., 2008) suggesting transcriptional co-regulation by the same regulatory cascade. CYP98 in addition provides the opportunity of a direct biochemical regulation. It has been proposed that shikimate has been selected for as a cofactor, because this allowed metabolic regulation of the rate limiting step into G and S lignin by the upstream shikimate pathway, which provides the aromatic amino acids including phenylalanine (Figure 1.12). If the shikimate pathway slows down, the shikimate levels are reduced. Reduced shikimate levels cease driving the major flux into lignin. This allows to maintain sufficiently high phenylalanine levels for other essential functions, such as protein biosynthesis (Schoch et al., 2006; Alber and Ehlting, 2012).

1.4. CYP98

Several cytochromes P450 (CYP) hydroxylases are involved in the phenylpropanoid pathway and are considered to catalyse the rate-limiting steps defining flow into the core pathway and into the respective branch pathways (Anterola and Lewis, 2002). As the gatekeeper to the phenylpropanoid pathway, cinnamate 4-hydroxylase (C4H) constitutes the CYP73 family and catalyses the *para*- or 4-hydroxylation of cinnamic acid. The 4-coumaroylshikimate/quinate 3'-hydroxylase (C3'H) belongs to the CYP98 family and catalyses the 3-hydroxylation of the

phenolic ring (shikimate has a ring system as well, so that carbon numbering starts from the most substituted carbon, the carboxylate of shikimate and the aromatic phenylpropanoid ring becomes annotated as the prime-ring). Further downstream in the monolignol pathway coniferaldehyde / coniferyl alcohol 5-hydroxylase (generally referred to as F5H for ferulate 5-hydroxylase) constitutes the CYP84 family of cytochrome P450s (Figure 1.12) (Humphreys and Chapple, 2002; Ehlting et al., 2006).

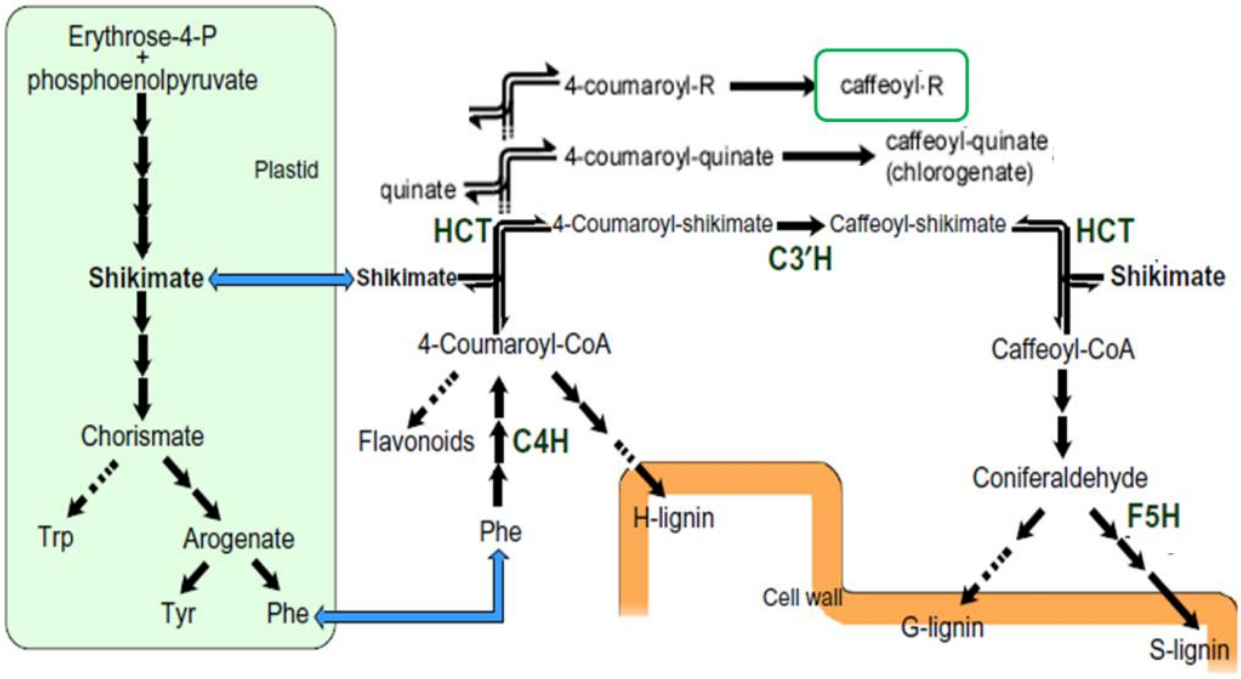


Figure 1.12 Connection between the shikimate and phenylpropanoid pathway

Shown is an outline of the shikimate pathway (pale green box) and the phenylpropanoids pathway. Only branch-point metabolites are given. Trp: Tryptophan; Tyr: Tyrosine; Phe: Phenylalanine; coumaroyl/caffeoyl-R: coumaroyl/caffeoyl-conjugates; HCT: hydroxycinnamoyl CoA:shikimate/quinate hydroxycinnamoyltransferase; C4H: cinnamate 4-hydroxylase; C3'H: *p*-coumaroylester 3'-hydroxylase; F5H: coniferaldehyde / coniferyl alcohol 5-hydroxylase (Alber and Ehlting, 2012).

As described above, my thesis focuses on the CYP98 family, involved in *meta*- or 3-hydroxylation of phenylpropanoid precursors. This hydroxylation step is necessary for the biosynthesis of G and S units of lignin, but also for the generation of UV-absorbing compounds such as sinapoyl malate, and for the formation of many other bioactive compounds, for example chlorogenic acid, rosmarinic acid or some coumarins (Vogt, 2010).

1.4.1. A surprising twist in the lignin pathway

Originally, it was believed that the 3-hydroxylation of the aromatic ring occurs on free 4-coumarate, or on the level of the corresponding CoA-thioesters yielding caffeate or caffeoyl-CoA, respectively. Multiple classes of enzymes were proposed to catalyse the reaction, but none had been characterized (for review, see (Ehlting et al., 2006). Among them, P450 enzymes have been suggested to catalyse the 3-hydroxylation of quinate and shikimate esters of 4-coumarate yielding chlorogenic acid and caffeoyl-shikimate, respectively (Heller and Kühnl, 1985; Kühnl et al., 1987). But only in the early 2000s these enzymes were characterized at the molecular level, and an involvement in lignin monomer biosynthesis became apparent: the *CYP98A3* gene from *A. thaliana* was identified independently by functional genomics and classical genetic approaches and shown to encode the 3-hydroxylase of the phenylpropanoid pathway. Schoch *et al.* (2001) and Nair *et al*. (2002) employed a candidate gene approach based on sequence and expression similarity to C4H, while Franke *et al*. (2002) identified *A. thaliana CYP98A3* via map-based cloning of the *reduced epidermal fluorescence 8* (*ref8*) mutant, which was selected for the lack of fluorescence caused by sinapate ester in wild-type *A. thaliana* leaves. The *A. thaliana CYP98A3* gene was shown to be expressed predominantly in lignifying tissues, similar to other phenylpropanoid genes. Recombinant protein expressed in yeast showed that the shikimate and quinate esters of 4-coumarate are the primary substrates for the 3-hydroxylation of the phenolic moiety. In contrast, 4-coumarate, 4-coumaroyl-CoA, or the corresponding aldehyde and alcohol were poorly or not converted (Schoch et al., 2001; Franke et al., 2002; Nair et al., 2002). The *Arabidopsis* CYP98A3 converts the shikimate ester most efficiently, but the quinate ester of 4-coumarate is also converted with high activity. This defined CYP98A3 as 4-coumaroyl-shikimate/quinate-3'-hydroxylase (C3'H). Thus, C3'H can also catalyse the final step of the biosynthesis of chlorogenic acid (caffeoyl-quinate) (Schoch et al., 2001). However, functional proof that C3’H is also the central 3-hydroxylase of the phenylpropanoid pathway came from a phenotypic analysis of *A. thaliana cyp98a3* mutants (Franke and Hemm, 2002; Abdulrazzak et al., 2006) (Figure 1.13).

Figure 1.13 8 week old *A. thaliana cyp98a3* knock-out mutant plant.

In the *ref8* mutant, soluble sinapoyl malate and sinapoyl choline levels are drastically reduced in leaves and seeds, respectively. Total lignin content is reduced to 20-40% of the wild type level, and both G and S units were found only in trace amounts (Franke and Hemm, 2002). Instead, the mutant accumulates almost exclusively 4-coumarate-derived H units, which are found only in minute amounts in wild-type *A. thaliana* lignin. Regular H lignin biosynthesis is taking place early in inflorescence stem development of the *ref8* mutant, while only small amounts of H monolignols are incorporated into walls that would normally produce S or G lignins later on (Patten et al., 2010). The inability of the *ref8* mutant to produce G and S lignin thus strongly suggested that the 3-hydroxylation of the monolignol pathway occurs at the level of the shikimate ester of 4-coumarate in *A. thaliana* rather than on the free acid or CoA-ester.

This hypothesis was further supported by the characterization of a transferase belonging to the BAHD superfamily in tobacco that was shown to catalyse the synthesis of 4-coumaroyl-shikimate (and -quinate) from 4-coumaroyl-CoA (Hoffmann et al., 2003). The same enzyme also efficiently catalysed the inter-conversion between caffeoyl-shikimate and caffeoyl-CoA, and was thus named hydroxycinnamoyl-CoA:shikimate/quinate hydroxycinnamoyltransferase (HCT). HCT down-regulation also causes reduction of G and S lignin in several plant species and leads to a lignin mainly composed of H units (Besseau et al., 2007; Shadle et al., 2007; Pu et al., 2009).

Together, the discovery of HCT and C3'H immediately suggested that caffeoyl-CoA is synthesized from 4-coumaroyl-CoA via coumaroyl-shikimate and caffeoyl-shikimate. From there it has been suggested that HCT also mediates the reaction to free caffeate. Recently caffeoyl shikimate esterase (CSE) has been described in *A. thaliana* (Vanholme et al., 2013). CSE mediates the reaction from caffeoyl-shikimate to caffeate and shikimate in *A. thaliana* (Figure 1.11). Caffeoyl-CoA can subsequently be formed from caffeic acid and CoA via 4CL. Knocking out CSE in *A. thaliana* shows a reduction of total lignin in the mutants and an increase in H lignin units (Vanholme et al., 2013). CSE loss of function mutants in *Medicago truncatula* show a strong lignin phenotype as well, with mutants reduced in total lignin, enriched in H lignin units and severe dwarfing (Ha et al., 2016). While strong phenotypes of *cse* mutants in *A. thaliana* and *M. truncatula* confirm the role of CSE in lignin biosynthesis, no orthologs of CSE can be found in *Brachypodium distachyon* or *Zea mays*. It may thus be possible that a CSE mediated reaction is not involved in lignin biosynthesis in all plant species (Ha et al., 2016).

An involvement of C3'H in the formation of both G and S lignin units has since been confirmed in other species by reverse genetic approaches. Down-regulation of C3'H in alfalfa (*Medicago sativa*) and hybrid poplar (*Populus grandidentata* x *alba*) resulted in strong reduction in total lignin and a drastic increase in H lignin units (Reddy and Chen, 2005; Coleman et al., 2008b). In alfalfa, both G and S lignin units were strongly reduced and differences in lignin unit coupling were apparent (Ralph et al., 2006). This is accompanied by reduced recalcitrance to saccharification and in consequence, positively impacts bioconversion of lignocellulosic material to ethanol (Chen and Dixon, 2007). In poplar, C3'H down-regulation leads to reduced total lignin, but an increase in H lignin units is mirrored by a decrease in G lignin units only, while S lignin units remain largely unchanged (Coleman et al., 2008b). Again, cell-type specific variation in down-regulation efficiency or species-specific control of fluxes into the distinct sub-branches may explain these apparent differences.

Neither the alfalfa nor the hybrid poplar C3'H targeted for down-regulation have been characterized biochemically, but close orthologs of both have been characterized. *CYP98A44* from red clover (*Trifolium pratense*), which shares 96% sequence identity with the *M. truncatula* C3'H, is able to hydroxylate 4-coumaroyl-shikimate (Sullivan and Zarnowski, 2010).

More detailed analyses have been performed with the C3'H from black cottonwood (*P. trichocarpa*): PtC3'H expressed in yeast hydroxylates 4-coumaroyl-shikimate, but not the free acid (4-coumarate). When PtC3'H was co-expressed with C4H in yeast, a drastic increase in catalytic activity and efficiency with 4-coumaroyl-shikimate was observed. In this case also a low activity with free 4-coumarate becomes detectable (Chen et al., 2011). Most other biochemically characterized CYP98 family members display a clear preference for 4-coumaroyl-shikimate as a substrate (see below for details). Together with the *A. thaliana* results described above and the effects of down-regulation on lignin composition in alfalfa and poplar, this supports the hypothesis that 4-coumaroyl-shikimate is the major intermediate and substrate of the 3-hydroxylation step towards G- and S-lignin.

1.4.2. More than 'just' lignin

CYP98 family members characterized biochemically to date catalyse the 3-hydroxylation of phenylpropanoid moieties. While the first P450 of the phenylpropanoid pathway, C4H, is highly specific for cinnamic acid, the CYP98 3-hydroxylases have less stringent substrate specificity and can accept multiple 4-coumaroyl-conjugates. The products, caffeoyl-conjugates and derivatives thereof, such as feruloyl- or sinapoyl-conjugates, are typical specialized plant natural products that come in hundreds of varieties and frequently accumulate in a lineage- or even species-specific manner (Figure 1.14). Chlorogenic acid, i.e. caffeoyl-quinate, and rosmarinic acid, i.e. caffeoyl-3,4-dihydroxyphenyllactate, are just two common examples (Figure 1.14) (Petersen et al., 2009).

Most CYP98 enzymes characterized to date have a substrate preference for 4-coumaroyl-shikimate, but can also metabolize the quinate ester to appreciable levels, thus producing chlorogenic acid. This holds true for the *A. thaliana* CYP98A3 (albeit *A. thaliana* is not known to accumulate chlorogenic acid *in vivo*), and also for CYP98s from wheat (*Triticum aestivum*), globe artichoke (*Cynara cardunculus*), sweet basil (*Ocimum basilicum*), and coffee (*Coffea canephora*) (Gang et al., 2002; Mahesh et al., 2007; Morant et al., 2007; Moglia et al., 2009). Both coffee isoforms, CYP98A35 and CYP98A36, converted *p*-coumaroyl shikimate at similar rates, but only CYP98A35 hydroxylates the chlorogenic acid precursor, *p*-coumaroyl quinate, with the same efficiency as the shikimate ester, indicating functional divergence within the gene family

(Mahesh et al., 2007). The sweet basil enzyme CYP98A13 was shown able to hydroxylate the phenolic moiety of 4-coumaroyl-4'-hydroxyphenyllactic acid, the rosmarinic acid precursor, albeit at a very low rate (Gang et al., 2002). Likewise, CYP98A6 from a different rosmarinic acid producing plant, *Lithospermum erythrorhizon*, catalyses the 3-hydroxylation of *p*-coumaroyl-4'-hydroxyphenyllactic acid, and was therefore implicated in rosmarinic acid biosynthesis (Matsuno et al., 2002), but other substrates were not tested.

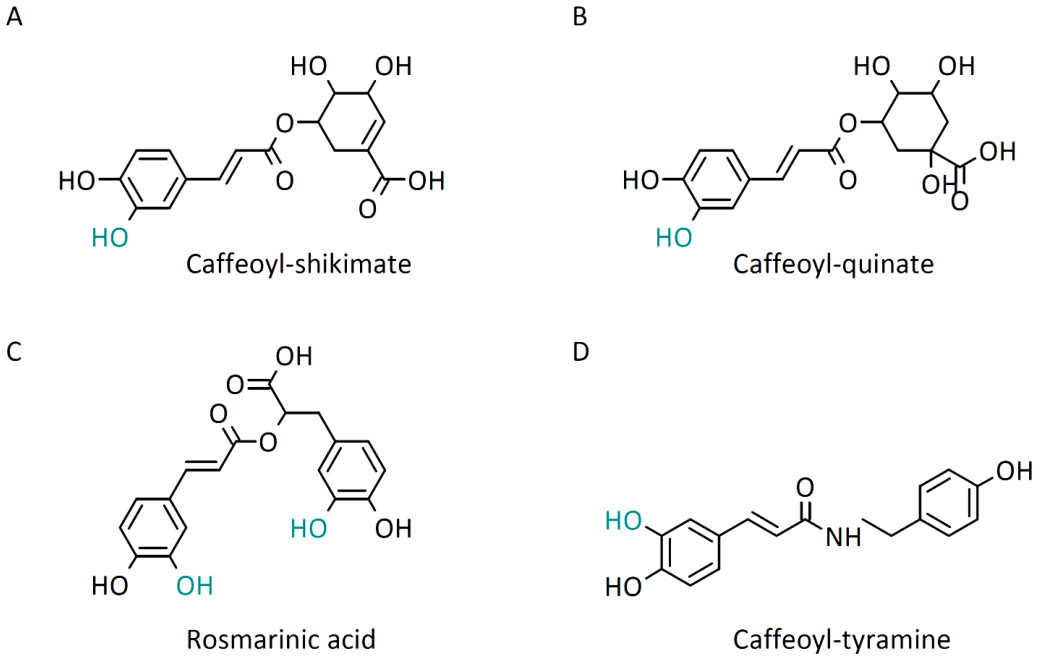


Figure 1.14 Structures of hydroxycinnamic conjugates described in the text.

The position of the CYP98 mediated hydroxylation is shown in blue.

Coleus, *Plectranthus scutellarioides,* also accumulates large amounts of rosmarinic acid and the corresponding CYP98A14 from this species was shown to catalyse the hydroxylations of rosmarinic acid precursors, forming rosmarinic acid. This was the first example of a CYP98 that has no apparent activity with 4-coumaroyl-shikimate or -quinate (Eberle et al., 2009). Likewise, the two CYP98A3 paralogs present in *A. thaliana*, CYP98A8 and CYP98A9, lack appreciable 4-coumaroyl-shikimate/quinate 3'-hydroxylase activity. They were shown to have evolved

recently through retroposition from C3'H and have gained a novel function, that is hydroxycinnamoyl-spermidine 3 and 5 ring-hydroxylations to form components of pollen coat and pollen wall (Matsuno, et al., 2009; Xu et al., 2014). In this case the substrate is an amide rather than an ester, highlighting that CYP98s can also be involved in hydroxycinnamoyl-*N*-conjugate metabolism. Likewise the wheat enzymes, CYP98A11 and CYP98A12, were capable of *meta*-hydroxylating an *N*-conjugate, namely 4-coumaroyltyramine and may thus be involved in the biosynthesis of feruloyltyramine (Morant et al., 2007), which is a common constituent of the cell wall and accumulates in response to wounding or pathogen penetration. However, in this case, the activity with the *N*-conjugate is much lower compared with the hydroxylation of the shikimate ester (Morant et al., 2007) shedding a doubt on an *in vivo* role of CYP98s in feruloyltyramine biosynthesis. Taken together, it is likely that species and isoform specific differences in substrate specificities of CYP98 enzymes are contributing to the wide array of hydroxycinnamoyl conjugates found in plants. Combined with substrate differences of HCT isoforms, which can act on either the substrates or the products of the 3-hydroxylase, it appears plausible that even a limited number of isoforms can create a large array of conjugates.

1.4.3. Are there alternative pathways to monolignols?

While there is now convincing evidence that the aromatic 3-hydroxylation in the lignin pathway occurs primarily on the shikimate-ester level of 4-coumarate in angiosperms, it is still a matter of debate if this is the sole pathway leading to coniferyl and sinapyl monolignols. 3-hydroxylation activity with free 4-coumarate has been described in crude extracts from several plants, for the *A. thaliana* CYP98A3 at very small levels, and in the case of poplar the recombinant C3'H was also capable to convert the free acid albeit only when co-expressed with C4H and only to minute levels (Franke et al., 2002; Chen et al., 2011). Furthermore, 4CL enzymes, which activate the free acids to the corresponding CoA-esters, have a fairly broad substrate range and generally can activate not only 4-coumarate, but also caffeate and ferulate with high efficiency (e.g. (Allina, 1998; Hu et al., 1998; Ehlting et al., 1999)) and 4CL isoforms have been characterized that specifically activate sinapate to the CoA ester (Lindermayr and Möllers, 2002; Hamada et al., 2004; Hamberger and Hahlbrock, 2004). Taken together, these observations suggest a role of the free acid in phenylpropanoid metabolism. More direct

evidence for an alternative pathway comes from the analysis of the *A. thaliana cyp98a3* T-DNA knock-out mutant: although shoot lignin is almost completely devoid of S and G lignins (confirming the *ref8* results described above), these plants still produce detectable levels of soluble sinapoyl-conjugates and display an aberrant lignification phenotype in roots with substantial amounts of G and S units (Abdulrazzak et al., 2006). The same plants also produce sinapoylated flavonoids. Thus, a CYP98A3-independent pathway exists in *A. thaliana*, but this alternative pathway cannot complement a defect in the "normal" developmental lignin pathway. One possibility is that *meta*-hydroxylated phenylpropanoids are released from the spermidine-conjugates produced via CYP98A8/A9 in *A. thaliana* (Matsuno, et al., 2009). These genes are expressed primarily in anthers during normal development, but are slightly induced by environmental stresses. CYP98A9 is expressed in seeds and root tips. It thus seems possible that hydroxycinnamoyl-conjugates not normally produced as lignin intermediates can be recruited in specific tissues or "at times of despair".

An alternative pathway towards G and S lignin units has been described in *Selaginella moellendorffii*. *S. moellendorffii* can bypass the hydroxylation step of CYP98 on phenolic esters, described for angiosperms, hydroxylating the positions 3- and 5- of the phenolic ring by the same enzyme, *Sm*F5H (DN837863). *Sm*F5H can 3-hydroxylate *p*-coumaraldehyde and *p*-coumaryl alcohol, forming caffealdehyde and caffeyl alcohol, respectively. Reaction steps between aldehyde and alcohol are mediated by (hydroxy)cinnamyl alcohol dehydrogenase (CAD) and methoxylations are mediated by caffeic acid *O*-methyl transferase (COMT). Subsequently the enzyme can 5-hydroxylate coniferaldehyde to 5-OH-coniferaldehyde and coniferyl-alcohol to sinapyl-alcohol. The reaction steps mediated by *Sm*F5H are shown in (Figure 1.11). When *SmF5H* is expressed in a *cyp98a3* knock-out or *F5H*-deficient *A. thaliana* mutant, it partially rescues the *cyp98a3* phenotype, and rescues the *F5H* deficient phenotype (Weng et al., 2010a). *Sm*F5H evolved independently of described F5H enzymes in the angiosperms. Nothing is known to date about the enzymes involved in 3-hydroxylation in gymnosperms or monilophytes.

1.4.4. CYP98 family member distribution

While CYP98 members are absent in green algal genomes, a single *CYP98* gene is present in the bryophyte *P. patens* and in the lycophyte *S. moellendorffii* (a basal tracheophyte). Bryophytes are able to synthesize flavonoids and lignans, but lack true lignin (Basile et al., 1999; Umezawa, 2003). Eight out of ten genes considered to encode core lignin pathway enzymes, can be found in *P. patens* (Xu et al., 2009). While *P. patens* does not produce lignin, *S. moellendorffii* can do so by bypassing the C3'H/CYP98 step (Weng et al., 2008a). Thus, in both bryophytes and lycopods the single copy CYP98 member is not involved in lignin biosynthesis. Conversely, in angiosperms, several species have multiple *CYP98* copies (typically not more than three to four) and initial phylogenetic reconstruction suggests that independent duplications may have occurred early in the monocot and eudicot lineages (Ehlting et al., 2006). This leads to the hypothesis that in angiosperms one *CYP98* copy is involved in lignin biosynthesis, while additional copies exist that may be involved in soluble HCC biosynthesis.

1.5. Hypotheses and objectives

The gene family distribution and functional properties of characterized CYP98s suggests that the CYP98 family was gained by plants during their conquest of land and that the ancestral function of CYP98 family members was to produce soluble, protective HCCs. Eventually, CYP98 was recruited to produce an intermediate in lignin biosynthesis in seed plants, while other copies continued to or were newly recruited to produce diverse bioactive compounds. This forms the starting hypotheses of this thesis.

In chapter 2, I will describe a phylogenetic reconstruction of the CYP98 gene family across the plant lineage to test whether gene duplications occurred prior to the separation of the major land plant lineages or if duplications happened only within lineages. This will be combined with biochemical and functional characterization of CYP98s from representatives of each lineage. Combining the functional data with the phylogenetic reconstruction will allow inferring ancestral functions of CYP98s and may guide timing the recruitment of CYP98 for lignin biosynthesis.

Following this, the focus of Chapter 3 will be on the fate of gene duplicates within the angiosperms. As mentioned above, one documented example of species-specific evolution of CYP98 functional diversity is found in *A. thaliana*, where gene duplicates generated through retroposition acquired novel functions under relaxed selection to become 3- and 5-hydroxylases of coumaroyl-spermidine involved in pollen wall biogenesis (Matsuno, et al., 2009). In this case, a fairly recent gene duplication event (in a common ancestor of Brassicaceae) occurred, and one of the duplicates was newly recruited for further conjugate hydroxylation. This leaves the possibility that, also in other angiosperms, independent gene duplications occurred to produce lineage-specific spectra of hydroxycinnamoyl conjugates. To this end, phylogenetic analyses will be performed to judge the number and timing of gene duplications within the angiosperms. Selecting two species with independently duplicated gene families, namely the basal angiosperm *Amborella trichopoda* and the eudicot *P. trichocarpa*, detailed functional characterizations will determine the fates of duplicates in terms of their relationship to lignin and soluble phenolic biosynthesis. This will test the hypothesis that gene duplication and recruitment to novel pathways is common in angiosperms and happened independently not only in the Brassicaceae, but in other lineages as well.

1.6. Acknowledgement

I thank Heather Down, Scientific Assistant at the University of Victoria, for kind help with technical equipment taking pictures of the *A. thaliana* knock-out mutants.

2. The evolution of CYP98s within land plants

2.1. Summary

When plants moved onto land about 480 million years ago, multiple protection mechanisms were needed to face novel environmental challenges. Plants adapted their metabolism and plant natural products became important for survival. One pathway giving rise to such compounds is the phenylpropanoid pathway. Here we describe the role of a cytochrome P450 enzyme family, CYP98, which is involved in the phenylpropanoid metabolism of land plants. CYP98s have been described in angiosperms to be involved in the biosynthesis of monolignols in the lignin biosynthetic pathway. They have further been described to be involved in the formation of natural products such as chlorogenic acid and rosmarinic acid. To reveal the evolution of the CYP98 family we investigated CYP98s from the moss *Physcomitrella patens*, the lycopod *Selaginella moellendorffii*, the fern *Pteris vittata*, the gymnosperm *Pinus taeda*, and from two angiosperms, namely the monocot *Brachypodium distachyon* and the eudicot *Arabidopsis thaliana*.

Phylogenetic reconstructions suggest that a single copy CYP98 founded each major land plant lineage and that gene duplications appear to have occurred only in angiosperms. Based on *in vitro* assays, the CYP98s from the angiosperms tested prefer *p*-coumaroyl-shikimate as their substrate, while CYP98s from non-seed plants have distinct substrate preferences. *P. patens*, *S. moellendorffii* and *P. vittata* CYP98s show preference for *p*-coumaroyl-anthranilate and produce only trace amounts of caffeoyl-shikimate *in vitro* or are even unable to produce caffeoyl-shikimate *in vitro*. An involvement of CYP98 in lignin biosynthesis of ferns is proposed. Metabolic profiling of a *P. patens cyp98a34* knock-out mutant revealed *p*-coumaroyl-threonate as the CYP98A34 substrate *in vivo*. The *P. patens* knock-out mutant showed a severe developmental phenotype. The *P. patens* CYP98A34 was unable to complement the *A. thaliana cyp98a3* loss of function phenotype. The gymnosperm CYP98A19 from *P. taeda* has a comparably broad substrate range. It overlaps with those observed for angiosperms and non-seed plants. This possibly represents a transitional stage between the biochemically and physiologically distinct functions of CYP98s in angiosperms and non-seed plants.

2.2. Introduction

Land plants evolved from freshwater green algae (Kranz et al., 1995) about 480 million years ago (Sanderson, 2003; Kenrick et al., 2012). Diverse new secondary metabolite-based protection mechanisms were acquired and established throughout evolution of early land plants to overcome new or enhanced stresses such as damaging UV-light, desiccation, rapid wide and extreme temperature fluctuations and the loss of structural support (Raven, 1984; Proctor, 2014). Coevolution of pests and pathogens made additional protection mechanisms necessary. Land plants are sessile organisms. They diversified and expanded their metabolism to adapt to the changing environment. The development of the phenylpropanoid metabolism was likely among the most critical processes during this time (Douglas, 1996; Weng and Chapple, 2010; Bhardwaj et al., 2014). Phenylpropanoids in plants today range widely between species both in their quantity and chemical diversity. While some phenylpropanoids can be found in almost all plants, some are taxon specific (Clifford, 2000; Dixon, 2001; Petersen and Simmonds, 2003; Petersen et al., 2009). Phenylpropanoids range from relatively simple hydroxycinnamic acid conjugates (HCCs) to complex polyphenols. Common to all phenylpropanoids is a six carbon aromatic ring which is bound to a three carbon skeleton. HCCs include for example caffeic or ferulic acid esterified to alcohols or amines. The characterization of these conjugates as anti-herbivory, antiviral, antibacterial, anti-inflammatory and antioxidant compounds implies a role in plant defence (Petersen and Simmonds, 2003; Gülçin, 2006; Chao et al., 2009; Barbehenn et al., 2010; Bassard et al., 2010; Macoy et al., 2015b; Corral-Lugo et al., 2016). For example caffeoyl-quinate, chlorogenic acid (CGA), is an anti nutrient for insects (Barbehenn et al., 2010). Due to its anti-inflammatory and antioxidant activities CGA is also subject to research in the medical field. It can reduce liver inflammation and fibrosis (Shi et al., 2013) and is used as a weight-control agent (Buchanan and Beckett, 2013).

It is known that one hydroxycinnamoyl ester, *p*-hydroxycinnamoyl-shikimate, functions as an intermediate in lignin biosynthesis in angiosperms (Schoch et al., 2001). Lignin is a complex polymer of covalently linked phenylpropanoids called monolignols and a major constituent of many secondary cell walls. It is found in the walls of vessels in xylem, and fiber cells of woody tissues (Boerjan et al., 2003). The content and composition of lignin varies between species,

different tissues and also the different layers of the plant cell wall (Campbell and Sederoff, 1996). As main units, *p*-coumaryl, coniferyl and sinapyl alcohols are incorporated into the lignin polymer where these monolignols produce *p*-hydroxyphenyl (H), guaiacyl (G) and syringyl (S) lignin units respectively (Boerjan et al., 2003). The biosynthesis of G and S monolignols as well as soluble HCCs involves hydroxylation of the phenolic ring. Cytochromes P450 (CYPs) are capable of fulfilling this task. CYPs are monooxygenases found in all organisms, from bacteria to humans (Nelson, 1999). Characterized CYPs in plants are membrane anchored and need to be coupled to an electron-donating protein to be active. Hydroxylation at the 3'-position of the aromatic ring of phenylpropanoid precursors is an important step for the biosynthesis of metabolites (Mizutani and Ohta, 2010; Vogt, 2010). This hydroxylation step is catalysed by 4-coumaroylshikimate 3'-hydroxylase (C3'H, CYP98). Diverse hydroxycinnamoyl esters or amides, for example 4-coumaroyltyramine and tri-coumaroyl-spermidine, can be substrates for enzymes belonging to the CYP98 family (Morant et al., 2007; Matsuno, et al., 2009).

Loss of function of this 3'-hydroxylation step leads to drastic inhibition of plant development, as shown in the *Arabidopsis thaliana cyp98a3* mutant. Plants have a stem lignin mainly composed of H units and show severely dwarfed phenotypes with rosette diameters not exceeding 1.5 cm (Franke et al., 2002; Abdulrazzak et al., 2006). *A. thaliana* has three CYP98 family members. Unlike CYP98A3, CYP98A8 and CYP98A9 do not show appreciable *p*-coumaroyl-shikimate hydroxylase activity *in vitro*. *A. thaliana* knock-out plants of *CYP98A8* and *CYP98A9* do not show the dwarf phenotype of the *cyp98a3* mutant. *CYP98A8* and *CYP98A9* were shown to have evolved recently through retroposition from *CYP98A3* and have gained a novel function, that is hydroxycinnamoyl-spermidine 3- and 5- hydroxylations involved in pollen development (Matsuno, et al., 2009). Characterized CYP98 families of other angiosperm species also show functional divergence within the family. This will be discussed in detail in Chapter 3. These findings point towards a functional divergence within the gene family in general. It appears that CYP98s related to lignin biosynthesis are specific for *p*-coumaroyl-shikimate, while other CYP98s with different substrate preferences are involved in the biosynthesis of soluble hydroxycinnamic conjugates in angiosperms.

While *CYP98* members are absent in green algal genomes (Nelson, 2006; Xu et al., 2009), a single *CYP98* gene is present in the bryophyte *Physcomitrella patens*. *P. patens* is a model moss, with a fully sequenced and annotated genome (www.cosmoss.org). It has a dominant haploid life form, with gametophores consisting of leaf-like phyllids, only a single cell layer thick. Reverse genetics studies in *P. patens* are facilitated by high rates of homologous recombination, which can be used for the creation of knock-out plants (Hohe et al., 2004). Comparing functions of a gene family in bryophytes, which separated from vascular plants about 430 million years ago, to functions of the same gene family in separately evolving plant lineages, can provide insight into likely ancestral gene functions. Albeit it might be challenging to elucidate events which happened long ago, this can provide interesting insight on a genetic and functional level in the molecular evolutionary mechanisms that shape divergence of gene family members. Bryophytes are reported to be able to synthesize soluble phenylpropanoids such as flavonoids and lignans, but lack true lignin (Basile et al., 1999; Umezawa, 2003). Nevertheless, eight out of ten genes considered to be core monolignol biosynthesis genes can be found in the *P. patens* genome (Xu et al., 2009), among them one *CYP98*. A single *CYP98* exists in the lycopod *Selaginella moellendorffii* (a basal tracheophyte) (Goodstein et al., 2012). While *P. patens* does not produce lignin, *S. moellendorffii* can do so by bypassing the C3'H/CYP98 step (Weng et al., 2008a). *S. moellendorffii* instead employs a distinct P450, CYP788A1, which catalyses both the aromatic 3- and 5-hydroxylation on the level of the free aldehyde or alcohol (Weng et al, 2010). Thus, in neither the bryophytes or lycopods is the single CYP98 involved in lignin biosynthesis.

No CYP98 has been biochemically characterized in the fern group (monilophytes) or in the gymnosperms to date. Both groups are known to produce extensive amounts of lignin. The biosynthesis of lignin in ferns is largely uncharacterized, but several ferns have been investigated for their lignin presence and composition, and mainly G units have been found, but also traces of H units and relevant amounts of S lignin in some species (Espiñeira et al., 2011). G units are also the main component of gymnosperm lignin, with small proportions of H units (Baucher and Monties, 1998).

In summary, CYP98 members in angiosperms are necessary for the biosynthesis of soluble hydroxycinnamic conjugates and the biosynthesis of monolignols. Lignin related CYP98s appear

to be specific in their choice of substrate and prefer *p*-coumaroyl-shikimate. Lignin composed of all three major building blocks, H, G and S subunits, has been found in ferns, gymnosperms and lycopods. An involvement of CYP98s in lignin biosynthesis in ferns, gymnosperms and lycopods remains unclear. Lycopods and bryophytes possess CYP98 genes, but do not use these genes for lignin biosynthesis.

Here, we reconstruct the evolutionary history of the CYP98 family across the land plants and biochemically characterize CYP98 family members from representative species of each major land plant lineage. We further characterize the CYP98 from the bryophyte *P. patens* through reverse genetic approaches. We aim to reconstruct the functional evolutionary changes of enzymes belonging to the CYP98 family, to gain insight about the functional diversity of CYP98 and the recruitment of CYP98 for lignin biosynthesis

2.3. Material and methods

2.3.1. Phylogenetic analysis

Amino acid sequences of all characterized CYP98s and of CYP98s from bryophytes, lycopods, ferns gymnosperms and angiosperms were used in the phylogenetic analysis (List of species see supplemental Table 2.1). Sequences were retrieved from ncbi (https://www.ncbi.nlm.nih.gov/), Phytozome v. 11 (https://phytozome.jgi.doe.gov/pz/portal.html) and oneKP (https://sites.google.com/a/ualberta.ca/onekp/). The phylogenetic tree was based on an amino acid sequence alignment generated with DIALIGN (Morgenstern, 1999). Positions with DIALIGN similarities greater than one were kept in the alignment. The Maximum Likelihood phylogenetic tree was reconstructed using PhyML (Guindon and Gascuel, 2003). The chosen amino acid substitution model was LG (Le and Gascuel, 2008). Statistical branch support was obtained by bootstrapping 100 replicates.

2.3.2. Heterologous enzyme expression in *Saccharomyces cerevisiae*

The complete open reading frame of *CYP98A4* from *B. distachyon* was cloned from cDNA using appropriate primers. Open reading frames for *P. patens*, *S. moellendorffii*, *P. vittata* and *P. taeda* were chemically synthesized as yeast-codon optimized genes by GenScript, New Jersey,

USA. Genes were cloned into yeast expression vector pYeDP60USER by USER™ cloning (Nour-Eldin and Hansen, 2006; Nour-Eldin et al., 2010). For primer sequences see supplemental Table 2.2. The *A. thaliana CYP98A3* had been cloned into pYeDP60 by (Schoch et al., 2001). *S. cerevisiae* strain WAT11 was transformed by heat shock using salmon sperm DNA as a carrier (Gietz and Jean, 1992) or by electroporation (400 Ohm/250 µF/0,45 kV). The growth method for yeast cultures and the preparation of microsomal fractions, containing the recombinant enzyme, have been described in (Gavira et al., 2013). P450 quality control and quantification was performed by differential spectrophotometry as described in (Gavira et al., 2013) using the absorption coefficient at 450 nm: ε=91 mM^{-1} cm^{-1} (Omura and Sato, 1964).

2.3.3. CYP98 enzyme incubations with a library of potential substrates

Microsomal fractions of yeast transformed with CYP98 were used in incubations with various substrates. 10 pmol of P450 were added to a reaction volume of 400µl. Reactions were performed in 50mM potassium phosphate buffer (pH7.4), containing 100µM substrate and 500µM NADPH. For kinetic properties of the reductase ATR1, refer to (Urban et al., 1997) Reactions were started by addition of NADPH and incubated at 28°C for 30 min. Reactions were stopped by addition of 1/10 (v/v) 50% acetic acid and 4/10 (v/v) methanol. After centrifugation (10min; 15000g; 4°C) the supernatant was used for analysis on HPLC/DAD. Three independent incubations were performed for each enzyme/substrate combination. Substrate conversion was monitored. For this, the substrate peak area of the chromatogram was integrated using the Empower (Waters) software. The percentage of conversion was calculated from the peak areas of substrate after incubation, compared to the initial amount of substrate.

2.3.4. Expression of *P. patens* HCT (Phpat.002G119200)

The complete open reading frame of Phpat.002G119200 was cloned from cDNA using appropriate primers (Table 2.2). The gene was cloned into the Gateway™ entry vector pDONR207 (Invitrogen) first and then recombined into the pHGGWA expression vector. Recombinant protein was produced by the protein platform of the IBMP CNRS Strasbourg. Competent *E. coli* Rosetta2pLysS (Novagen) were transformed with the vector by heat shock transformation and selected on LB agar plates containing 100 µg/mL carbenicillin and 34 µg/mL

chloramphenicol. A preculture of cells was transferred to ZYP-5052 auto induction medium (Amresco) containing carbenicillin and chloramphenicol. After growth over night cells were harvested by centrifugation and cell pellets eluted in buffer containing 50mM Tris (pH8), 300mM NaCl, 5% glycerol. Before sonication, 0.2% Triton X100 and 1mM AEBSF were added and cells sonicated for 6 minutes (2 seconds sonification, 2 seconds break). After centrifugation for 20 minutes at 4°C 16,000g, the supernatant was filtered (0.22µm) and 25mM imidazole were added. The enzyme was eluted by IMAC on an AKTA Purifier 10, using a HisTrap column (IH1007).

2.3.5. HCT incubations

P. patens HCT (50 pmol) was incubated in 50mM potassium phosphate buffer at pH 6.6 with L-threonic acid (4mM) and *p*-coumaroyl-CoA (0.25mM) for one hour at 30°C. The reaction was stopped by addition of 1/10 (v/v) 50% acetic acid and 4/10 (v/v) methanol.

In a coupled reaction, first the *P. patens* HCT was incubated in 50 mM potassium phosphate buffer at pH 7.4 with L-threonic acid (4mM) and *p*-coumaroyl-CoA (0.25mM) for one hour at 30°C, and subsequently the *P. patens* CYP98A34 (10pmol/400µl) and NADPH (500µM) were added. After 30 min incubation at 28°C the reactions were stopped by addition of 1/10 (v/v) 50% acetic acid and 4/10 (v/v) methanol. After centrifugation (10min; 15000g; 4°C) the supernatant was used for analysis on UPLC-MS/MS.

2.3.6. Analysis on HPLC/DAD

Reverse phase HPLC with photo-diode array detection (Alliance 2695, Waters, Photodiode 2996, Waters, Software Empower) was performed using a NOVA-PAK C18 4.6 x 250 mm column (at 37°C). 50 µl of sample were injected. Solvents were composed of water with 0.2% formic acid (A) and acetonitrile with 0.2% formic acid (B). The gradient used was 5% to 100% B within 16 minutes, following curve 7, followed by 1 min of 100% B, with a flow of 1ml per minute.

2.3.7. Analysis on UPLC-MS/MS

Analyses on UPLC-MS/MS (Acquity UPLC; Quattro Premier XE^{TM} tandem-mass spectrometer, electrospray ionization source; Waters corp.) were performed by the metabolic platform of the IBMP CNRS, Strasbourg or by H. Renault.

Ten microliters of extract were injected onto a UPLC BEH C18 column (100 x 2.1 mm, 1.7 µm; Waters) outfitted with a pre-column and operated at 35°C. Metabolites chromatography was performed at a 0.35 mL/min flow rate with a mixture of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) according to the following program: initial, 98% A; 11.25 min, 0% A, curve 8; 12.75 min, 0% A, curve 6; 13.50 min, 98% A, curve 6; 15 min, 98% A. Nitrogen was used as the drying and nebulizing in-source gas. The nebulizer gas flow was set to 50 L/h, and the desolvation gas flow was set to 900 L/h. The interface temperature was set to 400°C, and the source temperature to 135°C. Capillary voltage was set to 3.4 kV. Data acquisition and analysis were performed with the MassLynx v4.1 software (Waters corp.). Metabolites were ionized in positive mode and detected using dedicated multiple reaction monitoring (MRM) methods. The QuanLynx module of MassLynx was executed to integrate peaks and to report corresponding areas. Peak areas were normalized to plant dry weight and internal standard level (morin), leading to relative levels.

2.3.8. Standards for incubations

Substrate and reference phenolic conjugates except *p*-coumaroyl-shikimate were provided by the group of M. Schmitt (CNRS, UMR 7200, Illkirch).

*p-C*oumaroyl-shikimate was produced enzymatically from *p*-coumarate, as described in (Morant et al., 2007). In brief, the tobacco HCT (Hoffmann et al., 2003) was expressed in the bacterial strain BL21-G612. Cells were mechanically lysed for enzyme extraction. Enzymes were extracted using GSH agarose beads (Sigma, ref G-4510) and lysis buffer (1% Triton X, 1mM EDTA pH8, 0.1% ß RSH, protease inhibitor in PBS). Enzyme was stored in 0.1M K phosphate buffer pH6.6 containing thrombin (1U thrombin/µl. Sigma ref T-7513).

The 4CL1 from *A. thaliana* was expressed in the bacterial strain BL21-G612. For extraction, lysis buffer (Na phosphate 20 mM / NaCl 500 mM / imidazole 20 mM, pH 7.4) was added to the cells and cells ground mechanically in liquid nitrogen. Enzyme was purified on His GraviTrap column following the manufacturer's instructions (Amersham Biosciences, ref 11-0033-99). Subsequently a buffer exchange was performed on a Sephadex G25 column K phosphate buffer with DTT and glycerol to a final concentration of 10% (v/v) was added for storage. SDS-PAGE gels of enzyme purification steps are presented in Figure 2.15 and Figure 2.16 in the supplement.

For the preparation of *p*-coumaroyl-CoA, 4CL1 (0.02mg/ml final) was incubated in K phosphate buffer (50mM) at pH 7 with *p*-coumaric acid (0.2mM), $MgCl_2$ (2.5 mM), DTT (1 mM), ATP (2.5 mM), and CoA-SH (0.2 mM) for 30 minutes at 25°C.

For the preparation of *p*-coumaroyl-shikimate, HCT (0.02mg/ml final) was incubated with unpurified *p*-coumaroyl-CoA (0.21mM) from previous reaction and shikimic acid (4mM) in K phosphate buffer pH 6.6 (50mM) at 30°C for one hour. Reaction was stopped by addition of HCl (1v/20) and extracted by ethyl acetate two times.

2.3.9. Plant material and growth conditions

A. thaliana plants were grown under 16h/8h day/night cycle, with a light intensity of 50µmol/m^{-2}/s^{-1}. The humidity was 70% and the day/night temperature cycle 21°C/18°C.

P. patens (Hedw., Bruch & Schimp., strain Gransden) was grown to gametophores on KNOP (pH 5.8) (Reski and Abel, 1985) plates containing 1.2% (w/v) agar. KNOP medium contained 250 mg/l KH_2PO_4, 250 mg/l KCl, 250 mg/l $MgSO_4$ x $7H_2O$, 1 mg/l $Ca(NO_3)_2$ x 4 H_2O, 12.5 mg/l $FeSO_4$ x 7 H_2O. 10 ml microelement solution was added per 1l KNOP medium. Microelement solution contained 309 mg/l H_3BO_3, 845 mg/l $MnSO_4$ x 1 H_2O, 431 mg/l $ZnSO_4$ x 7 H_2O, 41.5 mg/l KI, 12.1 mg/l Na_2MoO_4 x 2 H_2O, 1.25 mg/l $CoSO_4$ x 5 H_2O, 1.46 mg $Co(NO_3)_2$ x 6 H_2O. Liquid cultures were grown in KNOP medium (pH 5.8) supplemented with microelements as described above, in Erlenmeyer flasks under shaking (Reski and Abel, 1985). The moss was kept at protonema stage by disruption (90 sec at 18,000 rpm) using an Ultra-Turrax (IKA) device every 2 weeks. Plants were grown in a cycle of 16h light (70 µmol/m^2s) and 8h dark at 23°C.

2.3.10. *P. patens CYP98A34* knock-out generation by homologous recombination

To generate the *PpCYP98* knock-out mutants, two 750 bp genomic regions were PCR amplified from *P. patens* genomic DNA and assembled with the *nptII* selection cassette into a PCR-linearized pGEM-T vector (Promega) via GIBSON™ cloning (NEB). The *PpCYP98* disruption construct was excised from the vector backbone by BamHI digestion, using restriction sites introduced during PCR. 25 µg of linearized construct were used for PEG-mediated transfection of *P. patens* protoplast. Transformants were selected on Knop plates supplemented with 25 mg/L geneticin (G418).

Following the selection process, a previously established direct-PCR protocol (Schween et al., 2002) was implemented to identify transformants with proper genomic integration of the DNA construct. Both 5' and 3' integrations were verified using appropriate PCR strategy and primers (Table 2.2).

PpCYP98 knock-out lines identified by direct-PCR were further validated at the molecular level by conventional RT-PCR. To this end, total RNA was isolated from ~8 mg of lyophilized 3-day-old protonema material using TriReagent (Sigma-Aldrich). Twenty micrograms of RNA were treated with 5U of RQ1 DNaseI (Promega) and subsequently purified with the Nucleospin RNA clean-up XS kit (Macherey-Nagel). One microgram of DNaseI-treated RNA was reversetranscribed with oligo(dT)23V and the SuperScriptIII™ enzyme (Thermo Scientific) in 20 µl reaction. *PpCYP98* transcripts were amplified from one microliter of cDNA using the Phire II polymerase (Thermo Scientific) in 20 µl reaction. The constitutively expressed *L21* gene (Pp1s107_181V6.1), encoding a 60S ribosomal protein, was used as reference gene. The number of *PpCYP98* knock-out lines was narrowed down after transcript analysis; remaining lines were subjected to transgene copy analysis by qPCR as described before (Horst et al., 2016). Genomic DNA was isolated using a protocol adapted from Edwards et al., 1991. Briefly, nucleic acids were extracted from 5 mg of 3-day-old lyophilized protonema material with 500 µl of lysis buffer (200 mM Tris HCl pH 8.0, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) and thorough agitation. Nucleic acids were purified by addition of a 500 µl of a phenol:chloroform (1:1) solution (pH 8.0) followed by a precipitation with sodium acetate and isopropanol. DNA pellets were washed

with 75% ethanol before solubilization in a 5 mM Tris pH 8.5 solution. Samples were further treated with a RNase A/T1 mix (Thermo Scientific) to remove RNA. DNA was re-purified with a phenol:chloroform step as described above. Typical yields were ~0.5 µg DNA/mg dry plant material. Quantitative PCR reactions were run on a LightCycler 480 II device (Roche) in 10 µl comprising in 1 ng genomic DNA reaction, 250 nM of each primer and 1X LightCycler® 480 SYBR Green I Master mix (Roche). Reactions were performed in triplicate and crossing points (Cp) were determined using the manufacturer software. Both the 5'- and 3'-homologous regions were targeted using specific primers. The single copy gene *PpCLF* (Pp1s100_146V6.1) was amplified using two primer pairs and served as an internal standard for input amount normalization. Transgene copy number was determined by comparing relative values of the tested genomic segment in transgenic lines with those of the wild type. Three *PpCYP98* knock-out lines with a single transgene copy were kept for subsequent phenotypic and metabolic analyses.

2.3.11. *A. thaliana* Tn4 mutant complementation assay

The *A. thaliana* T-DNA insertion mutant knock-out for *CYP98A3* (Abdulrazzak et al., 2006) was used in a mutant complementation assay with the *P. patens CYP98A34* gene. As homozygous T-DNA lines of *cyp98a3* show dwarf morphology and are male sterile, heterozygous plants were used for transformation with *CYP98A34* under the promoter of the *A. thaliana* C4H gene (Bell-Lelong and Cusumano, 1997). The use of this promotor ensured enhanced expression in lignified tissues. The open reading frame of *CYP98A34* was cloned into the pDONR207 Gateway™ entry vector (Invitrogen). Recombination with the Gateway™ destination vector pCC0996 resulted in the expression construct. pCC0996 contains the C4H promoter sequence of *A. thaliana* (Weng et al., 2010b). *Agrobacterium tumefaciens* strain GV3101 was transformed with the expression construct and *A. thaliana* plants transformed by floral dip as described (Clough and Bent, 1998). Seed of control plants with pCC0996:CYP98A3 expression constructs were a courtesy of Z. Liu (John Innes Centre, Norwich, UK).

2.3.12. RT-PCR of *A. thaliana* Tn4 mutant complementation

A. thaliana RNA was extracted using the Macherey Nagel NucleoSpin™ RNA Plant kit, following the manufacturer's instructions. cDNA was synthesized from 1.5 µg total RNA of each sample, using SuperScriptIII™ and oligo(dT)23 primer, following the manufacturer's instructions. Primers used for RT-PCR see supplement, Table 2.2.

2.3.13. *P. patens* plant extract analysis

P. patens plant material was collected from 2 month-old gametophores growing on KNOP (pH 5.8) (Reski and Abel, 1985) plates supplemented with microelements, or from plants at the protonema stage growing in liquid culture, 4 days after tissue disruption and exchange of growth medium. Plant tissue was harvested, flash frozen in liquid nitrogen and lyophilized. Lyophilized material was ground to fine powder using a tissue lyzer (Qiagen). Five mg of plant dry material were extracted using methanol, chloroform and water (2:1:2) in subsequent extraction steps.

2.3.14. *p*-Coumaroyl-threonate isomerization

p-Coumaroyl-2-threonate was incubated in 0.1M K phosphate buffer (pH 7.4, optimized for subsequent P450 incubations) for 60 minutes at 90°C in the dark. Conversion of *p*-coumaroyl-2-threonate to *p*-coumaroyl-4-threonate was verified by HPLC/DAD. The isomers showed clearly distinguishable elution times. No *p*-coumaroyl-3-threonate was detectable in any of these reactions, although it is expected to be an intermediate.

2.4. Results and discussion

2.4.1. Genome mining and phylogenetic analysis

Extensive sequence similarity searches using the *A. thaliana* CYP98A3 protein sequence as bait were performed against available bryophyte, lycophyte, fern and gymnosperm genome and transcriptome databases. This identified a single CYP98 member in all species belonging to these groups. In total, sequence data from currently 58 genomes available on "Phytozome v11" (Goodstein et al., 2012) and transcriptomes of the 1000 plants transcriptomes project (1kp,

www.onekp.com) were analysed. Multiple CYP98 isoforms were found only in the angiosperms, both in monocots and eudicots. Gene family size in angiosperms ranged from one (examples are *B. distachyon*, *Carica papaya*) to 12 *(Malus* x *domestica*) members with a median of two gene family members across 43 angiosperm genomes analysed. A BLAST search (Altschul et al., 1990) against genomes of the green algae *Chlamydomonas reinhardtii*, *Volvox carteri*, *Coccomyxa subellipsoidea*, *Micromonas pusilla* and *Ostreococcus lucimarinus* did not result in the identification of any CYP98 members, neither did searches against any NCBI sequence database when excluding land plants (embryophytes). CYP98 families of representative species from each major land plant lineage (bryophytes, lycopods, ferns, gymnosperms, and angiosperms) were chosen for phylogenetic reconstruction. Species were selected to have similar representation in each major lineage and to cover the major orders in each group where data was available. CYP98A34 of the moss model *P. patens* and CYP98 members from additional mosses, as well as from hornworts were included. Available lycopod CYP98s have been included, among these the CYP98A38 of *S. moellendorffii*. Following the classification of extant ferns by Smith *et al*. (Smith et al., 2006b) CYP98s of the four classes Polypodiopsida, Marattiopsida, Equisetopsida and Psilotopsida were identified for the analysis. The CYP98 of *P. taeda*, CYP98A19, which has been previously described (Anterola, 2002), was included in the analysis, as well as other CYP98s of conifers, cycads, gnetales and ginkgo. Angiosperm monocot and eudicot CYP98s have been added to the analysis including all biochemically characterized CYP98s. Both CYP98s connected to lignin biosynthesis as well as CYP98s connected to the biosynthesis of soluble HCCs are contained within the angiosperm clade. The tree topology follows the expected land plant phylogeny and all major lineages form monophyletic clades with strong bootstrap support (Figure 2.1).

Focussing on CYP98s from vascular plants (i.e. lycopods, ferns, gymnosperms, and angiosperms), there is no evidence for a separation into two subclades prior to separation of the lineages, one related to lignin biosynthesis, the other(s) to soluble HCC biosynthesis, as may have been expected given that a central uniting feature of all these lineages is lignin biosynthesis.

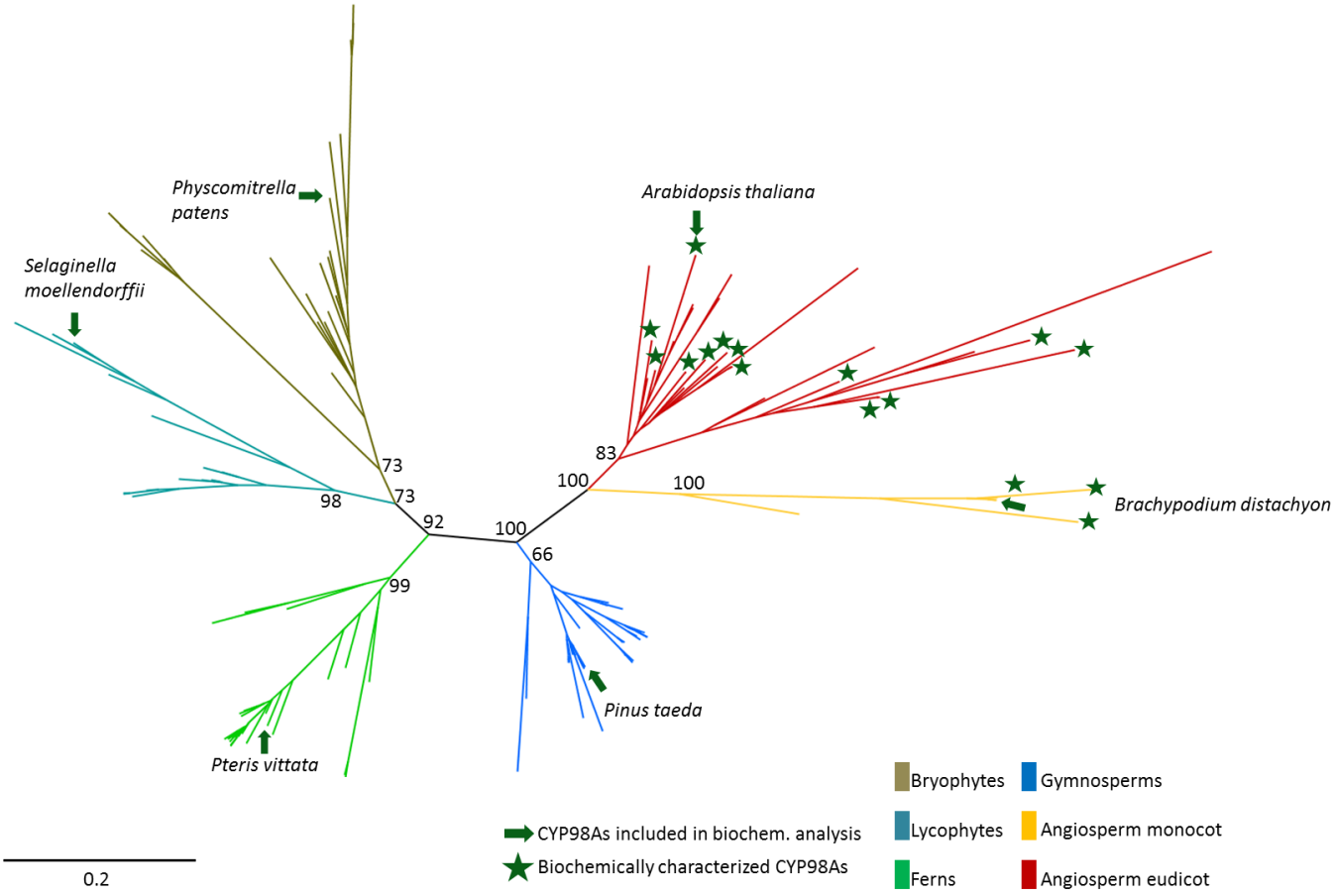


Figure 2.1 Phylogenetic reconstruction of CYP98s of land plants.

Radial phylogenetic tree. Amino acid sequences of bryophytes, lycophytes, ferns, gymnosperms and angiosperms were included (For a list of species see supplemental table Table 2.1) Only bootstrap support for the branching points of the major land plant lineages is shown in the radial phylogenetic tree. The phylogenetic tree was based on an amino acid sequence alignment generated with DIALIGN (Morgenstern, 1999). The Maximum Likelihood phylogenetic tree was reconstructed using PhyML (Guindon and Gascuel, 2003). Statistical branch support was obtained by bootstrapping 100 replicates. All characterized CYP98s belong to the angiosperm clade and were indicated by stars. CYP98 family members characterized here were indicated by green arrows.

Instead, gene duplications are lineage specific and occurred or were maintained only in the angiosperms. In consequence, all genes from a given lineage arose from a single common ancestor. Phylogenetic reconstruction of the CYP98 family is often challenging. Good statistical support and high resolution of subfamilies was obtained for bryophytes to gymnosperms, i.e.

lineages without duplicates. In contrast, within the angiosperms, statistical bootstrap support was generally poor and relationships within the group were inconsistent and frequently yielded poor bootstrap support. This points to an intense history of gene duplication and gene loss within the angiosperms, difficult to reconstruct. Nevertheless, it is clear that both lignin and soluble HCC related CYP98s are contained within the monophyletic angiosperm clade, suggesting that functional diversification within the angiosperms only occurred after their split from the gymnosperms. A detailed characterization of these gene duplication events within the angiosperms will be the focus of Chapter 3 of this thesis.

Assuming a recruitment of CYP98 for lignin only after the lycopod/euphyllophyte split (because lycopods utilize a CYP98-independent pathway), it remains unclear when, and how frequently that happened. Formulating alternative hypotheses (Figure 2.2), it is possible that one recruitment of CYP98 for lignin biosynthesis happened early in the euphyllophytes (ferns, gymnosperms, and angiosperms), and in consequence, the function of the single ancestral euphyllophyte CYP98 included lignin biosynthesis. As described above, all described CYP98s related to lignin biosynthesis prefer coumaroyl-shikimate as their substrate. In this scenario, a lignin related function is ancestral and the single-copy fern and gymnosperm CYP98s would also be expected to favour coumaroyl-shikimate (hypothesis 1). It is equally possible that CYP98 isoforms exist in ferns and also in gymnosperms that are rather versatile in function in order to fulfil both lignin and soluble phenolics related functions, and that subfunctionalization happened only within the angiosperms (hypothesis 2). An alternative third hypothesis would be the recruitment of CYP98 for the biosynthesis of lignin only in seed plants, or even only in angiosperms. Alternative pathways to monolignol biosynthesis might exist in ferns and maybe even in gymnosperms, as described for lycopods (hypothesis 3). Using coumaroyl-shikimate as a proxy for an involvement in lignin biosynthesis, we would then expect distinct substrate utilization profiles for non-lignin related CYP98s. A fourth scenario includes independent recruitments of CYP98s for lignin in each lineage of plants, although this hypothesis would be less parsimonious compared to a single ancestral recruitment (hypothesis 4).

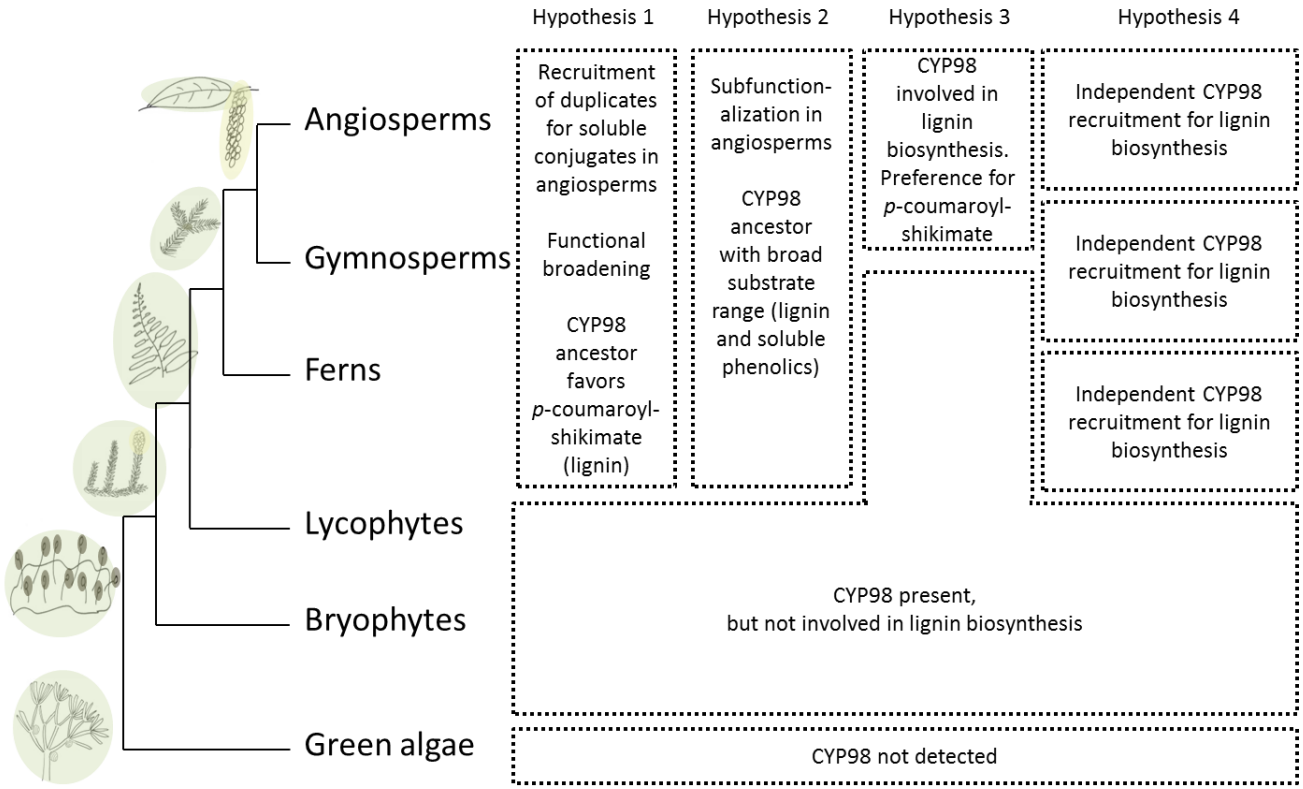


Figure 2.2 Four different hypotheses of the recruitment of CYP98 for lignin biosynthesis.

The groups of green algae and land plants are given in the cladogram on the left. Four different hypotheses are drawn on the right.

Towards testing these alternative hypotheses, we next characterized the substrate utilization profiles of CYP98s from representatives of each lineage.

2.4.2. Enzymatic diversity of CYP98s across the plant lineage

Representative plants of each major lineage have been chosen for enzymatic screening of their CYP98 family members. Species were selected with the following rationale: The moss model plant *P. patens* was the first bryophyte genome to be sequenced and provides extensive genome annotation. CYP98A34 is the only *P. patens* CYP98 family member. The only lycopod with a complete genome available, *S. moellendorffii*, has been subject to research in the context of lignin content and composition, but its single CYP98A38 has not been characterized biochemically to date. *P. vittata*, a species of leptosporangiate ferns, has transcriptome data available from the 1000 plant transcriptomes project. Generally, ferns were given little

appreciation, but *P. vittata* is under research related to phytoremediation, because it is a hyper accumulator of arsenic and capable of taking up antibiotics from water (Danh et al., 2014; Li et al., 2015a). Lignin analysis of *P. vittata* showed a lignin consisting of G units only, similar to gymnosperms (Weng et al., 2008b). Its single CYP98 member, designated PvCYP98, was included here. CYP98A19 of *P. taeda* has been cloned and its expression investigated in a *P. taeda* cell suspension culture experiment (Anterola, 2002). Upregulation of enzymes involved in lignin biosynthesis except for CYP98A19 and cinnamate 4-hydroxylase was detected when the cells were transferred to a medium containing sucrose and potassium-iodide. CYP98A19 has not been biochemically characterized to date. It was thus included in the functional enzyme screening of this work. The grass *B. distachyon* is an angiosperm monocot model plant and the focus of research in the field of grass cell walls and its use as a biofuel crop (Coomey and Hazen, 2015). We selected *B. distachyon* also because its genome only contains a single *CYP98*, *CYP98A4*. It can therefore be assumed to be involved also in lignin biosynthesis in this monocot. The eudicot angiosperms representative CYP98A3 of *A. thaliana* has been subject to research in the past and was the CYP98 member which revealed the involvement of CYP98s in monolignol biosynthesis (Schoch et al., 2001; Franke et al., 2002; Nair et al., 2002). It has been included in the enzymatic screen to provide reference to a member with known function, but also to evaluate substrate specificity/promiscuity of this reference CYP98.

The respective cDNA sequences were cloned after reverse transcriptase polymerase chain reaction (RT-PCR) from total RNA were plant material was available or were chemically synthesized commercially as yeast expression codon optimized DNA if plant material was not available. Enzymes were cloned into a yeast expression vector and transformed into *S. cerevisiae* strain WAT11, which contains the *A. thaliana* cytochrome P450 reductase ATR1 in its genome, for heterologous protein production. Yeast microsomes containing membrane anchored enzymes were prepared for enzymatic tests. Measured on a spectrophotometer, intact CYPs' Soret absorption band shift to a maximal absorption at 450 nm when the enzymes are reduced and complexed with CO. The amount of functional CYP enzyme can thus be determined in a differential spectrum of reduced CYP and CYP that is reduced and complexed

with CO (Figure 2.3). The CO spectra that were obtained for the species characterized here indicated the presence of intact CYP98 enzyme for all species.

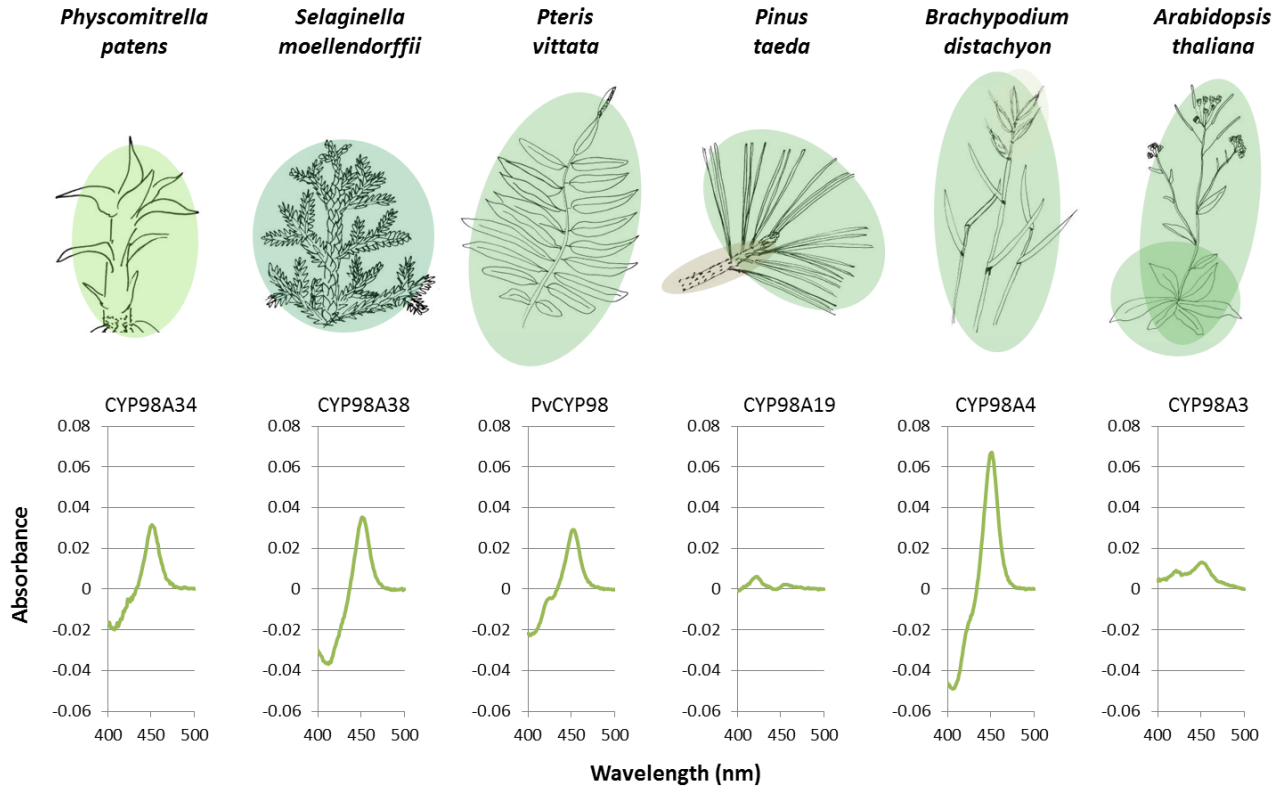


Figure 2.3 Differential CO spectra of CYP98 included in the biochemical analysis.

Microsomes for the CYP98 expressing yeast were isolated by differential centrifugation. An absorbance peak at 450 nm in CO bound reduced/reduced differential spectra is indicative of the expression of functional enzyme and allows its quantification.

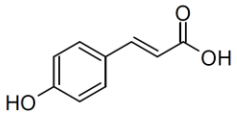
Yeast microsomes were incubated with a variety of substrates (Figure 2.4). Most of these substrates are not commercially available and were provided by the group of M. Schmitt, Laboratoire d'Innovation Thérapeutique, UMR 7200, Illkirch France. The range of substrates comprised known substrates of CYP98s, such as *p*-coumaroyl-shikimate (enzymatically synthesized) and *p*-coumaroyl-quinate (chemically synthesized). Free coumaric acid was included in the experiment to test for CYP98-mediated hydroxylation on the free acid. Several phenolamides were included because they are known to be utilized by some CYP98s, for

example *Triticum aestivum* CYP98A11 and CYP98A12 can hydroxylate *p*-coumaroyl-tyramine (Morant et al., 2007). Furthermore, several coumaric esters, potential precursors of caffeic esters that exist in nature, have been chemically synthesized for the experiment, for example prenyl-, isoprenyl- and benzyl-coumarate (Rubiolo et al., 2013). To assess the potential substrate range, some artificial substrates were tested as well.

Product formation was analysed via liquid chromatography for all reactions. An example is shown in Figure 2.5. Peak areas of substrates incubated with CYP98 containing microsomes, but without NADPH, were determined as control. Substrate peak area reduction after incubation with CYP98-containing microsomes and NADPH, relative to control, yielded relative metabolization of a given substrate in this end-point screening experiment (Figure 2.6).

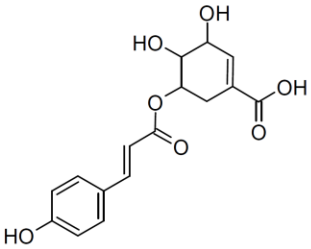
As general observation, we found that the end point screening data revealed different substrate preferences for CYP98s of angiosperms and of bryophytes, lycopods and ferns (Figure 2.6, Figure 2.7). CYP98s utilized naturally occurring, but also synthetic substrates. Phenolic esters and phenolamides both were substrates of the CYP98s tested. Free *p*-coumaric acid was not converted by any of the CYP98s tested. Cinnamoyl esters and amides were not substrates of CYP98s. While *p*-coumaroyl-shikimate and *p*-coumaroyl-quinate were the preferred substrates of the angiosperm CYP98s, they were not or only to small extent converted by the enzymes from bryophyte, lycopod and fern. The substrate range of the *P. taeda* CYP98A19 (gymnosperm) was intermediate between the two groups. CYP98A19 of *P. taeda* utilized many different substrates. CYP98s of bryophyte and lycopod showed substrate preference for *p*-coumaroyl-anthranilate. A hierarchical clustering analysis of the data showed that CYP98s of the bryophyte, lycopod, fern, and the gymnosperm and angiosperms cluster together, respectively (Figure 2.7). Based on the substrate conversion profiles of each CYP98, pairwaise Pearson Correlation factors were calculated. In an average linkeage clustering the substrates with the highest Pearson Correlation factor were arranged close to each other, as indicated in the dendrogram. The CYP98s with the most similar substrate utilization profile, thus the highest Pearson Correlation factor between them, were as well arranged close to each other and distances are indicated as dendrogram.

1



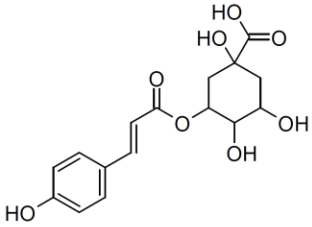
***p*-coumarate**

2



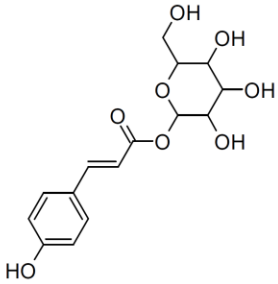
***p*-coumaroyl-shikimate**

3



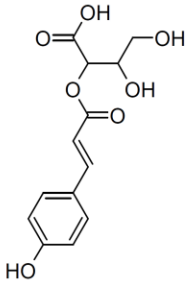
***p*-coumaroyl-quinate**

4



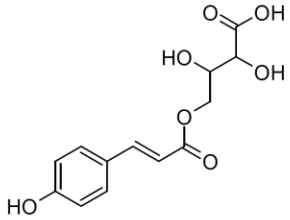
***p*-coumaroyl-β-D-glucose**

5



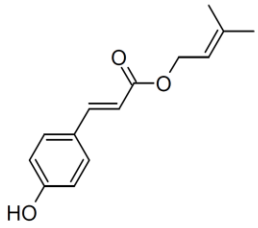
***p*-coumaroyl-2-threonate**

6



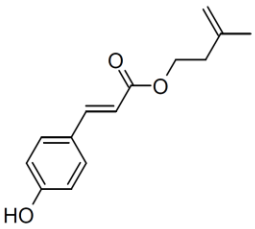
***p*-coumaroyl-4-threonate**

7



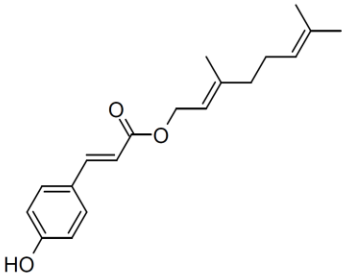
prenyl-*p*-coumarate

8



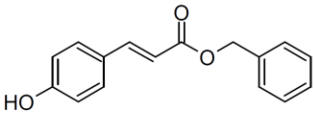
isoprenyl-*p*-coumarate

9



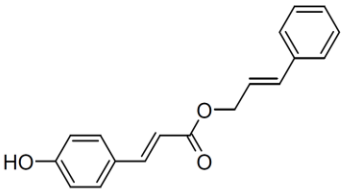
geranyl-*p*-coumarate

10



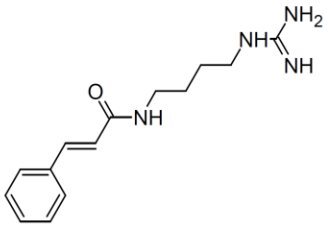
benzyl-*p*-coumarate

11



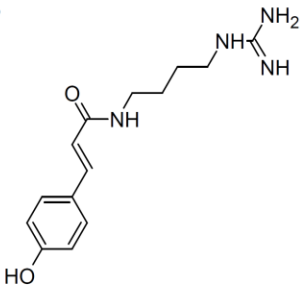
cinnamyl-*p*-coumarate (syn)

12



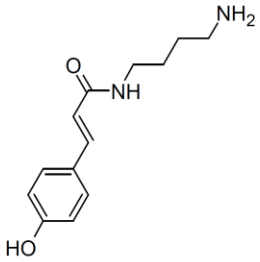
cinnamoyl-agmatine

13



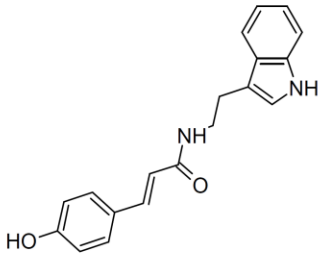
***p*-coumaroyl-agmatine**

14



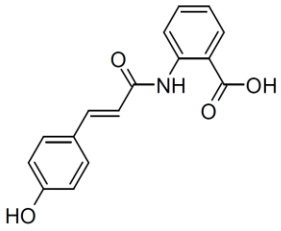
***p*-coumaroyl-putrescine**

15



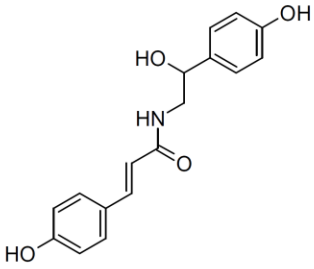
***p*-coumaroyl-tryptamine**

16



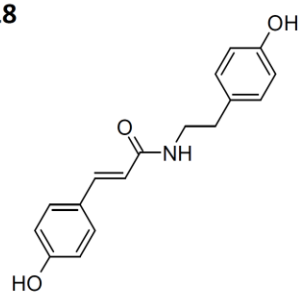
***p*-coumaroyl-anthranilate**

17



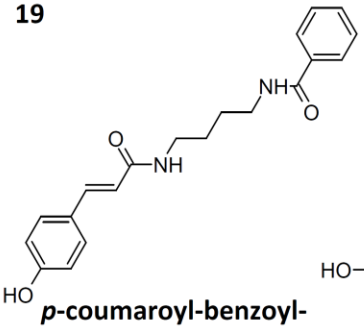
***p*-coumaroyl-octopamine**

18



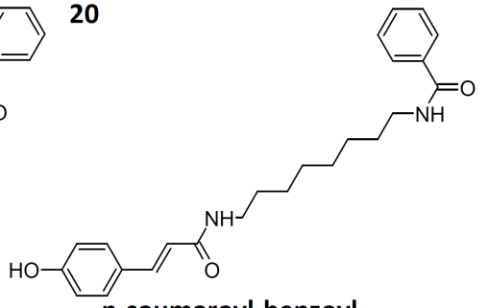
***p*-coumaroyl-tyramine**

19



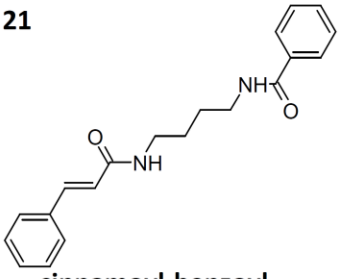
***p*-coumaroyl-benzoyl-putrescine (syn)**

20



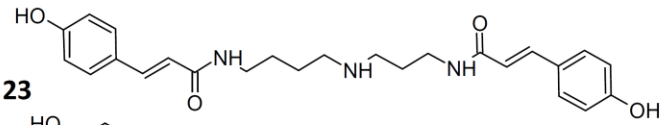
***p*-coumaroyl-benzoyl-octanediamine (syn)**

21



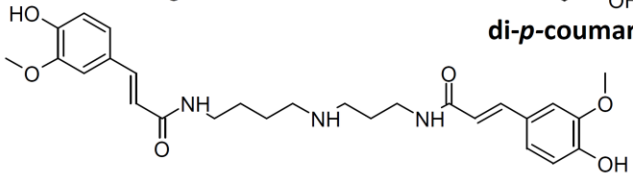
cinnamoyl-benzoyl-putrescine (syn)

22



di-*p*-coumaroyl-spermidine

23



di-feruloyl-spermidine

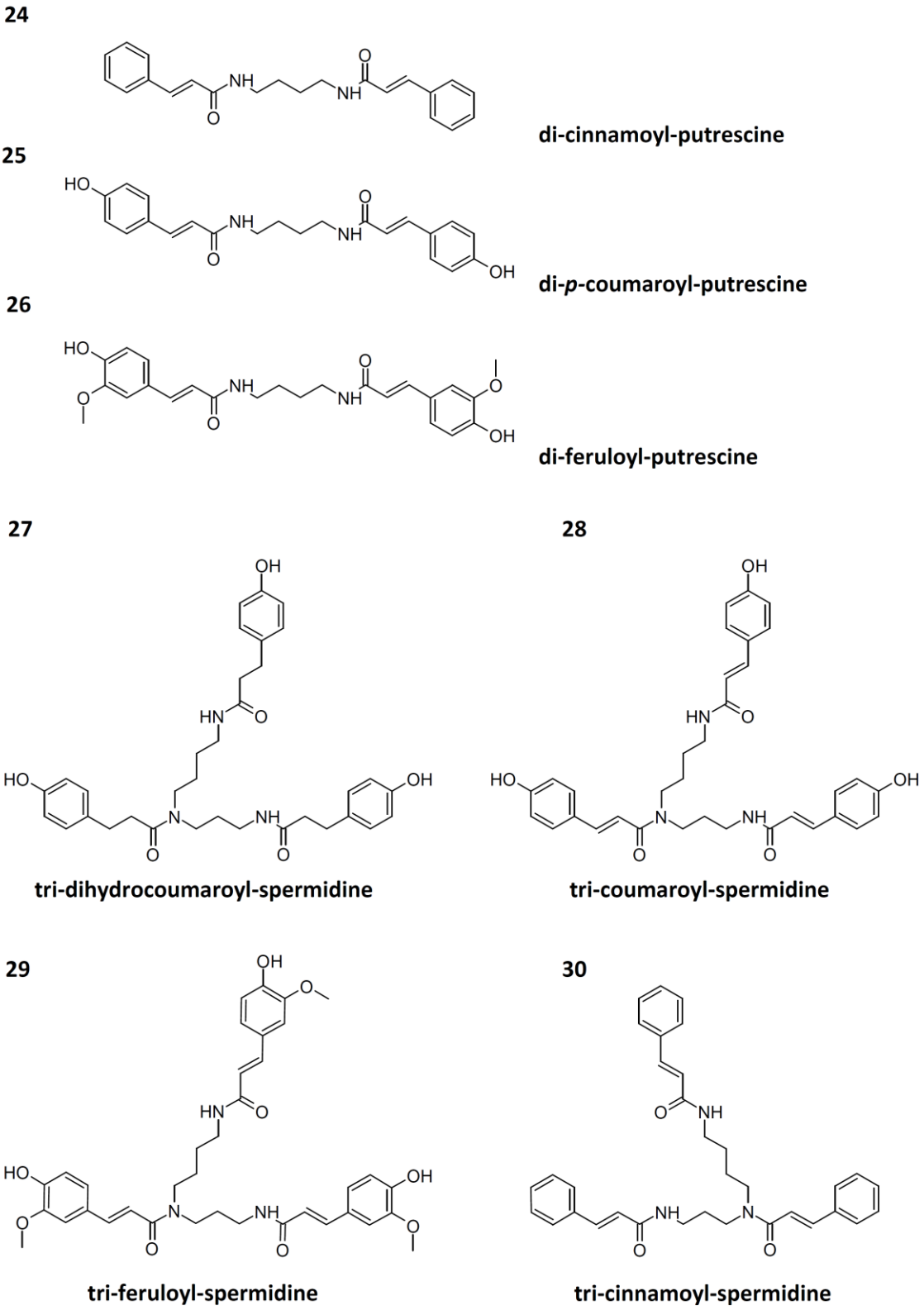


Figure 2.4 Chemical structures of substrates tested in the CYP98 end-point substrate screening.

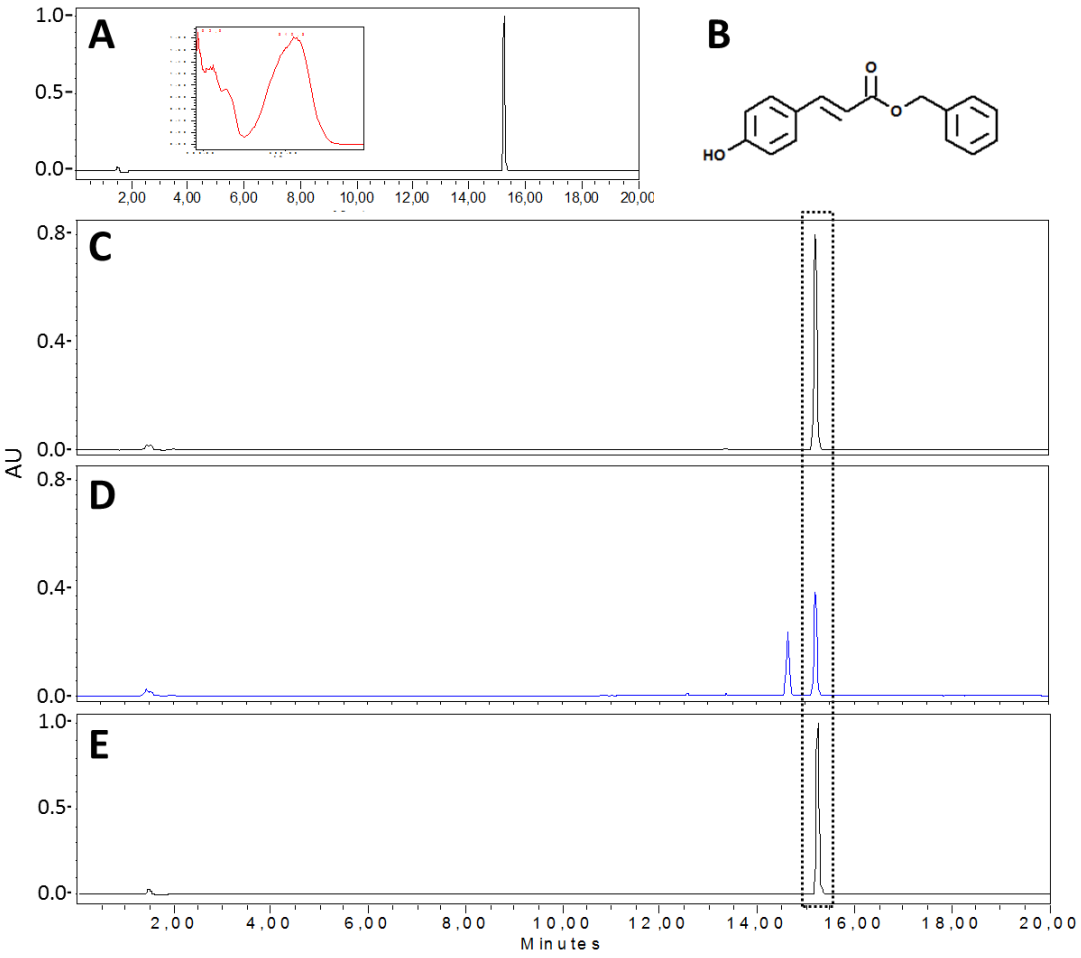
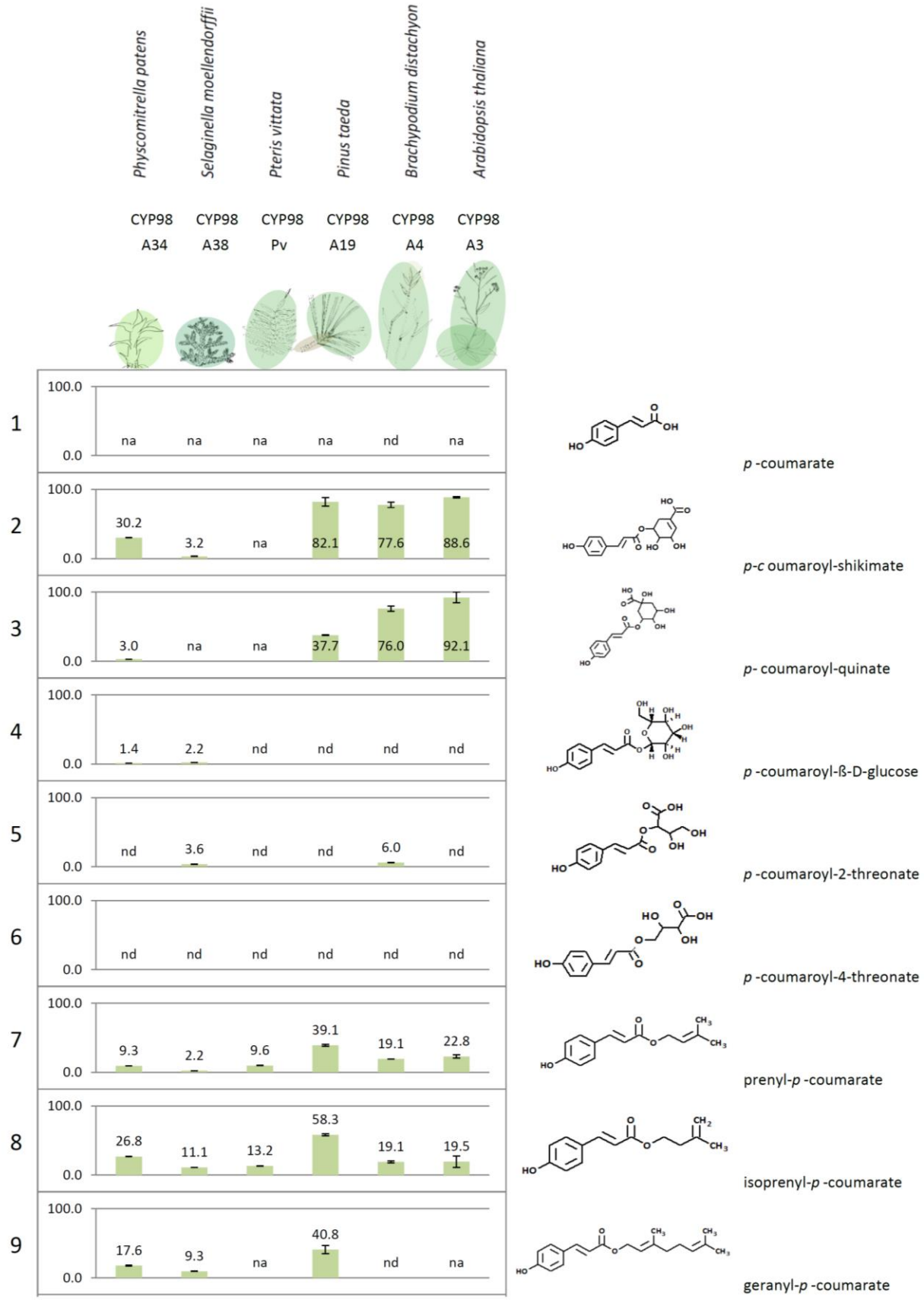
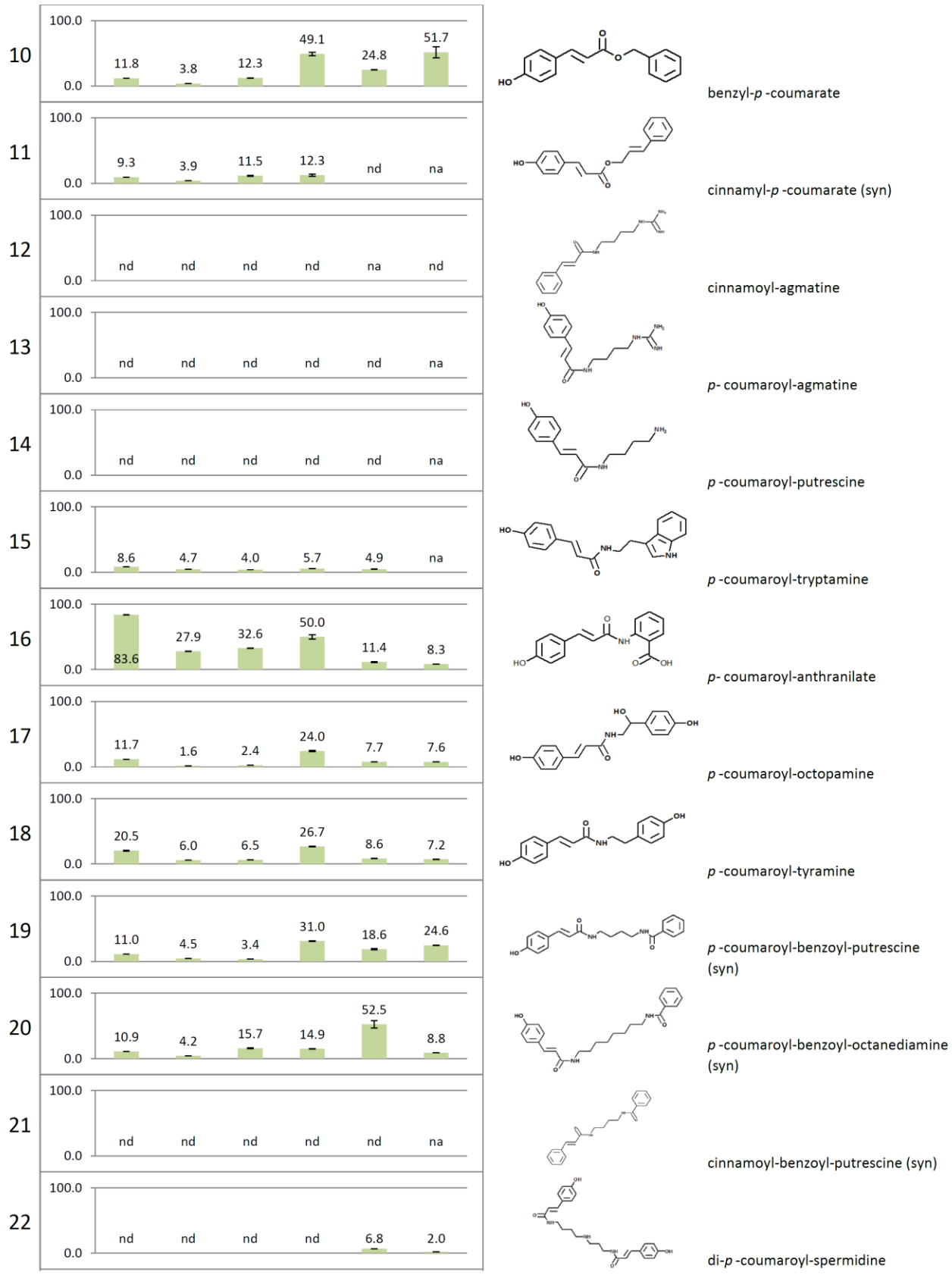


Figure 2.5 Incubation of *P. taeda* CYP98A19 with benzyl-*p*-coumarate (10) and analysis on HPLC/DAD.

In the end point screening assay, 10 pmol of P450 enzyme were incubated with 100µM of substrate (expected to be saturating) and 500µM NADPH. Reactions were run for 30 minutes under agitation, at 28°C, in the dark.

A: Chromatogram of the substrate benzyl-*p*-coumarate (10) and its corresponding UV spectrum when analysed on HPLC/DAD. **B**: Chemical structure of benzyl-*p*-coumarate (10). **C**: Control incubation of microsomes prepared from yeast transformed with the empty pYeDP60 transformation vector, with substrate and NADPH. The substrate peak on the chromatograms is framed by the dotted box. **D:** Incubation of CYP98A19 with benzyl-*p*-coumarate, with NADPH. **E**: Negative control of incubation of CYP98A19 with benzyl-*p*-coumarate: Incubation without NADPH.


Physcomitrella patens
Selaginella moellendorffii
Pteris vittata
Pinus taeda
Brachypodium distachyon
Arabidopsis thaliana
CYP98 A34
CYP98 A38
CYP98 Pv
CYP98 A19
CYP98 A4
CYP98 A3
1
100.0
0.0
na na na na nd na
p -coumarate
2
100.0
0.0
30.2 3.2 na 82.1 77.6 88.6
p-c oumaroyl-shikimate
3
100.0
0.0
3.0 na na 37.7 76.0 92.1
p- coumaroyl-quinate
4
100.0
0.0
1.4 2.2 nd nd nd nd
p -coumaroyl-ß-D-glucose
5
100.0
0.0
nd 3.6 nd nd 6.0 nd
p -coumaroyl-2-threonate
6
100.0
0.0
nd nd nd nd nd nd
p -coumaroyl-4-threonate
7
100.0
0.0
9.3 2.2 9.6 39.1 19.1 22.8
prenyl-p -coumarate
8
100.0
0.0
26.8 11.1 13.2 58.3 19.1 19.5
isoprenyl-p -coumarate
9
100.0
0.0
17.6 9.3 na 40.8 nd na
geranyl-p -coumarate


10
100.0
0.0
11.8
3.8
12.3
49.1
24.8
51.7
benzyl-p -coumarate
11
100.0
0.0
9.3
3.9
11.5
12.3
nd
na
cinnamyl-p -coumarate (syn)
12
100.0
0.0
nd
nd
nd
nd
na
nd
cinnamoyl-agmatine
13
100.0
0.0
nd
nd
nd
nd
nd
na
p- coumaroyl-agmatine
14
100.0
0.0
nd
nd
nd
nd
nd
na
p -coumaroyl-putrescine
15
100.0
0.0
8.6
4.7
4.0
5.7
4.9
na
p -coumaroyl-tryptamine
16
100.0
0.0
83.6
27.9
32.6
50.0
11.4
8.3
p- coumaroyl-anthranilate
17
100.0
0.0
11.7
1.6
2.4
24.0
7.7
7.6
p -coumaroyl-octopamine
18
100.0
0.0
20.5
6.0
6.5
26.7
8.6
7.2
p -coumaroyl-tyramine
19
100.0
0.0
11.0
4.5
3.4
31.0
18.6
24.6
p -coumaroyl-benzoyl-putrescine (syn)
20
100.0
0.0
10.9
4.2
15.7
14.9
52.5
8.8
p -coumaroyl-benzoyl-octanediamine (syn)
21
100.0
0.0
nd
nd
nd
nd
nd
na
cinnamoyl-benzoyl-putrescine (syn)
22
100.0
0.0
nd
nd
nd
nd
6.8
2.0
di-p -coumaroyl-spermidine

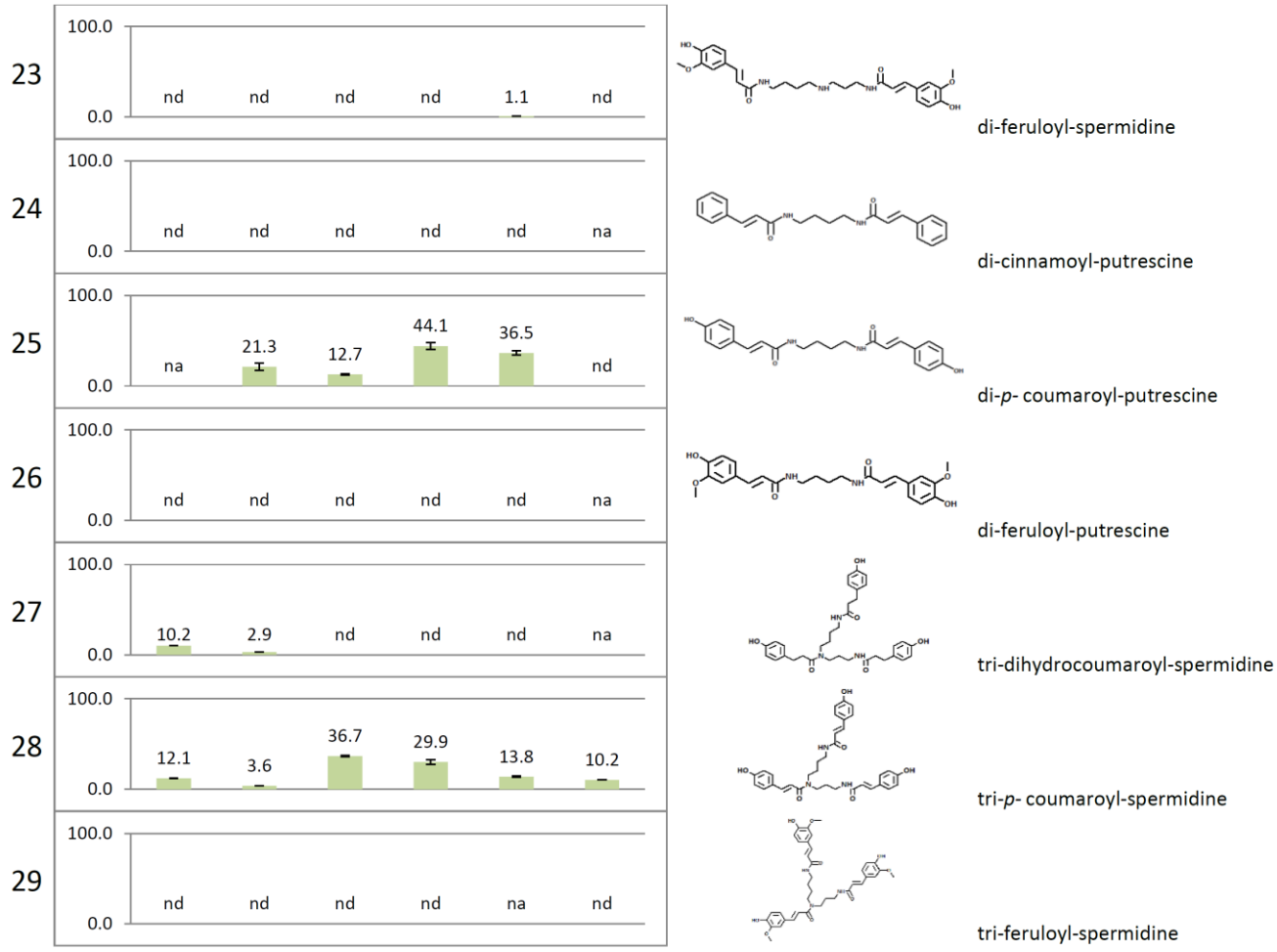


Figure 2.6 Substrate conversion rates obtained in end-point enzyme incubations.

Numbers of substrates correspond to Figure 2.4. Substrate structures and trivial names (if existing) are given on the right. No apparent conversion is indicated by na (no activity). Combinations that were not tested are indicated by nd (not determined). In the end point screening assay, 10 pmol of P450 enzyme were incubated with 100µM of substrate (expected to be saturating)and 500µM NADPH for 30 minutes under agitation at 28°C in the dark.

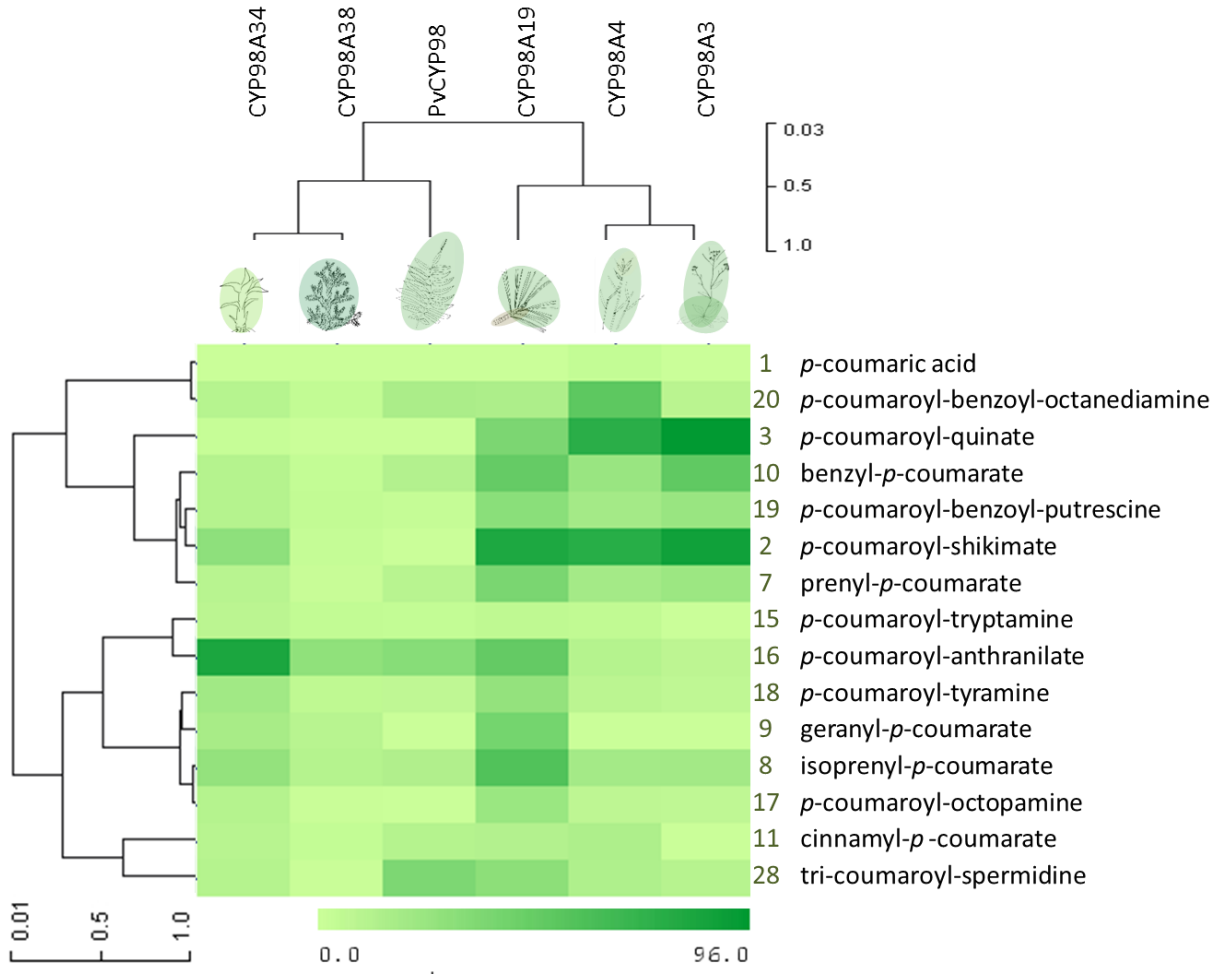


Figure 2.7 Hierarchical clustering of substrates and P450s tested in the substrate screening.

Average linkage clustering by Pearson Correlation.The corresponding substrate conversion rates from the end point screening assay are presented in detail in Figure 2.6. The end point screening was performed using 10 pmol P450 enzyme, incubated with 100 µM of substrate (expected to be saturating) and 500µM NADPH for 30 minutes under agitation, at 28°C, in the dark.

For CYP98A3 from *A. thaliana*, conversion of *p*-coumaroyl-shikimate (2; Numbers of substrates correspond to Figure 2.4) and coumaroyl-quinate (3) in our incubation over 30 minutes was close to complete. Free *p*-coumaric acid (1) was not converted to caffeic acid *in vitro. A. thaliana* CYP98A3 was able to convert prenyl- (7) and isoprenyl-*p*-coumarate (8) to about 20% of the initial amount of substrate. Conversion of benzyl-*p*-coumarate (10) was close to 50% and the phenolamides *p*-coumaroyl-anthranilate (16) and tri-coumaroyl-spermidine (28) were

converted to about 8% and 10% respectively. None of the remaining natural compounds were utilized to appreciable levels by *A. thaliana* CYP98A3.These results are consistent with previous biochemical characterizations of *A. thaliana* CYP98A3, which showed that CYP98A3 utilizes the shikimate ester (2) with high efficiency and the quinate ester (3) fairly well (Schoch et al., 2001). We extended the list of *A. thaliana* CYP98A3 substrates here but note that *p*-coumaroyl-shikimate and *p*-coumaroyl-quinate remain the best substrates for this lignin-related CYP98. Mutants of *A. thaliana CYP98A3* show a severe developmental phenotype, with a stem lignin that is almost devoid of G and S lignin (Franke et al., 2002; Abdulrazzak et al., 2006). The single CYP98A4 of the monocot *B. distachyon* was expected to be involved in lignin biosynthesis, as no other CYP98 family member exists in this species. Consistent with this expectation, enzymatic reactions showed a clear preference for *p*-coumaroyl-shikimate (2) and *p*-coumaroyl-quinate (3) as substrates *in vitro* with a conversion of more than 75% (Figure 2.6). Lower levels of substrate conversion were apparent with all other substrates tested: about 20% of benzyl-*p*-coumarate (10), prenyl-*p*-coumarate (7) and isoprenyl-*p*-coumarate (8) were converted and 37% and 14% of the phenolamides di-coumaroyl-putrescine (25) and tri-coumaroyl-spermidine (28), respectively. The conversion of other phenolamides such as *p*-coumaroyl-tyramine (18) and coumaroyl-octopamine (17) was less than 10%. Compared to other CYP98s tested, *B. distachyon* CYP98A4 showed unusually high activity with a synthetic compound (20), which was converted to about 50%. Overall, the two angiosperm CYP98s showed very similar substrate utilization profiles as determined by hierarchical cluster analysis (Figure 2.7).

Conifers typically do not to produce S-Lignin, but do produce a G-rich lignin and therefore also require 3-hydroxylation at the phenolic ring (Wagner et al., 2015). Most steps of the monolignol pathway have been characterized on the gene level in conifers (Wadenbäck et al., 2008; Wagner et al., 2009; Wagner et al., 2015). An involvement of CYP98 in phenolic 3-hydroxylation has been generally assumed, but has not experimentally been tested yet. In our enzymatic *in vitro* screen we investigated the single-copy CYP98A19 of *P. taeda* (Anterola, 2002). The conifer CYP98 converted *p*-coumaroyl-shikimate (2) to about 82%. *p*-Coumaroyl-quinate (3) was utilized less, to a level of about 37%. A variety of the substrates tested showed high conversion

rates when incubated with *P. taeda* CYP98A19: 40 – 60% of prenyl-*p*-coumarate (7), isoprenyl-*p*-coumarate (8), benzyl-*p*-coumarate (10), geranyl-*p*-coumarate (9), *p*-coumaroyl-anthranilate (16) and di-*p*-coumaroyl-putrescine (25) were converted. Compared to the angiosperm CYP98s tested, the conifer CYP98 showed a broader range of substrates accepted and also higher conversion rates of many of these substrates. These findings suggest that the single copy gene in conifers has broader functions. It may hydroxylate *p*-coumaroyl-shikimate (2) in the context of lignin biosynthesis, but has a broader substrate range, possibly to accommodate biosynthesis of soluble HCCs. Little is known about HCCs in conifers, but some species are known to accumulate chlorogenic acid or caffeoyl-shikimate (Radwan, 1975; Herrmann, 1978).

Comparing the *in vitro* enzymatic data of angiosperms and the conifer, the ancestor of both could have been an enzyme with *p*-coumaroyl-shikimate specificity. This would have been followed by substrate broadening of the enzyme in conifers (as exemplified in pine). Another possibility is that the ancestor of both had a broad substrate range realizing both lignin and soluble HCC synthesis and that consequently a specialization towards *p*-coumaroyl-shikimate, related to lignin biosynthesis, occurred after gene duplication in angiosperms only. It would be equally possible that the pine CYP98A19 is not involved in the biosynthesis of monolignols at all. However we cannot exclude the possibility that there might be additional *CYP98s* in conifer genomes not covered by the available transcriptome and genome data sets.

In an attempt towards distinguishing these possibilities we next investigated the enzymatic substrate utilization range of a fern CYP98, PvCYP98 of *P. vittata*. PvCYP98 utilized the phenolamides *p*-coumaroyl-anthranilate (16) with about 32% substrate conversion and tri-*p*-coumaroyl-spermidine (28) with about 36% of substrate conversion. The enzyme showed appreciable activity with some additional esters, but was devoid of any detectable activity with *p*-coumaroyl-shikimate (2) or *p*-coumaroyl-quinate (3) (Figure 2.6). This strikingly different substrate utilization profile of the single-copy PvCYP98 from *P. vittata* compared to those from *B. distachyon* and *A. thaliana* CYP98s strongly indicates that CYP98-mediated *p*-caffeoyl-shikimate biosynthesis is not part of the monolignol biosynthetic pathway in ferns, or at least not in *P. vittata*. Another ester or amide could be the intermediate in the monolignol

biosynthetic pathway if this step is CYP98-mediated. It is also possible that another distinct enzyme is responsible for the 3'-hydroxylation in the monolignol biosynthetic pathway of *P. vittata*. This holds true for the lycopod *S. moellendorffii*, which has been shown to possess a distinct enzyme, SmCYP788A1. Using an alternative route and not the canonical lignin biosynthesis path to perform 3'- and 5'-hydroxylation of the phenolic ring, *S. moellendorffii* is capable of making monolignols for S- and G-lignin (Weng et al., 2008b). A BLAST search with CYP788A1 as bait against fern transcriptome data revealed sequences with 38% amino acid identity in *P. vittata*. Reciprocal BLAST searches recovered back CYP788A1 (Phytozome ID 166299) from *S. moellendorffii*, but the most similar CYP from *A. thaliana* is a flavonoid 3'-monooxygenase (CYP75B1; AT5G07990) with a similarly low sequence identity of 38%, indicating that the fern sequence defines yet another CYP family. *P. vittata* does not produce S lignin (Weng et al., 2008b). It likely has another enzyme involved in lignin biosynthesis or uses another substrate that is converted by PvCYP98 leading to caffeoyl-units.

CYP98A38, the CYP98 of *S. moellendorffii,* was characterized next. In general, although CYP98A38 expressed well in yeast, indicating functional P450 enzyme, the highest rate of substrate conversion obtained was only 28%. Similar to the fern PvCYP98, CYP98A38 from *S. moellendorffii* showed almost no conversion of *p*-coumaroyl-shikimate (2) or *p*-coumaroyl-quinate (3) *in vitro*. The hydroxylation of isoprenyl-*p*-coumarate (8) yielded about 11% of substrate conversion. Other phenolic esters showed either no conversion or less than 10%. The best substrate utilized *in vitro* was *p*-coumaroyl-anthranilate (16), with about 28% of substrate conversion. Thus, the substrate utilization profile of the lycopod CYP98A38 is more similar to the fern CYP98 than to the angiosperm homologs. Especially the near absence of activity with *p*-coumaroyl-shikimate (2) in both lineages is consistent with a CYP98 independent pathway towards monolignols in both lineages. It is noteworthy that none of the substrates tested yielded conversion rates comparable to that of *p*-coumaroyl-shikimate (2) conversion by angiosperm CYP98s and thus we cannot exclude the possibility that a yet unidentified substrate is primarily utilized *in vivo* by fern and lycopod CYP98s. Based on these results, it is not parsimonious and thus not likely that hydroxylation of *p*-coumaroyl-shikimate has been the

ancestral function of CYP98s in vascular plants. A preference for *p*-coumaroyl-shikimate as substrate only starts to emerge within the gymnosperms and it is only in the angiosperms that *p*-coumaroyl-shikimate becomes the predominant substrate for CYP98 isoforms related to lignin biosynthesis.

The next step was to investigate the CYP98 of a non-vascular plant, the moss *P. patens*. CYP98A34 was capable of converting several substrates to appreciable amounts, but the by far best utilized substrate was again *p*-coumaroyl-anthranilate (16), with 84% of substrate conversion (Figure 2.6). The substrate preference of the moss, lycopod and fern CYP98s are similar based on hierarchical clustering (Figure 2.7), with *p*-coumaroyl-anthranilate (16) being the best of all tested substrates for the moss and lycopod CYP98s, and the second best for the fern CYP98 *in vitro* (highest activity with the fern CYP98 was measured with tri-*p*-coumaroyl-spermidine (28) with about 37% conversion, *p*-coumaroyl-anthranilate (16) with 33% conversion). Caffeoyl-anthranilates have been described in plants, for example as phytoalexins in oats (Ishihara et al., 1999a; Okazaki et al., 2004; Ahuja et al., 2012). Synthetically produced *N*-(3'4'- dimethoxycinnamoyl)-anthranilic acid, known as Tranilast™, is used as an antihistamine drug. Tranilast™ further shows anti-inflammatory and antiproliferative characteristics (Isaji et al., 1998; Rogosnitzky et al., 2012). Due to its beneficial molecule properties for humans, a metabolic engineering approach towards the production of hydroxycinnamoyl-anthranilates in *E. coli* has been made (Eudes et al., 2013). Coumaroyl-anthranilate is an attractive candidate for an ancestral CYP98 substrate, as anthranilate is an intermediate in tryptophan biosynthesis in primary metabolism and it is thus likely that *p*-coumaroyl-anthranilate was present in ancestral species. Like most phenolics, *p*-coumaroyl-anthranilate absorbs light in the UV range and a function in UV protection of early land plants is thus plausible.

Nevertheless these data have to be interpreted carefully, as the substrates investigated here obviously do not cover the complete breadth of possible substrates and likely not even all natural occurring substrate classes. The results are based solely on *in vitro* endpoint activity measurements, excluding the possibility to properly compare kinetic properties. The *A. thaliana* CYP98A3, for example, was capable of hydroxylating *p*-coumaroyl-quinate (3) to an equally high

extent as *p*-coumaroyl-shikimate (2) by the end-point in our *in vitro* assay, even though it has a lower catalytic efficiency with *p*-coumaroyl-quinate compared to *p*-coumaroyl-shikimate based on K_{cat}/K_M measurement (Schoch et al., 2001) and *A. thaliana* is not known to produce chlorogenic acid *in vivo*.

2.4.3. *In vivo* characterization of *CYP98* in the bryophyte *P. patens*

To probe deeper into the biological role of non-seed plant CYP98s, a reverse genetic approach was employed. This approach was aimed on the generation and analysis of a *P. patens CYP98A34* knock-out mutant. We focussed on *P. patens* because this was the only non-seed plant where such an approach was experimentally feasible. The lifecycle of *P. patens* alternates between two generations: the haploid gametophyte and the diploid sporophyte. Spores give rise to protonema, which is a filamentous structure with fast tip growth. Protonema can grow side-branches and buds. The protonema buds can subsequently develop into gametophores. *P. patens* gametophores show leaf-like structures, called phyllids.

A *CYP98A34* knock-out mutant of *P. patens* was generated by G. Wiedemann and H. Renault, taking advantage of the plant's natural high rate of homologous recombination. The selection cassette was inserted towards the 5' end of the gene, ensuring the functional P450 domains are deleted (Figure 2.8A). The putative ko-lines that were identified by screening according to antibiotic resistance were further verified by genotyping to validate the presence of the insert (Figure 2.8B). Three identified lines were further validated by RT-PCR to ensure the absence of the *CYP98A34* transcript (Figure 2.8C). To validate the single integration of the *nptII* cassette, copy numbers of both the 3' and 5' fragments used for homologous recombination were compared to wild type *CYP98A34* using quantitative PCR (Figure 2.8D).

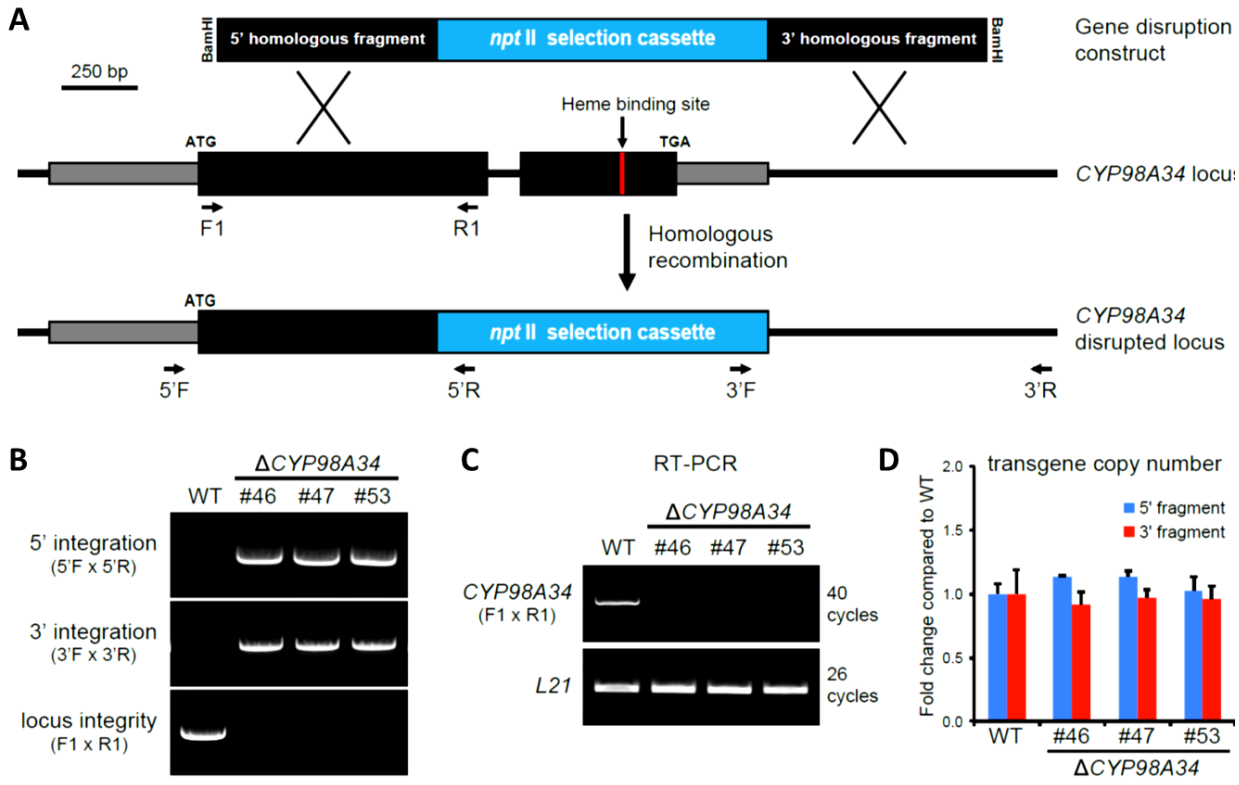


Figure 2.8 ***CYP98A34*** **knock-out construct and moss mutant validation.**

Generated by H. Renault. **A:** Homologous recombination-mediated strategy for *PpCYP98* gene disruption: a *PpCYP98* genomic fragment encompassing the critical heme-binding site was excised with simultaneous insertion of the *nptII* selection cassette conferring resistance to geneticin (G418). **B:** PCR validation of the correct integration of the construct in the *PpCYP98* genomic locus of the G418-selected lines. **C:** RT-PCR analysis of selected *PpCYP98* knock-out mutants confirming the absence of *PpCYP98* transcripts. **D:** qPCR-based evaluation of transgene copy number indicating a single integration event in the three selected mutant lines. Primer hybridization sites are indicated in (A). WT, wild type.

Compared to wildtype plants, knock-out mutants showed a severe developmental phenotype (Figure 2.9). Growth of the gametophore stage was impaired. Instead of gametophores with phyllids, knock-out mutant plants developed a more compact structure: phyllids initiate, but do not expand. As a result, pin-like stems developed, surrounded by “scale”-like phyllids, resembling a fused organ phenotype. Knock-out mutant plants at the protonema stage grown in liquid medium were compared to wild type and no obvious phenotypic difference was apparent.

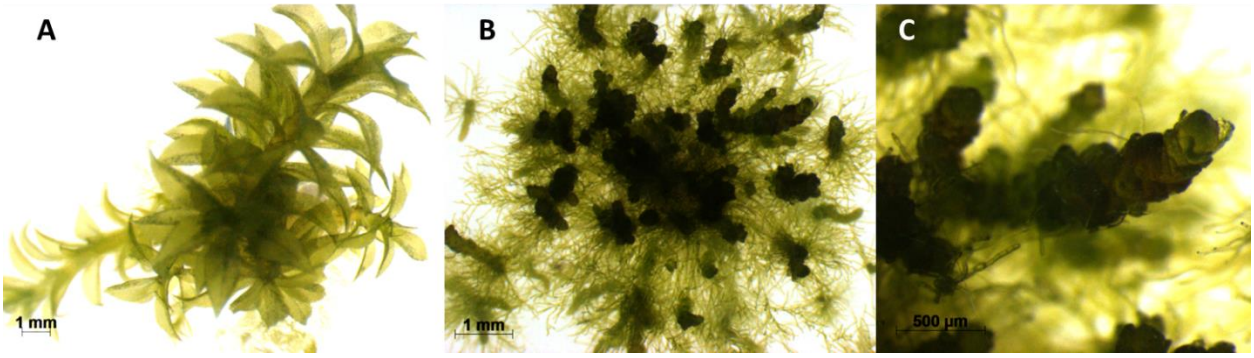


Figure 2.9 ***P. patens cyp98a34*** **mutant phenotype.**

Generated by H. Renault. **A:** 8 week old wild type gametophores. **B:** 8 week old *cyp98a34* knock-out mutant showing severe phenotype. Pin-like stems are surrounded by "scale"-like phyllids. **C:** Close up of the knock-out mutant phenotype, resembling known fused organ phenotypes.

An analysis of plant metabolite extracts of gametophores, grown on solid medium and of plants at the protonema stage, grown in liquid culture, showed distinct metabolic differences between wildtype and *cyp98a34* plants. Several peaks differed in peak area when comparing knock-out with wild type plant methanol extracts. *p*-Coumaroyl- or caffeoyl-anthranilate, the best substrate/product for CYP98A34 tested *in vitro*, was not among these peaks and was not detectable in extracts of either mutant or wild-type plants based on comparison with the authentic standards.

Instead, quantitatively the most important difference of mutant and wild type plant extracts was a near absence of two compound peaks in the knock-out plants which are highly abundant in wild type plant extracts (Figure 2.10a).

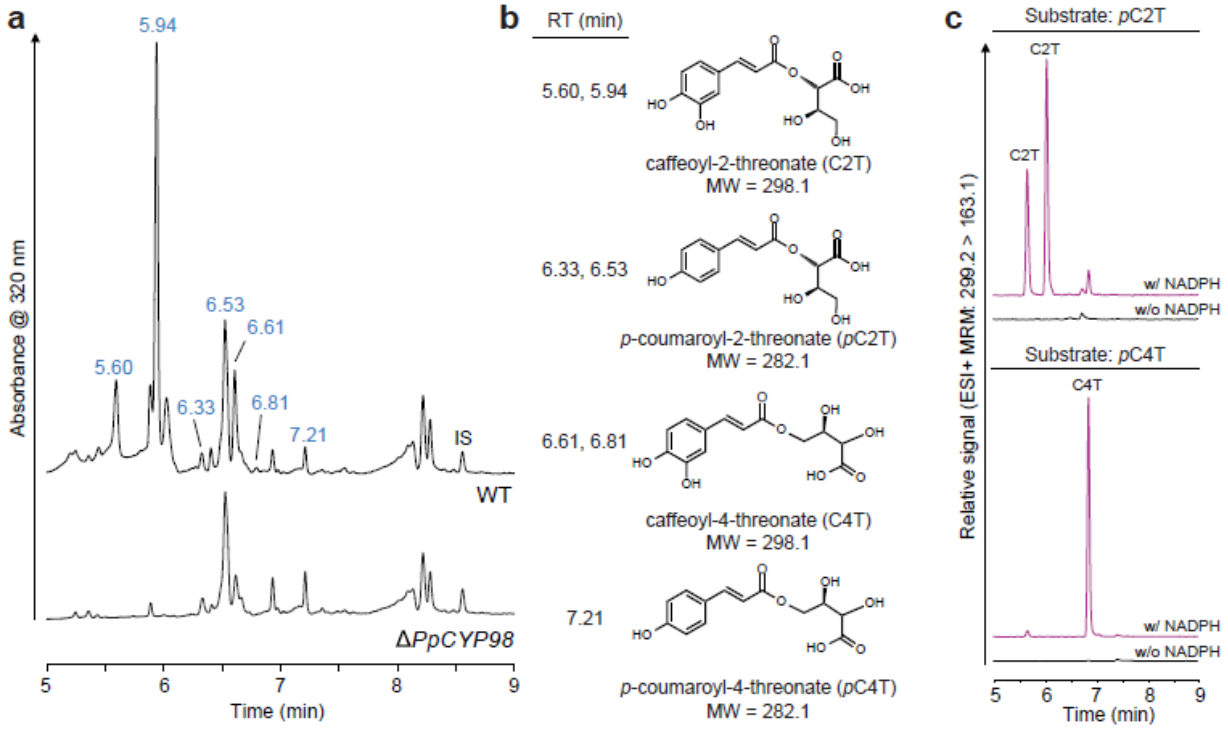


Figure 2.10 HPLC/DAD chromatogram of wild type *P. patens* gametophore extracts and *cyp98a34* knock-out gametophore extracts.

UPLC analysis by H. Renault. **a:** UV chromatogram showing the absence of major peaks in the *PpCYP98* mutant gametophore crude extract. IS, internal standard (morin). **b:** Names and structures of molecules at the indicated retention times (RT). **c:** PpCYP98-dependent conversion of *p*-coumaroyl-2-threonate (pC2T) and *p*-coumaroyl-4-threonate (pC4T) esters into corresponding caffeoyl threonate esters (C2T and C4T). Control reactions without NADPH were concurrently analyzed. Molecules were detected using dedicated multiple reaction monitoring (MRM) methods.

A UPLC-MS/MS analysis was subsequently performed by H. Renault, to determine the masses of the compounds of interest. Based on the mass spectrum and in comparison to published literature the peaks were tentatively identified as caffeoyl-threonic acid (Hahn and Nahrstedt, 1993; Richter et al., 2012; Kuczkowiak et al., 2014). Little is known about the function of caffeoyl-threonic acid in plants. It has been detected in a variety of plants such as *Fagus sylvatica*, *P. patens*, *Saniculiphyllum guangxiense*, *Miscanthus sacchariflorus*, *Miscanthus giganteus* and *Cornus controversa* (Lee et al., 1995; Richter et al., 2012; Parveen et al., 2013; Cadahía et al., 2014; Geng et al., 2014). Caffeoyl-threonic acid has been identified as substrate

of a polyphenol oxidase in orchard grass (*Dactylis glomerata*) (Parveen et al., 2008). Polyphenol oxidases (PPOs) are thought to be involved in plant defence (Constabel and Barbehenn, 2008). Thirteen PPOs have been identified in the *P. patens* genome (Tran et al., 2012).

The absence of caffeoyl-threonic acid in *cyp98a34* mutants immediately suggested *p*-coumaroyl-threonic acid as potential *in vivo* substrate for CYP98A34. Thus, *p*-coumaroyl-threonate was chemically synthesized (M. Schmitt and coll., UMR CNRS 7200) and used as a substrate for *in vitro* tests with recombinant yeast-expressed CYP98A34. CYP98A34 was able to hydroxylate *p*-coumaroyl-threonate, when analysed using dedicated MRM analysis. However, under standard photometric enzyme conditions, conversion was close to the background and only detectable when increased amounts of enzyme were used.

Because the comparison of *P. patens cyp98a34* mutant plant extract to wild type plant extract showed two peaks with the same mass in UPLC-MS/MS (Figure 2.10a), the possibility of *p*-coumaroyl- and caffeoyl-threonate isomers in the plant extracts (Figure 2.10b+c; Figure 2.11) was investigated.

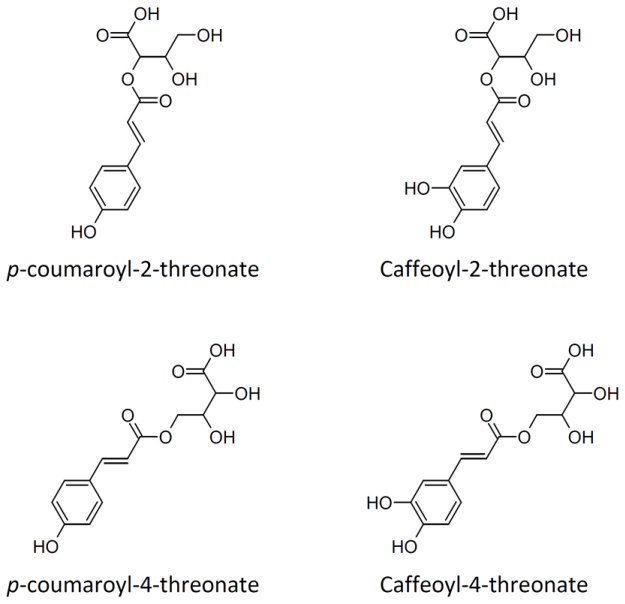
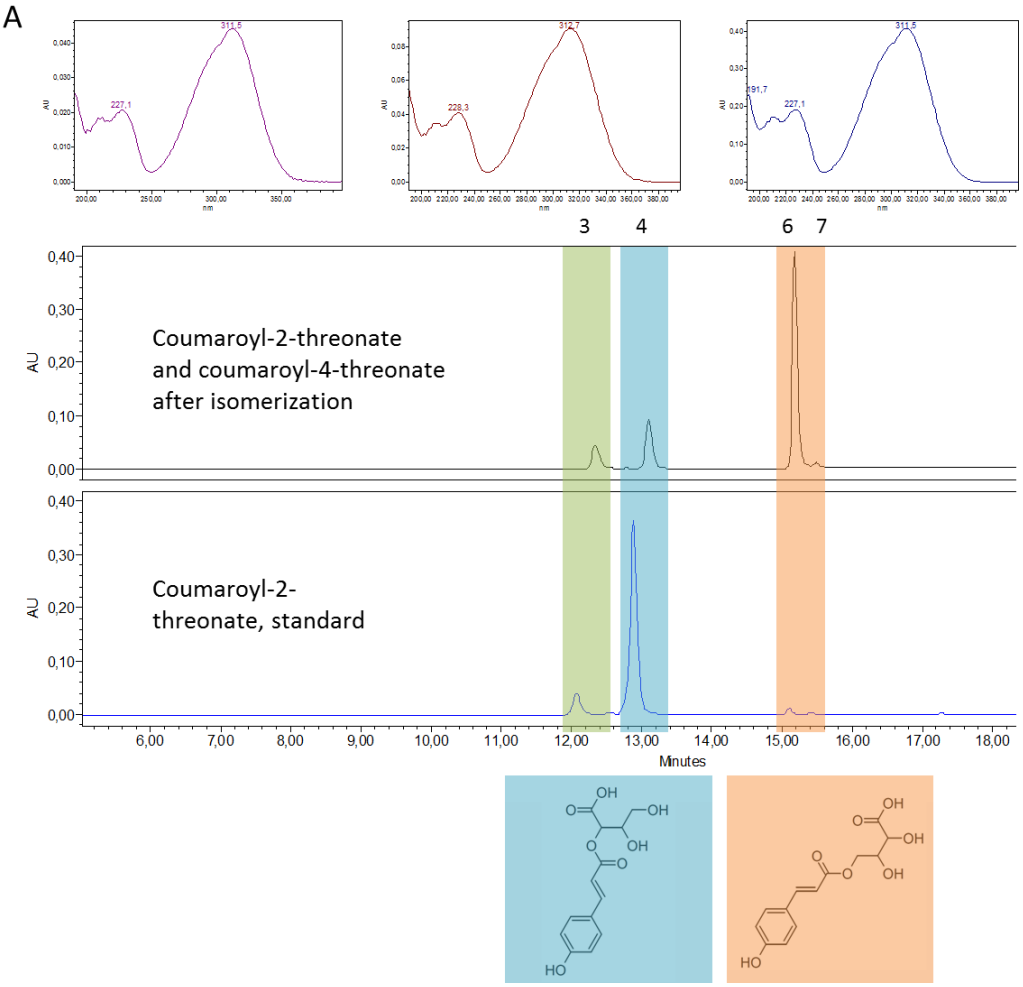


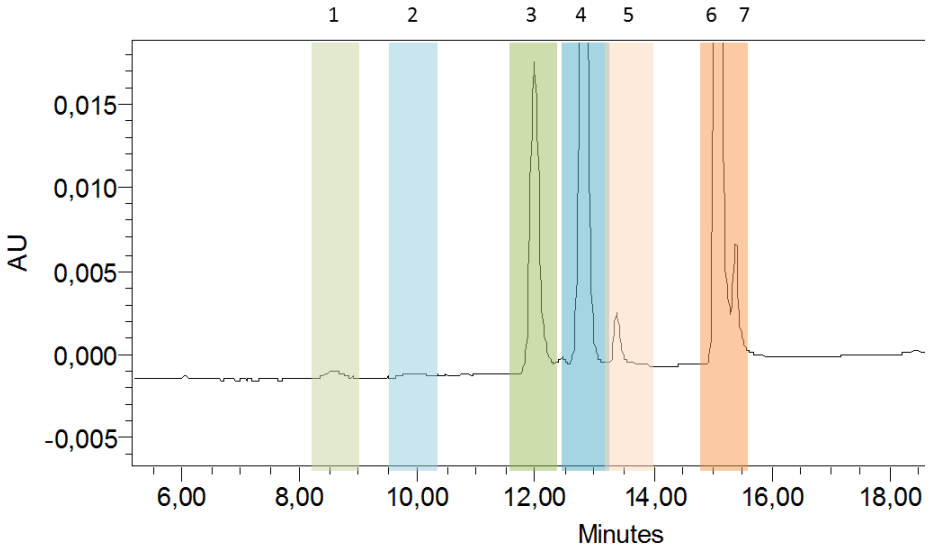
Figure 2.11 *p*-Coumaroyl-threonate and corresponding caffeoyl-threonate isomers.

The chemical structures of *p*-coumaroyl-2-threonate and *p*-coumaroyl-4-threonate are shown on the left. The corresponding caffeoyl-2-threonate and caffeoyl-4-threonate structures are shown on the right.

A



B Incubation of isomerized substrate with *P. patens* CYP98A34



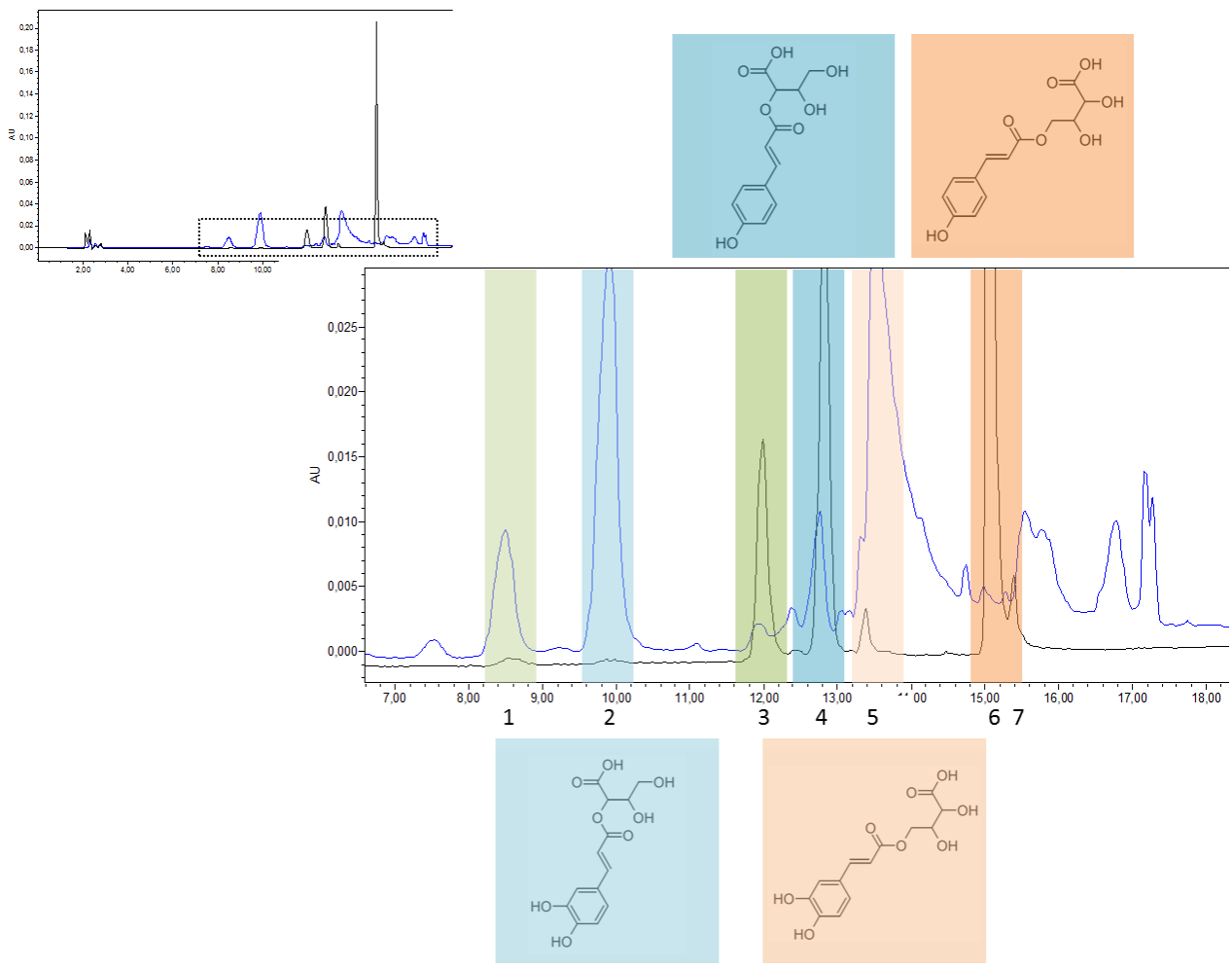


Figure 2.12 Isomerization of *p*-coumaroyl-2-threonate to obtain *p*-coumaroyl-4-threonate.

A: Coumaroyl-2-threonate and coumaroyl-4-threonate peaks after isomerization, analysed by HPLC/DAD. UV Spectra of the three peaks are shown above the chromatograms. **B:** The obtained isomers were incubated with yeast-expressed *P. patens* CYP98A34. Peaks 3, 4, 6, 7 are *p*-coumaroyl-threonate isomers (substrates), peaks 1, 2, 5 are caffeoyl-threonate isomers (products). **C:** A comparison with the products of incubation with a *P. patens* plant extract.

For this purpose, the *p*-coumaroyl-2-threonate substrate was isomerized and analysed on HPLC/DAD (Figure 2.12).

The subsequent NMR analysis verified the *p*-coumaroyl-4-threonate isomer. Further incubations with CYP98A34 and *p*-coumaroyl-4-threonate were performed *in vitro*, but did not result in higher levels of substrate hydroxylation compared to *p*-coumaroyl-2-threonate, in the end point experiment with fixed amounts of P450 enzyme. An incubation of the substrates with an increased enzyme concentration led to detectable peaks of caffeoyl-threonate isomers on HPLC/DAD (Figure 2.12B).

To validate the results and to determine which isomer(s) is (are) produced *in vivo*, we cloned and heterologously expressed the *P. patens* HCT homolog in *E. coli*. A single putative HCT is present in the *P. patens* genome, presumably synthesizing the precursor of the CYP98 mediated reaction. The His_6-tagged HCT protein was purified via Ni-affinity chromatography (protein production carried-out by the dedicated IBMP platform), and the native protein was incubated with *p*-coumaroyl-CoA and L-threonic acid. The 2-and 4-isomers of *p*-coumaroyl-threonate were formed *in vitro* (Figure 2.13; Figure 2.18) with the 4-isomer as main product. We next performed coupled enzymatic activity tests of purified *P. patens* HCT and yeast microsomes containing CYP98A34, incubated with *p*-coumaroyl-CoA and L-threonic acid as substrates. In these coupled enzymatic activity tests we obtained both the caffeoyl-2-threonate and caffeoyl-4-threonate isomers, in the same proportions as the HCT products (Figure 2.13). Together, this suggests that multiple *p*-coumaroyl-threonate isomers can be produced by the *P. patens* HCT and utilized by CYP98A34 and that the 4-isomer is preferentially formed *in vitro*. The hydroxylation of both isomers is still slower than the conversion of *p*-coumaroyl-anthranilate, but an incubation of *p*-coumaroyl-CoA and anthranilinic acid with the purified *P. patens* HCT did not result in the formation of *p*-coumaroyl-anthranilate.

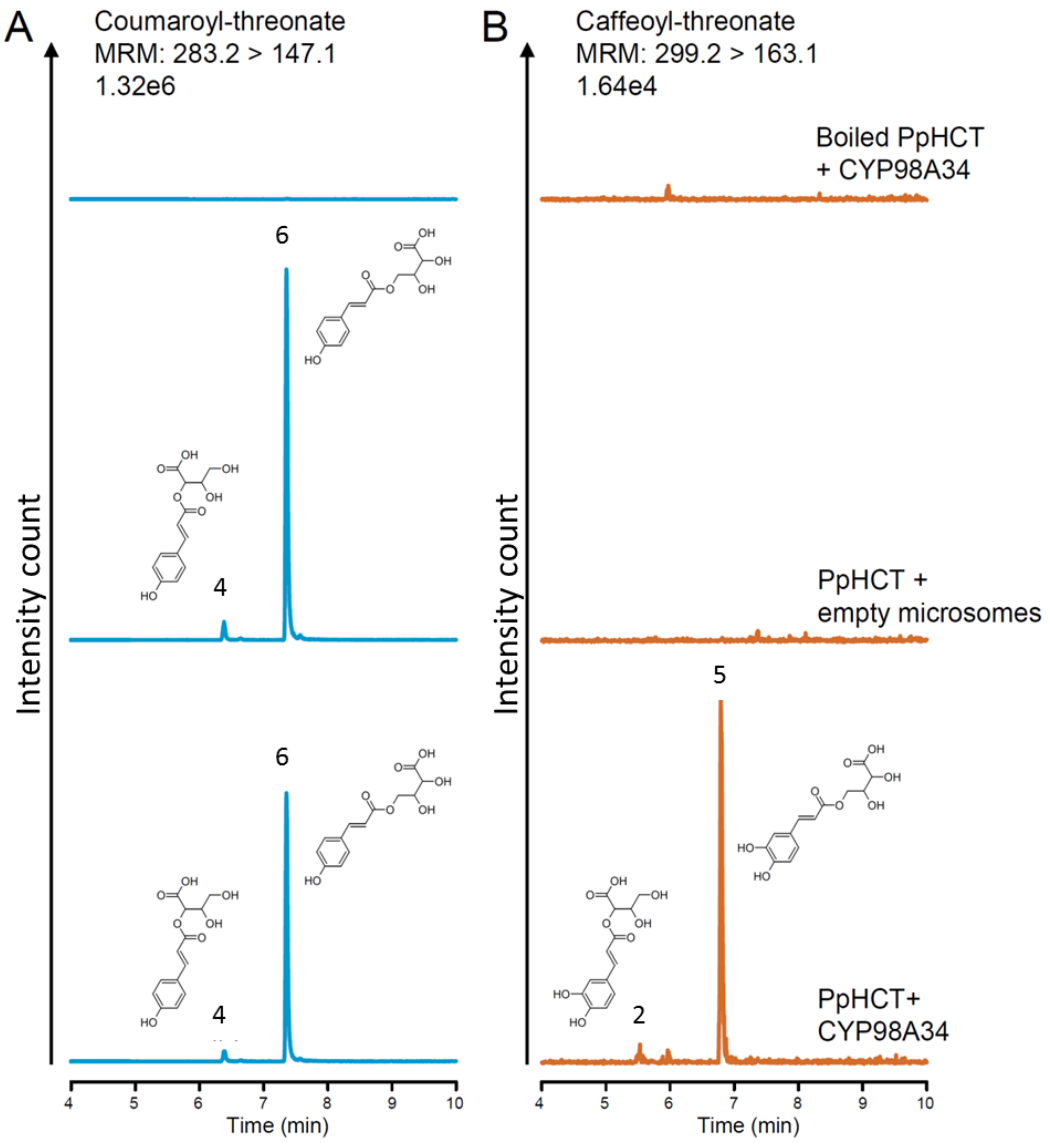


Figure 2.13 ***P. patens*** **HCT and CYP98A34 incubations.**

Incubations A. Alber, UPLC-MS/MS analysis and data analysis H. Renault.

An incubation of the *P. patens* HCT and CYP98A34 in a coupled reaction. Enzymes are incubated with *p*-coumaroyl-CoA and L-threonate. **A**: MRM targeted to detection of coumaroyl-threonate. **B**: MRM targeted to detection of caffeoyl-threonate. **Lower panels:** In the presence of both HCT and CYP98 coumaroyl-threonate and caffeoyl-threonate isomers are produced. **Middle panels:** HCT alone produces coumaroyl-threonate isomers but not caffeoyl-threonate. **Upper panels:** Incubation of CYP98 alone with coumaroyl-CoA and L-threonate does not lead to coumaroyl-threonate nor caffeoyl-threonate production. Peak labels correspond to Figure 2.12: *p*-coumaroyl-2-threonate (4) and *p*-coumaroyl-4-threonate (6). Caffeoyl-2-threonate (2) and caffeoyl-4-threonate (5).

Loss of CYP98A34 function in *P. patens* causes both a severe phyllid developmental phenotype and a lack of caffeoyl-threonate accumulation. However, it remains unclear if the lack of caffeoyl-threonate and the severe developmental phenotype of the *cyp98a34* mutant are directly causally related. Although caffeoyl-threonate accumulates to high levels in wild-type gametophores, it could also be an intermediate for other products in *P. patens.* The lack of these products may in turn be the cause of the developmental aberrations observed. Alternatively, diverted flow into alternative pathways branching upstream of CYP98 may lead to over-production of the causal agent.

The organ fusion phenotype observed on the gametophore is reminiscent of the phenotype of mutants of cutin biosynthesis (Buda et al., 2013). The composition of the cutin of *P. patens* was recently determined and is characterized by the presence of significant amounts of hydroxycinnamic units, including caffeate (Buda et al., 2013). To determine if an altered cutin composition is responsible for the *cyp98a34* knock-out mutant phenotype, an analysis of the cutin of the plant would need to be performed. Analogous to the situation in angiosperms, and consistent with the absence of activity with free coumaric acid observed for CYP98A34, free caffeic acid may be released from caffeoyl-threonate for integration into cutin. The action of a caffeoyl-shikimate-esterase (CSE; AT1G52760) has recently been described in *A. thaliana* (Vanholme et al., 2013). Using the *A. thaliana* CSE sequence as bait in a BLAST search against the *P. patens* genome, a gene similar to CSE was found (Pp3c19_14430; ~51% amino acid sequence identity to AT1G52760). Another possibility in this scenario would be the hydroxylation of coumaroyl-threonate, followed by transesterification directly or via caffeoyl-CoA to hydroxylated fatty acids during cutin biosynthesis (Rautengarten et al., 2012).

Alternatively, blocking CYP98 may cause an increased flux into other pathways branching off the phenylpropanoid pathway, prior to coumaroyl-esters. This increased flux in other pathways may result in the accumulation of compounds, which eventually cause the developmental phenotype of the knock-out mutant. One possible pathway the flux could be redirected to is the flavonoid pathway. *P. patens* is known to possess a rich repertoire of flavonoids (Jiang et al., 2006). This could be analogous to the situation described in *A. thaliana*, where an over-accumulation of flavonoids in the *cyp98a3* knock-out plant has been suggested to cause the

dwarf phenotype. However in *A. thaliana* it was also recently shown that the dwarf phenotype of the knock-out mutant can be uncoupled from the flavonoid over-accumulation phenotype (Li et al., 2010). Interestingly, also the lack of G- and S- lignin in *ref8* mutants can be uncoupled from the dwarf phenotype of the plants (Bonawitz et al., 2014), indicating that lignin composition is not causing the dwarf phenotype (although the rescued mutant produces higher amounts of H-lignin).

While *in vivo* data showed a lack of *p*-caffeoyl-threonate as the major metabolic *cyp98a34* phenotype, *in vitro* data suggests that metabolization of *p*-coumaroyl-threonate to produce caffeoyl-threonate is not optimal. It remains consequently possible that another compound, not detected by our analyses, may be the substrate of CYP98A34 *in vivo*. As a promiscuous enzyme, CYP98A34 might use this substrate and produce a transient product, precursor for caffeoyl-threonate, only in particular cells or tissues. A lack of this additional or locally restricted compound may be causing the developmental phenotype of the knock-out mutant.

The high amount of caffeoyl-threonate found in the wild type *P. patens* plant extracts could come from storage of caffeoyl-threonate. Hydroxycinnamic esters, such as chlorogenic acid, are usually stored in the central vacuole of plant cells (Hutzler, 1998). Storage could provide a reservoir of carbon, which could easily be accessed at times of stress. As described for flavonoids, caffeoyl-threonate storage could help the plant in scavenging H_2O_2, in a reaction involving hydroxycinnamic esters and a peroxidase (Yamasaki et al., 1997).

Further work is required to pinpoint the biological function of CYP98 in bryophytes. Both the *in vitro* and the *in vivo* data indicate that its role is not the biosynthesis of caffeoyl-shikimate as it is the case for lignin-related CYP98s in angiosperms. To further elaborate on the distinct biochemical and thereby implied broader biological roles of CYP98s in angiosperms and non-vascular plants, we next tested whether the *P. patens* CYP98A34 is capable of complementing the strong developmental phenotype caused by the lack of CYP98A3 in the angiosperm *A. thaliana*.

2.4.4. *CYP98A34* cannot complement the *cyp98a3* T-DNA knock-out mutant

In vitro enzymatic tests showed a low, but appreciable hydroxylation of *p*-coumaroyl-shikimate by CYP98A34 of *P. patens* (Figure 2.6). The severely dwarfed *cyp98a3* T-DNA insertion mutant of *A. thaliana* (Abdulrazzak et al., 2006) was used for a complementation assay. Heterozygous *A. thaliana* plants of this mutant were transformed with *CYP98A34* of *P. patens* driven by the promoter of the *A. thaliana* cinnamate-4-hydroxylase gene (*C4H*). Putative transgenic plants containing both the *P. patens* expression construct and the *cyp98a3* T-DNA were first identified through kanamycin and BASTA resistance conferred by the *cyp98a3* T-DNA and the *CYP98A34* overexpression constructs, respectively. Homozygosity of the T-DNA insertion was validated by PCR. The presence and transcription of the *P. patens CYP98A34* were validated using RT-PCR. After screening and genotyping our results showed that *CYP98A34* did not rescue the severe phenotype of the *cyp98a3* knock-out mutant. Homozygous *cyp98a3* plants, expressing the *P. patens* transgene, showed the same dwarf phenotype as non-complemented *cyp98a3* knock-out plants (Figure 2.14). In contrast, using the same promoter to drive the wild-type *A. thaliana CYP98A3* did complement the *cyp98a3* loss of function phenotype. While the *CYP98A34* transgene is transcribed (Figure 2.14D), future experiments need to be performed to demonstrate that the CYP98A34 protein is present and functional in the homozygous *cyp98a3* knock-out plants. Due to the dwarf phenotype of this mutant all further analyses require the growth and genotyping of large numbers of mutant plants. *In planta* tests for activity with *p*-coumaroyl-shikimate or *p*-coumaroyl-anthranilate could be performed as soon as enough plant material would be available for extraction. The conversion of *p*-coumaroyl-shikimate or *p*-coumaroyl-anthranilate could then also be tested in a whole plant extract.

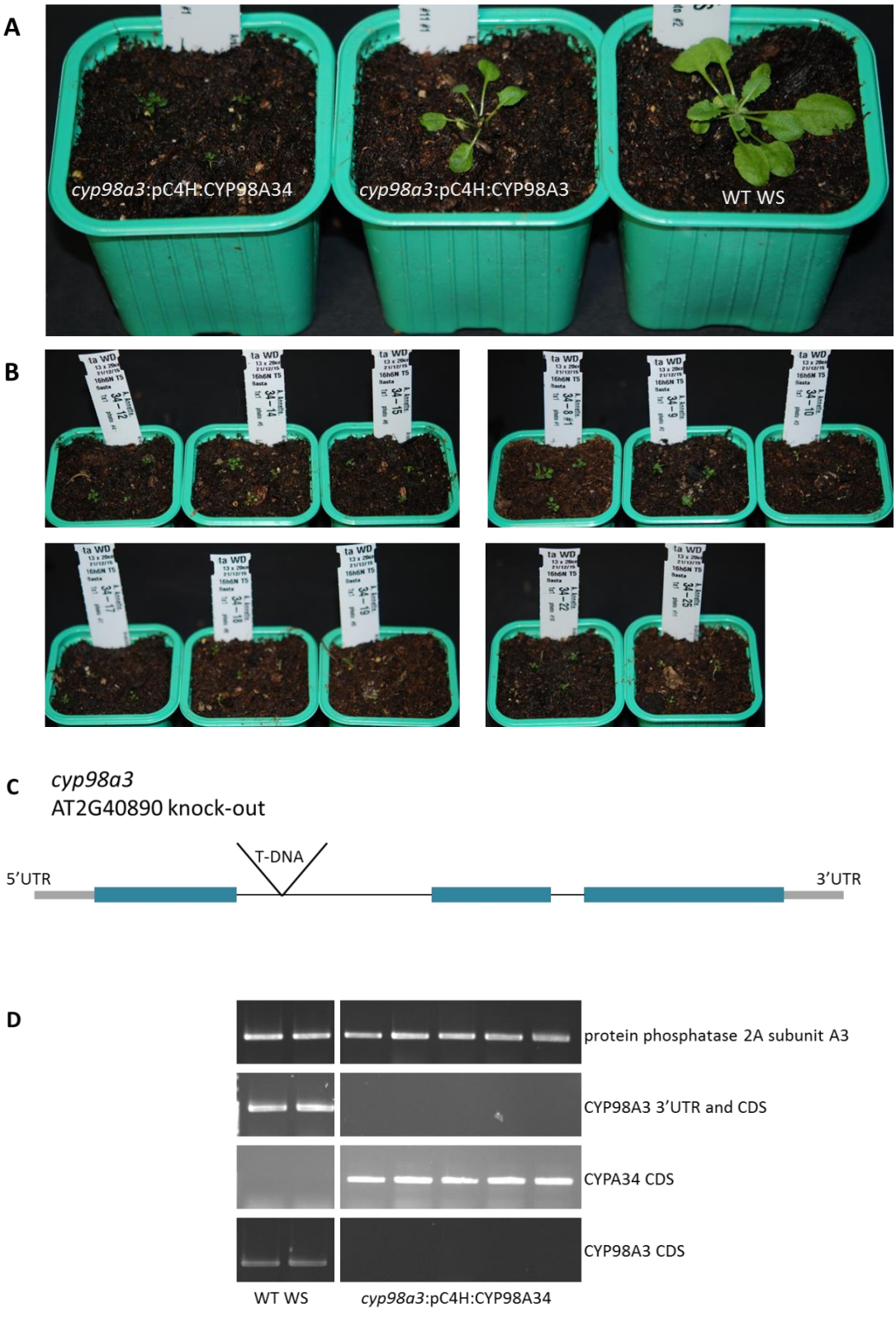


Figure 2.14 ***A. thaliana cyp98a3*** **mutant complementation by** ***P. patens CYP98A34*****.**

A: Close up of 4 weeks old *A. thaliana* plants. On the left a line which is homozygous for the T-DNA *cyp98a3* knock-out and expresses *CYP98A34*; in the middle, a homozygous *cyp98a3* knock-out plant, back-complemented with *CYP98A3*; on the right a wild type *A. thaliana* Wassilewskija plant. **B:** Eleven further lines of plants homozygous for the T-DNA *cyp98a3* knock-out and expressing *CYP98A34*. **C:** Schematic view of *CYP98A3* locus in *A. thaliana* and the location of T-DNA insertion to create the *cyp98a3* knock out. **D:** RT-PCR of WT *A. thaliana* and lines homozygous for the T-DNA *cyp98a3* knock-out and expressing *CYP98A34*.

2.5. Conclusion

Substrate specificity of CYP98 changed during land plant evolution. Isoforms that are specific for *p*-coumaroyl-shikimate appeared only in seed plants. The single copy CYP98 from the bryophyte *P. patens*, but also those from the lycopod *S. moellendorffii* and from the fern *P. vittata* essentially lack activity with *p*-coumaroyl-shikimate and instead appear to produce distinct caffeoyl-esters or -amides. Likewise *in vivo*, distinct, non-complementary functions of the moss and angiosperm CYP98s must be assumed, since the *P. patens CYP98A34* cannot complement the *cyp98a3* loss of function mutant in *A. thaliana*. Nevertheless, loss of function of *CYP98* in both angiosperms and in the bryophyte *P. patens* have severe, albeit distinct, developmental defects that go beyond the expectations of losing secondary metabolic activities only. Indeed, in *A. thaliana* the reduced lignin and also the enhanced flavonoid accumulation phenotype can be uncoupled from the developmental dwarf phenotype caused by *cyp98a3* loss of function (Li et al., 2010; Gallego-Giraldo et al., 2011; Kim et al., 2014), which indicates that it is not changes in major secondary metabolites (lignin or flavonoids) that are causing the dwarf phenotype. By analogy, the same may be true in *P. patens* were the lack of caffeoyl-threonate accumulation in the *cyp98a34* mutant may either be causal or coincidental with the developmental phenotype. In either case, the data presented here demonstrate a crucial role of CYP98s and 3,4-dihydroxylated HCCs in the development of both bryophytes and angiosperms, but also that distinct and non-complementary esters are produced in bryophytes and angiosperms to fulfil these developmental roles.

Ferns produce lignin extensively, but *P. vittata*'s single CYP98 does not show a similar substrate profile as seed plant CYP98s connected to lignin biosynthesis. In particular, no activity with *p*-coumaroyl-shikimate was detectable. It thus appears that ferns do not use caffeoyl-shikimate produced by CYP98 for lignin biosynthesis. The same is also evident for *S. moellendorffii*, whose CYP98A38 is also incapable of producing caffeoyl-shikimate. *S. moellendorffii* possesses a distinct enzyme, SmF5H / CYP788A1 (DN837863), which is capable of catalyzing the 3- and 5-hydroxylation steps on the aldehyde and alcohol level, contrary to what has been described in angiosperms (Weng et al., 2008a). This is consistent with a non-lignin-related function of the CYP98 from *S. moellendorffii*. The CYP98A38 substrate profile was more similar to that of the CYP98s from *P. patens* and *P. vittata* than it was to lignin-related angiosperm CYP98s from both *A. thaliana* and *B. distachyon*. Together this indicates that caffeoyl-shikimate specific CYP98s were likely recruited for lignin biosynthesis in seed plants. Gymnosperms contain a single CYP98, with CYP98A19 from *P. taeda* displaying fairly broad substrate range, but capable of synthesizing caffeoyl-shikimate *in vitro*. This could mean that it is involved both in the biosynthesis of lignin precursors, and in the biosynthesis of other soluble compounds. However, it also remains possible that even in gymnosperms CYP98s do not contribute to monolignol biosynthesis at all. In an experiment with cell suspension cultures, the addition of Phe only slightly upregulated the expression of *CYP98A19* and cinnamate 4-hydroxylase, contrary to other genes in the lignin biosynthetic pathway (Anterola, 2002). This was interpreted as the CYPs being rate-limiting steps and thus under distinct transcriptional control compared to the remainder of the phenylpropanoid pathway genes. However, *CYP98s* in angiosperms are tightly co-regulated with most other monolignol biosynthetic genes as shown through gene co-expression analysis in *A. thaliana*, poplar and rice (Ehlting et al., 2005; Hirano et al., 2012; Chen et al., 2014). In contrast, the single *CYP98* from the gymnosperm *Picea glauca* is notably absent from a monolignol biosynthesis pathway gene co-expression network (Porth et al., 2011). It thus appears that gymnosperm CYP98s are not only biochemically distinct from angiosperm, lignin related CYP98s, but that they are also under distinct transcriptional control. Only angiosperms possess several CYP98 copies and thus only in this group the CYP98 family may have subdivided roles via gene duplication.

Based on the phylogenetic analysis provided in this chapter, independent duplications of this enzyme family occurred only during angiosperm evolution. The molecular evolutionary history and functional divergence of gene duplicates will be further investigated and focused on in the following chapter.

2.6. Acknowledgements

We gratefully acknowledge funding from the Agency Nationale de la Recherche to the Phenowall project and funding from the University of Strasbourg Institute for Advanced Study and the Freiburg Institute for Advanced Studies to the MetabEvo project.

2.7. Contributions

Annette Alber: performed gene cloning and enzyme expression for yeast expression (except *B. distachyon* CYP98A4), enzyme incubations, coumaroyl-2-threonate substrate isomerization, *A. thaliana* mutant complementation assay, gene cloning for *P. patens* transformation, phylogenetic analyses, soluble enzyme expression and synthesis of coumaroyl-shikimate, and wrote the original manuscript. **Hugues Renault:** performed gene cloning and enzyme expression of *B. distachyon* CYP98A4, *P. patens* knock-out constructs and mutant generation, analysis of enzyme incubations on UPLC/MS-MS, analysis of *P. patens* mutant extracts on UPLC/MS/MS, supervised and coordinated moss work. **Alexandra Basilio Lopes Martine Schmitt** and **Frédéric Bihel:** performed the chemical synthesis of hydroxycinnamoyl conjugates. **Pascaline Ullmann:** provided assistance in soluble enzyme expression and biosynthesis of *p*-coumaroyl-shikimate, supervision of enzyme incubations. **Ralf Reski and Gertrud Wiedemann:** provided laboratory material and equipment, and technical support. **Jürgen Ehlting:** edited the manuscript, provided supervision, funding and laboratory material / equipment. **Danièle Werck-Reichhart:** edited the manuscript, provided supervision, funding and laboratory material / equipment.

2.8. Supplement

2.8.1. List of species included in the land plant phylogeny of Figure 2.1

Family	Species	Group
Andreaea	*rupestris*	Bryophytes
Anthoceros	*agrestis*	Bryophytes
Atrichum	*angustatum*	Bryophytes
Ceratodon	*purpureus*	Bryophytes
Diphyscium	*foliosum*	Bryophytes
Encalypta	*streptocarpa*	Bryophytes
Hedwigia	*ciliata*	Bryophytes
Leucobryum	*glaucum*	Bryophytes
Megaceros	*vincenti*	Bryophytes
Neckera	*douglasii*	Bryophytes
Paraphymatoceros	*hallii*	Bryophytes
Phaeoceros	*carolinianus*	Bryophytes
Phaeoceros	*carolini*	Bryophytes
Phaeomegaceros	*coriaceus*	Bryophytes
Phagnum	*fallax*	Bryophytes
Physcomitrella	*patens*	Bryophytes
Rhynochstegium	*serrulatum*	Bryophytes
Scouleria	*aquatica*	Bryophytes
Takakia	*lepidozioides*	Bryophytes
Tetraphis	*pellucida*	Bryophytes
Thuidium	*delicatum*	Bryophytes
Timmia	*austriaca*	Bryophytes
Dendrolycopodium	*obscurum*	Lycopods
Diphasiastrum	*digitatum*	Lycopods
Huperzia	*selago*	Lycopods
Huperzia	*lucidula*	Lycopods
Huperzia	*myrisinites*	Lycopods
Huperzia	*squarrosa*	Lycopods
Isoetes	*tegetiformans*	Lycopods
Lycopodiella	*apressa*	Lycopods
Lycopodium	*annotinum*	Lycopods
Lycopodium	*deuterodensum*	Lycopods
Phylloglossum	*drummondii*	Lycopods
Pseudolycopodiella	*caroliniana*	Lycopods
Selaginella	*stauntoniana*	Lycopods
Selaginella	*moellendorffii*	Lycopods

Selaginella	*wildenowii*	Lycopods
Selaginella	*apoda*	Lycopods
Selaginella	*wallacei*	Lycopods
Adiantum	*tenerum*	Ferns
Adiantum	*aleuticum*	Ferns
Argyrochosma	*nivea*	Ferns
Athyrium	*filix femina*	Ferns
Blechnum	*spicant*	Ferns
Botrypus	*virigianus*	Ferns
Ceratopteris	*thalictroides*	Ferns
Cheilanthes	*arizonica*	Ferns
Cibotium	*glaucum*	Ferns
Cystopteris	*protrusa*	Ferns
Danaea	*sp*	Ferns
Davallia	*fejeensis*	Ferns
Deparia	*lobato-crenata*	Ferns
Diplazium	*wichurae*	Ferns
Equisetum	*diffusum*	Ferns
Equisetum	*hymale*	Ferns
Gymnocarpium	*dryopteris*	Ferns
Hymenophyllum	*bivalve*	Ferns
Marattia	*attenuata*	Ferns
Myriopteris	*eatonii*	Ferns
Notholaena	*montieliae*	Ferns
Osmundastrum	*cinnamomeum*	Ferns
Psilotum	*nudum*	Ferns
Pteris	*vittata*	Ferns
Tmesipteris	*parva*	Ferns
Woodsia	*scopulina*	Ferns
Woodsia	*ilvensis*	Ferns
Abies	*lasiocarpa*	Gymnosperms
Acmopyle	*pancheri*	Gymnosperms
Arucaria	*sp*	Gymnosperms
Cathaya	*agryrophylla*	Gymnosperms
Cedrus	*libani*	Gymnosperms
Cupressus	*dupreziana*	Gymnosperms
Cycas	*micholitzii*	Gymnosperms
Encephalartos	*barteri*	Gymnosperms
Ephedra	*sinica*	Gymnosperms
Ginkgo	*biloba*	Gymnosperms
Juniperus	*scopulorum*	Gymnosperms
Keteleeria	*evelyniana*	Gymnosperms

Larix	*speciosa*	Gymnosperms
Metasequoia	*glyptostroboides*	Gymnosperms
Nothotsuga	*longibracteata*	Gymnosperms
Picea	*abies*	Gymnosperms
Picea	*engelmanii*	Gymnosperms
Pinus	*jeffreyi*	Gymnosperms
Pinus	*radiata*	Gymnosperms
Pinus	*taeda*	Gymnosperms
Podocarpus	*coriaceus*	Gymnosperms
Podocarpus	*rubens*	Gymnosperms
Pseudolarix	*amabilis*	Gymnosperms
Pseudotaxus	*chienii*	Gymnosperms
Pseudotsuga	*menziesii*	Gymnosperms
Retrophyllum	*minus*	Gymnosperms
Stangeria	*eriopus*	Gymnosperms
Taxus	*baccata*	Gymnosperms
Thuja	*plicata*	Gymnosperms
Welwitschia	*mirabilis*	Gymnosperms
Wollemia	*nobilis*	Gymnosperms
Arabidopsis	*thaliana*	Angiosperm eudicot
Brachypodium	*distachyon*	Angiosperm monocot
Coffea	*canephora*	Angiosperm eudicot
Cynara	*cardunculus*	Angiosperm eudicot
Eucalyptus	*grandis*	Angiosperm eudicot
Heliothropum	*greggii*	Angiosperm eudicot
Ilex	*vomitoria*	Angiosperm eudicot
Lithospermum	*erythrorhizon*	Angiosperm eudicot
Lonicera	*japonica*	Angiosperm eudicot
Musa	*acuminata*	Angiosperm monocot
Nicotiana	*tabacum*	Angiosperm eudicot
Ocimum	*basilicum*	Angiosperm eudicot
Populus	*trichocarpa*	Angiosperm eudicot
Ruta	*graveolens*	Angiosperm eudicot
Solenostemon	*scutellarioides*	Angiosperm eudicot
Trifolium	*pratense*	Angiosperm eudicot
Triticum	*aestivum*	Angiosperm monocot
Vitis	*vinifera*	Angiosperm eudicot

Table 2.1 List of speciesincluded in the land plant phylogeny Figure 2.1

2.8.2. Table of primers used in the work presented in this chapter

Species	Gene identifier	Primer sequence	Primer purpose
P. patens	Pp1s22_138V6	GGGGACAAGTTTGTACAAAAAAGCAGGC TTCATGGCCGCCGCAAGTCAAG	Cloning *P. patens* HCT for Gateway™ entry
P. patens	Pp1s22_138V6	GGGGACCACTTTGTACAAGAAAGCTGGGT CTTAGAAGGATGCCACTAGTTTGG	Cloning *P. patens* HCT for Gateway™ entry
P. patens	Pp1s22_138V6	ATGGCCGCCGCAAGTCAAG	Cloning *P. patens* HCT
P. patens	Pp1s22_138V6	TTAGAAGGATGCCACTAGTTTGG	Cloning *P. patens* HCT
A. thaliana	AT2G40890	CCGATCGTCGGTAACCTCTA	genotyping fw
A. thaliana	AT2G40890	AAATGCTGTTTCGCTCCACT	genotyping rv
T-DNA insertion *CYP98A3*	-	TTGCTTTCGCCTATAAATACGACGGATCG	genotyping fw
T-DNA insertion *CYP98A3*	-	AAATGCTGTTTCGCTCCACT	genotyping rv
A. trichopoda	CYP98A84	GGATCCATGGACTTTCTCTCTCCACTCTC	TA cloning includes bamh1
A. trichopoda	CYP98A84	GGTACCTCACATTTGTGTGGGCACAC	TA cloning includes kpn1
P. patens	CYP98A34	GGATCCATGGCAGTCATGTGGGAGA	TA cloning includes bamh1
P. patens	CYP98A34	GGTACCTCACGAAGGGGATGATCC	TA cloning includes kpn1
P. patens	CYP98A34 exon 1	ATGGCAGTCATGTGGGAGA	exon 1 for knock-out generation
P. patens	CYP98A34 exon 1	GGTACCCTGATCCAGCTCTTGTTGTGC	exon 1 for knock-out generation. contains kpn1
	nptII	CGGAATTCaagcttgcatgcctgca	nptII selection cassette and EcoRI restriction site
	nptII	CGGAATTCcccagtcacgacgttgtaaa	nptII selection cassette and EcoRI restriction site
A. trichopoda	CYP98A85	GGATCCATGGAGTCTCTCTTCCTACTTGC	TA cloning includes bamh1
A. trichopoda	CYP98A85	GGTACCTCACACTTTCATGGACTGACA	TA cloning includes kpn1
P. vittata	PvCYP98	ATGGCAGAAATGCTAACTGGA	cloning
P. vittata	PvCYP98	TCATTTTGTACTTCTGGCTTCTCTT	cloning
P. vittata	PvCYP98	GGCTTAAUATGGCAGAAATGCTAACTGGA	USER™ cloning
P. vittata	PvCYP98	GGTTTAAUTCATTTTGTACTTCTGGCTTCTCT T	USER™ cloning

P. taeda	CYP98A19	ATGTCTGTTCCTGAAATGGGTC	cloning
P. taeda	CYP98A19	TCAATTGAGTGGTTGTCGCT	cloning
P. taeda	CYP98A19	GGCTTAAUATGTCTGTTCCTGAAATGGGTC	cloning
P. taeda	CYP98A19	GGTTTAAUTCAATTGAGTGGTTGTCGCT	cloning

Table 2.2 Primers used in the experiments described.

2.8.3. Purification of *A. thaliana* 4CL1 and *Nicotiana tabacum* HCT

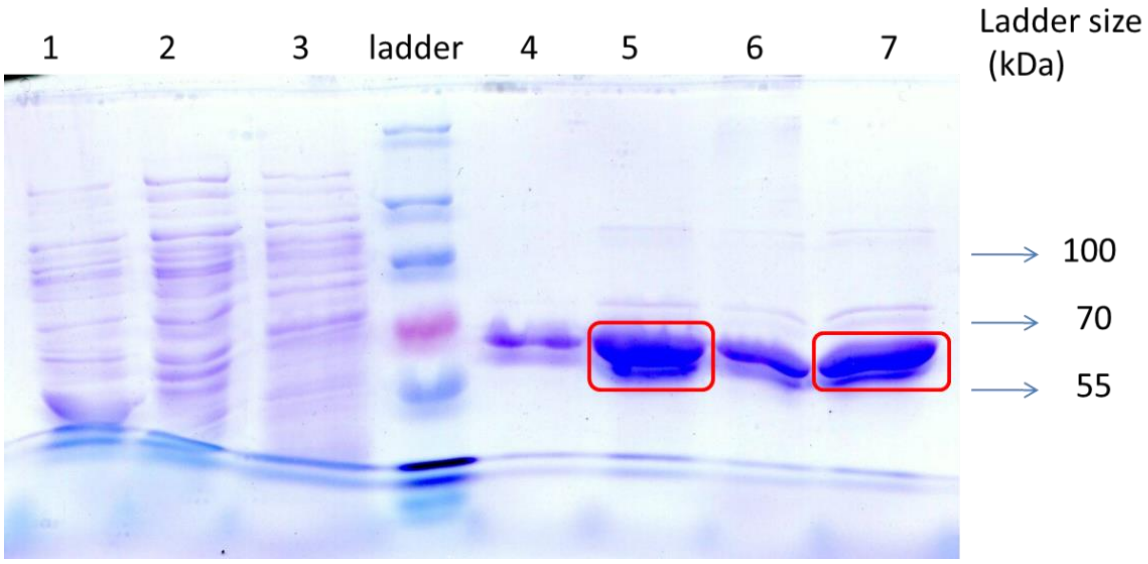


Figure 2.15 Purification of *A. thaliana* 4CL1.

SDS PAGE analysis of different purification steps of the expression of *A. thaliana* 4Cl1 in the bacterial strain BL21-G612. 1;2;3 on a Ni column, 4;5;6;7 on a Sephadex column. 1: Elution buffer, 2: Flow through, 3: Wash. 4: Flow through 1; 5: Elution 1, containing 4CL1 at a size of ~ 66 kDa; 6: Flow through 2; 7: Elution 2, containing 4 CL1 at a size of ~ 66 kDa. SDS PAGE gel stained with coomassie blue, PageRuler prestained protein ladder plus (Fermentas SM1811).

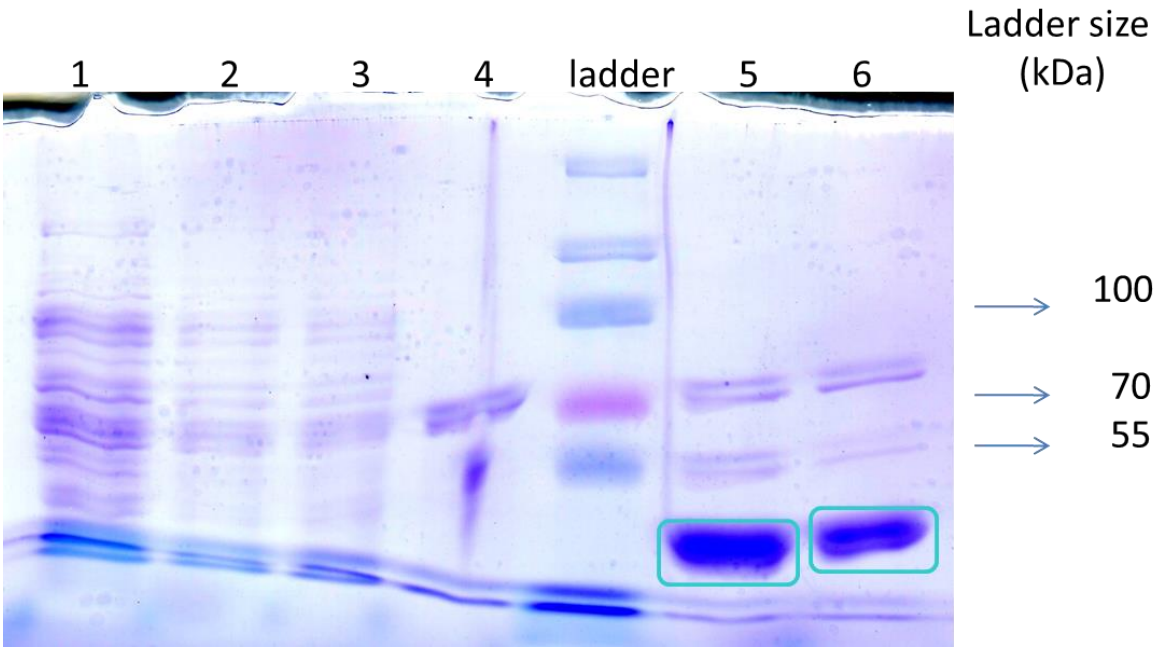


Figure 2.16 Purification of *N. tabacum* HCT (Hoffmann et al., 2003)**.**

SDS PAGE analysis of different purification steps of the expression of the *N. tabacum* HCT in the bacterial strain BL21-G612. 1: Flow through; 2: Wash 1; 3: Wash 2; 4: Wash PK; 5: Elution 1 containing HCT at a size of ~ 51 kDa; 6: Elution 2 containing HCT at a size of ~ 51 kDa. SDS PAGE gel stained with coomassie blue, PageRuler prestained protein ladder plus (Fermentas SM1811).

2.8.4. Incubation of the *A. thaliana* HCT with *p*-coumaroyl-CoA and L-threonic acid

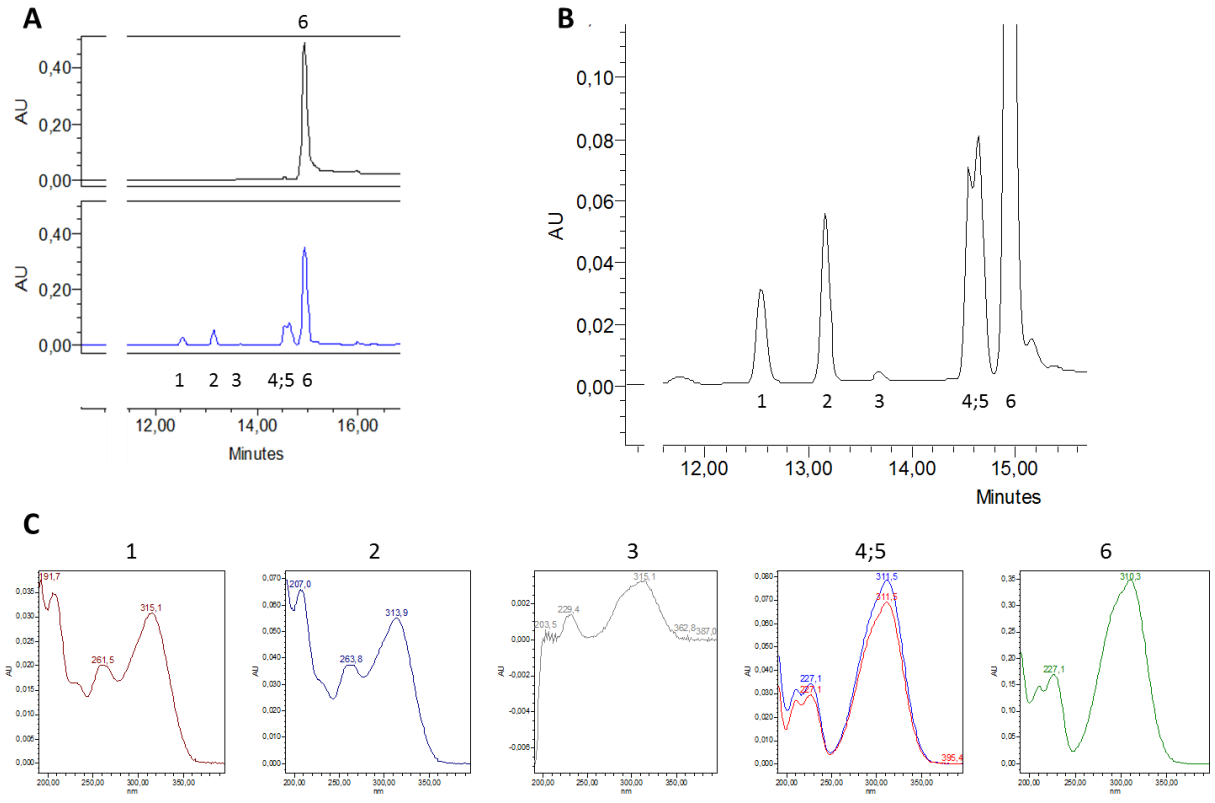


Figure 2.17 Incubation of *A. thaliana* HCT (courtesy of Pascaline Ullmann) with L-threonic acid and *p*-coumaroyl-CoA.

Analysis of reaction products was performed on HPLC/DAD. **A:** negative control with boiled HCT, L-threonic acid and *p*-coumaroyl-CoA (in black) and reaction of *A. thaliana* HCT with *p*-coumaroyl-CoA and L-threonic acid (in blue below). **B:** Close up of the reaction products. Peaks 1 to 5 are potentially coumaroyl-threonic acid, peak 6 remaining coumaric acid. **C:** UV/DAD spectra of all peaks shown in **B**.

2.8.5. Incubation of *P. patens* HCT with *p*-coumaroyl-CoA and L-threonate, shikimate, quinate.

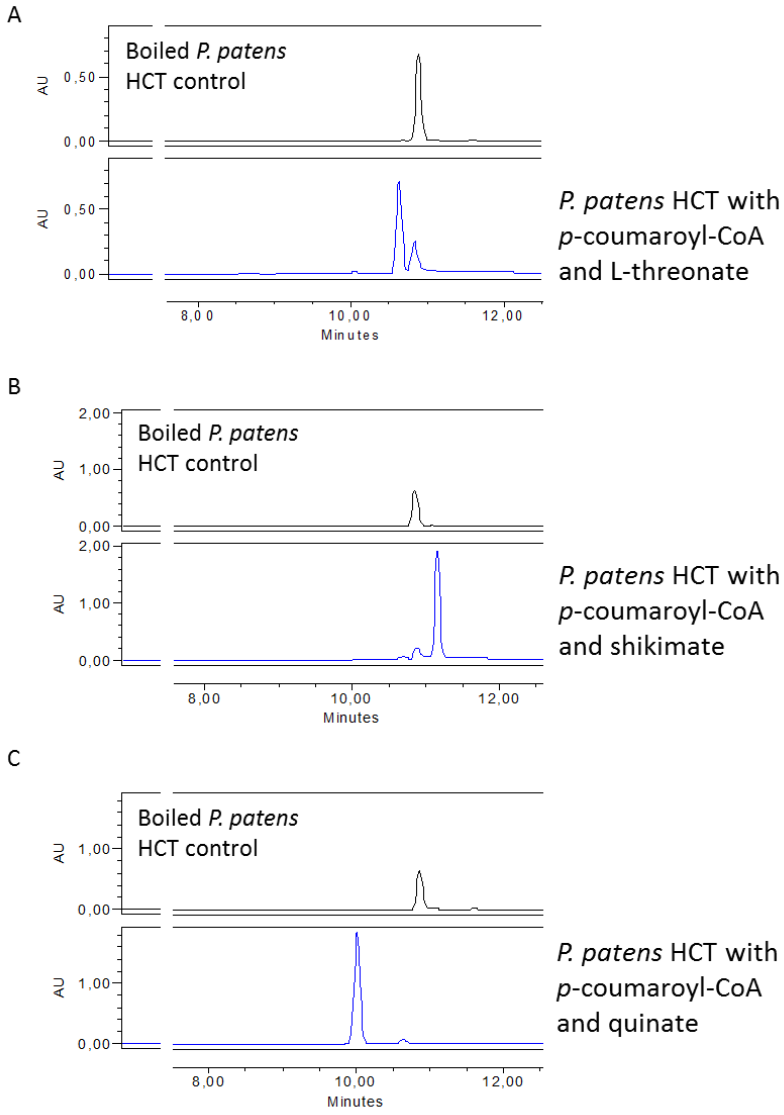


Figure 2.18 Incubation of *P. patens* HCT with *p*-coumaroyl-CoA and L-threonic acid, shikimic acid and quinic acid.

A: Incubation of *Pp*HCT with *p*-coumaroyl-CoA and L-threonic acid. Boiled enzyme control is shown in black above; the reaction is shown in blue below. **B:** *Pp*HCT with *p*-coumaroyl-CoA and shikimic acid. Boiled enzyme control in black above, reaction in blue below. **C:** *Pp*HCT incubation with *p*-coumaroyl-CoA and quinic acid. Boiled enzyme control in black above and reaction in blue below. HPLC/DAD analysis, chromatograms were taken at 310nm.

3. *CYP98* gene duplication and diversification within the angiosperms

3.1. Summary

A rich variety of angiosperm species can be found in almost all living environments today. Here we describe the evolution and function of the *CYP98* family within the angiosperms. Even though many plant gene families are highly conserved and can be found throughout the embryophytes, these gene families can vary widely in family size in different plant lineages. Adapting to specific environmental challenges, basic inherited sets of gene families expanded and functionally diversified in different lineages in response to environmental constraints. Gene duplications provide the opportunity to these adaptive events, creating additional gene copies, which can diversify. The involvement of *CYP98s* in the monolignol biosynthetic pathway in angiosperms has been described and *CYP98s* are also known to be involved in the pathways leading to soluble phenolic compounds in plants. Multiple *CYP98* member families exist only in the angiosperms. All angiosperms investigated possessed at least one *CYP98* member. A phylogenetic reconstruction of the *CYP98* family across angiosperm orders showed no distinct clades associated to distinct biochemical or *in vivo* functions. Instead, results suggested that independent *CYP98* duplications happened many times within the angiosperms. Two *CYP98* families which underwent independent duplication events, in *Populus trichocarpa* and *Amborella trichopoda,* have been biochemically characterized. In each species one CYP98 favoured *p*-coumaroyl-shikimate as substrate and is presumably involved in the biosynthesis of monolignols, while another isoform showed a broad range of accepted substrates. A third isoform in *P. trichocarpa* did not show biochemical activity *in vitro* or in a mutant complementation assay in *Arabidopsis thaliana*. Kinetic data of two *P. trichocarpa* isoforms together with end point enzyme assay data, gene expression and co-expression data supported the hypothesis of one CYP98 being involved in lignin biosynthesis and the other in the biosynthesis of soluble phenolic compounds. The two *P. trichocarpa* isoforms both complemented the severe *A. thaliana cyp98a3* knock-out mutant phenotype. In the Salicaceae, the first duplication of the *CYP98* gene family happened before the salicoid whole genome

duplication. The duplication event in *Populus trichocarpa* leading to *CYP98A23* and *CYP98A25* happened after the salicoid WGD, as tandem duplication.

3.2. Introduction

Angiosperms have diversified to a rich variety. About 300,000 species can be found today - in almost all environments (Soltis and Soltis, 2014). Most of our human and animal nutrition is provided by angiosperms. Considering their quantity and distribution, angiosperms are massive actors in photosynthesis and carbon sequestration (Smith et al., 2006a). Charles Darwin was fascinated and frustrated at the same time by the early evolution of angiosperms. His notion that nature does not make a leap (natura non facit saltum), was questioned by the immense diversity he observed in angiosperms of the mid-Cretaceous, as known at that time. His theory did not allow for an abrupt origin or for diversification at highly accelerated rates (Friedman, 2009). While Darwin looked at evolution on the level of species, methods available today help us to look at a different scale, the level of genes. The increasing number of sequenced genomes allows us to compare genes of closer or farther related species or families, deciphering the evolution of these genes. Combining sequence analysis, phylogenetic reconstruction and functional biochemical data helps us to redraw the history of evolution of given gene families within the angiosperms.

Many plant gene families are highly conserved and can be found throughout the embryophytes (Rensing et al., 2008). These gene families can vary widely in family member count in different plant lineages. It was suggested that gene families expanded and functionally diversified in different lineages, in response to specific environmental challenges. Gene duplications provide the opportunity for these adaptive events, by creating additional gene copies, which can then diversify (Hurles, 2004; Soltis and Soltis, 2016). Comparing and investigating genomes of different species or taxa helps us to identify gene duplications. These duplications can occur by various mechanisms, such as tandem duplication or the duplication of segments during DNA replication. Transposable elements can cause transduplication and retropositioning of segments can happen, in which reverse transcribed mature RNAs can be reintegrated into genomic DNA (Hurles, 2004). Whole genome duplication (WGD) events are frequent in the angiosperms (Jiao et al., 2011; Zheng et al., 2015). These gene duplications and subsequent

evolutionary events enabled the rapid diversification within angiosperm species (Soltis et al., 2009).

Enzymes involved in plant natural product metabolism are especially suited for evolutionary studies, because they can allow a direct linkage between gene evolution and adaptive traits. Pathways such as the phenylpropanoid pathway are involved in the production of natural products. A large chemical diversity in natural products coming from the phenylpropanoid pathway is found in angiosperm species. The pathway gives also rise to monolignols, the building blocks of lignin, which can be found in all angiosperms. The hydroxycinnamic ester coumaroyl-shikimate has been discovered as an intermediate in the biosynthesis of these monolignols in angiosperms (Schoch et al., 2001; Humphreys and Chapple, 2002). While lignin biosynthesis is common to all angiosperms, many, sometimes lineage specific, soluble hydroxycinnamic acid derivatives (HCCs) produced by the same pathway exist in angiosperms (Petersen and Simmonds, 2003; Bassard et al., 2010; Parveen et al., 2011; El-Seedi et al., 2012; Kim et al., 2015; Macoy et al., 2015a). Among the many HCCs important examples are chlorogenic acid, rosmarinic acid and phenolamides such as tyramine derivatives. These HCCs are important for plant defence, for example as antioxidants or feeding deterrents.

An important family of enzymes in this context are cytochromes P450 (CYPs). CYPs are present in all plants and share common conserved motifs. In addition to other functions, CYPs frequently hydroxylate molecules. The hydroxylation reaction is NADPH consuming, involves the cleavage of dioxygen and oxygen insertion in the substrate, which makes reactions catalysed by CYPs irreversible. Therefore many CYP enzymes define rate limiting and critical channelling positions in biochemical pathways. Members of the cytochromes P450 CYP98 family perform 3-hydroxylation on the phenolic ring. This *meta*-hydroxylation of the phenolic ring is a critical step in the biosynthesis of HCCs. *Meta*-hydroxylation by CYP98 enzymes is typically performed on coumaric conjugates such as the esters coumaroyl-shikimate and coumaroyl-quinate, but not on free coumaric acid. The CYP98 family has been described in a variety of plant species, from the lycopod *Selaginella moellendorffii*, to the angiosperm *Populus trichocarpa* (Coleman et al., 2008a; Weng et al., 2008b). All biochemically characterized CYP98 family members to date belong to angiosperm species. Many of these biochemically

characterized CYP98s are involved in lignin biosynthesis. CYP98 family members that are linked to the biosynthesis of lignin show a preference for coumaroyl-shikimate as their substrate *in vitro* (Table 3.1).

The CYP98s of *Arabidopsis thaliana, Medicago sativa* and *Populus grandidentata* X *alba* are examples of CYP98s involved in lignin biosynthesis. These CYP98s show a substrate preference for coumaroyl-shikimate, producing caffeoyl-shikimate (Figure 3.1A). The *A. thaliana* CYP98A3 has been found highly expressed in inflorescence stems (Schoch et al., 2001). A T-DNA insertion mutant in *CYP98A3* shows a severely affected phenotype with dwarf morphology (Abdulrazzak et al., 2006). The lignin of the mutant consists almost completely of H units with only traces of G and S units. A downregulation of *CYP98A37* in alfalfa *Medicago sativa,* leads to similar effects as observed for *A. thaliana*. Mutant plants are smaller in growth compared to wild type control plants and the lignin composition changes to a high proportion of H units, at the expense of G and S units. Isolated microsomes from alfalfa stem tissue, containing CYP98A37, convert *p*-coumaroyl-shikimate to caffeoyl-shikimate *in vitro*. Three *CYP98s* have been identified in the hybrid poplar *Populus grandidentata* X *alba.* RNAi downregulation of one CYP98 enzyme in this hybrid poplar causes a reduction of total lignin and a change in lignin composition. Vessel elements are affected and not regular in shape. Soluble phenolic profiles show an accumulation of hydroxycinnamic esters in the transgenic lines including small amounts of coumaroyl-shikimate (Coleman et al., 2008a). The orthologous CYP98 from *P. trichocarpa*, referred to as PtrC3H3 thus presumably CYP98A27, has been described by (Chen et al., 2011) to act in a membrane protein complex, together with C4H, to catalyse the 4- and 3- hydroxylation of the phenolic ring in monolignol biosynthesis in differentiating stem xylem. When the *P. trichocarpa* C3H3 alone was expressed in yeast microsomes and incubated with coumaroyl-shikimate *in vitro*, conversion rates were reported to be very low.

Species	CYP98	Gene identifier	Biochemical or reverse genetics	Substrates in order of preference	Lignin phenotype / related	References
Arabidopsis thaliana	*CYP98A3*	AT2G40890; NP850337	both	Coumaroyl-shikimate; coumaroyl-quinate	yes	(Schoch et al., 2001; Franke et al., 2002; Franke and Hemm, 2002)
Arabidopsis thaliana	*CYP98A8*	AT1G74540; AAG52369	both	N1, N5, N10-tricoumaroyl spermidine; N1, N5, N10- triferuloyl spermidine	no	(Matsuno, et al., 2009)
Arabidopsis thaliana	*CYP98A9*	AT1G74550; AAM67314	both	N1, N5, N10-tricoumaroyl spermidine	no	(Matsuno, et al., 2009)
Triticum aestivum	*CYP98A10*	CAE47489	biochemical	Coumaroyl-shikimate; coumaroyl-quinate	NT	(Morant et al., 2007)
Triticum aestivum	*CYP98A11*	CAE47490	biochemical	Coumaroyl-shikimate; coumaroyl-quinate; coumaroyl-tyramine	NT	(Morant et al., 2007)
Triticum aestivum	*CYP98A12*	CAE47491	biochemical	Coumaroyl-shikimate; coumaroyl-quinate; coumaroyl-tyramine	NT	(Morant et al., 2007)
Cynara cardunculus	*CYP98A49*	FJ225121	biochemical	Coumaroyl-shikimate; coumaroyl-quinate	NT	(Moglia et al., 2009)
Ocimum basilicum	*CYP98A13v1*	AY082611	biochemical	Coumaroyl-shikimate; Coumaroyl-quinate; coumaroyl 4-hydroxyphenyllactate; coumaric acid; coumaroyl-CoA (the	NT	(Gang et al., 2002)

				last two very little)		
Ocimum basilicum	*CYP98A13v2*	AY082612	biochemical	Coumaroyl-shikimate; Coumaroyl-quinate; coumaroyl 4-hydroxyphenyllactate; coumaric acid; coumaroyl-CoA (the last two very little)	NT	(Gang et al., 2002)
Coffea canephora	*CYP98A35*	DQ269126	biochemical	Coumaroyl-shikimate; coumaroyl-quinate (same rate)	NT	(Mahesh et al., 2007)
Coffea canephora	*CYP98A36*	DQ269127	biochemical	Coumaroyl-shikimate;	NT	(Mahesh et al., 2007)
Panicum virgatum	*PvC3'H1*	AB723823	biochemical	Coumaroyl-shikimate; coumaroyl-quinate	NT	(Escamilla-Trevino et al., 2014)
Panicum virgatum	*PvC3'H2*	AB723824	biochemical	Coumaroyl-shikimate; coumaroyl-quinate	NT	(Escamilla-Trevino et al., 2014)
Lonicera japonica	*LjC3H*	KC765076	biochemical	Coumaroyl-shikimate; coumaroyl-quinate	NT	(Pu et al., 2013)
Lithospermum erythrorhizon	*CYP98A6*	BAC44836	biochemical	4-coumaroyl-4'-hydroxyphenyllactic acid	NT	(Matsuno et al., 2002)
Solenostemon scutellarioides	*CYP98A14*	CAD20576	biochemical	4-coumaroyl-3',4'-dihydroxyphenyllactate; caffeoyl-4'-hydroxyphenyllactate. Not coumaroyl-shikimate, not coumaroyl-quinate	NT	(Eberle et al., 2009)
Populus grandidentata		EU391631	both	Coumaroyl-shikimate; coumaroyl-quinate	yes	(Coleman et al., 2008a; Coleman et al., 2008b)

X alba						
Populus trichocarpa	*CYP98A27*		biochemical	Coumaroyl-shikimate, very low	NT	(Chen et al., 2011)
Ruta graveolens	*CYP98A22*	JF799117	both	Coumaroyl-quinate ; coumaroyl-shikimate	NT	(Karamat et al., 2012)
Salvia miltiorrhiza	*CYP98A78*	HQ316179.1	Non	Might be involved in rosmarinic acid biosynthesis	NT	(Di et al., 2013; Wang et al., 2015)
Trifolium pratense	*CYP98A44*	ACV91106.1	biochemical	Coumaroyl-shikimate	NT	(Sullivan and Zarnowski, 2010)
***Medicago sativa* L.**	*CYP98A37*	ABC59086.1	reverse genetics		yes	(Reddy and Chen, 2005)
Eucalyptus urophylla* x *E. grandis	*CYP98*	EC 1.14.13.36	*in vivo*	NT	yes	(Sykes et al., 2015)

Table 3.1 Overview of characterized *CYP98* genes from literature.

Note that additional CYP98s may exist in individual species.

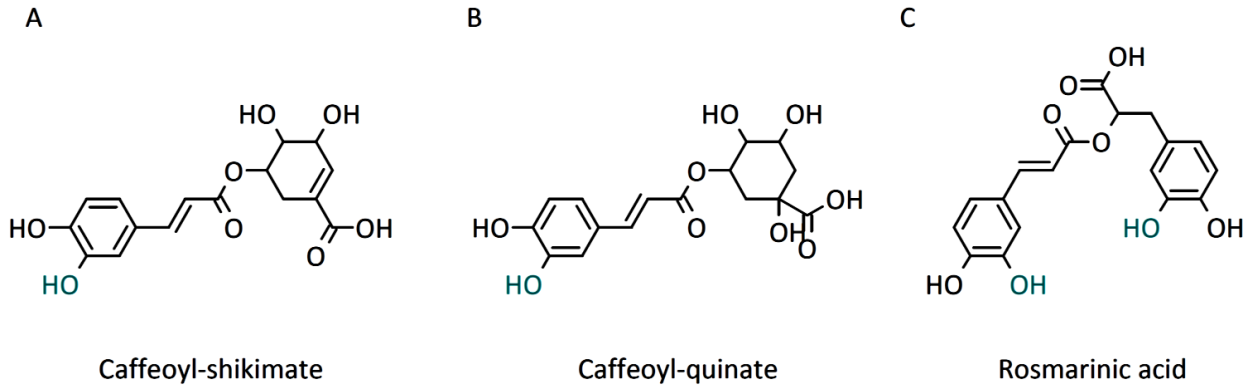


Figure 3.1 Hydroxycinnamic conjugates described in the text.

The position hydroxylated by CYP98 is marked in blue.

Coumaroyl-quinate is often used as a substrate by the above described CYP98s to appreciable levels, producing caffeoyl-quinate or chlorogenic acid (Figure 3.1B). Chlorogenic acid is an important factor for plant defence. In some cases, chlorogenic acid production is increased in plant leaves upon herbivory attack, and a correlation of reduced growth of the herbivore to the presence of chlorogenic acid has been made (Kessler and Baldwin, 2004). The capability to hydroxylate *p*-coumaroyl-quinate to chlorogenic acid has also been demonstrated for the *A. thaliana* CYP98A3, albeit *A. thaliana* is not known to produce chlorogenic acid *in vivo* (Schoch et al., 2001). Further examples of CYP98s that use coumaroyl-shikimate and also coumaroyl-quinate as a substrate, are the *CYP98s* from wheat (*Triticum aestivum*), globe artichoke (*Cynara cardunculus*), sweet basil (*Ocimum basilicum*), coffee (*Coffea canephora*), switchgrass (*Panicum virgatum*) and *Lonicera japonica* (Gang et al., 2002; Mahesh et al., 2007; Morant et al., 2007; Moglia et al., 2009; Pu et al., 2013; Escamilla-Trevino et al., 2014). Two CYP98 isoforms have been described in *C. canephora*, CYP98A35 and CYP98A36. While both isoforms utilize coumaroyl-shikimate at similar rates, only CYP98A35 hydroxylates the chlorogenic acid precursor coumaroyl-quinate with the same efficiency as the shikimate ester (Mahesh et al., 2007).

Similar to coffee, the monocot *Panicum virgatum* has two characterized *CYP98s*. Both of these CYP98s have the capacity to hydroxylate coumaroyl-shikimate and coumaroyl-quinate, forming caffeoyl-shikimate or chlorogenic acid (Escamilla-Trevino et al., 2014). While these *P. virgatum*

CYP98 enzymes are biochemically very similar, differences in transcript levels of the two genes have been found. In general one *CYP98* shows higher transcript levels in highly lignified tissues, while the other *CYP98* shows higher transcript levels in tissues with less lignin (except for one tissue). The authors suggest that one *CYP98* could thus be involved in the biosynthesis of monolignols, while the other could be involved in the biosynthesis of chlorogenic acid (Escamilla-Trevino et al., 2014). *Lonicera japonica* is a plant used in Chinese medicine, known for its antioxidants. Several *CYP98s* exist in *L. japonica* and one CYP98 (accession: KC765076) has been expressed in *E. coli* and functionally characterized. As mentioned above, this enzyme prefers coumaroyl-shikimate over coumaroyl-quinate. An increase in gene expression was recorded, when *L. japonica* plants were either exposed to UV-B light, or treated with methyl jasmonate (MeJA). A positive correlation between the detected amount of CGA in the leaves of the plant and the transcript abundance of the gene has been made (Pu et al., 2013).

Not all CYP98s prefer coumaroyl-shikimate or coumaroyl-quinate, which suggests the existence of isoforms specialized in the formation of more distinct soluble phenolic natural products. CYP98A22 of *Ruta graveolens* prefers coumaroyl-quinate over coumaroyl-shikimate and expression of *CYP98A22* is responsive to UV-B light in leaves (Karamat et al., 2012). Recently, rosmarinic acid (Figure 3.1C) has been identified to be a homoserine lactone mimic, playing a role in plant defence, by activating a bacterial quorum sensing regulator (Corral-Lugo et al., 2016). The characterized CYP98A13 of the rosmarinic acid producing plant sweet basil (*O. basilicum*) was shown to be able to hydroxylate the phenolic moiety of the rosmarinic acid precursor in addition to coumaroyl-shikimate. However, hydroxylation of the rosmarinic acid precursor takes place at a very low rate (Gang et al., 2002). Likewise, another C3'H (CYP98A6) from a different rosmarinic acid producing plant, *Lithospermum erythrorhizon*, has been characterized. CYP98A6 also catalyses the synthesis of rosmarinic acid, but other substrates were not tested with CYP98A6 (Matsuno et al., 2002). The first example of a CYP98 that has no apparent activity with coumaroyl-shikimate, or -quinate, was CYP98A14 from coleus (*Plectranthus scutellarioides* previously referred to as *Coleus blumei* or *Solenostemon scutellarioides*). Coleus accumulates large amounts of rosmarinic acid. The coleus CYP98A14 was shown to catalyse both the 3-hydroxylation of 4-coumaroyl-3',4'-dihydroxyphenyllactate

and the 3'-hydroxylation of caffeoyl-4'-hydroxyphenyllactate, in both cases forming rosmarinic acid (Eberle et al., 2009). Another CYP98 has been identified in *Salvia miltiorrhiza, CYP98A78*, which is also likely involved in rosmarinic acid biosynthesis. *S. miltiorrhiza CYP98A78* is more closely related to the coleus CYP98 in phylogenetic analyses than to *CYP98A3* from *A. thaliana*. Together with C4H of *S. miltiorrhiza,* the enzyme is most highly expressed in the roots of the plant and its expression is inducible by methyl jasmonate (MeJA) treatment (Wang et al., 2015). Several CYP98s have been identified in a further rosmarinic acid producing plant, *Eritrichium sericeum.* Overexpressing the *rolC* gene of a phytopathogen transcriptionally activates a *CYP98* in callus cultures of *E. sericeum.* The activation of *CYP98s* correlates with an increased accumulation of rosmarinic acid. The sequences of the activated *CYP98s* are similar to the sequence of *CYP98A6* of *L. erythrorhizon*. Transcript levels of further *CYP98* genes found in the *E. sericeum* plant were contrary not increased in the transformed *rolC* callus culture (Inyushkina et al., 2009).

A further connection between *CYP98* expression and plant defence was made in a study with the common bean, *Phaseolus vulgaris*. When bean leaves have been treated with 3,5-dichlorosalicylic acid (DC-SA) and 2,6- dichloroisonicotinic acid (DC-INA), the expression of *P. vulgaris CYP98A5* was upregulated. DC-SA is an agent which primes plant defence and DC-INA is an agent which induces systemic acquired resistance (Basson and Dubery, 2007). An increase of *CYP98* (GenBank: HM585369) transcript levels was also detected in *Withania somnifera* upon treatment of the plants with MeJA and salicylic acid (SA). The increase of *CYP98* transcripts correlates with an increase in levels of triterpenoids (withanolides) of *W. somnifera in vitro* cultures (Rana et al., 2014). However, an involvement of CYP98 in the biosynthesis of withanolides seems not likely, and the *W. somnifera* CYP98 activity has not been characterized biochemically. Nevertheless, the increase of *W. somnifera CYP98* transcript upon MeJA and SA treatment suggests an involvement of *CYP98* in plant defence also in *W. somnifera.*

CYP98s have also been connected to mechanisms that prevent plant proteins from degradation during harvest or storage, by providing phaselic acid, or 2-*O*-caffeoyl-L-malate. CYP98A44 of red clover (*Trifolium pratense)*, a species known to accumulate large quantities of phaselic acid, was characterized, to elucidate the possibility of this CYP98 to hydroxylate coumaroyl-malate.

However, *T. pratense* CYP98A44 expressed in yeast can hydroxylate coumaroyl-shikimate, but not coumaroyl-malate. Still, the presence of other CYP98s which can hydroxylate coumaroyl-malate cannot be ruled out (Sullivan and Zarnowski, 2010).

In summary, several *CYP98s* have been described or characterized in various angiosperm species. CYP98s that are shown to be involved in the biosynthesis of monolignols favour coumaroyl-shikimate as their substrate. Apart from being involved in the biosynthesis of structural building blocks of plants, CYP98s were shown to be involved in the biosynthesis of protective compounds: CYP98 isomers can utilize distinct substrates to produce soluble, protective HCCs and *CYP98* gene expression can be increased upon treatment with UV light, MeJA and DC-SA. This points to a function of *CYP98* in angiosperms that is dual: structural molecules and protective compounds such as chlorogenic acid, rosmarinic acid and phaselic acid. In some cases impacts on both insoluble lignin and soluble phenolics can be apparent when a single CYP98 is manipulated. However, where multiple family members have been described, isoforms can be differentially controlled on the transcription level. The expression of one *CYP98* family member can be increased upon treatment, while other *CYP98* family members do not show differences in expression level under the same treatment. Multiple CYP98 family members within a given species can have non-redundant functions and biochemical properties of several CYP98s suggest a specialized involvement in soluble bioactive compounds in some plants.

An important species for deciphering angiosperm gene evolution is *Amborella trichopoda*. Together with some aquatic herbs, *A. trichopoda* is considered as the sister of all other extant flowering plants (Goremykin et al., 2013; Goremykin et al., 2015; Chase et al., 2016). *A. trichopoda* is a small tree or shrub, endemic to New Caledonia. The *A. trichopoda* genome has been sequenced ((Amborella Genome Project, 2013) Amborella Genome Database, www.amborella.org). Another angiosperm tree species with a sequenced genome is *Populus trichocarpa,* black cottonwood (Tuskan et al., 2006) (available on Phytozome version 11, https://phytozome.jgi.doe.gov/pz/portal.html). *P. trichocarpa* is a forest species, providing a rich repertoire of chemical defence metabolites. The most abundant natural products synthesized in the genus *Populus* derive from the shikimate-phenylpropanoid pathway (Chen et

al., 2009). The work of Greenaway, English and Whatley in the early 1990s investigated bud exudates of various *Populus* species by GC-MS and found more than 40 different hydroxycinnamic acid conjugates, described in several publications, for example in English et al., 1991; Greenaway et al., 1991b; English et al., 1992; Greenaway and English, 1992. Poplar trees are good model trees, as their comparably fast growth and established manipulation methods allow for reverse genetic approaches. Poplar trees are a promising source for biofuel production and subject to manifold approaches of elucidating and altering secondary cell wall biosynthetic pathways.

3.2.1. Hypotheses and objectives

It remains unknown how this complex system of functional diversity and divergence within the *CYP98* family evolved. Functional divergence can be detected in many angiosperm species with several *CYP98* isoforms. A general trend throughout the angiosperms seems to be that CYP98s involved in the biosynthesis of monolignols favour coumaroyl-shikimate as substrate, while distinct isoforms provide the, sometimes lineage-specific, chemical diversity of soluble phenolic conjugates. It may be plausible to assume that an ancient separation, e.g. through a WGD early in the angiosperm lineage of lignin-specific and soluble phenolic related CYP98s took place, where lignin-related isoforms gained or maintained coumaroyl-shikimate specificity, while soluble phenolic isoforms evolved to provide the chemical diversity currently observed (Hypothesis 1).

However, lineage specific evolution within the *CYP98* gene family has been observed, for example within the Brassicaceae: Three *CYP98* members exist in *A. thaliana.* One member, *CYP98A3*, is involved in lignin biosynthesis as described above, while the other two members, *CYP98A8* and *CYP98A9*, are gene duplicates generated through retroposition from *CYP98A3*. These duplicates acquired new functions to become 3- and 5- hydroxylases of coumaroyl-spermidine, involved in pollen wall biogenesis (Matsuno et al., 2009). This recent duplication event happened only in the Brassicaceae. This gives an example for duplicates which are newly recruited to produce a distinct hydroxycinnamoyl conjugate. Contrary to the hypothesis above, that *CYP98* functional divergence is ancient, it could thus be possible that multiple independent *CYP98* duplications in distinct angiosperm lineages led to the production of lineage-specific

CYP98 isoforms and subsequently hydroxycinnamoyl conjugates (Hypothesis 2). In this case gene duplication and independent recruitment to novel pathways could be common in angiosperms. It is the primary objective of this chapter to distinguish between these alternative hypotheses.

3.3. Material and methods

3.3.1. Genome mining and phylogenetic analysis

The 43 angiosperm genomes available on phytozome v 11 (February 2016, https://phytozome.jgi.doe.gov/pz/portal.html) (Goodstein et al., 2012) have been searched by the BLAST algorithm (Altschul et al., 1990), using the *A. thaliana* CYP98A3 (AT2G40890) as bait sequence. Sequences were assigned CYP98 family members when their sequence identity to the CYP98A3 was above 40%. Further confirmation of CYP98 family membership was obtained by phylogenetic analysis using phyml, implementing the maximum likelihood algorithm (Guindon and Gascuel, 2003).

Sequences of characterized CYP98s have been included in phylogenetic reconstructions and the identifiers are listed in Table 3.5.

Sequences of CYP98 enzymes have been included from transcriptome data available on the 1000 Plant Transcriptomes project (onekp.com), under exclusion of species listed in the table of sample source and purity issues (https://pods.iplantcollaborative.org/wiki/display/iptol/Sample+source+and+purity).

If not stated otherwise, sets of sequences were either manually aligned or aligned by DIALIGN, based on segment to segment comparison (Morgenstern, 1999). Phylogenetic reconstructions were calculated using the maximum likelihood algorithm, implemented in phyml (Guindon and Gascuel, 2003). Programs were accessed through the mobyle V1.5 platform (http://mobyle.pasteur.fr/cgi-bin/portal.py#welcome). Model testing was performed by ProtTest (Abascal et al., 2005) for protein alignments and by Smart Model Selection (Lefort V, Longueville JE, Gascuel O; atgc-montpellier.fr/sms/) for nucleic acid sequences.

3.3.2. Heterologous enzyme expression in *Saccharomyces cerevisiae*

The complete open reading frame of *P. trichocarpa* CYP98A23 (Poptr_0016s03090), *CYP98A25* (Poptr_0016s03080) and *CYP98A27* (Poptr_0006s03180) and *A. trichopoda CYP98A84* (evm_27.TU.AmTr_v1.0_scaffold00101.79) and *CYP98A85* (evm_27.TU.AmTr_v1.0_scaffold00040.62) was cloned from cDNA. *P. trichocarpa* Nisqually 1 cDNA was obtained from RNA of young leaves, harvested from the Forest Biology tree collection, Victoria BC. RNA was extracted following the method described in Kolosova et al., 2004. One µg of total RNA was used for cDNA synthesis by SuperScriptIII™ polymerase and an oligo dT23 pimer, following the manufacturer's instructions. *A. trichopoda* cDNA was kindly provided by Charles P. Scutt, Laboratoire RDP, CNRS, ENS de Lyon, DNA library from Fourquin et al., 2005). Open reading frames were cloned using appropriate primers (Table 3.4 supplement). Genes were cloned into yeast expression vector pYeDP60USER by USER™ cloning (Nour-Eldin and Hansen, 2006; Nour-Eldin et al., 2010). *S. cerevisiae* strain WAT11 was transformed by heat shock using salmon sperm DNA as a carrier (Gietz and Jean, 1992) or by electroporation (400 Ohm/250 µF/0,45 kV). The growth method for yeast cultures and the preparation of microsomal fractions, containing the recombinant enzyme, have been described in (Gavira et al., 2013). P450 quality control and quantification was performed by differential spectrophotometry as described in (Gavira et al., 2013) using the absorption coefficient at 450 nm: ε=91 mM−1 cm−1 (Omura and Sato, 1964).

3.3.3. CYP98 enzyme incubations with a library of potential substrates

Microsomal fractions containing CYP98 were used in incubations with various substrates. 10 pmol of P450 were added to a reaction volume of 400µl. Reactions were performed in 50mM K phosphate buffer (pH7.4), containing 100µM substrate and 500µM NADPH. For kinetic properties of the reductase ATR1, refer to (Urban et al., 1997) Reactions were started by addition of NADPH and incubated at 28°C for 30 min. Reactions were stopped by addition of 1/10 (v/v) 50% acetic acid and 4/10 (v/v) methanol. After centrifugation (10min; 15000g; 4°C) the supernatant was used for analysis on HPLC/DAD. Three independent incubations were performed for each enzyme/substrate combination. Substrate conversion was monitored. For

this, the substrate peak area of the chromatogram was integrated using the Empower™ (Waters) software. The percentage of conversion was calculated from the peak areas of substrate after incubation, compared to the initial amount of substrate.

3.3.4. Standards for enzyme incubations

Compare to Chapter 2, Material and methods, 2.3.8.

Substrate and reference phenolic conjugates except *p*-coumaroyl-shikimate were provided by the group of M. Schmitt (CNRS, UMR 7200, Illkirch).

p-*C*oumaroyl-shikimate was produced enzymatically from *p*-coumarate, as described in (Morant et al., 2007).

3.3.5. Enzyme kinetics for *P. trichocarpa* CYP98A23 and CYP98A27

Enzyme kinetics performed for *P. trichocarpa* CYP98A23 and CYP98A27 with *p*-coumaroyl-shikimate, *p*-coumaroyl-quinate, benzyl-*p*-coumarate and isoprenyl-*p*-coumarate were modelled as non-linear regression of the Michaelis-Menten equation (vmax*x/(x+Km)) by defining a user specific function in the program SciDavis (Free application for Scientific Data Analysis and Visualization; Benkert T, Franke K, Standish R, 2007; scidavis.sourceforge.net). Non-linear regression was fitted under the Nelder-Mead-Simplex algorithm and statistical error (Poisson) source. Enzyme concentrations, derived from CO spectra measurements, were as follows: 0.05 µM CYP98A23 incubated with *p*-coumaroyl-shikimate, 0.05 µM CYP98A23 incubated with *p*-coumaroyl-quinate, 0.2 µM CYP98A23 incubated with isoprenyl-*p*-coumarate and benzyl-*p*-coumarate and 0.25 µM CYP98A27 incubated with *p*-coumaroyl-shikimate, 0.2 µM CYP98A27 incubated with *p*-coumaroyl-quinate, 10 µM CYP98A27 incubated with isoprenyl-*p*-coumarate and benzyl-*p*-coumarate. Reactions were performed in 50mM K phosphate buffer (pH7.4) containing 500 µM NADPH. Substrate concentrations were 1; 5; 10; 20; 50; 100; 150; 200 µM. Reactions were incubated for 5 min at 28°C und agitation in the dark. Reactions were stopped by addition of 1/10 (v/v) 50% acetic acid and 4/10 (v/v) methanol. After centrifugation (10min; 15000g; 4°C) the supernatant was used for analysis on HPLC/DAD. A standard curve was included in each run, using chlorogenic acid at different concentrations. The formation of product was determined by the area of the peak in the HPLC chromatogram. The equation of

the standard curve of known product concentrations was used to determine the corresponding product concentrations by peak area.

3.3.6. *A. thaliana* Tn4 mutant complementation assay with *P. trichocarpa CYP98s*

The *A. thaliana* T-DNA insertion mutant knock-out for *CYP98A3* (Abdulrazzak et al., 2006) was used in a mutant complementation assay with the *P. trichocarpa CYP98A23; CYP98A25* and *CYP98A27* genes. As homozygous T-DNA lines of *cyp98a3* show dwarf morphology and are male sterile, heterozygous plants were used for transformation with the *CYP98s* under the promoter of the *A. thaliana* C4H gene (Bell-Lelong and Cusumano, 1997). The use of this promotor ensured enhanced expression in lignified tissues. The open reading frames of the *CYP98s* were cloned into the pDONR207 Gateway™ (Invitrogen) entry vector. Recombination with the Gateway™ destination vector pCC0996 resulted in the expression construct. pCC0996 contains the C4H promoter sequence of *A. thaliana* (Weng et al., 2010b). *Agrobacterium tumefaciens* strain GV3101 was transformed with the expression construct and *A. thaliana* plants transformed by floral dip as described (Clough and Bent, 1998). Seed of control plants with pCC0996:*CYP98A3* expression constructs were a courtesy of Dr. Z. Liu (John Innes Centre, Norwich, UK).

3.3.7. Real-time quantitative PCR on gypsy moth treated *P. trichocarpa* leaves

Total RNA was provided by Jan Günther and Tobias Köllner (Max Planck Institute for Chemical Ecology, Jena) from undamaged and gypsy moth (*Lymantria dispar*) damaged *P. trichocarpa* leaves in 5 biological replicates for each treatment. A single tree was harvested for each replicate, five undamaged control trees and five herbivore damaged trees. Leaves for each sample were harvested in a leaf pool of 5 to 11 leaves from top of the tree.

Total RNA was extracted using the Invitek Plant RNA Kit and eluted in pure water. The concentration of the RNA and the ratios of 260/280 and 260/230 nm were measured on a Nano-Drop2000™ spectrophotometer (Thermo Scientific). The integrity of the RNA was confirmed by electrophoresis on an agarose gel.

One µg of total RNA was used for cDNA synthesis by SuperScriptIII™ polymerase (Invitrogen) and an oligo dT23 pimer, following the manufacturer's instructions. The primers for qPCR

amplification were designed using Primer3 plus (Table 3.4) and a specificity test performed by BLAST search against the *P. trichocarpa* genome on Phytozome version 11 (Goodstein et al., 2012). qPCR reactions were performed with 250nM of each forward and reverse primer and 1x SYBR™ Green Master Mix (Roche). Samples were run on a LightCycler 480 (Roche). Melting curve analyses were included in the run (Figure 3.26). After verification of four reference genes by the GeNorm algorithm (Vandesompele et al.) two reference genes POPTR_0009s02370 and POPTR_0001s35630 were chosen for the normalization of the qPCR data (Figure 3.27). Transcript amplification efficiency was calculated for all primer sets using the LinRegPCR computer program (Ramakers et al., 2003). The relative expression ratio of the transcripts was calculated as described in Livak and Schmittgen, 2001.

3.3.8. Transient overexpression of P. trichocarpa CYP98s in *Nicotiana benthamiana*

All three *P. trichocarpa CYP98s* have been transiently overexpressed in *Nicotiana benthamiana* under the CaMV-35s promotor. *N. benthamiana* leaves were coinfiltrated with the viral protein P19 to avoid post transcriptional gene silencing. Leaf discs of ~1 cm diameter were incubated in petri dishes containing 20mM pKi buffer at pH 7.4 with or without (control) 100µM *p*-coumaroyl-shikimate. After a short vacuum application, leaf discs were incubated for four hours. The medium of the petri dishes was collected and extracted by ethyl acetate. 0.35g of fresh weight leaf discs of each (rinsed with milliQ water and dried carefully) construct were frozen in liquid nitrogen and ground with metal beads in a tissue-lyzer™ (Qiagen). The samples were extracted with methanol. All samples were analysed on UPLC-MS/MS for their amount of *p*-coumaroyl-shikimate, caffeoyl-shikimate and chlorogenic acid.

3.4. Results and discussion

3.4.1. Genome mining and phylogenetic analysis

Species with multiple *CYP98* isoforms exist only in angiosperms. Phylogenetic reconstructions show that all angiosperm *CYP98s* are derived from a single ancestor (see Chapter 2 of this

thesis). To infer information about the evolutionary gene duplication and loss events that shaped the CYP98 gene family, complete gene families from angiosperms were targeted for phylogenetic reconstructions.

43 sequenced angiosperm genomes were publically available on Phytozome version 11 as of February 2016 (https://phytozome.jgi.doe.gov/pz/portal.html) (Goodstein et al., 2012). 123 *CYP98* sequences were found in these angiosperm genomes. At least one CYP98 existed in each angiosperm species. The size of the *CYP98* family in these species ranges from 1 to 12 members per species, with a median of 2. Following the latest Angiosperm Phylogeny Group (APG IV) classification for orders and families of flowering plants (Chase et al., 2016), an angiosperm CYP98 dataset containing two species for each order (where available) was created. A schematic overview of the orders included and their relationships is presented in Figure 3.2. In the choice of representatives, first all characterized CYP98 isoforms were included. Subsequently isoforms of species with complete genome sequencing were included. The orders were completed by sequences from transcriptome data, available from the 1000 plant transcriptomes project (onekp.com). Species samples from the 1000 plants transcriptomes project with recorded identity and purity problems were excluded from the analysis.

The maximum likelihood phylogeny reconstructed based on the resulting alignment of the CYP98 family members from across all available angiosperm orders, showed a general branching pattern by orders (Figure 3.3). The two CYP98 genes of *A. trichopoda* are at the base of all angiosperm CYP98s. Both isoforms are in the same clade. Orders belonging to the monocots, austrobaileyales and eudicots form distinct clades. Only orders belonging to the magnoliids form two distinct clades. A functional separation into two clades, as described as a possible evolution model above, was not detected. Instead, *CYP98* isoforms of the same family often were located on close branches in the same clade.

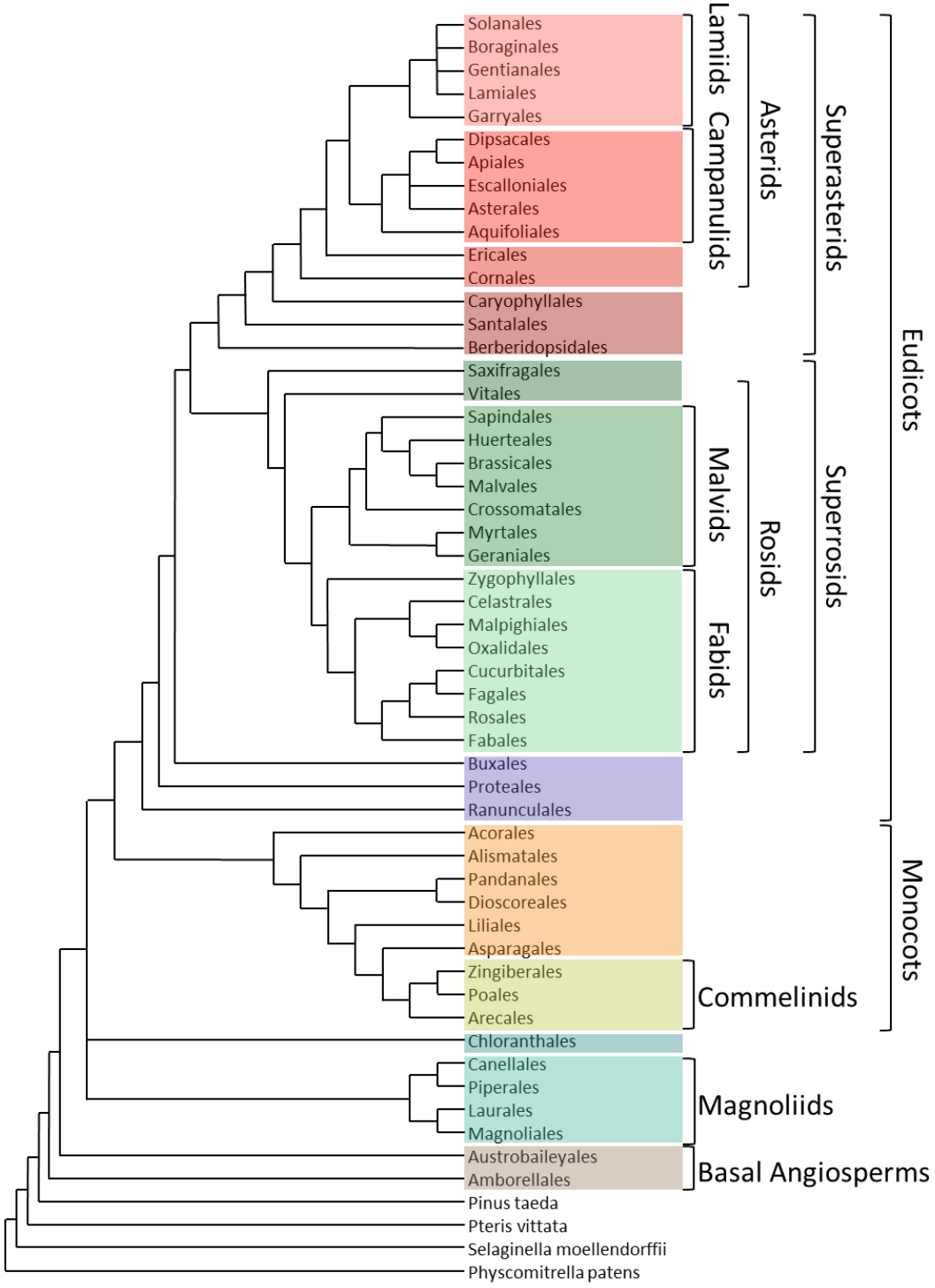


Figure 3.2 Schematic overview of angiosperm order interrelationships.

Adapted from Chase et al., 2016**.** Representatives of all orders displayed are included in the following phylogenetic reconstructions of this chapter and the colour scheme is maintained, individual branches are coloured accordingly.

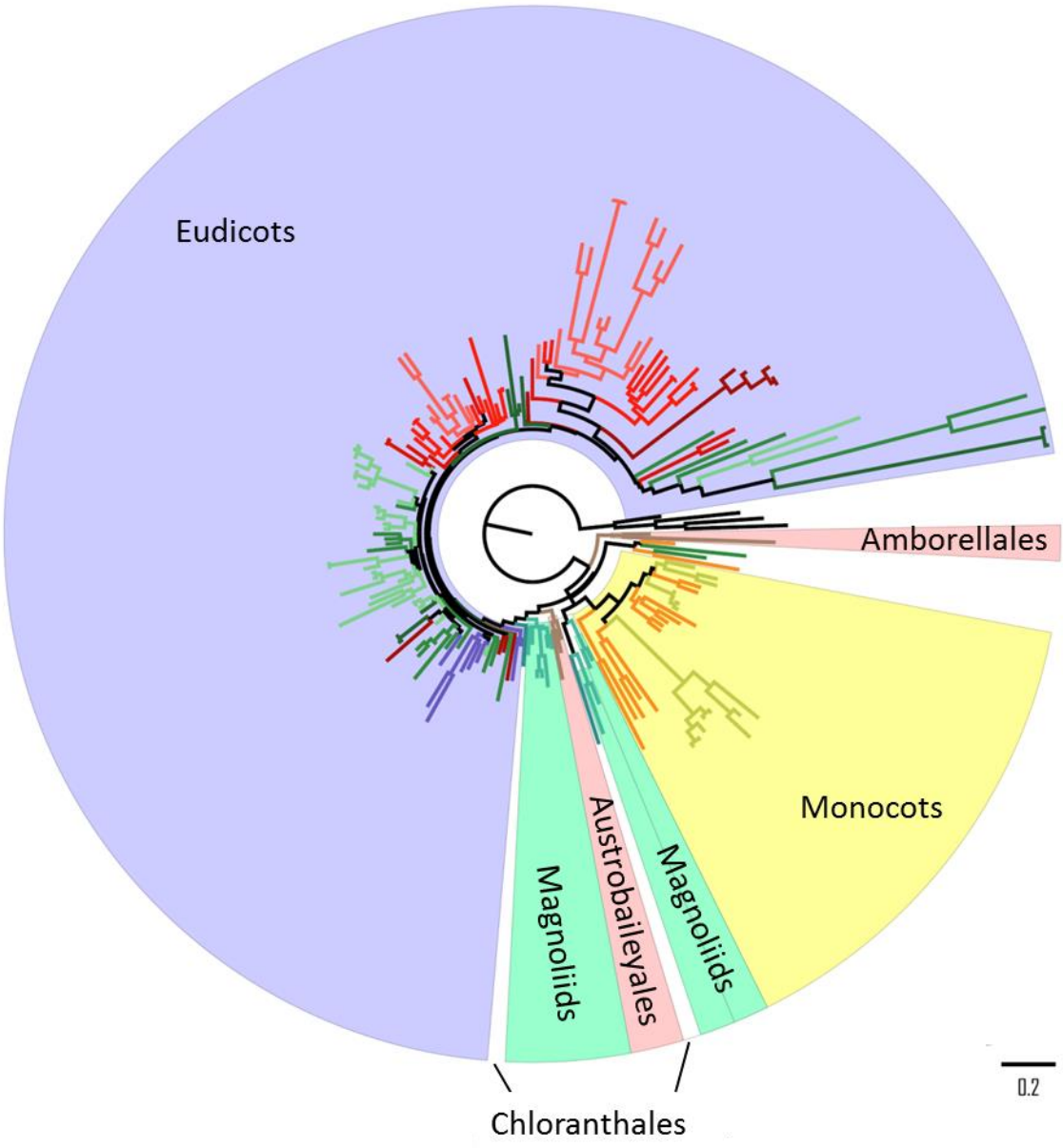


Figure 3.3 Phylogenetic reconstruction of the CYP98 family across angiosperm orders.

A. trichopoda is at the base of all angiosperms. Two species of each angiosperm order (where available) were chosen and amino acid sequences aligned by DIALIGN (Morgenstern, 1999). Only alignment positions with diagonal similarities greater than zero were maintained in the alignment. A maximum likelihood based phylogenetic reconstruction was performed by phyml (Guindon and Gascuel, 2003) under assumption of the JTT model. The branches of the tree are coloured according to the orders in Figure 3.2. All orders of Figure 3.2 are represented in the phylogenetic reconstruction.

Statistical bootstrap support for this branching pattern within the angiosperm CYP98 family was high for some clades such as *A. trichopoda,* the monocots, and a clade containing the *A. thaliana CYP98A8/CYP98A9* genes. Branches of CYP98 isoforms within families also showed very good bootstrap support. However, the support was very low for clades of several orders. To refine the analysis, a further phylogenetic reconstruction has been performed. In this reconstruction, characterized CYP98s were included as well as complete CYP98 families of species with sequenced genomes. Contrary to the angiosperm order phylogeny described above, orders were not equally represented and the number of species per order was not restricted to two. As transcriptome sequence data depends highly on the expression of genes in the analysed tissue, CYP98 families might not be complete. In some cases, the annotation of the CYP98s included introns that were eliminated in the sequence alignment. The work with available CYP98 family sequences of species with sequenced genomes might provide reliably accurate annotations.

The most obvious characteristic of the resulting phylogeny (Figure 3.4) was a distinct clade, containing CYP98 isoforms of the Brassicaceae. This clade contained the *A. thaliana CYP98A8* and *CYP98A9* genes. As described above, CYP98A8 and CYP98A9 have been characterized and shown to have a distinct function, which they acquired after a retroposition event through subsequent subfunctionalization.

A second very distinct clade was formed by two *CYP98s* of *Kalanchoe marnieriana (Kalma1 and Kalma3)*. This clade showed strong bootstrap support as well. Four *CYP98* sequences were found in the *K. marnieriana* genome on Phytozome (Goodstein et al., 2012). Two of these *K. marnieriana CYP98* sequences formed the separate clade in the phylogenetic analysis, while two other *CYP98s* were found with short branch length together with other angiosperm *CYP98s*. Kalma1 and Kalma3 formed a sister clade to the Brassicaceae CYP98A8/9-like clade, but belonging to the Saxifragales, Kalanchoe does not share a close taxonomic relationship to the Brassicales. It appears possible that these two clades were placed together with comparably high bootstrap support owing to long-branch attraction, where sequences are placed together simply because they all differ from the remainder of the sequences included.

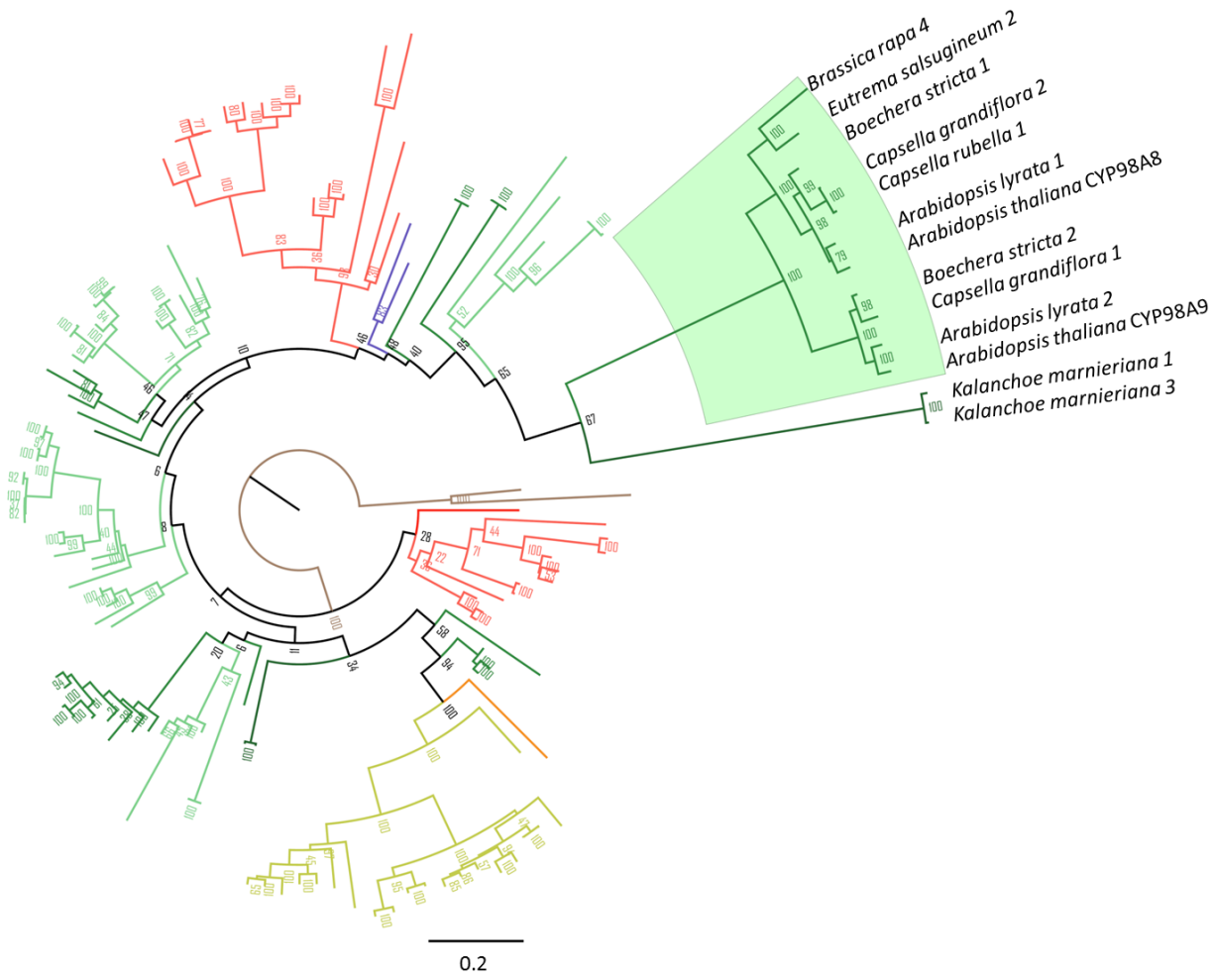


Figure 3.4 Phylogenetic reconstruction of CYP98 sequences from angiosperms with sequenced genomes and characterized CYP98s.

Amino acid sequences of CYP98s from angiosperm genomes and characterized CYP98s were included in this phylogenetic reconstruction. A distinct clade of Brassicaceae, containing the *A. thaliana CYP98A8* and *CYP98A9* is highlighted in green and species names are indicated. The amino acid alignment was performed by DIALIGN (Morgenstern, 1999), keeping positions above zero diagonal similarity. The maximum likelihood phylogenetic reconstruction was performed by phyml (Guindon and Gascuel, 2003), under consideration of the JTT model. Bootstrap support for 100 replicates is displayed at the branches of the phylogenetic reconstruction and also available in the supplement (Figure 3.23)

The *CYP98A8/9* clade was likely generated through gene duplication only within the Brassicaceae, despite its apparent distal location relative to *CYP98A3* in phylogenies including more distant outgroups (Matsuno, et al., 2009). Similar to the duplication and evolution of the *A. thaliana CYP98A8/CYP98A9* and presumably further Brassicaceae species, a duplication and evolution of the two *CYP98*s towards distinct functions in *K. marnieriana* might be possible. Indeed, *Kalanchoe* species are known to produce many bioactive natural plant products, including hydroxycinnamic conjugates (Gaind and Gupta, 1973; Pattewar, 2012). The Brassicaceae *CYP98A8/CYP98A9* and *K. marnieriana CYP98* clades were excluded from further phylogenetic analysis. These genes presumably provide unique evolutionary events that cannot be easily reconstructed using current phylogenetic models and algorithms.

The realisation of a rooted phylogenetic analysis excluding the above described clades resulted in the formation of distinct clades with strong bootstrap support, but only on the level of species or families (Figure 3.5). The statistical support for deciphering the relationship between these different clades within the angiosperms remained poor. Nevertheless, for some taxonomic groups, multiple clades were apparent in the phylogeny. Some of these clades split the isoforms from the same species, for example apple (“Maldo”) and citrus (“Citcl”) or tobacco (“Nicta”) and coffee (“Cofca”). Colour coding of the tree visualizes other clades, which include all gene family members from a subset of species, for example the Brassicales or the Populus/Salix clade.

CYP98 trees, which show the evolutionary history of the family, were very difficult to reconstruct. It is likely that molecular clocks in different species or families are running at extremely different paces. It is indeed known that trees, e.g. *P. trichocarpa* evolve slower than species with shorter life cycle and different reproduction strategy, e.g. *A. thaliana* (Tuskan et al., 2006)

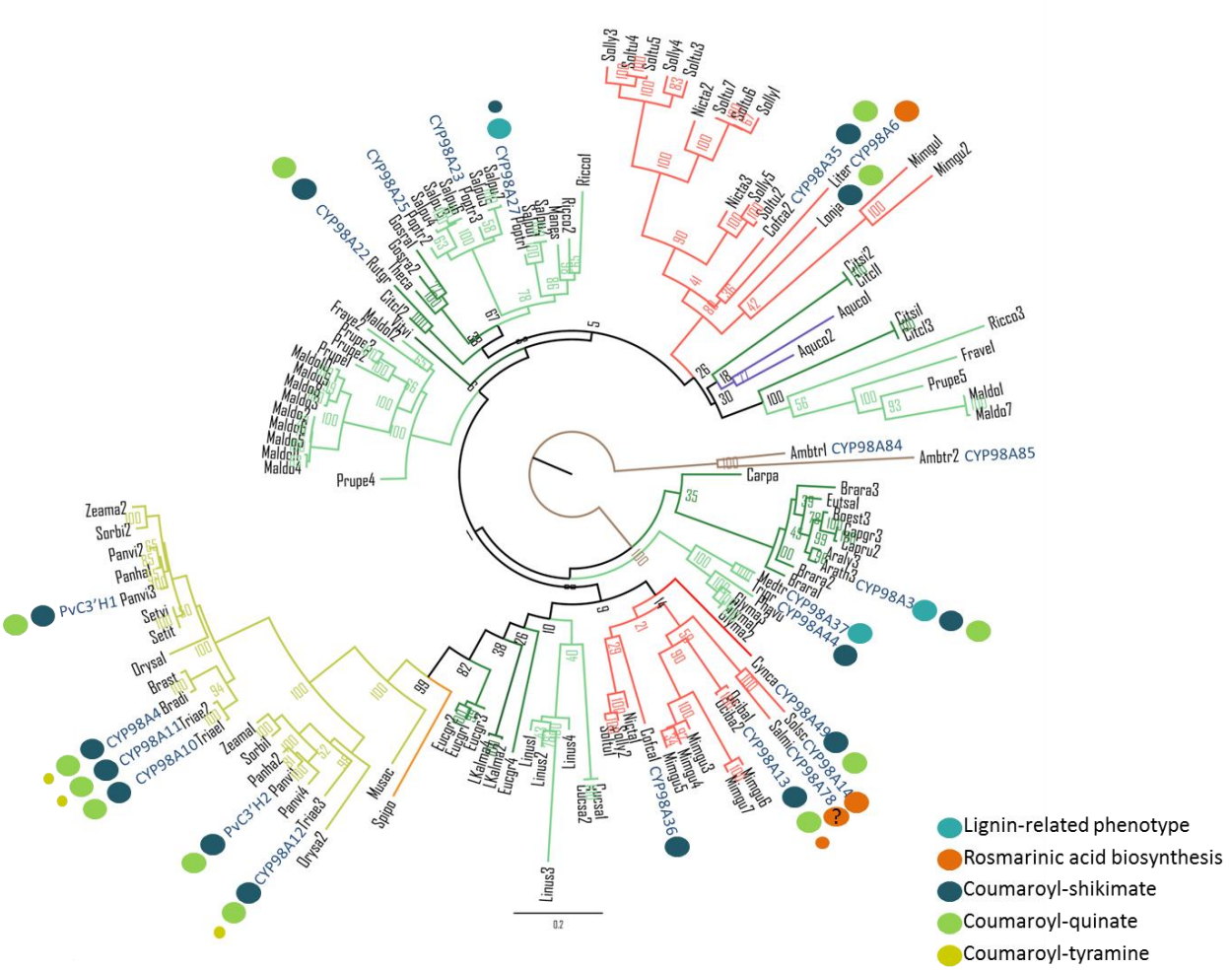


Figure 3.5 Phylogenetic reconstruction of characterized *CYP98* genes and *CYP98* genes of species with sequenced genomes.

A nucleotide alignment of *CYP98* sequences was generated using DIALIGN (Morgenstern, 1999), keeping alignment positions with diagonal similarities above zero. Maximum likelihood phylogenetic reconstruction was performed by phyml (Guindon and Gascuel, 2003), based on the HKY85 model. Statistical support of the phylogenetic reconstruction was obtained by bootstrapping in 100 replicates. The color coding of the branches follows the APG IV order classification as displayed in Figure 3.2. Species names abbreviations are given on the branches and a detailed species list can be found in the supplemental Table 3.5. The alignment is given in the appendix. Assigned CYP98 numbers are displayed in dark blue after the species names. Functions of characterized CYP98s as found in literature and listed in Table 3.1 are indicated by a coloured dot.

.

For most species, there was one isoform in a clade with very short branch length, while other members resided in clades characterized by much longer branch length, thus have undergone more changes since the separation from the last common ancestor. This can be due to varying selection pressures acting on different isoforms. It is possible that the apparent differences in branch length within different clades support this. As some CYP98s have been described to be involved in reactions with a variety of substrates, their substrate active sites might be larger, reflecting the relative substrate promiscuity of the enzymes, which might explain the observed phylogenetic modalities of the CYP98 family.

Nevertheless, there was no major separation into two clades, which would have been expected if there had been an ancient separation of lignin- and soluble-phenolic specific isoforms within the angiosperms. Instead, duplications of *CYP98* are lineage specific and appear to have occurred in multiple lineages. Thus, within the angiosperms the gene family has most likely undergone an intense history of gene duplications and gene losses following a rapid diversification early in angiosperm evolution. Good statistical support exists for monophyly of the *CYP98* families in *A. trichopoda* and *P. trichocarpa*, respectively. These duplications clearly happened independently.

Polyphenol oxidases (PPOs) in plants are enzymes that are involved in the browning reaction of tissues upon damage. The enzyme family is thought to be involved in plant defense. An *in silico* study across 25 land plant genomes identified *PPOs* throughout all land plant genomes investigated, except in the *A. thaliana* genome (Tran et al., 2012). Similar to what we found here for the *CYP98* family, Tran et al., 2012 describe varying size of the *PPO* family between species. They further describe several lineage-specific *PPO* gene family expansions and also gene loss. They conclude that the *PPO* family play many potential roles in the adaptions of plants to their environment. This variety of potential functions is consistent with the dynamic nature of the *PPO* family

Another example of a gene family with evolutionary plasticity is the terpene synthase (TPS) family. TPS gives rise to terpenes such as sterols and carotene (primary metabolism), but mainly to terpenes that are natural products (“secondary metabolites”). Similar to the *CYP98* gene family, the *TPS* family shows lineage and even species specific family expansion. Variation of

family size in the *TPS* family is large, from one gene in *P. patens* (Hayashi et al., 2006) to potentially 152 in *V. vinifera* (Martin et al., 2010). TPS family members which are closely related according to their amino acid sequence, can vary in their function *in planta* (Nagegowda et al., 2008).

An investigation of *CYP98* duplications within the Salicaceae family was performed. One WGD event is described specific to the Salicaceae (Tuskan et al., 2006), shared by *Populus* and *Salix* species, referred to as the "salicoid" duplication event. A phylogenetic reconstruction of the *CYP98* homologs across the Salicaceae included *Azara*, a species that separated prior to the salicoid WGD (Cronk et al., 2015). As found after aTRAM (Allen et al., 2015) assembly of RNAseq data (Allen et al., 2015), *Azara* features a single orthologue of each *CYP98A23/25* and *CYP98A27* (Figure 3.6). The duplication giving rise to CYP98A27 and to the common CYP98A23/A25 ancestor must thus have happened prior to the salicoid WGD. This is further confirmed by the analysis of synteny in *CYP98* regions. While *CYP98A27* is located on chromosome 6, *CYP98A23* and *CYP98A25* are located on chromosome 16, in close proximity of each other. These regions do not correspond to paralogous blocks generated by the salicoid WGD (Tuskan et al., 2006). Considering the tandem location of *CYP98A23* and *CYP98A25*, also this pair is not expected to result from the WGD, but instead from a local segmental duplication. Based on the phylogeny (Figure 3.6), the duplication event giving rise to *CYP98A23* and *CYP98A25* must have happened after the salicoid WGD. Similar to the case in poplar, very recent gene duplications occurred in its sister genus, *Salix*. As the duplication giving rise to *CYP98A23/25* and *CYP98A27* happened before the salicoid WGD, both *CYP98s* likely duplicated in the salicoid WGD. However neither the *CYP98A27* nor the *CYP98A23/25* duplicates were retained. Instead, a subsequent local tandem duplication then gave rise to *CYP98A23* and *CYP98A25*.

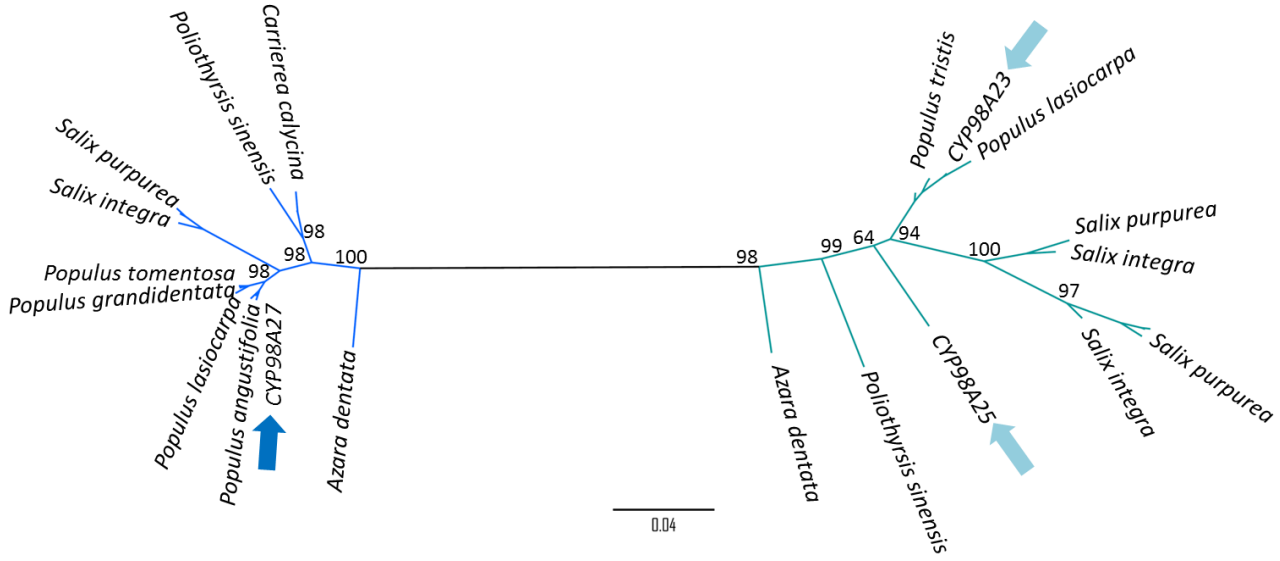


Figure 3.6 Phylogenetic reconstruction of CYP98 nucleotide sequences of the Salicaceae.

Nucleotide sequences of CYP98s of the Salicaceae were manually aligned and maximum likelihood phylogenetic reconstruction was performed using phyml, under the GTR substitution model. Statistical support was calculated by 100 bootstrap replicates. Lowest bootstrap support in the phylogenetic analysis was 64. *P.trichocarpa CYP98A27* is indicated by dark blue arrow, the *CYP98A27* containing clade is marked in dark blue. *P. trichocarpa CYP98A23* and *CYP98A25* are indicated by light blue arrows, and the *CYP98A23* and *CYP98A25* containing clade is marked in light blue. *Azara dentata* is representing a species that separated prior to the salicoid WGD (Cronk et al., 2015).

The Salicaceae *CYP98* sequences retrieved after aTRAM are highly similar based on sequence alignment. The Azara *CYP98A27* homologue showed 94% amino acid sequence identity and 92% nucleotide sequence identity to the *P. trichocarpa CYP98A27*. The Azara *CYP98A23* homologue showed 92% amino acid sequence identity and 90% nucleotide sequence identity to the *P. trichocarpa CYP98A23*. The alignment of sequences was therefore performed on the nucleotide sequences, where sequences were more divergent. Full sequence coverage of the CYP98 genes was not obtained. After manual alignment and inspection for identical sequences a reliable alignment of 800 nucleotide positions was obtained. Simply based on the observations from the phylogenetic tree, in the clade containing *CYP98A27*, the lignin related *P. trichocarpa* isoform, a higher degree of conservation was apparent, with shorter branch length

and therefore fewer changes since the split from the last common ancestor. Within this clade the median identity of sequences was 97%, compared to a median identity of 90% in the CYP98A23/25 clade. The *P. trichocarpa CYP98A23/CYP98A25* containing clade was characterized by longer branch length, more divergence and higher variability (Figure 3.6). Even though the alignment did only cover about half of the length of the gene, a calculation of non-synonymous substitutions per non-synonymous site (d_N/d_S) was performed across all branches of the tree. This global analysis showed that the clade of the tree with shorter branches, containing the lignin-related *P. trichocarpa CYP98A27,* showed less non-synonymous substitutions per non-synonymous site (ω=0.1) and the clade of the tree including the *P. trichocarpa CYP98A23* and *CYP98A25* genes showed comparably more non-synonymous substitutions per non-synonymous site (ω=0.4). These results suggest strong purifying selection pressures to act on the *CYP98A27*-like clade, while the selection pressure on the *CYP98A23/25*-like clade appeared more relaxed when considering the full sequence, indicating either overall relaxed selection or positive selection at only a few sites. These results are in accordance with the phylogenetic tree, but nevertheless have to be interpreted carefully, as the sequences are not complete, the sequences are highly similar and the analysis includes the complete *CYP98A27*-like and *CYP98A23/25*-like clades respectively.

Taken together this suggests a minimum of three duplications and two gene losses in the lineage giving rise to the three *P. trichocarpa CYP98s*. Frequent, independent duplications and gene losses seem to be a main characteristic of *CYP98* evolution within the angiosperms. This was supported by mapping previously published functional genetic and enzyme activity data onto the phylogenetic tree. There was no separation of “lignin” and “soluble phenolic” related enzyme activities on the phylogeny across angiosperm species (see coloured dots in Figure 3.5).

As mentioned above, *P. trichocarpa* has three *CYP98* members. *CYP98A27* is located on chromosome 6 and *CYP98A23* and *CYP98A25* are located on chromosome 16 in tandem. *A. trichopoda* has two *CYP98* isoforms, *CYP98A84* and *CYP98A85*, located on two different genome sequence scaffolds. For sequence similarities, also in comparison with *CYP98A3* of *A. thaliana*, see Table 3.2. *A. trichopoda* CYP98A84 and CYP98A85 share only 65% sequence identity, but

nevertheless were more closely related to each other than to any other angiosperm CYP98 based on the phylogenetic reconstructions.

	A.th CYP98A3	*P.tr* CYP98A27	*A.tr* CYP98A84	*P.tr* CYP98A23	*P.tr* CYP98A25	*A.tr* CYP98A85
A. th CYP98A3	ID	81%	73%	75%	76%	63%
P. tr CYP98A27	81%	ID	78%	81%	83%	66%
A. tr CYP98A84	73%	78%	ID	72%	74%	65%
P. tr CYP98A23	75%	81%	72%	ID	92%	61%
P. tr CYP98A25	76%	83%	74%	92%	ID	62%
A. tr CYP98A85	63%	66%	65%	61%	62%	ID

Table 3.2 Amino acid sequence identities of *A. thaliana* CYP98A3, *P. trichocarpa* CYP98s and *A. trichopoda* CYP98s.

Amino acid sequences were aligned by ClustalW (Thompson et al., 1994) and a sequence identity matrix calculated in the program BioEdit (Hall, 1999).

A search against the Phytozome version 11 database revealed two *Salix purpurea* CYP98s, followed by the *P. trichocarpa* CYP98A27, as most similar protein homologs of *A. trichopoda* CYP98A84. The closest protein homolog to *A. trichopoda* CYP98A85, however, is *A. trichopoda* CYP98A84, consistent with their placement in the phylogeny.

As the duplications in *A. trichopoda* and *P. trichocarpa* clearly occurred independently, the CYP98s of these two species were selected for further analysis.

3.4.2. Enzymatic diversity of CYP98 duplicates in Amborella and poplar

The two CYP98s of *A. trichopoda* CYP98A84 and CYP98A85, and the three CYP98s of *P. trichocarpa* CYP98A23, CYP98A25 and CYP98A27, were heterologously expressed in *Saccharomyces cerevisiae.* Recording differential spectra of reduced/CO associated and reduced microsomes containing the CYP98 indicated the presence of a functional P450 enzyme for four CYP98s: *A. trichopoda* CYP98A84 and CYP98A85, and *P. trichocarpa* CYP98A23 and CYP98A27. CYP98A25 of *P. trichocarpa* only showed an absorption peak at 420 nm, which suggested that the enzyme was poorly expressed and most likely not stable (Imai and Sato, 1967)(Figure 3.7).

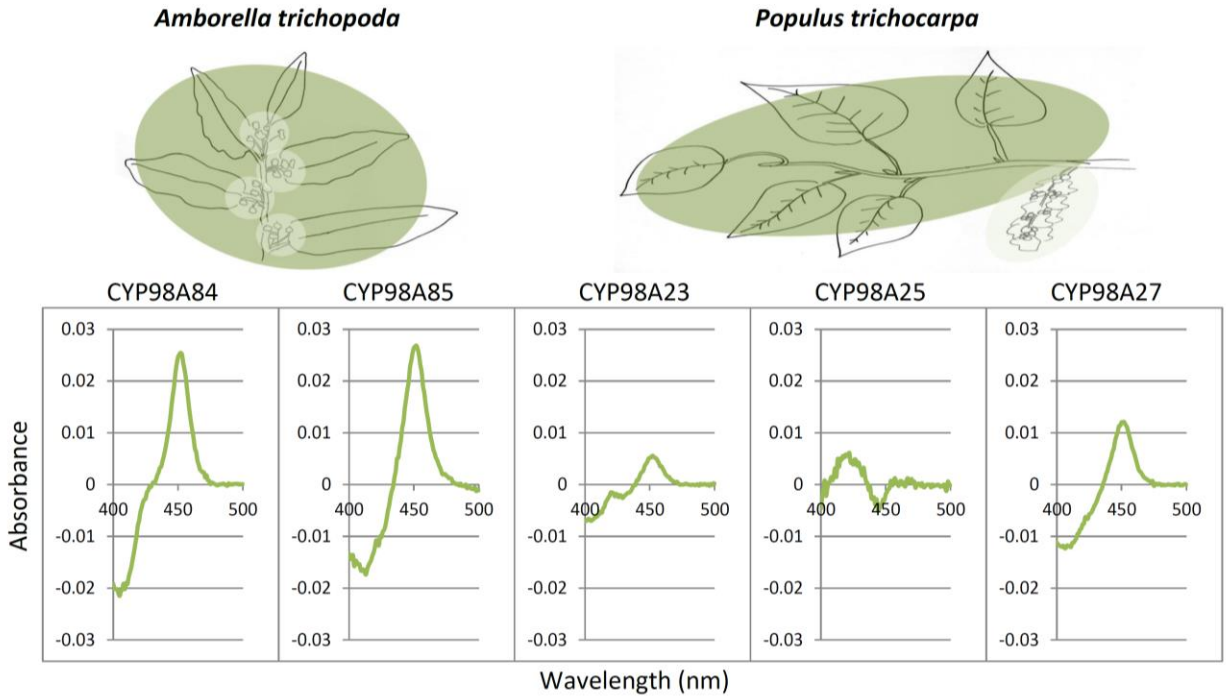


Figure 3.7 Differential CO spectra of CYP98s included in the biochemical analysis.

CYP98 enzymes were expressed in yeast and yeast microsomes isolated. When reduced and associated with CO, a peak at 450nm absorbance in a differential spectrum is indicative of functionally expressed enzyme and allows for its quantification.

P. trichocarpa CYP98A25 was subsequently expressed from several different transformation constructs (*P. trichocarpa CYP98A25* in pYeDP60) and various independent transformations, with no amelioration of the P450 CO differential spectrum (Figure 3.20, Supplement).

Yeast microsomes containing the *A. trichopoda* and *P. trichocarpa* CYP98s were investigated in an enzyme substrate screening. *Populus* species are known for their very rich soluble ester repertoire (English et al., 1991). Some of the substrates, tested in the substrate screening (compare Chapter 2, Figure 2.4) and/or their hydroxylated form, are known to occur in *P. trichocarpa*. Little is known about phenolic compounds existing in *A. trichopoda* to date. This is partly due to its endemic occurrence in New Caledonia.

Endpoint screening assays were performed as described in Chapter 2 (Figure 2.5) using the same set of 30 substrates (Figure 2.4). As may have been expected from the differential spectrum analyses, the *P. trichocarpa* CYP98A25 did not show any detectable substrate

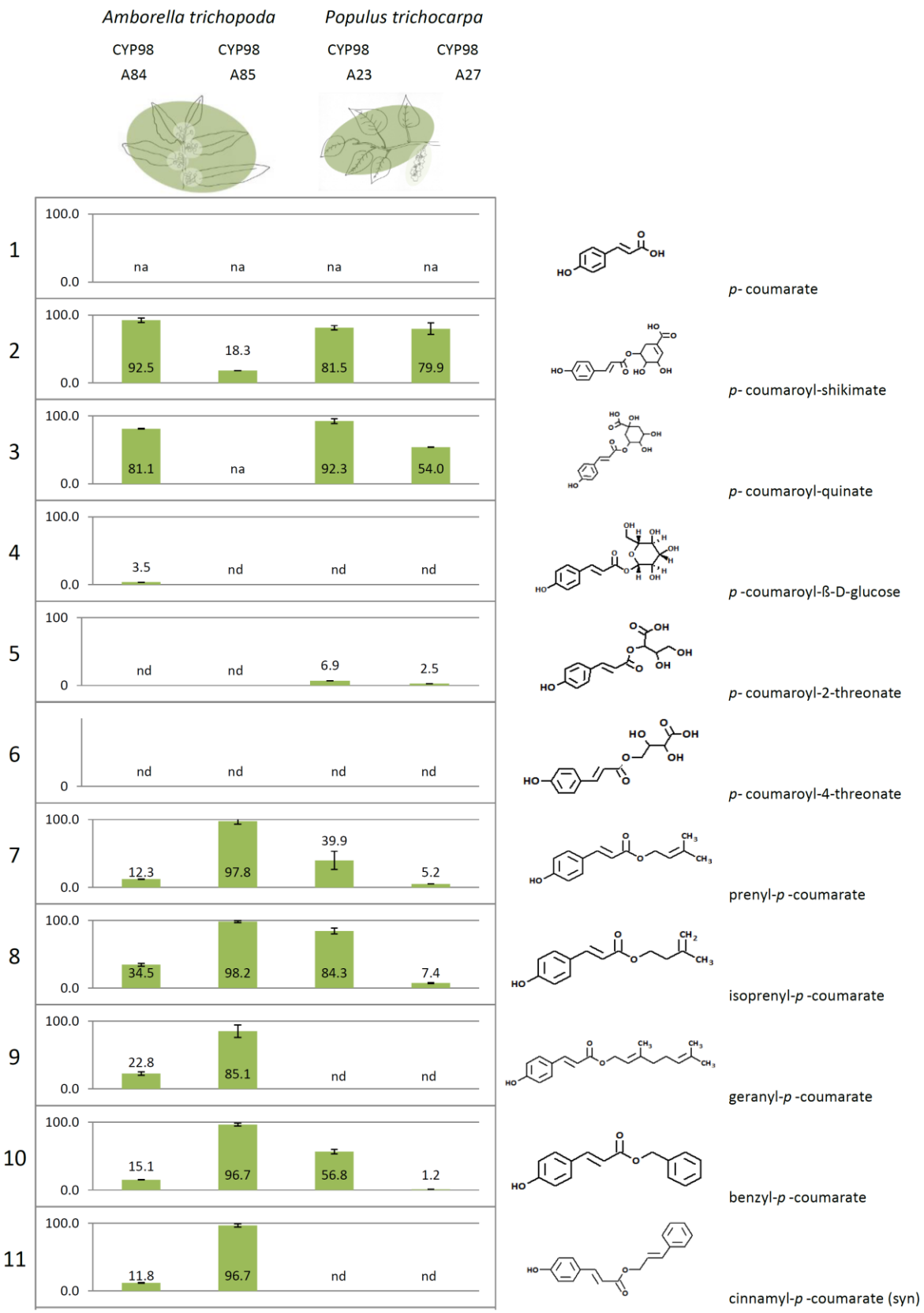
conversion with any of the substrates tested. In an attempt to express the enzyme in a different system, a pre-test for a transient over-expression in *Nicotiana benthamiana* leafs was performed. Leaf discs were cut and incubated with potential substrates. However, the presence of endogenous substrate in *N. benthamiana* leaves and the presences of endogenous *N. benthamiana* CYP98s did not allow for a proper set up of this assay (Figure 3.21, Supplement).

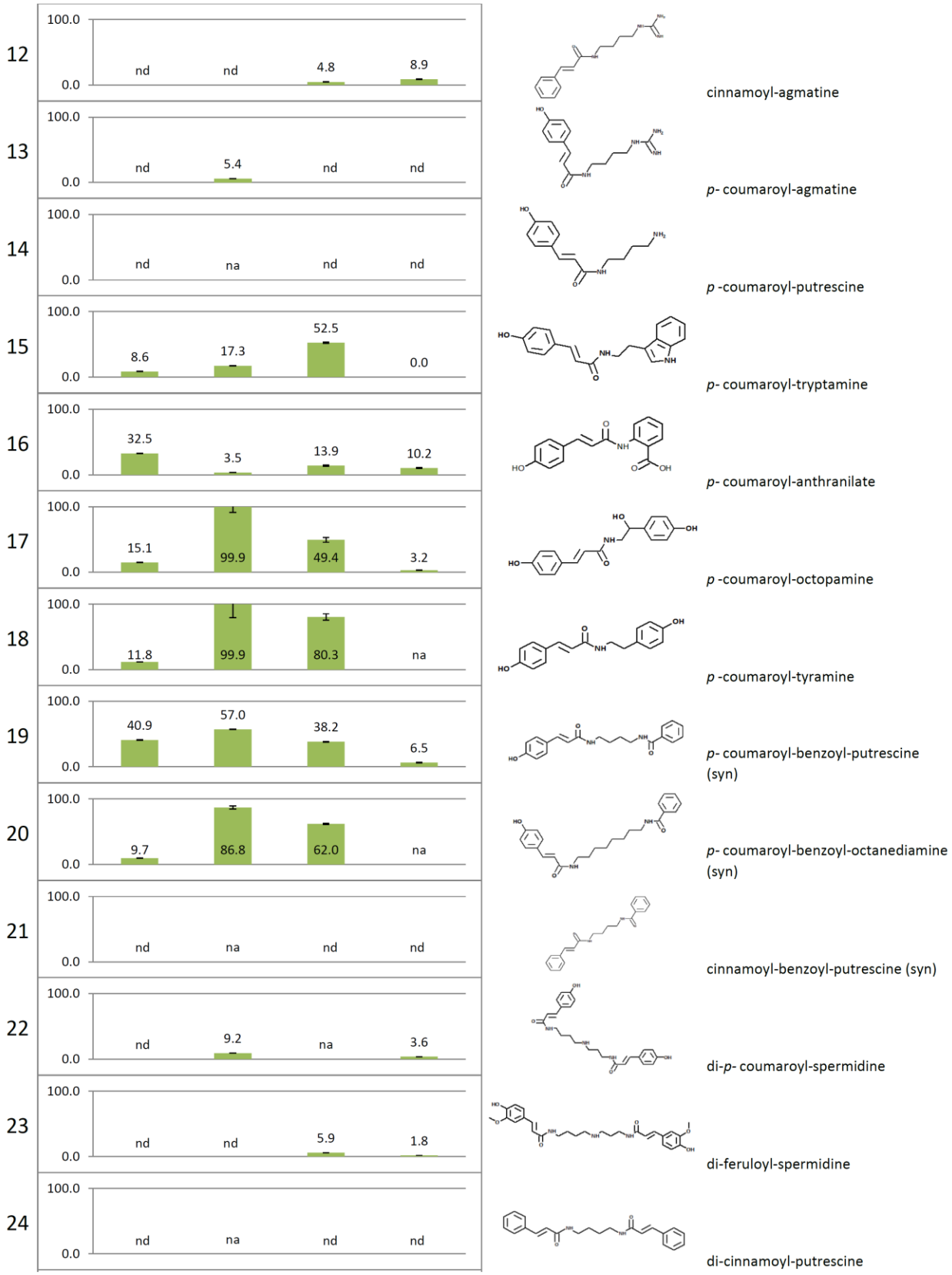
Both naturally occurring and synthetic substrates were utilized by the four remaining CYP98s *in vitro* (Figure 3.8; Figure 3.9). For none of the enzymes conversion of free coumaric acid (1; numbers refer to numbers of substrates as listed in Figure 2.4) was detected. Cinnamoyl-agmatine (12) showed very low conversion rates when incubated with the *P. trichocarpa* CYP98s, but generally cinnamoyl-conjugates were not converted by any of the enzymes tested. CYP98A27 of *P. trichocarpa* and CYP98A84 of *A. trichopoda* were rather specific to *p*-coumaroyl-shikimate and *p*-coumaroyl-quinate, whereas the CYP98A23 of *P. trichocarpa* and the CYP98A85 of *A. trichopoda* showed much broader substrate ranges.

Nevertheless all four CYP98s were capable of converting *p*-coumaroyl-shikimate (2), but with different apparent conversion rates. While *P. trichocarpa* CYP98A23 and CYP98A27 converted approximately the same amount of *p*-coumaroyl-shikimate to caffeoyl-shikimate in the end point screening experiment (Figure 3.8) (about 80% conversion rate), differences were found for the *A. trichopoda* CYP98s. *A. trichopoda* CYP98A84 showed similar rates as the poplar CYP98s, while *A. trichopoda* CYP98A85 only converted about 18% of *p*-coumaroyl-shikimate. The latter enzyme also did not show any activity with *p*-coumaroyl-quinate (3). In contrast, *P. trichocarpa* CYP98A23 and *A. trichopoda* CYP98A84 converted more than 80% of *p*-coumaroyl-quinate to chlorogenic acid. The lignin-related CYP98A27 of *P. trichocarpa* converted less *p*-coumaroyl-quinate than *P. trichocarpa* CYP98A23 and *A. trichopoda* CYP98A84, about 50%. Prenyl-*p*-coumarate (7), isoprenyl-*p*-coumarate (8) and benzyl-*p*-coumarate (10), which are all known to occur naturally in poplar trees, were converted efficiently by *P. trichocarpa* CYP98A23, while *P. trichocarpa* CYP98A27 only showed very low rates of conversion with these substrates, below 7%. The same substrates were also efficiently converted by the CYP98A85 of *A. trichopoda*, while the CYP98A84 of *A. trichopoda* showed only low conversion rates of these substrates. The phenolamide *p*-coumaroyl-anthranilate (16), the best substrate for a lycopod

and bryophyte CYP98 (refer to Chapter 2), was only converted by CYP98A84 of *A. trichopoda* in appreciable amounts (32.5%). Conversion of *p*-coumaroyl-anthranilate to caffeoyl-anthranilate by the other three CYP98s stayed below 14%. Further phenolamides, such as *p*-coumaroyl-octopamine (17), *p*-coumaroyl-tyramine (18) and tri-coumaroyl-spermidine (28) showed the same conversion patterns: *A. trichopoda* CYP98A85 converted the substrate completely or to high extent, while *A. trichopoda* CYP98A84 only converted small amounts of substrate. Likewise the *P. trichocarpa* CYP98A23 converted these substrates efficiently, while the conversion rates of *P. trichocarpa* CYP98A27 stayed close to the detection limit.

A hierarchical cluster analysis of the substrate conversion rates of the four CYP98s further supported the trend already seen above (Figure 3.8): based on the biochemical functions of the enzymes, the *P. trichocarpa* CYP98A27, which was shown to be involved in the biosynthesis of monolignols (Chen et al., 2011), and the CYP98A84 of *A. trichopoda*, formed a tight group with very similar substrate preferences (Pearson Correlation Coefficient r=0.92. For an overview of Pearson Correlation Coefficients see Supplement Table 3.6). Compared to the other two CYP98s investigated, their substrate range was rather narrow, with emphasis on the *p*-coumaroyl-shikimate and *p*-coumaroyl-quinate ester conversion. While the second *P. trichocarpa* isoform, CYP98A23, was also able to convert *p*-coumaroyl-shikimate efficiently *in vitro*, it had a much broader range of substrates accepted. Even broader was the range of substrates converted by the *A. trichopoda* CYP98A85, except for the notable absence of appreciable *p*-coumaroyl-shikimate and *p*-coumaroyl-quinate conversion. The substrate utilization profiles of *P. trichocarpa* CYP98A23 and *A. trichopoda* CYP98A85 and the two *A. trichopoda* isoforms to each other (r=0.22 and -0.39, respectively) are not correlated. In contrast, phylogenetic analysis showed no clade separation by function of the CYP98s, but by species instead. This suggests that either the broad-range substrate utilization ability of *P. trichocarpa* CYP98A23 and *A. trichopoda* CYP98A85 evolved independently or that the *p*-coumaroyl-shikimate specificity for the lignin-related isoforms evolved independently in Amborella and in the lineage leading to Populus.


Amborella trichopoda
Populus trichocarpa
CYP98 A84
CYP98 A85
CYP98 A23
CYP98 A27
1
100.0
0.0
na na na na
2
100.0
0.0
92.5 18.3 81.5 79.9
3
100.0
0.0
81.1 na 92.3 54.0
4
100.0
0.0
3.5 nd nd nd
5
100
0
nd nd 6.9 2.5
6
0
nd nd nd nd
7
100.0
0.0
12.3 97.8 39.9 5.2
8
100.0
0.0
34.5 98.2 84.3 7.4
9
100.0
0.0
22.8 85.1 nd nd
10
100.0
0.0
15.1 96.7 56.8 1.2
11
100.0
0.0
11.8 96.7 nd nd
p- coumarate
p- coumaroyl-shikimate
p- coumaroyl-quinate
p -coumaroyl-ß-D-glucose
p- coumaroyl-2-threonate
p- coumaroyl-4-threonate
prenyl-p -coumarate
isoprenyl-p -coumarate
geranyl-p -coumarate
benzyl-p -coumarate
cinnamyl-p -coumarate (syn)


12
100.0
0.0
nd
nd
4.8
8.9
cinnamoyl-agmatine
13
100.0
0.0
nd
5.4
nd
nd
p- coumaroyl-agmatine
14
100.0
0.0
nd
na
nd
nd
p -coumaroyl-putrescine
15
100.0
0.0
8.6
17.3
52.5
0.0
p- coumaroyl-tryptamine
16
100.0
0.0
32.5
3.5
13.9
10.2
p- coumaroyl-anthranilate
17
100.0
0.0
15.1
99.9
49.4
3.2
p -coumaroyl-octopamine
18
100.0
0.0
11.8
99.9
80.3
na
p -coumaroyl-tyramine
19
100.0
0.0
40.9
57.0
38.2
6.5
p- coumaroyl-benzoyl-putrescine (syn)
20
100.0
0.0
9.7
86.8
62.0
na
p- coumaroyl-benzoyl-octanediamine (syn)
21
100.0
0.0
nd
na
nd
nd
cinnamoyl-benzoyl-putrescine (syn)
22
100.0
0.0
nd
9.2
na
3.6
di-p- coumaroyl-spermidine
23
100.0
0.0
nd
nd
5.9
1.8
di-feruloyl-spermidine
24
100.0
0.0
nd
na
nd
nd
di-cinnamoyl-putrescine

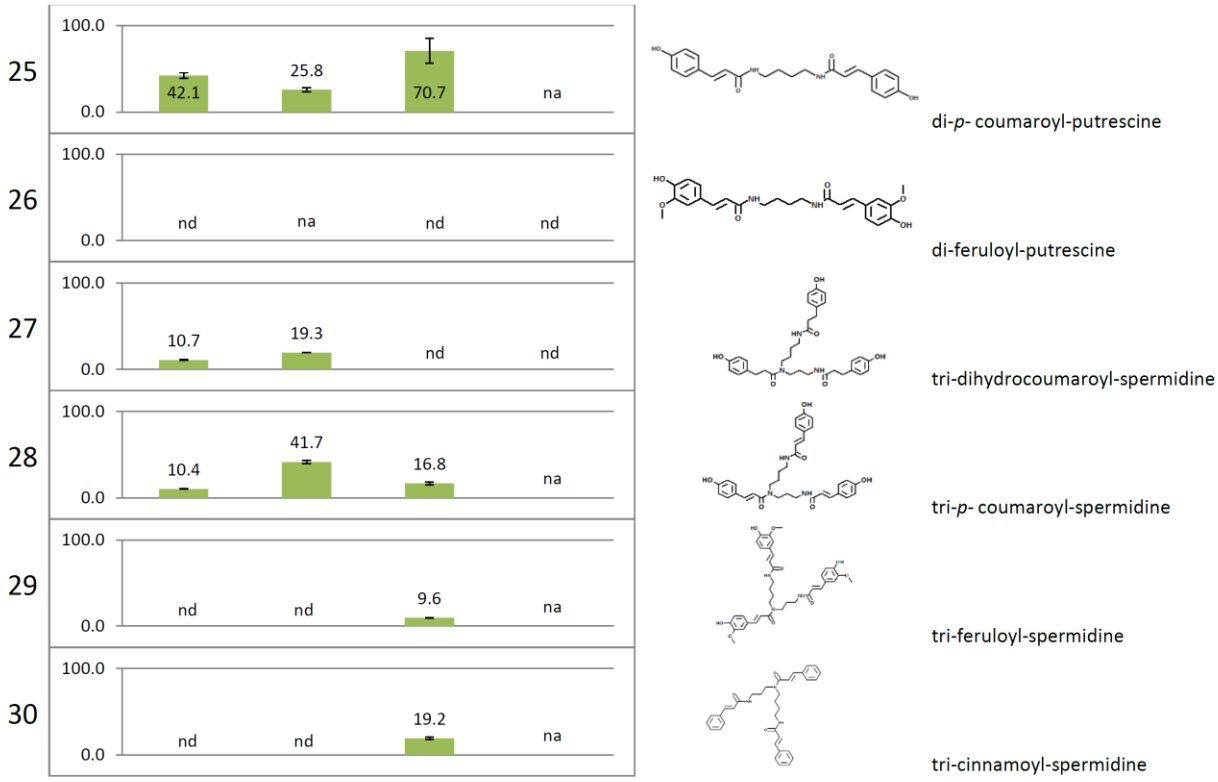


Figure 3.8 Substrate conversion rates obtained in end-point enzyme incubations.

The substrate conversion rates of *P. trichocarpa* and *A. trichopoda* CYP98s with 30 potential substrates are listed. Substrate numbering corresponds to Figure 2.4. Substrate structures and trivial names (if existing) are given on the right. No apparent conversion is indicated by na (no activity). Combinations that were not tested are indicated by nd (not determined). In the end point screening assay, 10 pmol of P450 enzyme were incubated with 100µM of substrate (expected to be saturating) and 500µM NADPH, for 30 minutes at 28°C under agitation in the dark.

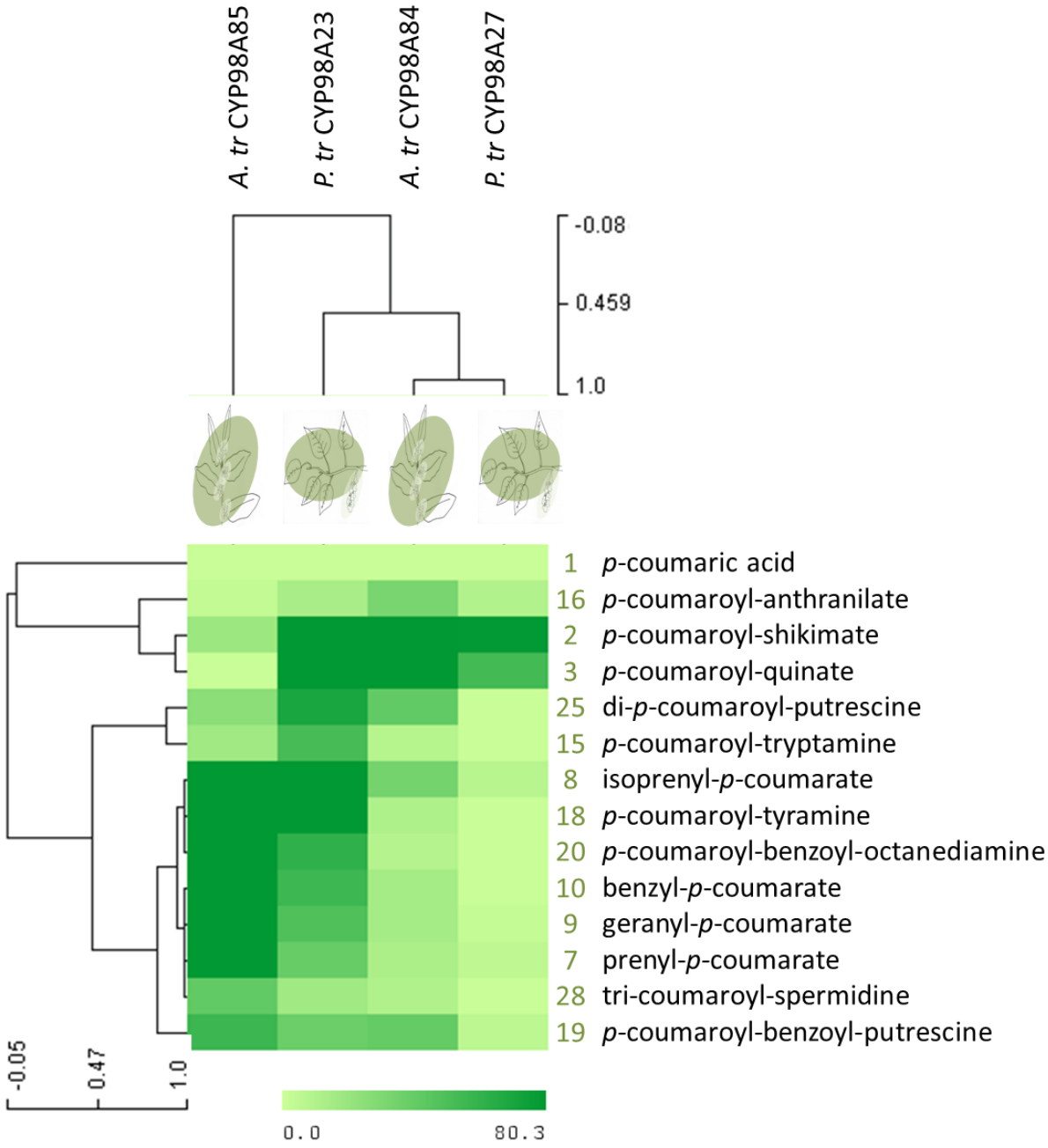


Figure 3.9 Hierarchical clustering of substrates and P450s tested in the substrate screening.

Average linkage clustering by Pearson Correlation. The corresponding substrate conversion rates are presented in detail in Figure 3.8. In the end point screening, 10 pmol of P450 enzyme were incubated with 100 µM of substrate (expected to be saturating) and 500µM NADPH for 30 minutes at 28°C, under agitation, in the dark.

The biochemical characterization of the *P. taeda* CYP98 showed a broad range of substrates accepted by this enzyme *in vitro* (refer to CYP98A19 of *Pinus taeda* in Chapter 2). Gene expression data is publically available for the conifer *Picea abies*. The *P. abies CYP98* showed high expression in two different organs, the vegetative shoots and the wood of the trunk of the conifer (Figure 3.10). The high expression of the CYP98 gene in the wood of the conifer might be linked to a role in lignin biosynthesis. The *P. taeda* CYP98 converted *p*-coumaroyl-shikimate (82% substrate conversion in the end-point enzymatic screen) and *p*-coumaroyl-quinate, albeit to a lower rate (38% substrate conversion rate). The substrate utilization profile of the pine CYP98A19 is more similar to CYP98A84 of *A. trichopoda* (r=0.71), does not correlate to CYP98A85 of *A. trichopoda* (r=0.03) or to CYP98A23 of *P. trichocarpa* (r=0.42) and only to lower extent to CYP98A27 (r=0.57). It is thus likely that the ancestral angiosperm CYP98 was also capable of converting *p*-coumaroyl-shikimate, or might have even been specific for *p*-coumaroyl-shikimate. This would also be in accordance with the expression of the single *CYP98* of *P. abies* in the wood of the tree. Substrate broadening thus seems to have evolved independently in the different lineages, with isoforms resulting from independent duplications. Coumaroyl-shikimate hydroxylation activity most likely emerged early, between the emergence of the seed plant CYP98 ancestor and the angiosperm CYP98 ancestor. The most parsimonious conclusion in this scenario would be, that after independent CYP98 duplications, repeatedly occurring relaxation of purifying selection pressure would have lead to the function of the non-lignin related CYP98 isomers.

The *A. trichopoda* CYP98A85 might have lost its *p*-coumaroyl-shikimate/quinate hydroxylating activity. No *A. trichopoda* gene expression data is publically available to data. It would be interesting to compare the expression of the two Amborella genes, which might be expressed in different tissues in correlation to their functions *in vivo*.

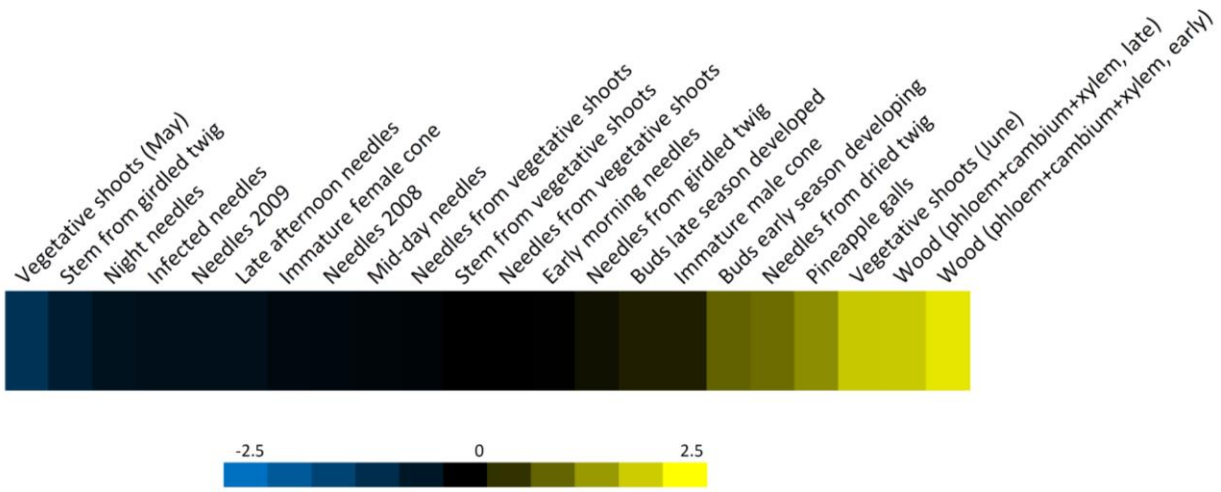


Figure 3.10 ***Picea abies CYP98*** **gene expression analysis.**

The expression of the *P. abies CYP98* gene MA_109548g0010 was identified from publically available gene expression data (Nystedt et al., 2013). Data is presented in expression of the gene relative to the mean expression across all samples.

3.4.3. *A. trichopoda* and *P. trichocarpa* CYP98 substrate recognition sites

To detect possible common structures or differences of phylogenetically distinct but functionally similar CYP98s, the putative substrate recognition sites of the *A. trichopoda* and *P. trichocarpa* CYP98s were highlighted in an amino acid sequence alignment. Putative substrate recognition sites have been described for CYP98 enzymes by (Matsuno, et al., 2009). The sequences of the three poplar and two Amborella enzymes were aligned manually, and the substrate recognition sites and the common P450 structural motifs highlighted according to the homologous sites of CYP98A3 from *A. thaliana* (Matsuno, et al., 2009) (Figure 3.11).

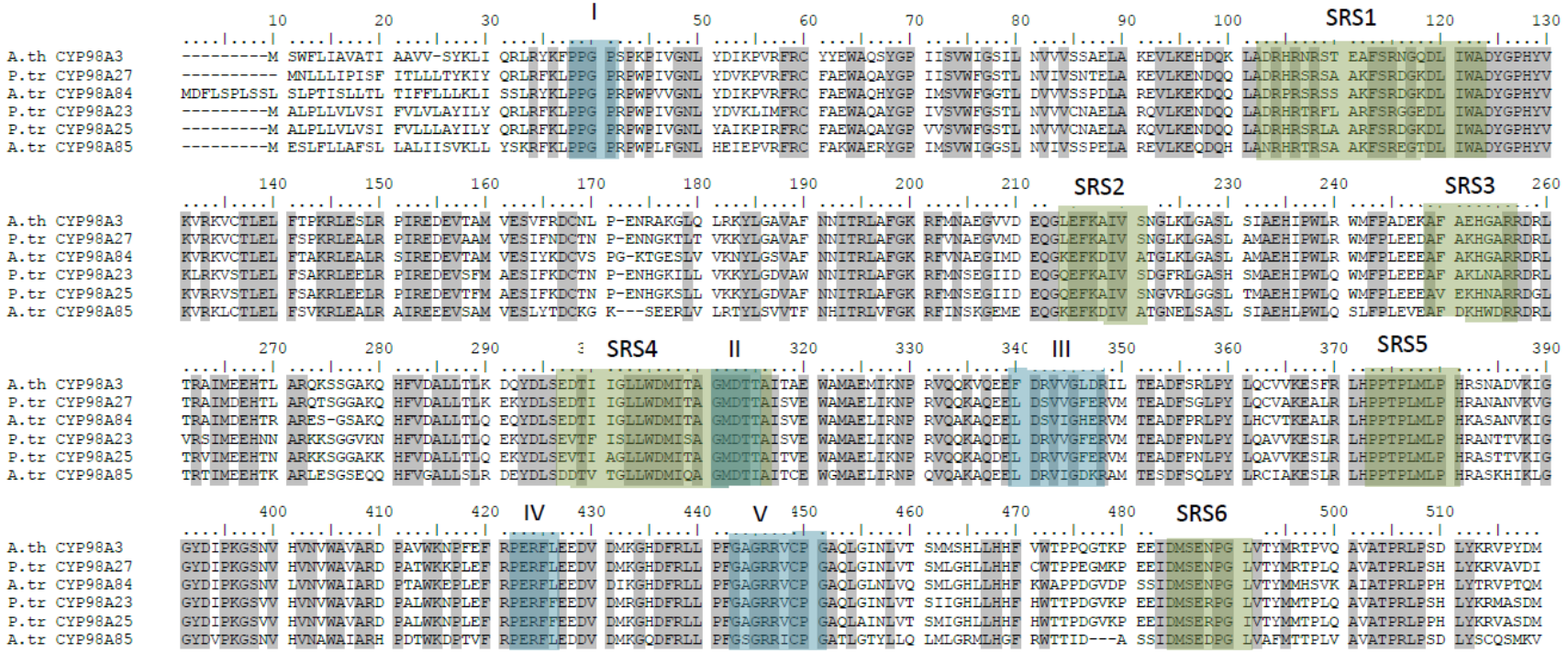


Figure 3.11 CYP98 putative substrate recognition sites and conserved P450 structural motifs.

Amino acid sequences of *A. thaliana* CYP98A3, *P. trichocarpa* CYP98A27*, A. trichopoda* CYP98A84, *P. trichocarpa* CYP98A23, *P. trichocarpa* CYP98A25 and *A. trichopoda* CYP98A85 were manually aligned and putative substrate recognition sites (SRS; green boxes) and common P450 structural motifs highlighted (I to V; blue boxes), according to identified homologous sites in the *A. thaliana* CYP98A3 (Matsuno, et al., 2009). Structural P450 motifs: **I**: Proline rich membrane hinge, **II**: I-helix, **III**: ERR triad, **IV**: clade signature, **V**: Heme binding region. Grey shading indicates identical amino acid residues in the six CYP98s. The alignment was generated in BioEdit (Hall, 1999).

No sequence variation was observed in the proline-rich membrane hinge (I) among the sequences, but the other common P450 structural domains showed variation. CYP98A85 of *A. trichopoda*, differed in the sequence of the I-helix (II) and heme binding region (V) compared to the other four CYP98s. Variation among the sequences was also observed in the ERR triad (III) and the clade signature (IV). Variation was observed as well for the putative substrate recognition site 1, (SRS1), but no trend between *p*-coumaroyl-shikimate versus broad-range isoforms was apparent. The same holds true for SRS2. Only SRS3 showed conserved amino acids for the *p*-coumaroyl-shikimate specific isoforms and varying amino acids for the broad-range isoforms, with the CYP98A23 of *P. trichocarpa* being intermediate (Figure 3.11). In a less differentiated form, this was observed for SRS4 as well. Matsuno, et al., 2009 described the SRS5 as site in the vicinity of the phenylpropane unit, based on CYP98A3 structure modelling. This SRS was highly conserved in all isoforms investigated. SRS6 showed only one amino acid varying from all others, which again belonged to CYP98A85 of *A. trichopoda*, the broad range enzyme *in vitro*.

Albeit differences in the substrate recognition sites of the six CYP98s were apparent, there were very few amino acids common to each functional group yet different between groups: a trend corresponding to the biochemical substrate preferences of the enzymes was not detected. The broad-range isoforms CYP98A85 and CYP98A23 did not show the same changes when compared to the putative lignin related isoforms CYP98A27, CYP98A3, and CYP98A84.

These findings correlate with the phylogenetic analysis performed above. In the phylogenetic analysis (Figure 3.5) *CYP98s* of poplar and Amborella clustered by species, rather than by their substrate preferences determined in the end point screening (Figure 3.8).

Most frequently, gene duplicates acquire random mutations turning them into pseudogenes which do not get fixed in the population. Instead, CYP98 gene duplications seen in poplar and Amborella might have provided an advantage by dosage effects first, in which latent occurring functions of the CYP98 could evolve, because one gene copy performed the known function, while relaxation of selection pressure on the duplicated form allowed for substrate broadening. This might have coincided with a larger active site and a slightly more unstable protein. With careful interpretation, the CO spectra of CYP98A23 shown in Figure 3.7 gave hints to a rather

unstable P450. It could be interpreted as substrate broadening which is still ongoing. In accordance with this, the CYP98A85, which seems to have undergone more changes since the duplication from CYP98A84, showed a rather stable CO spectrum and an even broader substrate acceptance range than CYP98A23 except for the absence of activity with *p*-coumaroyl-quinate and much reduced activity with *p*-coumaroyl-shikimate. A relaxation of selection pressure might have occurred several times, because the *CYP98* genes in poplar and Amborella come from independent gene duplications. This lead to functionally similar enzymes in poplar and Amborella, but different *CYP98* gene sequences between the two species.

3.4.4. Enzyme kinetics, focusing on *P. trichocarpa*

To confirm data of the substrate screening, and also to further differentiate between isoforms that show the same substrate activity in the end point screening experiment, enzyme kinetics were performed for chosen *P. trichocarpa* enzyme/substrate pairs. Four substrates were chosen due to their natural occurrence in poplar trees.

Michaelis-Menten based enzyme kinetic constants were determined for the two *P. trichocarpa* CYP98s, CYP98A23 and CYP98A27, and the four coumaric esters isoprenyl-*p*-coumarate (8), benzyl-*p*-coumarate (10), *p*-coumaroyl-quinate (3) and *p*-coumaroyl-shikimate (2) (Figure 3.12; Figure 3.13; Table 3.3).

Caffeoyl-shikimate occurs in four distinct isomers. An initial test was performed to identify the *p*-coumaroyl-shikimate isomer accepted by CYP98A23 and CYP98A27. The preferred isomer was the *trans*-4-*O*-(*p*-coumaroyl)shikimate (Figure 3.24, Supplement).

CYP98A23 and CYP98A27 showed different Michaelis Menten based enzyme kinetic constants for the substrates tested. The calculations were based on nonlinear regression analysis. The graphs for CYP98A27 and CYP98A23 velocity at different substrate concentrations are shown in Figure 3.12 and Figure 3.13 respectively.

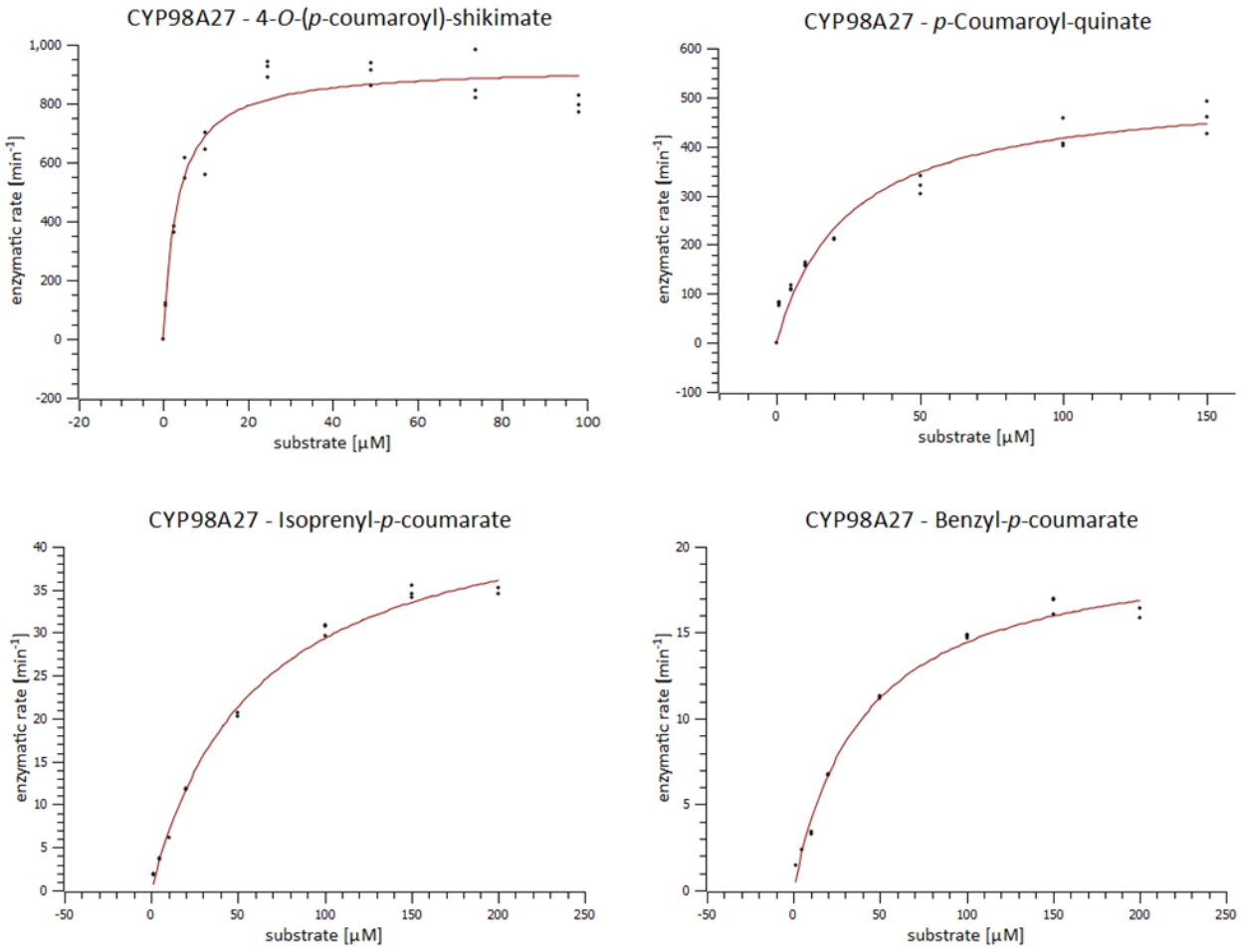


Figure 3.12 Michaelis Menten based enzyme kinetics for *P. trichocarpa* CYP98A27 with trans-4-*O*-(4-coumaroyl)shikimate, *p*-coumaroyl-quinate, isoprenyl-*p*-coumarate and benzyl-*p*-coumarate.

To obtain enzyme kinetic data for CYP98A27, 0.25 µM; 0.2 µM; 10 µM; 10 µM of CYP98A27 protein was incubated with *trans*-4-*O*-(4-coumaroyl)shikimate; *p*-coumaroyl-quinate, isoprenyl-*p*-coumarate and benzyl-*p*-coumarate, respectively. Reactions were analysed on HPLC/DAD. Product appearance was measured and linked to a standard curve. Non-linear regression of the Michaelis-Menten equation was fitted under the Nelder-Mead-Simplex algorithm in the program SciDavis. Shown are three independent incubation replicates.

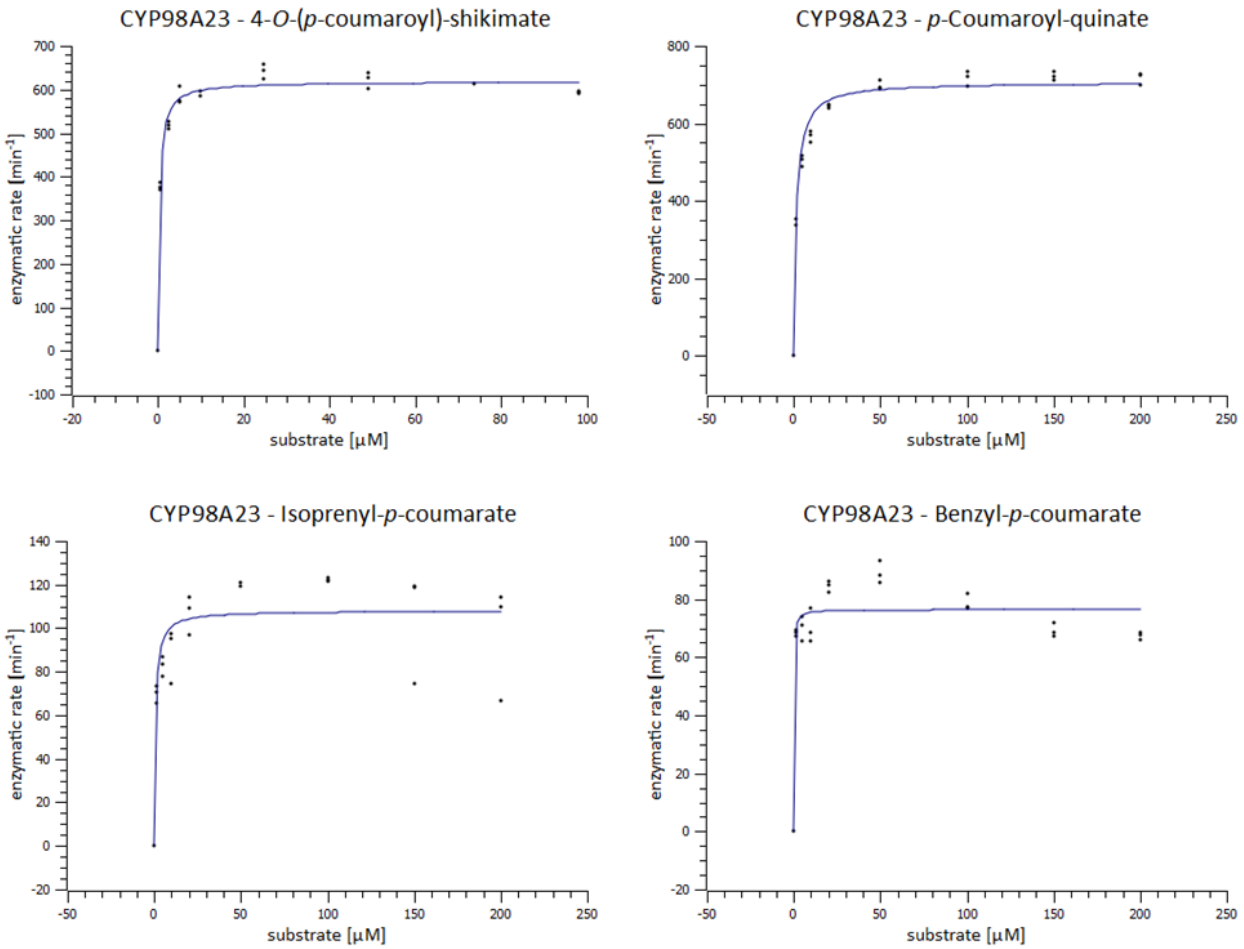


Figure 3.13 Michaelis Menten based CYP98A23 enzyme kinetics for *trans*-4-*O*-(4-coumaroyl)shikimate, *p*-coumaroyl-quinate, isoprenyl-*p*-coumarate, benzyl-*p*-coumarate.

To obtain enzyme kinetic data for CYP98A23, 0.05 µM; 0.05 µM; 0.2 µM; 0.2 µM of CYP98A23 protein was incubated with *trans*-4-*O*-(4-coumaroyl)shikimate; *p*-coumaroyl-quinate, isoprenyl-*p*-coumarate and benzyl-*p*-coumarate, respectively. Reactions were analysed on HPLC/DAD. Product appearance was measured and linked to a standard curve. Non-linear regression of the Michaelis-Menten equation was fitted under the Nelder-Mead-Simplex algorithm in the program SciDavis. Shown are three independent incubation replicates.

Enzyme	Substrate	K_M [µM]	± SE	K_{cat} [min^{-1}]	± SE	K_{cat} / K_M
CYP9A27	*p*-coumaroyl-shikimate	3.4	± 0.5	925.2	± 26.3	274
	p-coumaroyl-quinate	24.9	± 2.7	511.2	± 15.1	21
	benzyl-*p*-coumarate	40.4	± 3.1	20.3	± 0.5	1
	isoprenyl-*p*-coumarate	60.5	± 3.9	47.1	± 1.1	1
CYP9A23	*p*-coumaroyl-shikimate	[a]0.3	± 0.03	618.0	± 5.4	1819
	p-coumaroyl-quinate	1.4	± 0.2	706.6	± 10.0	488
	benzyl-*p*-coumarate	[a]0.1	± 0.1[b]	76.3	± 1.9	596
	isoprenyl-*p*-coumarate	[a]0.9	± 0.2	113.8	± 3.2	122
CYP9A27	3-*O*-(*p*-coumaroyl)-shikimate	[a]0.1	± 0.01	89.5	± 1.3	1146
CYP9A23	3-*O*-(*p*-coumaroyl)-shikimate	[a]0.02	± 0.01	310.3	± 3.2	13544

Table 3.3 Michaelis Menten based enzyme kinetics of *P. trichocarpa* CYP98A23 and CYP98A27 with trans-4-*O*-(4-coumaroyl) shikimate, *p*-coumaroyl-quinate, benzyl-*p*-coumarate and isoprenyl-*p*-coumarate.

To obtain enzyme kinetic data for CYP98A23, 0.05 µM; 0.05 µM; 0.2 µM; 0.2 µM of CYP98A23 protein was incubated with *trans*-4-*O*-(4-coumaroyl)shikimate; *p*-coumaroyl-quinate, isoprenyl-*p*-coumarate and benzyl-*p*-coumarate, respectively. To obtain enzyme kinetic data for CYP98A27, 0.25 µM; 0.2 µM; 10 µM; 10 µM of CYP98A27 protein was incubated with *trans*-4-*O*-(4-coumaroyl)shikimate; *p*-coumaroyl-quinate, isoprenyl-*p*-coumarate and benzyl-*p*-coumarate, respectively. Reactions were analysed on HPLC/DAD. Product appareance was measured and linked to a standard curve. Non-linear regression of the Michaelis-Menten equation was fitted under the Nelder-Mead-Simplex algorithm in the program SciDavis. Shown are three independent incubation replicates. Enzymatic rates for 3-*O*-(*p*-coumaroyl)-shikimate are shown and discussed in the supplement (Figure 3.25).

[a] note that K_M estimation is extrapolated since all tested substrate concentrations [S] were above K_M for technical reasons. [b] p-value of modelling t-statistic is not significant (p=0.144); all other estimates are significant at $p < 0.001$.

Even though CYP98A27 was shown to be involved in monolignol biosynthesis in *P. trichocarpa*, CYP98A23 had a lower K_M for *p*-coumaroyl-shikimate than CYP98A27 and showed high catalytic

efficiency (K_{cat}/K_M). CYP98A27 preferred *p*-coumaroyl-shikimate over *p*-coumaroyl-quinate, benzyl-*p*-coumarate and isoprenyl-*p*-coumarate. CYP98A23 also showed a high preference for *p*-coumaroyl-shikimate, but the difference to the other substrates tested was less pronounced. The lignin related CYP98A27 had a higher K_M and also a lower turnover number (K_{cat}) for isoprenyl-*p*-coumarate and benzyl-*p*-coumarate compared to CYP98A23 indicating that, especially at low substrate concentrations, CYP98A23 is clearly more efficient in catalysing this soluble phenylpropanoid ester than is the lignin related CYP98A27. In summary, the data obtained in the kinetic analysis correlates well with data obtained in the end point enzyme screening assay. While CYP98A27 seems more specific and shows a lower K_M for the lignin biosynthesis substrates *p*-coumaroyl-shikimate and *p*-coumaroyl quinate than for isoprenyl-*p*-coumarate and benzyl-*p*-coumarate, the broader substrate range of CYP98A23 is supported by a lower K_M for other soluble hydroxycinnamic conjugates.

Beyond catalytic properties, gene expression is also important in defining physiological functions. The genes might be under different transcriptional regulation, which ensures their expression in different tissues or under different environmental situations. The expression patterns of the three CYP98 genes of *P. trichocarpa* were subsequently studied across two different exhaustive data sets. No public expression data were available for *A. trichopoda*.

3.4.5. Poplar Gene expression

Mining publically available Affymetrix microarray data sets (compiled by (Guo et al., 2014)), *CYP98A27* and *CYP98A23/25* showed distinct gene expression profiles (Figure 3.14). The probe sets available on the Affymetrix microarray do not allow a distinction between *CYP98A23* and *CYP98A25* owing to high cross hybridization potential (the probe set Ptp.5940.1.s1_at recognizes both genes).

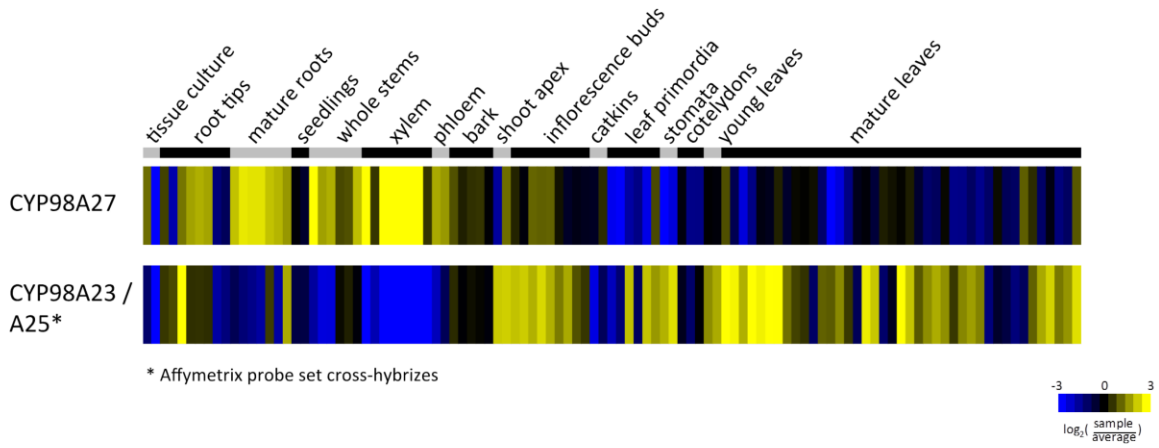


Figure 3.14 ***CYP98A23/25*** **(combined) and** ***CYP98A27*** **gene expression in publically available** ***P. trichocarpa*** **Affymetrix microarray organ and tissue sets.**

Shown are the median centred expression values of *CYP98A27* and *CYP98A23/25* across an Affymetrix dataset comprising various *P. trichocarpa* tissues, organs and developmental stages. For a description of the dataset and its normalization see Guo et al., 2014.

CYP98A27 was strongly expressed in tissues undergoing high lignification. This comprised organs such as the roots and stems and tissues such as the xylem of the poplar trees. *CYP98A23/25* in contrary was highly expressed in inflorescence buds, young leaves and mature leaves. Global gene expression thus is consistent with *CYP98A27* being involved in the biosynthesis of monolignols, expressed to highest levels in highly lignified tissues. Likewise, *CYP98A23/25* expression is consistent with an involvement in the biosynthesis of soluble, possibly protection related phenolic compounds, with the enzymes being strongly expressed in inflorescence buds and young leaves, which are known to accumulate high amounts of phenolic compounds (Greenaway and Whatley, 1990b; English et al., 1991).

A data set comprising young leave and developing xylem RNAseq based information from the POPCAN project (Corea O., Douglas C.J., Ehlting J., unpublished) was mined for expression of the three poplar genes across multiple natural accessions of *P. trichocarpa*. Expression data of 371 individual trees from 197 accessions in the young leaves data set, and of 390 individual trees of 194 accessions were available in the developing xylem data set (Figure 3.15). No expression was found in either of the young leaves or developing xylem datasets for *CYP98A25*

of *P. trichocarpa*. *CYP98A23* was highly expressed in the samples of the young leave tissue set. *CYP98A27* was expressed in the young leave data set, but to far lower abundance than *CYP98A23*. *CYP98A27* was, in contrast, highly expressed in the developing xylem data set. Only few individuals also showed detectable expression of *CYP98A23* in the developing xylem (Figure 3.15).

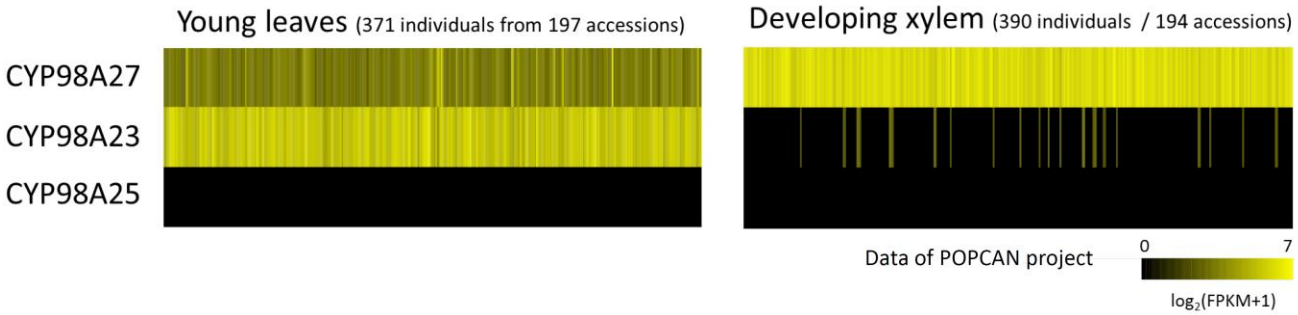


Figure 3.15 ***P. trichocarpa CYP98* gene expression in young leaves and developing xylem.**

RNAseq data set containing young leaves of 371 individual trees and developing xylem samples of 390 individual trees. Shown is a comparison of gene expression data of *CYP98A27*, *CYP98A23* and *CYP98A25* of *P. trichocarpa* in two data sets, one established by RNA sequencing of young leaves and one established by RNA sequencing from developing xylem. The expression data of the *CYP98s* of 371 individuals from 197 accessions are represented in the young leaves data set. The expression data of the *CYP98s* of 390 individuals from 194 accessions are represented in the developing xylem data set. These data sets are part of the POPCAN project and have been mined by O. Corea.

3.4.6. Poplar *CYP98* Co-expression analyses

To further exploit the large-scale gene expression data towards elucidating possible biological functions in poplar trees, co-expression analyses of the *P. trichocarpa* CYP98s were performed. Co-expression analysis was performed on the poplar developmental data set based on pairwise comparison of the target gene expression profiles with all other genes represented on the microarray. Using a Pearson correlation threshold of r>0.75 for *CYP98A27* (Ptp.1996.1.S1_s_at) 3,060 genes were found to be co-expressed (Figure 3.16). The top 50 co-expressed genes included many core monolignol pathway biosynthesis genes, indicated by a yellow bar in Figure 3.16, for example ferulate-5-hydroxylase, caffeic-acid-*O*-methyltransferase, cinnamate-4-

hydroxylase, hydroxycinnamoyl-CoA quinate/shikimate hydroxycinnamoyltransferase, among others. Known secondary cell wall carbohydrate biosynthesis related genes (in turquoise) such as cellulose synthase and laccase were among the co-expressed genes. Several transcription factors, including MYB transcription factors, are co-expressed with *CYP98A27*, indicated by an orange bar. MYB85, the most similarly co-expressed transcription factor is a known positive regulator of monolignol biosynthesis (Zhong et al., 2008). Together with the functional data obtained in the enzymatic screening, this further substantiates the distinct role of CYP98A27 in monolignol biosynthesis.

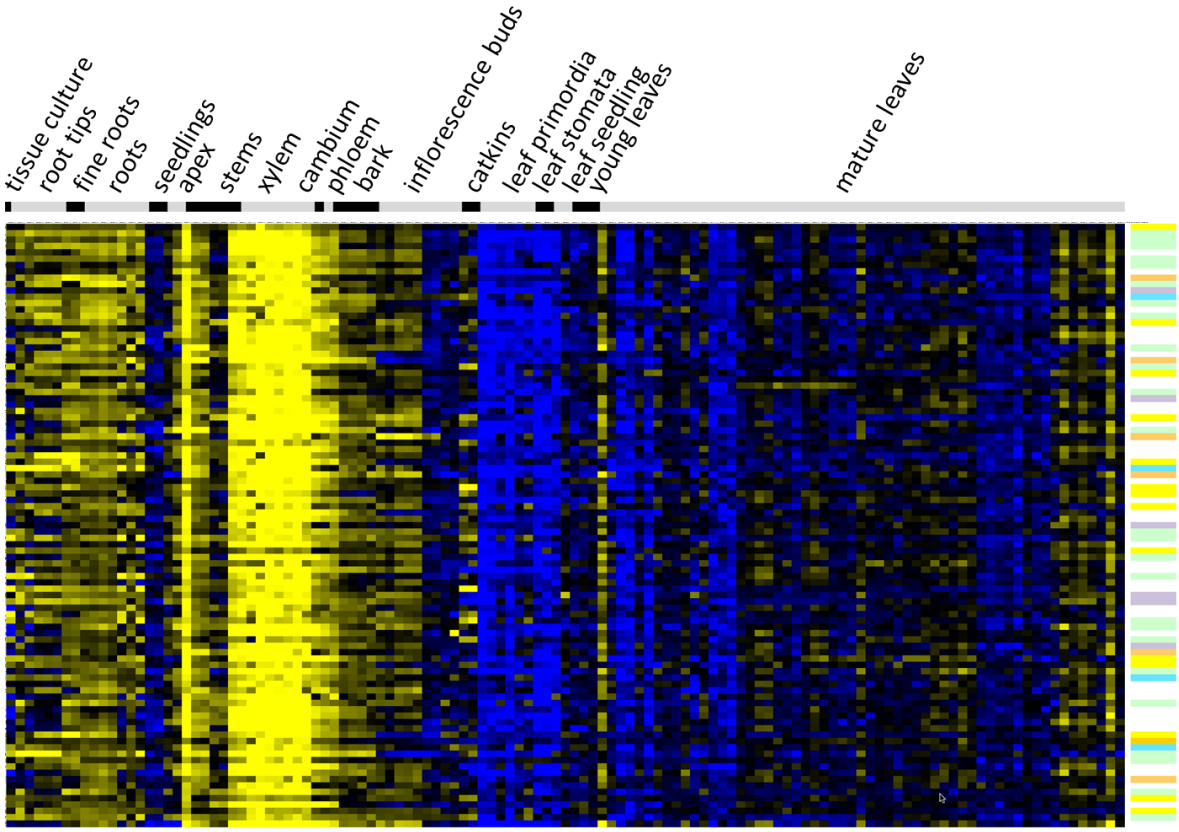


Figure 3.16 Co-expression analysis for *CYP98A27* in an Affymetrix microarray organ and tissue dataset.

The functions of co-expressed genes are based on described functions in *A. thaliana*. Yellow bars indicate lignin biosynthesis pathway genes. Secondary cell wall carbohydrate biosynthesis related genes are indicated by a turquoise bar. Transcription factors are indicated by an orange bar. Putative cell wall related proteins are marked by a purple bar. Cytoskeleton related genes are marked by a light blue bar. The dataset has been mined by A. Alber and O. Corea.

In contrast to *CYP98A27*, only very few co-expressed genes were found for *CYP98A23/*25 with a Pearson correlation coefficient larger than 0.75 in this data set. This threshold is indicated by the white line in Figure 3.17. A transcriptional regulator (orange bar), a homolog of MYB4, appears within these few co-expressed genes. MYB4 is known to be a repressor of monolignol biosynthetic genes, and to be involved in UV-B response regulation and cold tolerance in *A. thaliana* and *O. sativa* (Jin et al., 2000; Vannini et al., 2004; Schenke et al., 2011).

By extending the cut point of the co-expression analysis from r>0.75 to r>0.65, additional genes of interest have been found (Figure 3.17; below white line). Genes related to flavonoid biosynthesis are marked by a brown bar. One of these genes is a *4CL3* homolog. *4CL3* represents a distinct *4CL* isoform, which is related to the biosynthesis of flavonoid and other non-lignin soluble phenolic compounds (Ehlting et al., 1999).

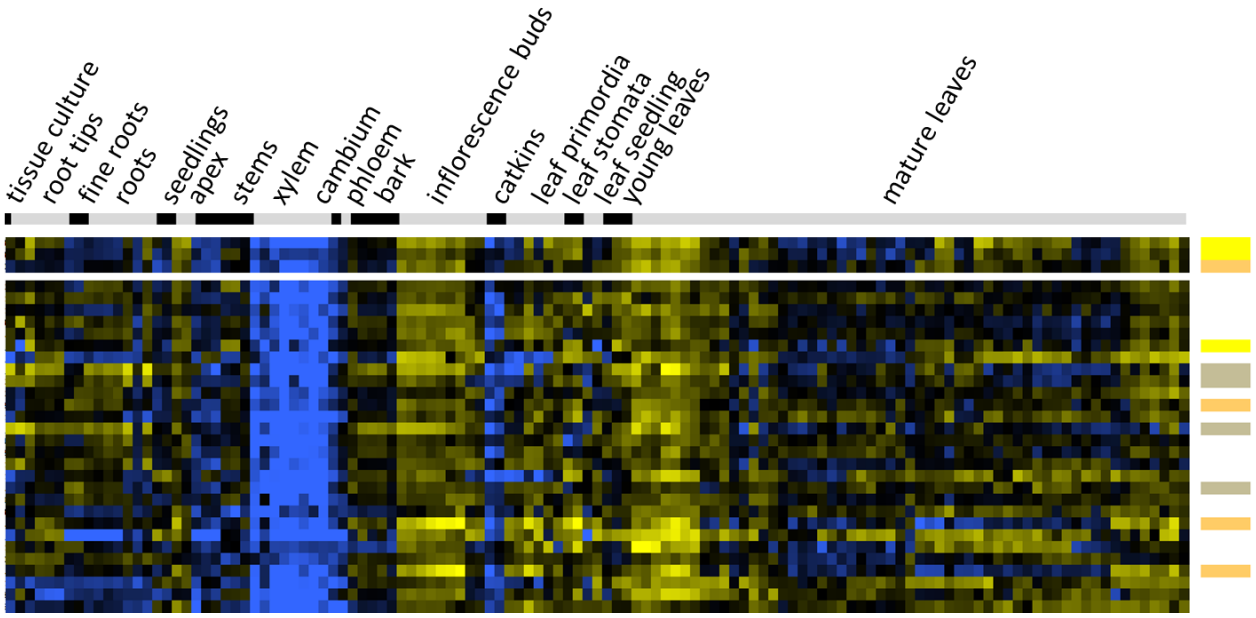


Figure 3.17 Co-expression analysis for *CYP98A23/25* in an Affymetrix microarray organ and tissue dataset.

The functions of co-expressed genes are based on described functions in *A. thaliana*. Yellow bars indicate lignin biosynthesis pathway genes. Transcription factors are indicated by an orange bar. Flavonoid biosynthesis related genes are indicated by a brown bar. The white line represents the cut off for a Pearson correlation coefficient greater than 0.75. Co-expression data below the white line can be found for a Pearson correlation coefficient greater than 0.65. The dataset has been mined by A. Alber and O. Corea.

The low number of co-expressed genes with *CYP98A23/25* might be due to the probe set recognizing both *CYP98A23* and *CYP98A25*, or the expression of the gene(s) in very specific tissues or under special circumstances, such as for example in response to herbivory, temperature or in a strict temporal pattern. In the co-expression analysis with a single dataset comprising all organ and tissue data, the gene might not be clearly co-expressed with one single pathway. If co-expression exists with many pathways included in the set, the results would not show clearly but might be equalised. A *P. trichocarpa* gene expression stress dataset exists, but is focused on nutrient and water stress and did not add conclusive information when mined for *CYP98A23/25* expression.

3.4.7. *P. trichocarpa CYP98*s expression in poplar leaves after gypsy moth feeding

An involvement in the hydroxylation of *p*-coumaroyl-shikimate to caffeoyl-shikimate, as seen *in vitro,* might be triggered by gene expression upon herbivory attack. Lignification can be one of the plant responses to pathogen attack. This also fits the observed increased transcription of *CYP98* genes under MeJA treatment (as reported for *L. japonica*, *S. miltiorrhiza*, *W. somnifera*, described above). It is known that chlorogenic acid can be stored in plant cells (Mølgaard and Ravn, 1988). In case of a strictly temporal CYP98 gene expression, caffeoyl conjugates of the reaction catalysed by *CYP98A23/25* could thus be stored in the plant.

A possible involvement of CYP98A23 in the response to herbivory attack in *P. trichocarpa* was investigated by quantitative real time PCR of *CYP98* transcripts in gypsy moth treated poplar leaves. The relative gene expression of the three *P. trichocarpa* genes has been monitored in *P. trichocarpa* leaves, treated by gypsy moth, *Lymantria dispar*, feeding. Non parametrical statistical analysis (Mann-Whitney U) showed significant increase in gene expression upon *L. dispar* feeding for *CYP98A23* (Z ratio: Z-score 2.2978; p-value 0.02144; U value 1; critical U 2), albeit the increase was very low. *CYP98A25* showed a stronger, approximately 10 fold induction on average, which was also significant (Z ratio: Z-score 2.5067; p-value 0.01208; U value 0; critical U 2). However, absolute levels of *CYP98A25* transcripts were very low, both in control and in gypsy moth treated samples. No significant difference was found for transcript levels of *CYP98A27* (Z ratio: Z-score 1.88; p-value 0.0601; U value 3; critical U 2) despite a similar fold-change average as found for *CYP98A23* (Figure 3.18). Due to the low overall increase in

expression of only two fold for both *CYP98A23* and *CYP98A27* statistical results need to be interpreted carefully. The difference in expression is minimal, but leads to a significant change in the Mann-Whitney-U test performed. An induction of only two fold of *CYP98A23* and *CYP98A27* does not strongly suggest enhanced gene expression upon herbivory treatment with *L. dispar*.

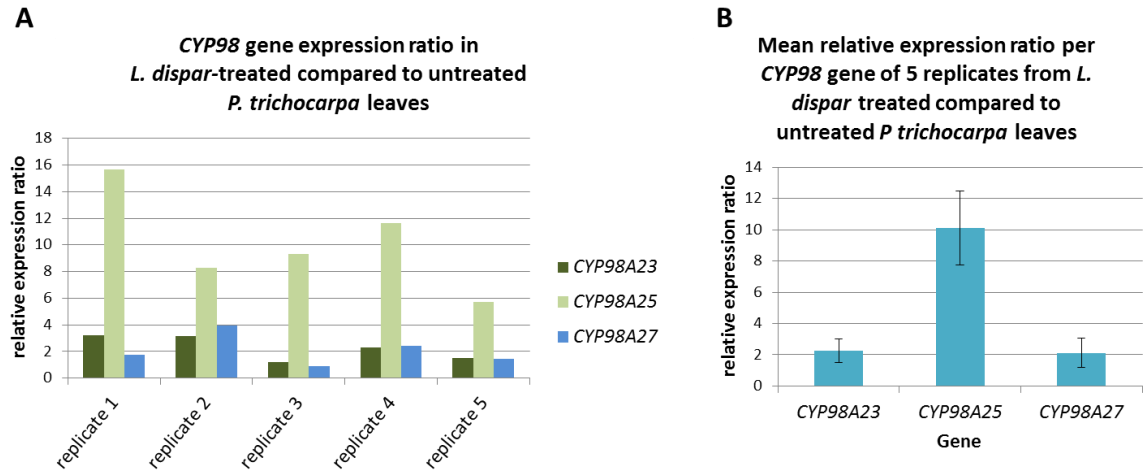


Figure 3.18 Relative gene expression of *P. trichocarpa CYP98A23*, *CYP98A25* and *CYP98A27* in *P. trichocarpa* leaves after *L. dispar* feeding, compared to gene expression in untreated *P. trichocarpa* leaves.

The relative gene expression of *P. trichocarpa CYP98s* in *P. trichocarpa* leaves after *L. dispar* feeding, compared to untreated *P. trichocarpa* leaves is shown. **A**: The relative expression ratio of the three *CYP98s* is shown for all 5 replicates. **B**: The average relative expression ratio of the 5 replicates of each *CYP98* is shown. The relative expression ratio was calculated with the 2^ ΔΔCT method (Livak and Schmittgen, 2001) for 5 biological replicates (3 technical replicates for each of the 5 biological replicates).

3.4.8. *A. thaliana cyp98a3* knock out mutant complementation with poplar *CYP98* genes

As *CYP98A23, CYP98A25* and *CYP98A27* showed distinct expression and co-expression patterns as well as distinct biochemical properties we next analysed if the function of CYP98A23 and CYP98A27 could complement the *A. thaliana cyp98a3* knock-out mutant *in vivo*. We further

investigated if *CYP98A25* would be able to complement the *A. thaliana* mutant *in vivo*, even though no function could be detected for the heterologous expressed enzyme *in vitro*.

Under control of the cinnamate-4-hydroxylase (C4H) promoter, both poplar *CYP98* isoforms, *CYP98A23* and *CYP98A27*, were able to complement the *A. thaliana cyp98a3* knock out mutant phenotype (Figure 3.19). This correlates to the substrate utilization profile of the enzymes *in vitro*, where both enzymes were shown to convert *p*-coumaroyl-shikimate efficiently (Figure 3.8;Figure 3.9). Consistent with data obtained from the *in vitro* assay was the observation made for the complementation assay with *CYP98A25*: no complementation was observed for the *CYP98A25* of *P. trichocarpa*. *A. thaliana* plants that were homozygous for the *cyp98a3* knock out that carried the *CYP98A25* expression construct, showed the same dwarf phenotype as non-complemented *cyp98a3* knock-out plants (Figure 3.19).

The promoter chosen for this experiment was the *C4H* promoter from *A. thaliana* (Bell-Lelong and Cusumano, 1997). Under this promoter, expression is assumed to be enhanced in lignified tissues. A general promoter such as the cauliflower mosaic virus 35s promoter might have led to CYP98A23 hydroxylating other conjugates potentially used in non-lignified tissues. Thus a complementation of the *cyp98a3* phenotype of both CYP98s of *P. trichocarpa* has to be interpreted carefully. Considering the broad range function of CYP98A23 observed *in vitro*, it is also possible that other caffeoyl-esters may be able to act as monolignol intermediates. A detailed lignin composition analysis of the complemented *A. thaliana* plants could give further information of molecules utilized and integrated into the lignin polymer under strong *CYP98A23* expression. For this, it might be interesting to repeat the complementation assay with *CYP98A23* under the cauliflower mosaic virus 35s promoter. An incorporation of unusual hydroxylated hydroxycinnamic conjugates into the lignin, which might be easier to cleave in treatments of feedstock processing, is an option for decreasing recalcitrance in biofuel production highly researched to date.

To further test the potential functionality of the CYP98A25, which did not express well in isolated yeast microsomes and did not show function when these microsomes were incubated with various substrates *in vitro*, transgenic plant material could be collected and analysed for the presence of functional CYP98A25 enzyme. To do this, plant extracts could be incubated with

the substrates and substrate conversion monitored. Conversely, the absence of complementation may also be indicative of an occurring *CYP98A25* pseudogeneization.

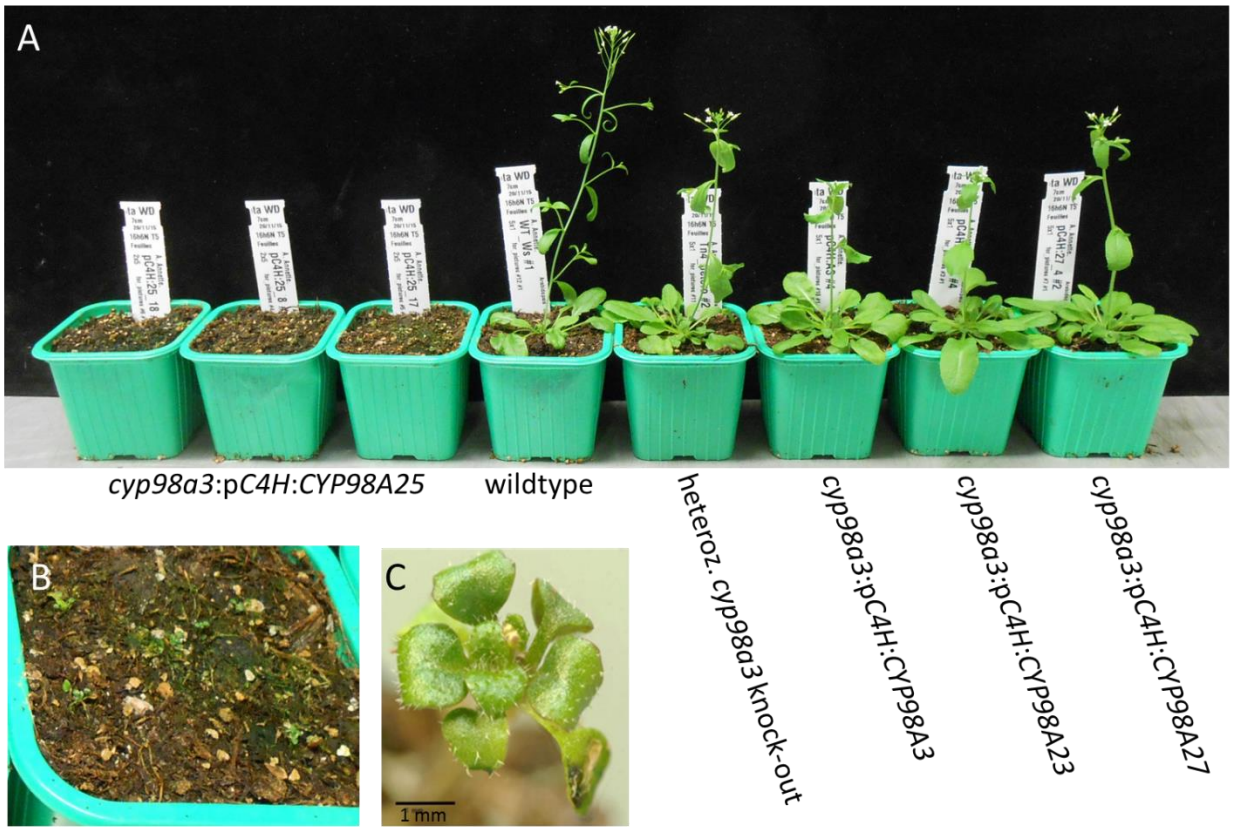


Figure 3.19 ***A. thaliana cyp98a3*** **knock-out mutant complementation assay with the three** ***P. trichocarpa CYP98*** **genes**.

The *A. thaliana* T-DNA insertion mutant knock-out for *CYP98A3* (Abdulrazzak et al., 2006) was used in a mutant complementation assay with the *P. trichocarpa CYP98A23*; *CYP98A25* and *CYP98A27* genes. As homozygous T-DNA lines of *cyp98a3* show dwarf morphology and are male sterile, heterozygous plants were used for transformation with the *P. trichocarpa CYP98s* under the promoter of the *A. thaliana C4H* gene (Bell-Lelong and Cusumano, 1997). **A**) Overview of wild type and transformed *A. thaliana* plants. **B**) Close up of *A. thaliana cyp98a3* plants carrying the *P. trichocarpa CYP98A25* expression construct. **C**) *A. thaliana cyp98a3* homozygous knock-out mutant close up. Genotyping see Figure 3.28.

3.5. Conclusion

Only angiosperms possess several *CYP98* paralogs. Several CYP98 family members were shown to be involved in the biosynthesis of monolignols in the angiosperms. All lignin-related CYP98s

biochemically characterized in angiosperms (here and in literature), show a marked preference for *p*-coumaroyl-shikimate/quinate esters as substrates. Our initial hypothesis, that an ancient duplication early in angiosperm evolution generated a *p*-coumaroyl-shikimate metabolizing and lignin-related clade, and a soluble phenolic related clade, was rejected. Instead, a complex series of gene duplications and gene losses shape the *CYP98* family in angiosperms suggesting frequent and independent recruitment of duplicates to specific functions. This was previously observed for Arabidopsis (Matsuno, et al., 2009), to some extend for coffee (Mahesh et al., 2007), and here now for poplar and Amborella. Duplications leading to the CYP98 isoforms present in poplar Amborella today clearly happened independently.

It would be most parsimonious to assume that lignin-related functions, common to all angiosperms, were maintained throughout angiosperm evolution. In addition, independent broadening of substrate preference of duplicates occurred multiple times and might have led to the, sometimes lineage specific, soluble phenolics repertoires. *Malus domestica* is an example for a species with many (twelve) CYP98 genes. A recent investigation of free and bound phenolic acids in the pulp and peel of four apple varieties identified eleven hydroxycinnamic acids, five hydroxyphenylacetic acids and one hydroxyphenylpropanoic acid, amongst others (Lee et al., 2017).

The biochemically characterized conifer CYP98 showed a broad range of accepted substrates, including conversion of *p*-coumaroyl-shikimate *in vitro*. However a role of CYP98s in gymnosperms in the biosynthesis of monolignols has yet to be shown *in vivo*. Only one *CYP98* was found in the genome of *P. abies* and an analysis of available expression data showed high expression of this *CYP98* in the wood of the tree and also in vegetative shoots. This could support the broad CYP98 substrate range observed for *P. taeda in vitro*. CYP98 could be involved in the biosynthesis of monolignols and therefore highly expressed in the wood of the tree, but also in the biosynthesis of soluble phenolic compounds, presumably for protective functions, in the vegetative shoots. As described above, *CYP98s* were not co-expressed with lignin biosynthetic pathway genes in *P. glauca* (Porth et al., 2011) and *CYP98A19* gene expression not induced upon addition of Phe in *P. taeda* cell suspension cultures (Anterola, 2002). As phylogenetic analysis revealed no distinct lignin-related CYP98 clade, another, albeit

less parsimonious, scenario of CYP98 evolution in the angiosperms could be possible. The ancestor of the angiosperm CYP98s could have been a broad substrate range enzyme, including activity with *p*-coumaroyl-shikimate. Subsequent duplications, expanding *CYP98* families in Amborella, Populus, and Arabidopsis, clearly occurred independent of each other. These duplications gave rise to isoforms, which were able to acquire/refine preference for *p*-coumaroyl-shikimate, independently. A repeated independent acquisition of *p*-coumaroyl-shikimate hydroxylating activity and recruitment for monolignol biosynthesis would require a strong selective advantage of the involvement of *p*-coumaroyl-shikimate in the lignin-biosynthetic pathway. The other duplicate would have been able to maintain/refine the broad substrate range.

Gene expression of the *CYP98A25* of *P. trichocarpa* was only found to low levels and in few individuals in a set of transcriptomes of young leaves and developing xylem of over 700 *P. trichocarpa* individuals. The enzyme was poorly expressed in yeast and most likely not stable. When incubated with a library of different substrates *in vitro,* no apparent activity with any potential substrate was detected. The *A. thaliana cyp98a3* knock-out mutant was not complemented by *CYP98A25*. Taken together, these results are indicative of a gene "on its way out", becoming a pseudogene. However, across the over 700 individual poplar genomes investigated, not a single *CYP98A25* showed any nonsense mutations. The open reading frame of the gene was conserved across the population. *CYP98A25* showed the highest induction in gene expression upon gypsy moth feeding on poplar leaves.

3.6. Acknowledgement

We thank Charles P. Scutt (CNRS Laboratoire de Reproduction et Développement des Plantes, Lyon) for providing *Amborella trichopoda* cDNA.

We thank Zhenhua Liu (John Innes Centre, Norwich) for pCC0996:CYP98A3 Arabidopsis seeds.

We thank Tobias Köllner and Jan Günther (Max Planck Institute for Chemical Ecology, Jena) for providing *P. trichocarpa* RNA.

We thank Daisie Huang for providing Salicaceae *CYP98* sequences.

We thank the members of the POPCAN project for providing *P. trichocarpa* expression data.

3.7. Contributions

Annette Alber: Phylogenetic analysis, gene cloning and microsome preparation, preparation of coumaroyl-shikimate, enzyme incubations, enzyme kinetics, transient overexpression in *N. benthamiana*, *A. thaliana* mutant complementation, manuscript writing. **Hugues Renault**: Supervision of bench work. **Pascaline Ullmann**: Supervision and guidance of enzymatic synthesis of substrates and enzyme incubations. **Alexandra Basilio Lopes** and **Martine Schmitt**: Synthesis of substrates. **Oliver Corea**: *P. trichocarpa* RNAseq expression analysis, Affymetrix dataset mining. **Danièle Werck-Reichhart**: Supervision, funding, manuscript editing. **Jürgen Ehlting**: Supervision, funding, manuscript editing.

3.8. Supplement

3.8.1. List of primers

Species	Gene	Gene identifier	Forward primer sequence	Reverse primer sequence	purpose
P. trichocarpa	CYP98A23	POPTR_0016s03090	CTG GCC AAG CAA GTT CTC AAA G	GGT CCT TGC CGT CTC TAC TAA AC	qPCR target gene
P. trichocarpa	CYP98A25	POPTR_0016s03080	TGG CTG ATA GGC ATA GGA CTA GA	TCC AAG CCA CAT CTC CCA AAT A	qPCR target gene
P. trichocarpa	CYP98A27	POPTR_0006s03180	ATG ATT ACT GCG GGC ATG GA	ACT CTT CCT GAG CCT TCT GTT G	qPCR target gene
P. trichocarpa	Elongation factor 1ß	POPTR_0009s02370	ACC TGG TCG TGA TTT CCC TAA TG	GCC ACA AAT GCT TAC ACC AAC A	qPCR reference gene
P. trichocarpa	Ribosomal protein	POPTR_0001s35630	TGT TGT GAC CGC TGA TTG TTT G	CCA CCT GTT CTT GCC TGT CTT A	qPCR reference gene
P. trichocarpa	CYP98A23	POPTR_0016s03090	GGCTTAAUATGGCTCTGC CTCTTTTAG	GGTTTAAUTCACATAT CTGAAGCCA	Gene cloning, USER™
P. trichocarpa	CYP98A25	POPTR_0016s03080	GGC TTA AUA TGG CTC TGC CTC TG	GGT TTA AUT CAC ATA TCT GAA GCC ATC	Gene cloning, USER™
P. trichocarpa	CYP98A27	POPTR_0006s03180	GGC TTA AUA TGA ATC TCC TTC TGA TTC	GGT TTA AUT TAA ATA TCA ACA GCA ACA C	Gene cloning, USER™
A. trichopoda	CYP98A84	evm_27.model.AmTr_v1.0_scaffold00101.79	GGATCCATGGACTTTCTC TCTCCACTCTC	GGTACCTCACATTTGT GTGGGCACAC	Gene cloning, TA
A. trichopoda	CYP98A85	evm_27.model.AmTr_v1.0_scaffold00040.62	GGATCCATGGAGTCTCTC TTCCTACTTGC	GGTACCTCACACTTTC ATGGACTGACA	Gene cloning, TA
A. thaliana	CYP98A3	AT2G40890	CCGATCGTCGGTAACCTC TA	AAATGCTGTTTCGCTC CACT	Genotyping
A. thaliana	T-DNA		TTGCTTTCGCCTATAAATA	AAATGCTGTTTCGCTC	Genotyping

	insertion		CGACGGATCG	CACT	
P. trichocarpa	CYP98A23	POPTR_0016s03090	aattgatgagcaaggccaag	gatttgggaaatctgcctca	Genotyping
P. trichocarpa	CYP98A25	POPTR_0016s03080	caatccagaaaaccatggaaa	tgcaggtaagggagatttgg	Genotyping
P. trichocarpa	CYP98A27	POPTR_0006s03180	gaaatatttgggggcagttg	gagaaatcagcctcggtcat	Genotyping

Table 3.4 Primer sequences used in gene cloning, quantitative real-time PCR and genotyping

3.8.2. CYP98A25 expression conditions

In an attempt to increase the amount and quality of expressed CYP98A25 enzyme in yeast, CYP98A25 was expressed from independent yeast transformations with CYP98A25 containing pYeDP60 plasmids. Yeast cultures were grown in YPGE medium. No amelioration in P450 expression after microsome isolation was recorded in CO differential spectra.

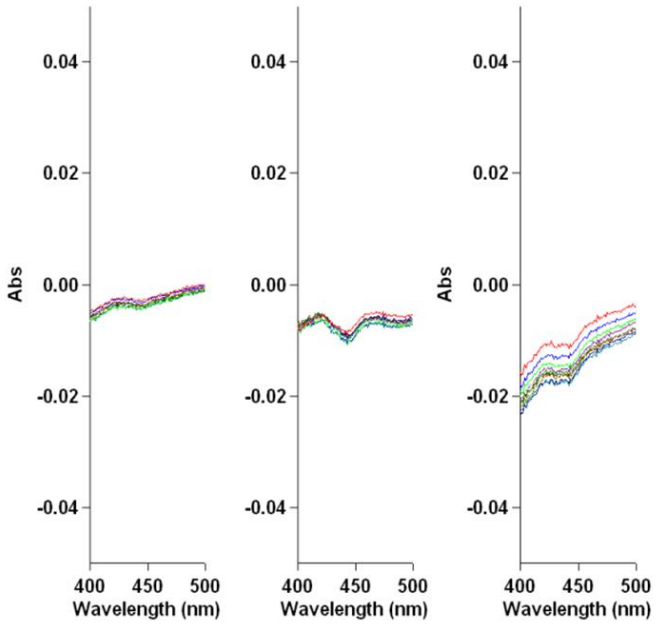


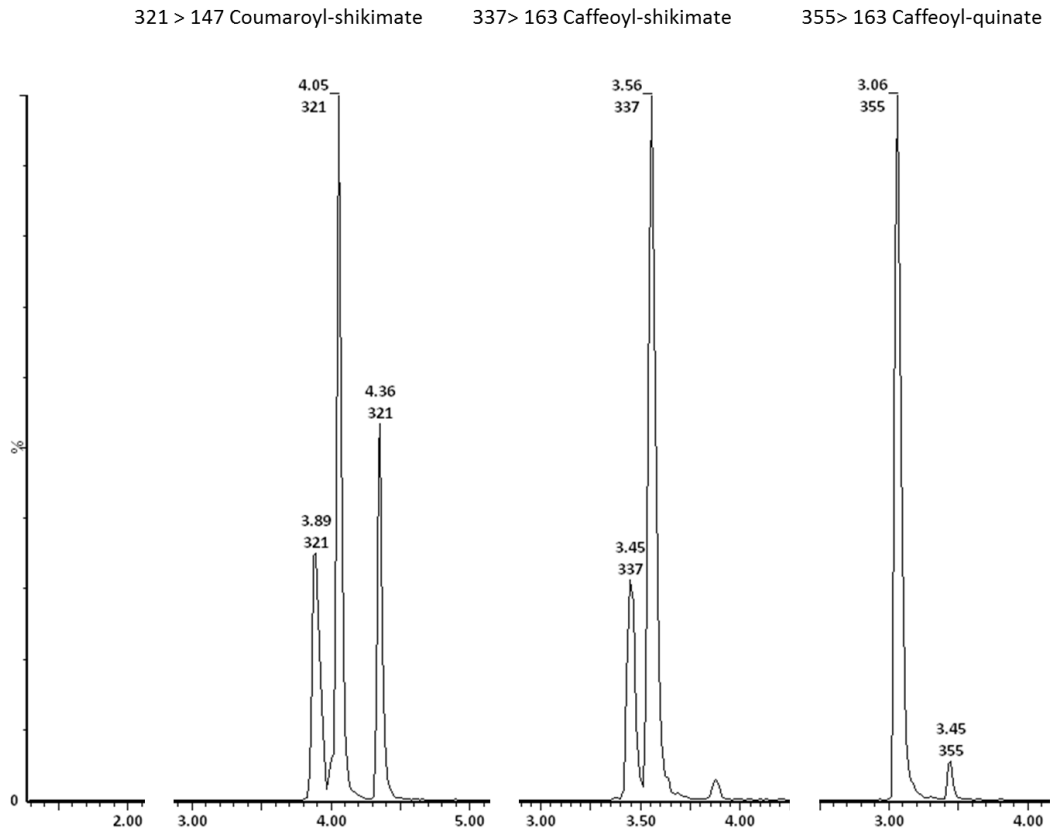
Figure 3.20 *P. trichocarpa* CYP98A25 expression from independent yeast transformations

Three independeltly created *CYP98A25* containing pYeDP60 plasmids were used for the three independent yeast transformations.

3.8.3. Transient overexpression of *P. trichocarpa CYP98s*

All three *P. trichocarpa CYP98s* were transiently overexpressed in *Nicotiana benthamiana* under the CaMV-35s promotor.

The assay was declared not conclusive under the set up used, as high substrate conversion was also detected in plants transformed by an empty-vector control (Figure 3.21). *N. benthamiana* possesses endogenous *CYP98* genes and also endogenous substrates and products.



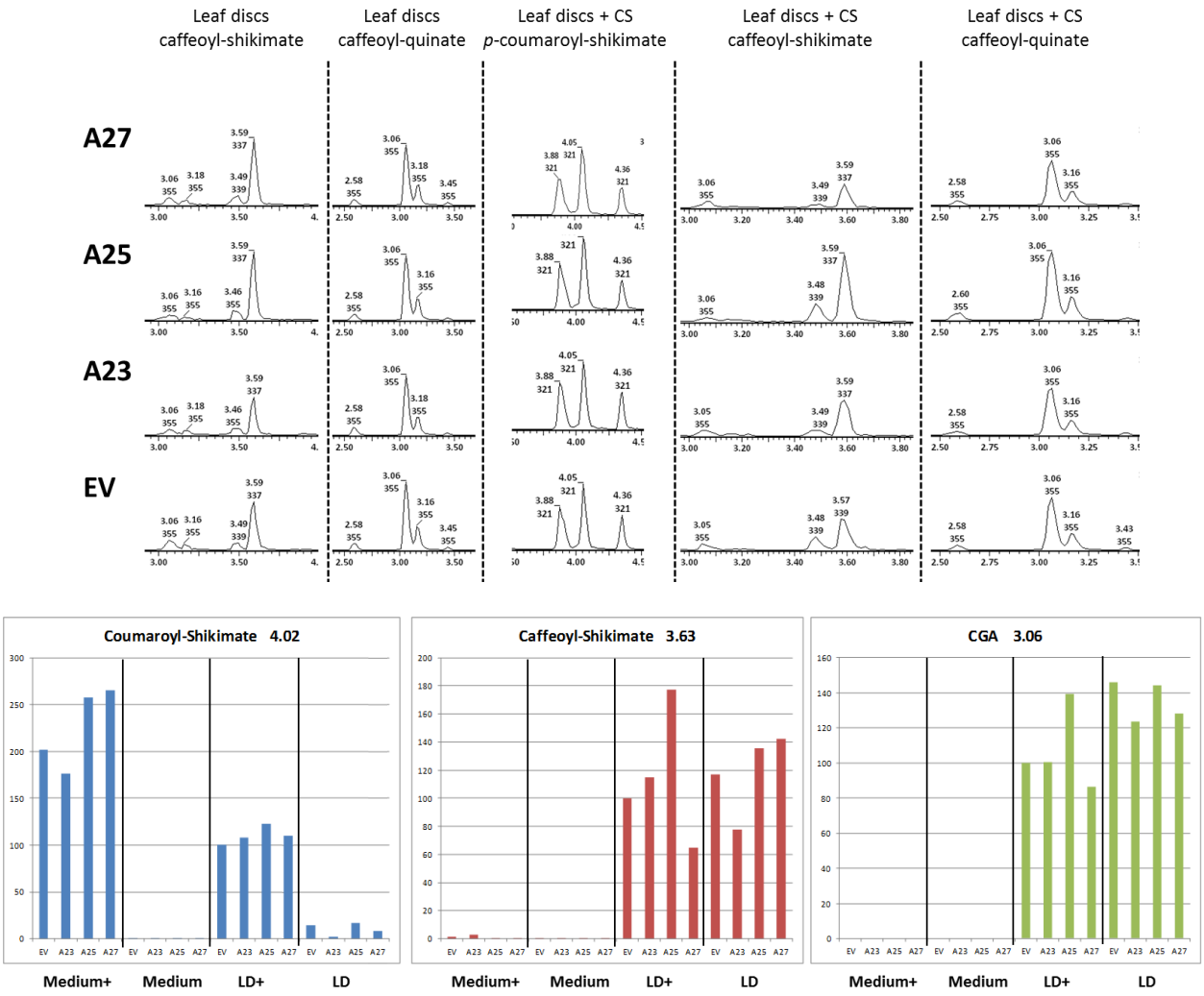
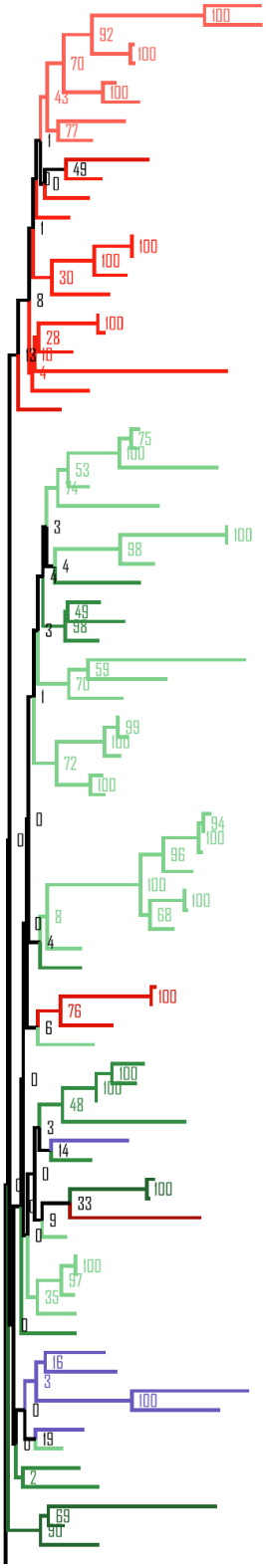


Figure 3.21 Transient overexpression of *P. trichocarpa* CYP98s in *N. benthamiana* and *N. benthamiana* leaf disc incubation in medium containing *p*-coumaroyl-shikimate.

The UPLC-MS/MS analysis was performed by the metabolic platform of the IBMP CNRS Strasbourg (Raphael Lugan). **Upper panel**: Chromatograms and ion fragmentation for the standards coumaroyl-shikimate, caffeoyl-shikimate and caffeoyl-quinate on UPLC-MS/MS. **Mid panel**: Analysis of extracted leaf discs with and without *p*-coumaroyl-shikimate supplementation. Detection of *p*-coumaroyl-shikimate, caffeoyl-shikimate and caffeoyl-quinate for each transformation construct CYP98A23, CYP98A25, CYP98A27 and for the empty-vector control group (EV). **Lower panel**: Data summary of extracted leaf discs and incubation media analysed by UPLC-MS/MS. Peak area of detected substrates and products for treatment (Medium+; Leaf Disk (LD)+), control (Medium; LD) and different constructs. The substrates are shown: in blue: coumaroyl-shikimate, in red: caffeoyl-shikimate, in green: caffeoyl-quinate or chlorogenic acid (CGA).

3.8.4. Phylogenetic reconstruction of CYP98s across angiosperm orders. Bootstrap support for Figure 3.3



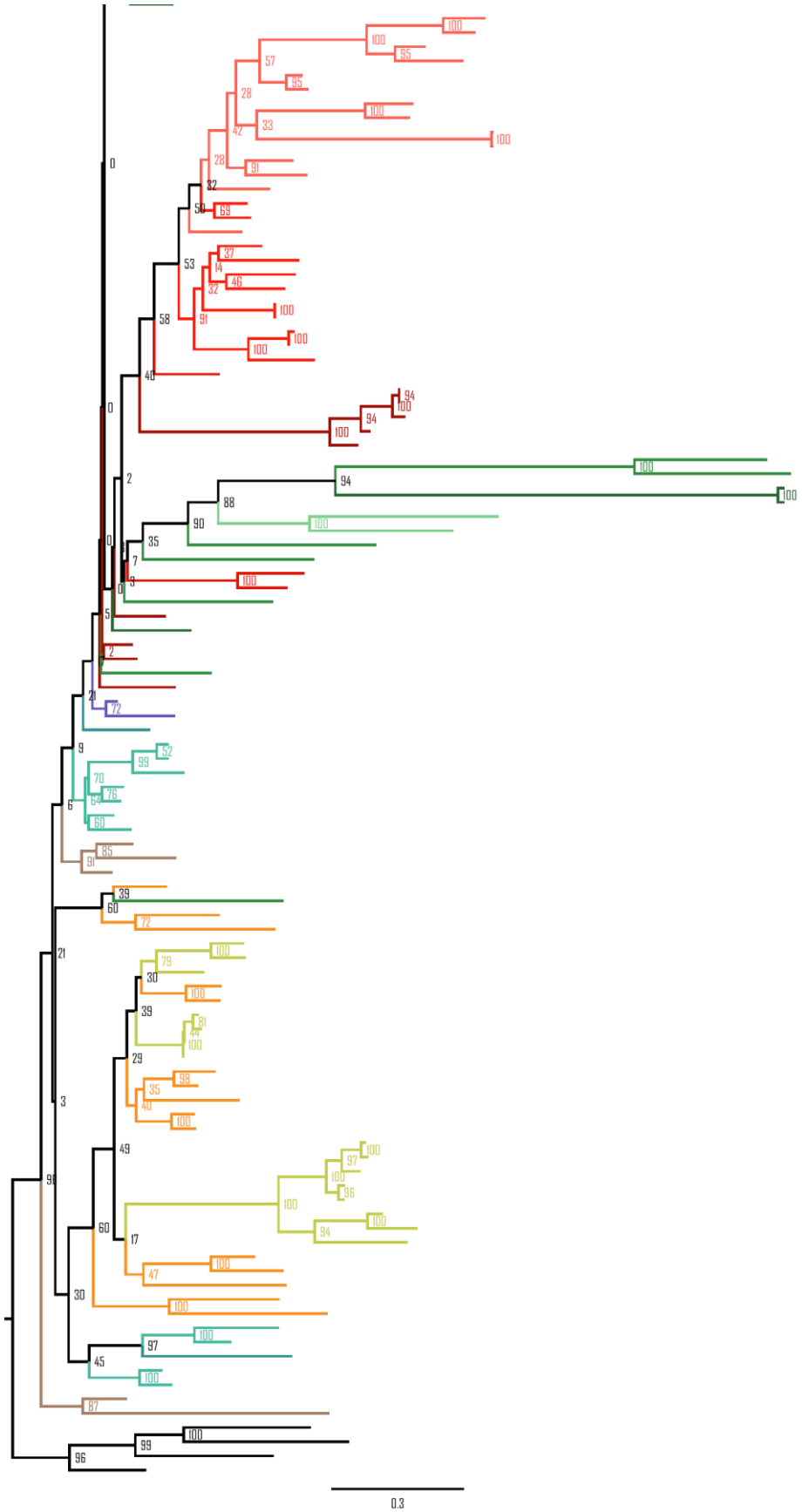
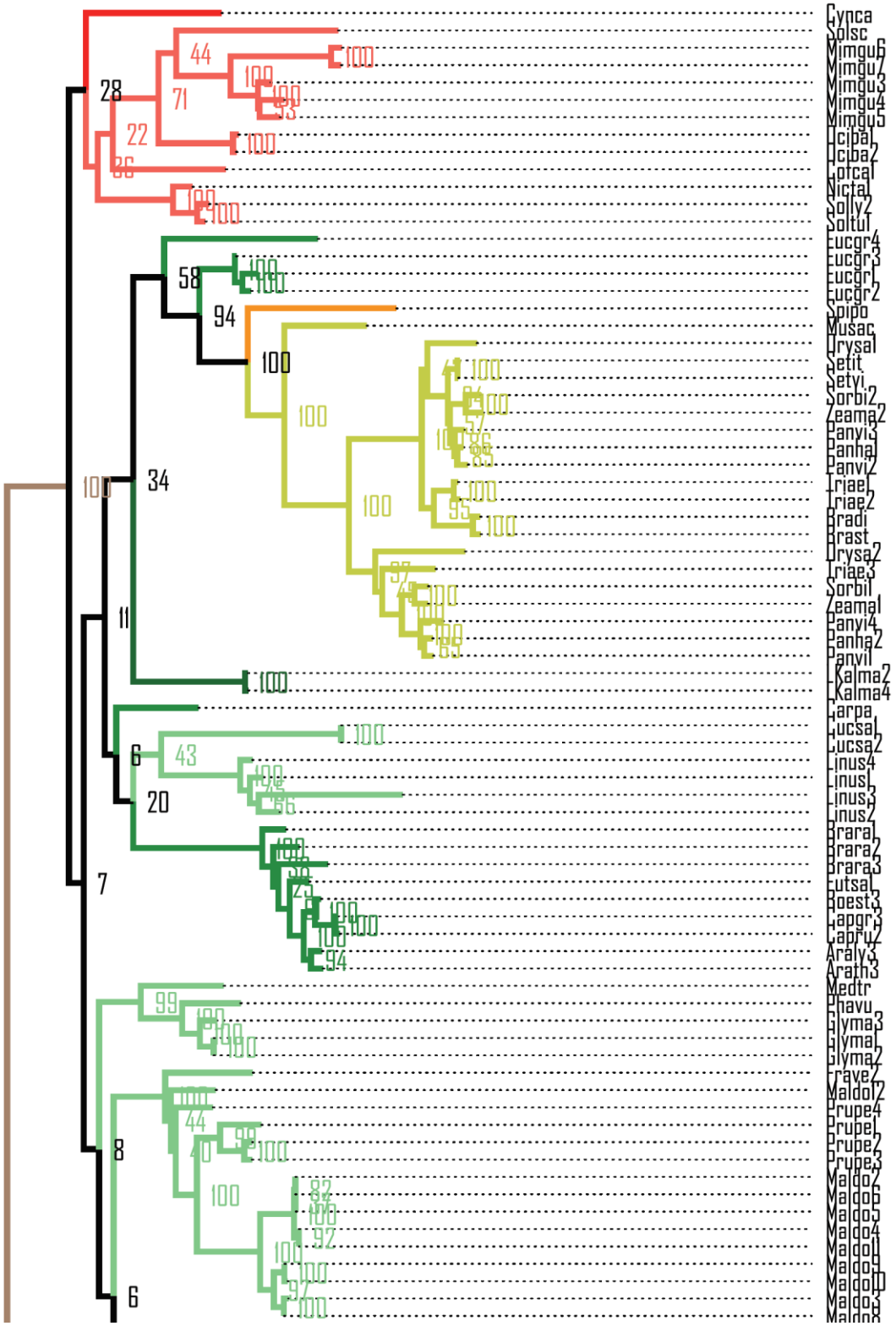

0.3

Figure 3.22 Phylogenetic reconstruction Figure 3.3 with bootstrap support.

The linear phylogenetic reconstruction shows the bootstrap support for the polar tree reconstruction in Figure 3.3: *A. trichopoda* is at the base of all angiosperms. Two species of each angiosperm order (where available) were chosen and amino acid sequences aligned by DIALIGN (Morgenstern, 1999). Only alignment positions with diagonal similarities greater than zero were maintained in the alignment. A maximum likelihood based phylogenetic reconstruction was performed by phyml (Guindon and Gascuel, 2003) under assumption of the JTT model. The branches of the tree are coloured according to the orders in Figure 3.2. All orders of Figure 3.2 are represented in the phylogenetic reconstruction.

3.8.5. Phylogenetic reconstruction of angiosperm CYP98s from sequenced genomes and characterized CYP98s. Bootstrap support for Figure 3.4



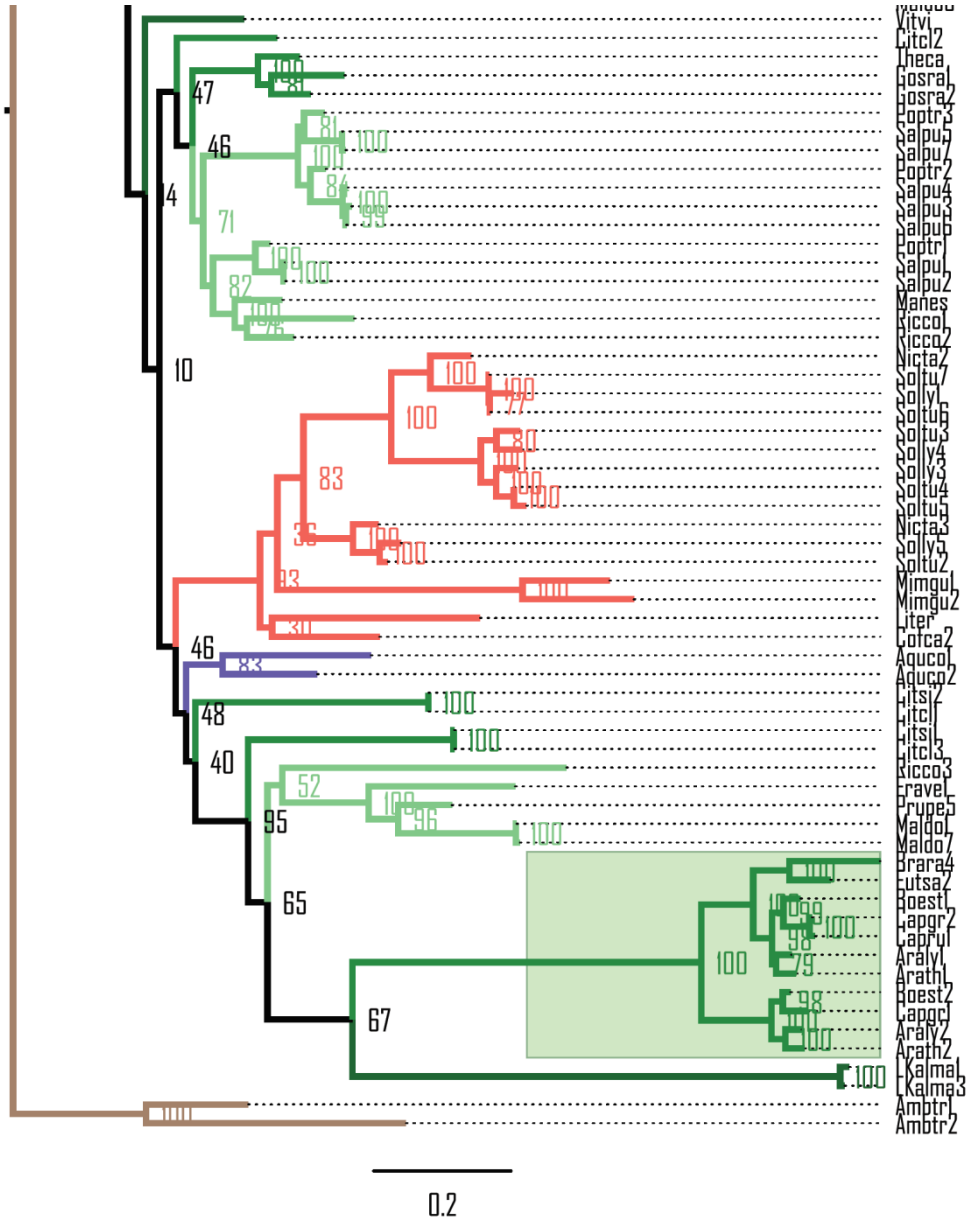


Figure 3.23 **Phylogenetic reconstruction** Figure 3.4 **with bootstrap support.**

The linear phylogenetic reconstruction shows the bootstrap support and species identifiers for the polar tree reconstruction in Figure 3.4: Amino acid sequences of CYP98s from angiosperm genomes and characterized CYP98s were included in this phylogenetic reconstruction. A distinct clade of Brassicaceae, containing the *A. thaliana CYP98A8* and *CYP98A9* is highlighted in green and species names are indicated. The amino acid alignment was performed by DIALIGN (Morgenstern, 1999), keeping positions above zero diagonal similarity. The maximum likelihood phylogenetic reconstruction was performed by phyml (Guindon and Gascuel, 2003), under consideration of the JTT model. Bootstrap support for 100 replicates is displayed at the branches of the phylogenetic reconstruction.

3.8.6. Species and identifiers used in phylogeny Figure 3.5

Name	Species	CYP98	Identifier
Ambtr1	Amborella trichopoda	CYP98A84	>evm_27.model.AmTr_v1.0_scaffold00101.79 CDS
Ambtr2	Amborella trichopoda	CYP98A85	>evm_27.model.AmTr_v1.0_scaffold00040.62 CDS
Aquco1	Aquilegia coerulea		PAC:22034145
Aquco2	Aquilegia coerulea		PAC:22027880
Araly3	Arabidopsis lyrata		PAC:16039267
Arath3	Arabidopsis thaliana	CYP98A3	PAC:19638264
Boest3	Boechera stricta		PAC:30667254
Bradi	Brachypodium distachyon	CYP98A4	PAC:32774693
Brara1	Brassica rapa		PAC:30618196
Brara2	Brassica rapa		PAC:30622649
Brara3	Brassica rapa		PAC:30625266
Brast	Brachypodium stacei		PAC:32860291
Capgr3	Capsella grandiflora		PAC:28898809
Capru2	Capsella rubella		PAC:20903099
Carpa	Carica papaya		PAC:16428634
Citcl1	Citrus clementina		PAC:20786454
Citcl2	Citrus clementina		PAC:20797703
Citcl3	Citrus clementina		PAC:20810082
Citsi1	Citrus sinensis		PAC:18118595
Citsi2	Citrus sinensis		PAC:18128133
Cofca1	Coffea canephora	CYP98A36	ABB83677.1
Cofca2	Coffea canephora	CYP98A35	ABB83676.1
Cucsa1	Cucumis sativus		PAC:16978632
Cucsa2	Cucumis sativus		PAC:16978633
Cynca	Cynara cardunculus	CYP98A49	ACO25188.1
Eucgr1	Eucalyptus grandis		PAC:32046321
Eucgr2	Eucalyptus grandis		PAC:32046538
Eucgr3	Eucalyptus grandis		PAC:32048884
Eucgr4	Eucalyptus grandis		PAC:32073169
Eutsa1	Eutrema salsugineum		PAC:20179996
Frave1	Fragaria vesca		PAC:27271707
Frave2	Fragaria vesca		PAC:27272839
Glyma1	Glycine max		PAC:30517847
Glyma2	Glycine max		PAC:30517848
Glyma3	Glycine max		PAC:30513655
Gosra1	Gossypium raimondii		PAC:26795734
Gosra2	Gossypium raimondii		PAC:26789405
Kalma2	Kalanchoe marnieriana		PAC:32589504

Kalma4	Kalanchoe marnieriana		PAC:32575868
Linus1	Linum usitatissimum		PAC:23178007
Linus2	Linum usitatissimum		PAC:23178006
Linus3	Linum usitatissimum		PAC:23143427
Linus4	Linum usitatissimum		PAC:23143381
Liter	Lithospermum erythrorhizon	CYP98A6	BAC44836.1
Lonja	Lonicera japonica	LjC3H	KC765076
Maldo1	Malus domestica		PAC:22626407
Maldo10	Malus domestica		PAC:22623485
Maldo11	Malus domestica		PAC:22643901
Maldo12	Malus domestica		PAC:22642863
Maldo2	Malus domestica		PAC:22639963
Maldo3	Malus domestica		PAC:22624965
Maldo4	Malus domestica		PAC:22627196
Maldo5	Malus domestica		PAC:22678318
Maldo6	Malus domestica		PAC:22639003
Maldo7	Malus domestica		PAC:22673017
Maldo8	Malus domestica		PAC:22635232
Maldo9	Malus domestica		PAC:22619971
Manes	Manihot esculenta		PAC:32331265
Medtr1	Medicago truncatula	CYP98A37	ABC59086.1
Mimgu1	Mimulus guttatus		PAC:28941784
Mimgu2	Mimulus guttatus		PAC:28941309
Mimgu3	Mimulus guttatus		PAC:28923814
Mimgu4	Mimulus guttatus		PAC:28922569
Mimgu5	Mimulus guttatus		PAC:28925852
Mimgu6	Mimulus guttatus		PAC:28924805
Mimgu7	Mimulus guttatus		PAC:28950611
Musac	Musa acuminata		PAC:32311599
Nicta1	Nicotiana tabacum	CYP98A31	XP_016482834
Nicta2	Nicotiana tabacum	CYP98A30	XP_016446948
Nicta3	Nicotiana tabacum	CYP98A33	ABC69384.1
Ociba1	Ocimum basilicum	CYP98A13	AAL99200.1
Ociba2	Ocimum basilicum	CYP98A13	AAL99201.1
Orysa1	Oryza sativa		PAC:33156152
Orysa2	Oryza sativa		PAC:33099057
Panha1	Panicum hallii		PAC:32512385
Panha2	Panicum hallii		PAC:32500223
Panvi1	Panicum virgatum		PAC:30310238
Panvi2	Panicum virgatum		PAC: 30197318
Panvi3	Panicum virgatum		PAC:30306346
Panvi4	Panicum virgatum		PAC:30190933
Phavu	Phaseolus vulgaris		PAC:27164879

Poptr1	Populus trichocarpa	CYP98A27	POPTR_0006s03180
Poptr2	Populus trichocarpa	CYP98A25	POPTR_0016s03080
Poptr3	Populus trichocarpa	CYP98A23	POPTR_0016s03090
Prupe1	Prunus persica		PAC:32118198
Prupe2	Prunus persica		PAC:32116027
Prupe3	Prunus persica		PAC:32120868
Prupe4	Prunus persica		PAC:32117279
Prupe5	Prunus persica		PAC:32098812
Ricco1	Ricinus communis		PAC:16814269
Ricco2	Ricinus communis		PAC:16814270
Ricco3	Ricinus communis		PAC:16820131
Rutgr	Ruta graveolens	CYP98A22	JF799117
Salmi	Salvia miltiorrhiza	CYP98A78	HQ316179.1
Salpu1	Salix purpurea		PAC:31411315
Salpu2	Salix purpurea		PAC:31411601
Salpu3	Salix purpurea		PAC:31432705
Salpu4	Salix purpurea		PAC:31432701
Salpu5	Salix purpurea		PAC:31432709
Salpu6	Salix purpurea		PAC:31394880
Salpu7	Salix purpurea		PAC:31394879
Setit	Setaria italica		PAC:32712107
Setvi	Setaria viridis		PAC:32676231
Solly1	Solanum lycopersicum		PAC:27303282
Solly2	Solanum lycopersicum		PAC:27301652
Solly3	Solanum lycopersicum		PAC:27281880
Solly4	Solanum lycopersicum		PAC:27279734
Solly5	Solanum lycopersicum		PAC:27281166
Solsc	Solenostemon scutellarioides	CYP98A14	CAD20576.2
Soltu1	Solanum tuberosum		PAC:24419832
Soltu2	Solanum tuberosum		PAC:24410676
Soltu3	Solanum tuberosum		PAC:24412859
Soltu4	Solanum tuberosum		PAC:24410863
Soltu5	Solanum tuberosum		PAC:24413573
Soltu6	Solanum tuberosum		PAC:24419962
Soltu7	Solanum tuberosum		PAC:24418017
Sorbi1	Sorghum bicolor		PAC:32731282
Sorbi2	Sorghum bicolor		PAC:32734921
Spipo	Spirodela polyrhiza		PAC:31506259
Theca	Theobroma cacao		PAC:27424773
Triae1	Triticum aestivum	CYP98A10	AJ583530.1
Triae2	Triticum aestivum	CYP98A11	AJ583531.1
Triae3	Triticum aestivum	CYP98A12	AJ583532.1
Tripr	Trifolium pratense	CYP98A44	ACV91106.1

Vitvi	Vitis vinifera	XM_002283302.2
Zeama1	Zea mays	PAC:31029252
Zeama2	Zea mays	PAC:31021991

Table 3.5 Names of species used in phylogeny Figure 3.5.

The identifiers given are GenBank identifiers or PAC identifiers to be used in Phytozome. For *A. trichopoda* the scaffold identifiers are given. For *P. trichocarpa* the identifiers of the version2 genome annotation are given.

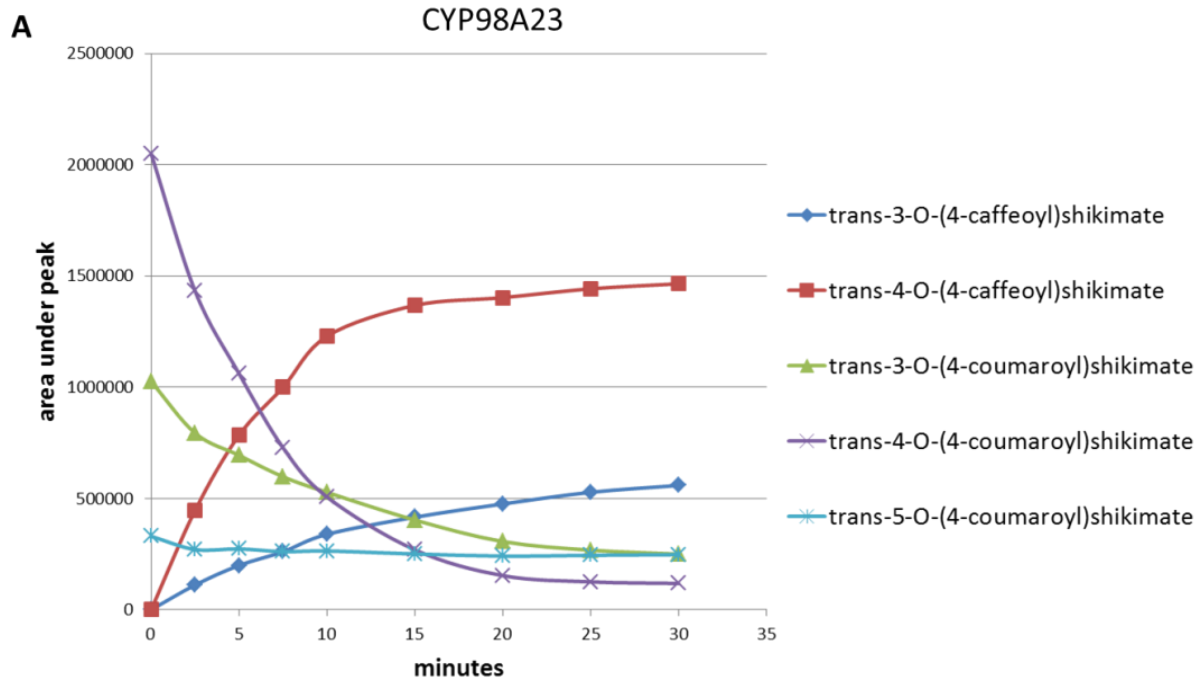
3.8.7. Pearson Correlation of substrate conversion rates

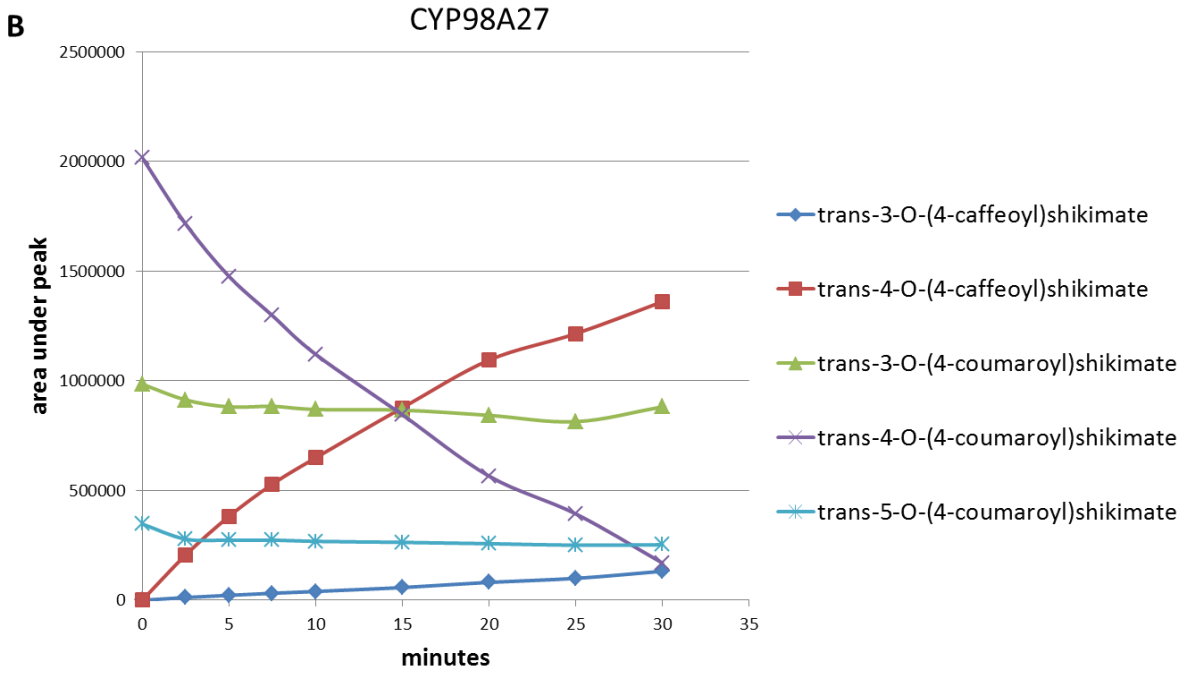
	CYP98A34 *P. patens*	CYP98A38 *S. moellendorffii*	CYP98A *P. vittata*	CYP98A19 *P. taeda*	CYP98A84 *A. trichopoda*	CYP98A4 *B. distachyon*	CYP98A27 *P. trichocarpa*	CYP98A85 *A. trichopoda*	CYP98A23 *P. trichocarpa*	CYP98A3 *A. thaliana*
CYP98A34 *P. patens*	ID	0.94	0.51	0.49	0.21	-0.05	0.09	-0.22	-0.14	-0.05
CYP98A38 *S. moellendorffii*	0.94	ID	0.50	0.32	0.02	-0.25	-0.15	-0.13	-0.20	-0.25
CYP98A *P. vittata*	0.51	0.50	ID	0.08	-0.27	-0.17	-0.32	-0.05	-0.40	-0.26
CYP98A19 *P. taeda*	0.49	0.32	0.08	ID	0.71	0.47	0.57	0.03	0.42	0.63
CYP98A84 *A. trichopoda*	0.21	0.02	-0.27	0.71	ID	0.80	0.92	-0.39	0.55	0.87
CYP98A4 *B. distachyon*	-0.05	-0.25	-0.17	0.47	0.80	ID	0.83	-0.27	0.61	0.87
CYP98A27 *P. trichocarpa*	0.09	-0.15	-0.32	0.57	0.92	0.83	ID	-0.49	0.45	0.86
CYP98A85 *A. trichopoda*	-0.22	-0.13	-0.05	0.03	-0.39	-0.27	-0.49	ID	0.22	-0.30
CYP98A23 *P. trichocarpa*	-0.14	-0.20	-0.40	0.42	0.55	0.61	0.45	0.22	ID	0.60
CYP98A3 *A. thaliana*	-0.05	-0.25	-0.26	0.63	0.87	0.87	0.86	-0.30	0.60	ID

Table 3.6 Pearson Correlation coefficients of substrate conversion rates of CYP98s.

The Pearson Correlation coefficient was assessed for a data set containing the conversion rates of 15 substrates for each CYP98.

3.8.8. Determination of *p*-coumaroyl-shikimate isomers and preferred isoforms utilized by *P. trichocarpa* CYP98s for enzyme kinetic analysis





C

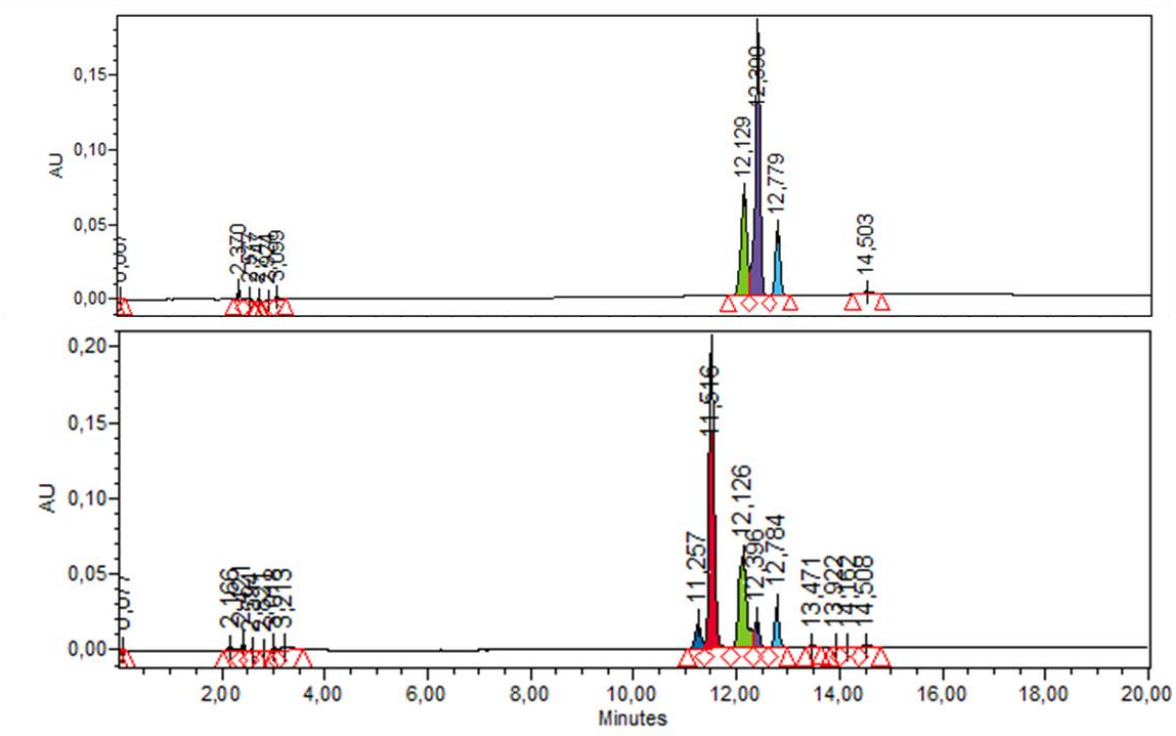


Figure 3.24 ***p*-Coumaroyl-shikimate and caffeoyl-shikimate isomer determination and testing of isomer preference by *P. trichocarpa* CYP98 isoforms.**

A: Time course of CYP98A23 enzyme incubations with *p*-coumaroyl-shikimate from enzymatic synthesis. 0.5 pmol P450 were incubated in 50mM KPi buffer at pH 7.4 with 100µM *p*-coumaroyl-shikimate for 5; 10; 15; 20; 25 and 30 min at 28°C under agitation in the dark. Plotted are peak areas from HPLC/DAD analysis of the reaction products. **B:** Time course of CYP98A27 enzyme incubations with *p*-coumaroyl-shikimate from enzymatic synthesis. 0.5 pmol P450 were incubated in 50mM KPi buffer at pH 7.4 with 100µM *p*-coumaroyl-shikimate for 5; 10; 15; 20; 25 and 30 min at 28°C under agitation in the dark. Plotted are peak areas from HPLC/DAD analysis of the reaction products. **C:** Chromatogram of HPLC/DAD analysis of *p*-coumaroyl-shikimate before and after incubation with CYP98 enzyme. Colors correspond to **A** and **B**.

3.8.9. Kinetics for CYP98s from *P. trichocarpa* with *trans*-3-*O*-(4-coumaroyl)shikimate

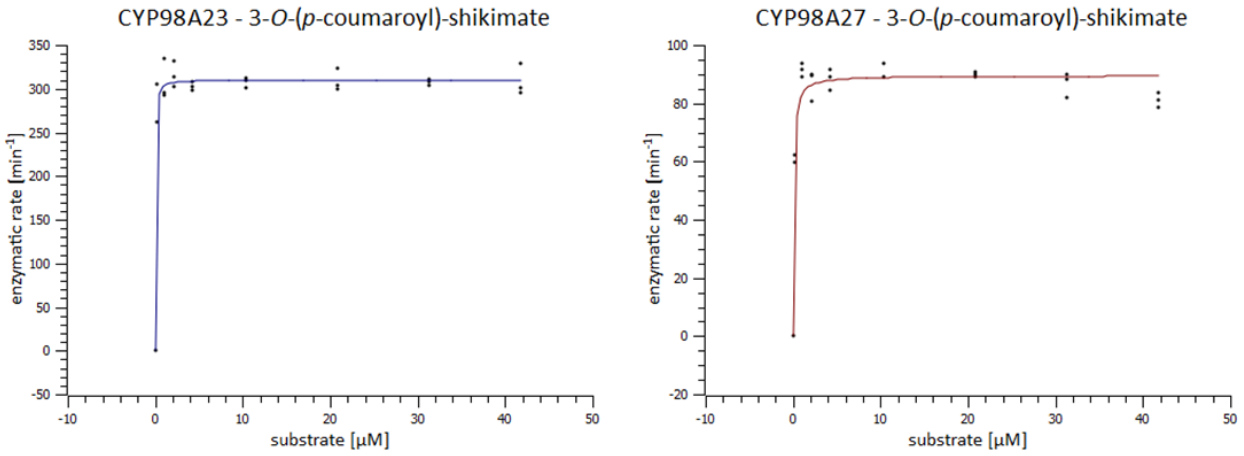


Figure 3.25 Kinetics of CYP98A23 and CYP98A27 with *trans*-3-*O*-(*p*-coumaroyl)shikimate.

To obtain enzyme kinetic data for CYP98A23, 0.05 µM of CYP98A23 protein was incubated with *trans*-3-*O*-(*p*-coumaroyl)shikimate. To obtain enzyme kinetic data for CYP98A27, 0.25 µM of CYP98A27 protein was incubated with *trans*-3-*O*-(*p*-coumaroyl)shikimate. Reactions were analysed on HPLC/DAD. Product appareance was measured and linked to a standard curve. Non-linear regression of the Michaelis-Menten equation was fitted under the Nelder-Mead-Simplex algorithm in the program SciDavis. Shown are three independent incubation replicates

A pre-test (Figure 3.24) showed the activity with *trans*-3-*O*-(4-coumaroyl)shikimate and *trans*-4-*O*-(4-coumaroyl)shikimate as substrates for CYP98A23 and CYP98A27 *in vitro*. The measurement of the *trans*-3-*O*-(4-coumaroyl)shikimate as substrate in enzyme kinetics is difficult. Coumaroyl-shikimate can only be tested as a mix of isomers in kinetic assays. A separation of the isomers is difficult due to close retention times and leads to high loss of substrate. The kinetics of *p*-coumaroyl-shikimate thus have to be interpreted carefully, as the isomers are competing substrates used in the same assay. As result of the isomer preference test (Figure 3.24), the *trans*-4-*O*-(4-coumaroyl)shikimate isomer is the preferred substrate of CYP98A23 and CYP98A27 *in vitro*. Kinetic analysis of *trans*-3-*O*-(4-coumaroyl)shikimate were performed for both enzymes, but because *trans*-3-*O*-(4-coumaroyl)shikimate is a non-major substrate of the enzyme, the results might not represent the results obtained from kinetics performed using the isolated isomer.

3.8.10. Melting curve analyses for products in qPCR

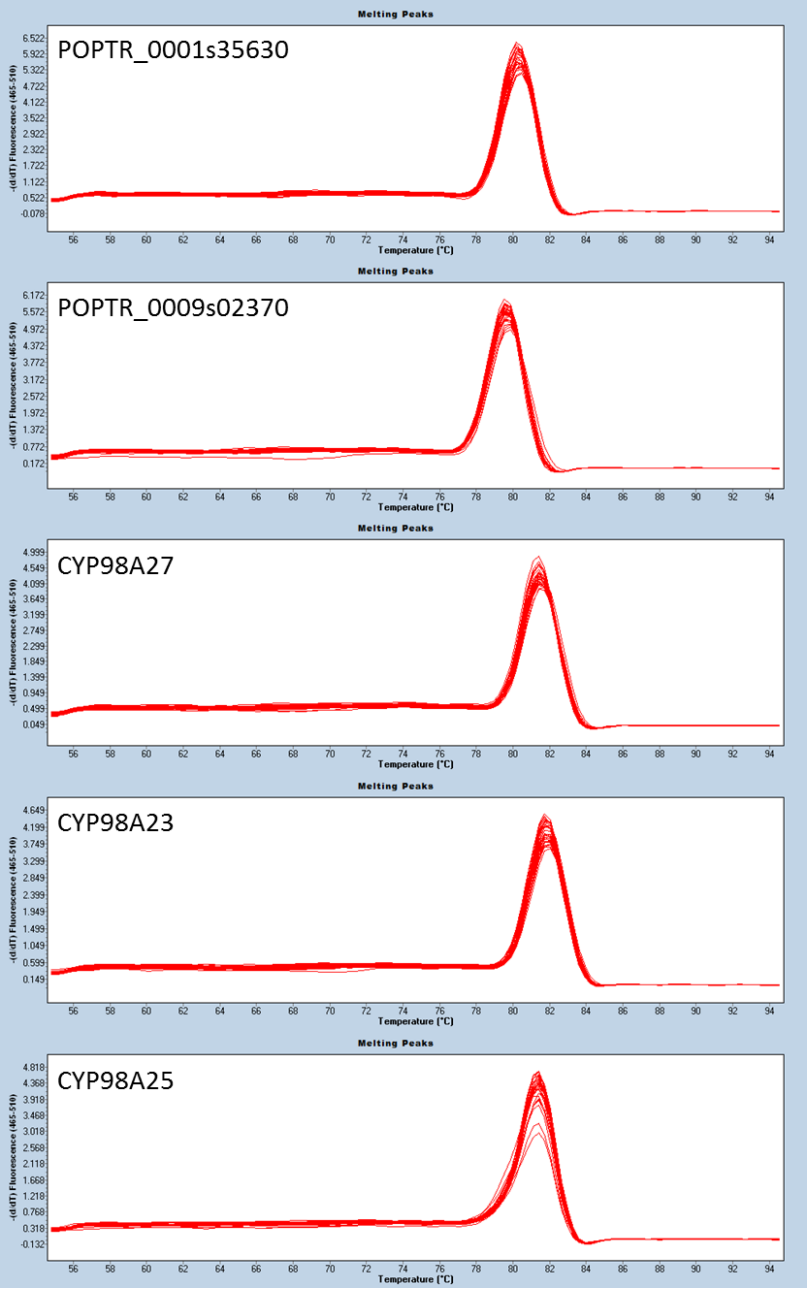


Figure 3.26 Melting curve analysis of product amplified by primer pairs used in qPCR analysis.

The primers for qPCR amplification were designed using Primer3 plus (Table 3.4) and a specificity test performed by BLAST search against the *P. trichocarpa* genome on Phytozome V11 (Goodstein et al., 2012). qPCR reactions were performed with 250nM of each forward and reverse primer and 1x SYBR® Green Master Mix (Roche). Samples were run on a LightCycler 480 (Roche). Melting curve analyses were included in the run and results are displayed here.

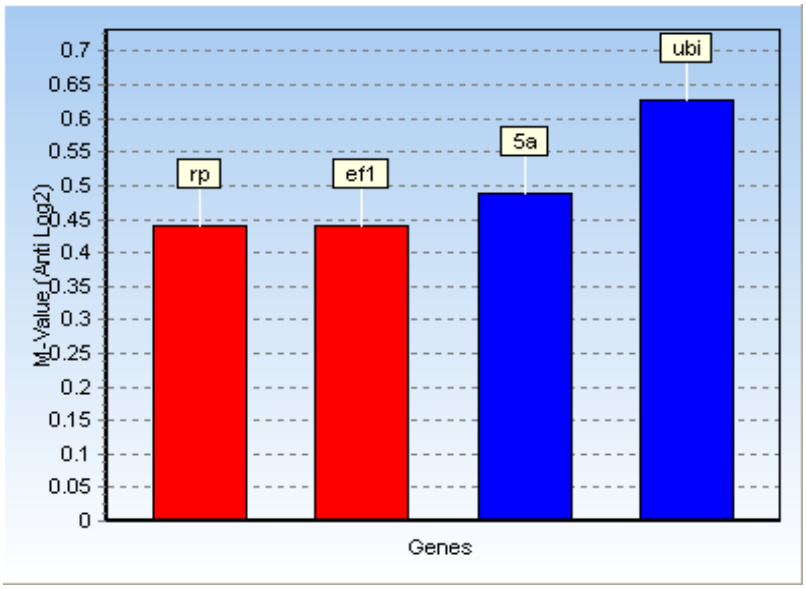


Figure 3.27 M-values of reference genes tested in qPCR analysis.

The primers for qPCR amplification were designed using Primer3 plus (Table 3.4) and a specificity test performed by BLAST search against the *P. trichocarpa* genome on Phytozome V11 (Goodstein et al., 2012). qPCR reactions were performed with 250nM of each forward and reverse primer and 1x SYBR® Green Master Mix (Roche).The M-values were calculated in the GeNorm software (Vandesompele et al.). Ribosomal Protein: 0.4; elongation factor 1ß: 0.4; Elongation factor 5 α: 0.5; Ubiquitin: 0.6.

3.8.11. Genotyping of *A. thaliana* mutant complementation lines

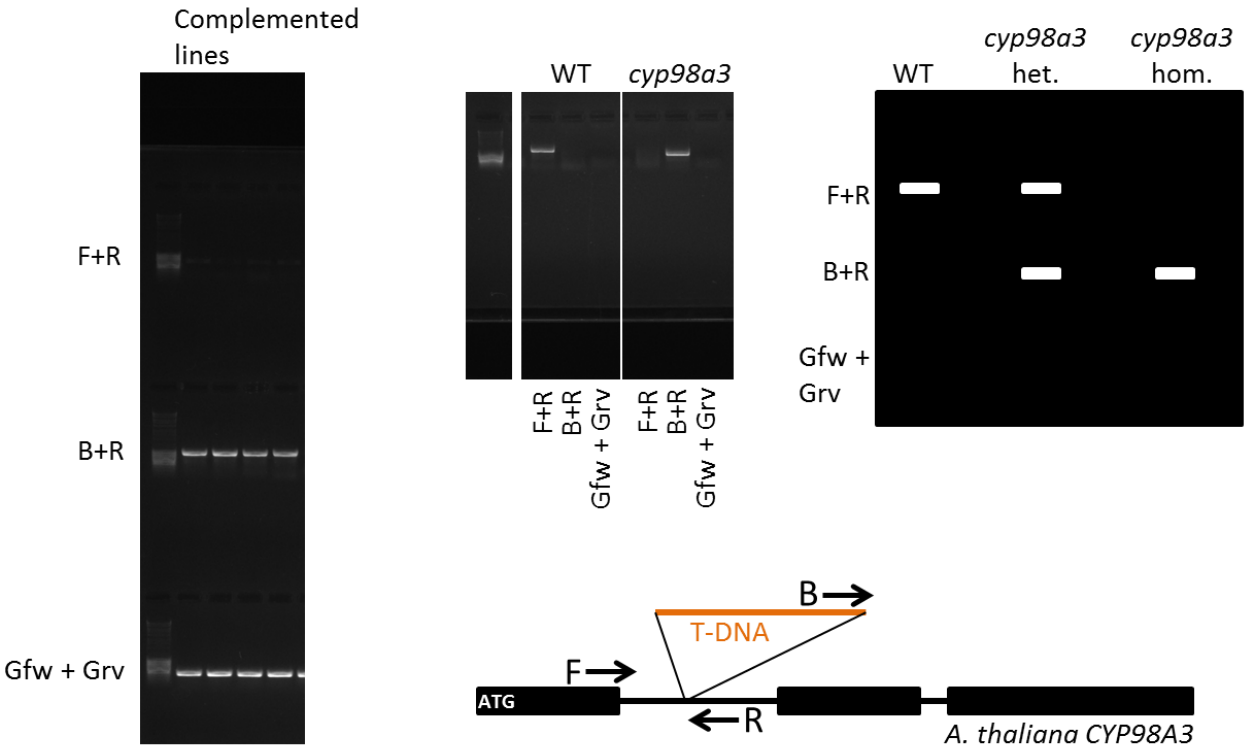


Figure 3.28 Genotyping scheme for *A. thaliana cyp98a3* mutant complementation assay with *P. trichocarpa CYP98s*.

The *A. thaliana* T-DNA insertion mutant knock-out for *CYP98A3* (Abdulrazzak et al., 2006) was used in a mutant complementation assay with the *P. trichocarpa CYP98A23*; *CYP98A25* and *CYP98A27* genes. As homozygous T-DNA lines of *cyp98a3* show dwarf morphology and are male sterile, heterozygous plants were used for transformation with the *P. trichocarpa CYP98s* under the promoter of the *A. thaliana C4H* gene (Bell-Lelong and Cusumano, 1997). Refer to Figure 3.19. **F**: Forward primer on *CYP98A3*, **R**: reverse primer on *CYP98A3*, **B**: Primer on the T-DNA insertion, **Gfw**: forward poplar gene specific primer, **Grv**: reverse poplar gene specific primer.

4. General Conclusion

In this study I characterized the *CYP98* family across the land plants and within the angiosperms. I worked with the *CYP98* members of *Physcomitrella patens*, *Selaginella moellendorffii*, *Pteris vittata*, *Pinus taeda*, *Amborella trichopoda*, *Brachypodium distachyon*, *Populus trichocarpa* and *Arabidopsis thaliana*. Genome mining and phylogenetic analysis of the *CYP98* family across the land plants and within various angiosperm orders showed that the *CYP98* family was represented by single copy members from mosses to gymnosperms. Only in angiosperms did the *CYP98* family expand, counting one to twelve members per plant genome. The phylogenetic reconstruction of the *CYP98* family was often challenging. However, bryophyte to gymnosperm *CYP98s* were all single member sets and formed monophyletic groups each with good statistical support when their relationship was reconstructed in a phylogenetic tree. Clades formed by *CYP98*s within the angiosperms usually obtained good statistical support only on the level of species or families. Relationships between angiosperm orders or isoform clades remained largely unresolved owing to poor statistical support of phylogenetic reconstructions. Nevertheless, I identified clearly independent *CYP98* duplication events within the angiosperms that did not result from the ancestral angiosperm whole genome duplication or another single event early during angiosperm radiation. The difficulty in reconstructing the molecular evolution history of the *CYP98* family with confidence may be explained by multiple factors: angiosperms underwent a rapid radiation early in their evolution and molecular clocks of different species can tick at very different paces. An example has been described by (Tuskan et al., 2006), where the *A. thaliana* clock ticks about six times as fast as the *P. trichocarpa* clock. This is possibly owed to the arborescent life style of *P. trichocarpa*. In addition, the *CYP98* family has apparently undergone numerous gene amplification events continuously throughout angiosperm evolution (Figure 4.1) and duplicates underwent evolutionary changes at very different paces, and under distinct selection pressures. The combination of frequent gene birth with vastly different molecular clocks makes it very challenging for any phylogenetic model to reconstruct the actual timing of molecular evolutionary events.

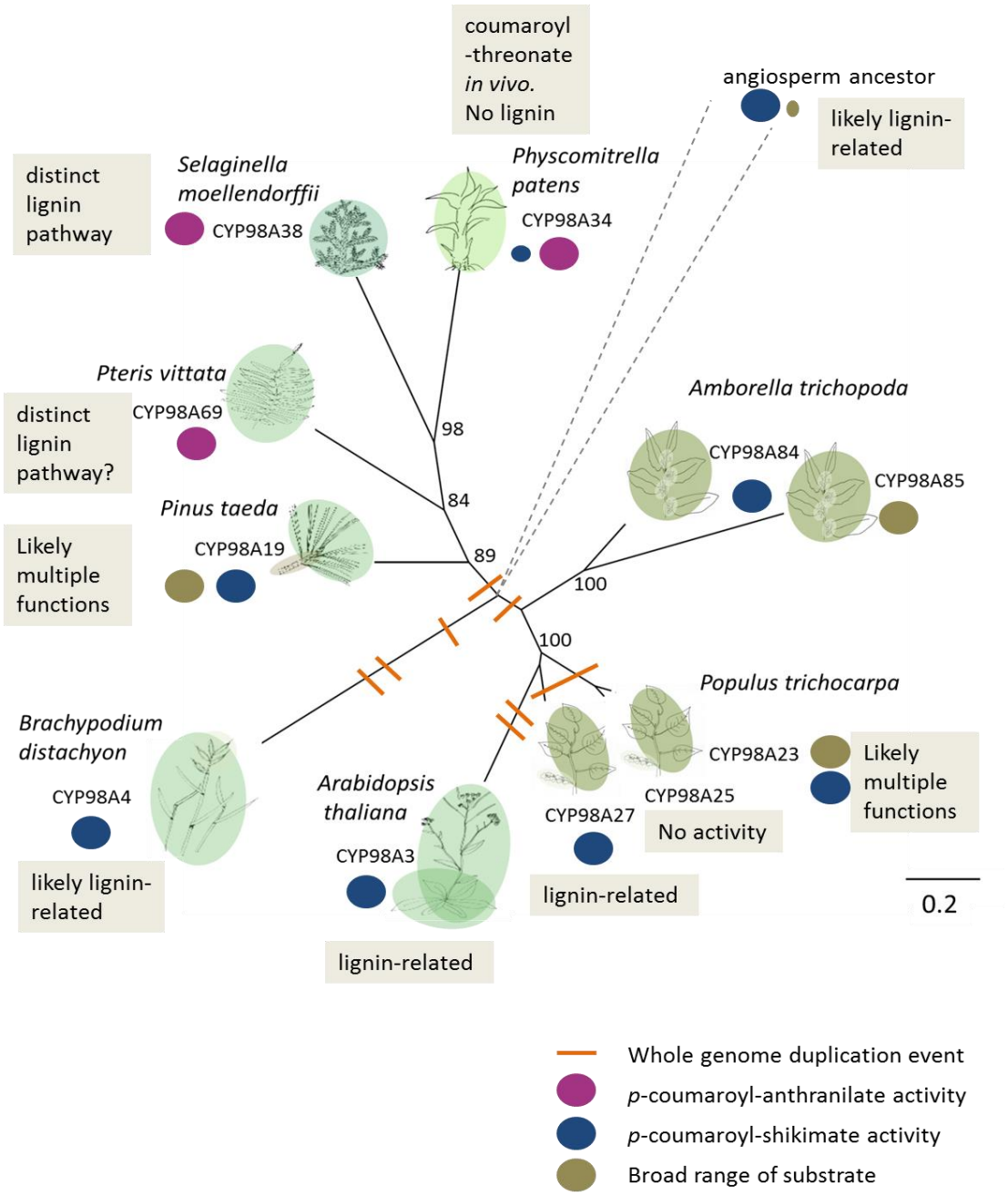


Figure 4.1 Phylogenetic reconstruction of CYP98s included in the work of this thesis and their substrate preferences *in vitro*.

Nucleotide based alignment by DIALIGN, positions with DIALIGN similarities greater than zero were kept. Maximum likelihood alignment using the phyml algorithm (model GTR). Bootstrap support by 100 replicates. Known whole genome duplication events in the angiosperms have been added schematically as orange bars (Jiao et al., 2011; Li et al., 2016). Activity with *p*-coumaroyl-anthranilate is indicated by a purple dot, *p*-coumaroyl-shikimate is indicated by a blue dot, the activity with a broad range of substrates is indicated by a brown dot.

Nevertheless, consistent across numerous phylogenetic reconstructions, the duplication events leading to the *CYP98* isoforms in *A. trichopoda* and *P. trichocarpa* clearly occurred independently within their respective lineages, and the same holds true for other duplications, for example those that occurred within the monocots.

I determined the modification of the substrate utilization profiles in a set of representative plants by screening CYP98s from all species investigated with a diverse library of possible substrates. In these biochemical analyses I found that the substrate specificity of CYP98s changed during the evolution of land plants (Figure 4.2). CYP98s from a moss, lycopod and fern preferred *p*-coumaroyl-anthranilate among the substrates tested *in vitro*, but these CYP98s from *P. patens*, *S. moellendorfii*, and *P. vittata* essentially lacked activity with *p*-coumaroyl-shikimate. A preference for *p*-coumaroyl-shikimate/quinate substrates was apparent for some of the angiosperm CYP98s tested, including those that have been previously connected to the biosynthesis of monolignols. The generation of a bryophyte *CYP98* knock-out mutant revealed *p*-coumaroyl-threonic acid as the most likely *in vivo* substrate of the enzyme. *In vivo* distinct, non-complementary functions of the moss and angiosperm CYP98s must be assumed, since the *P. patens CYP98A34* did not complement the *cyp98a3* loss of function mutant in *A. thaliana*. Loss of function of *CYP98* had severe, albeit distinct, developmental defects in *A. thaliana* and in *P. patens* beyond the expectations of losing secondary metabolic activities only. This suggests crucial roles of CYP98s and 3,4-di-hydroxylated HCCs in the development of bryophytes and angiosperms. Distinct and non-complementary esters are produced in bryophytes and angiosperms that might fulfil these developmental roles. Further analysis of the *P. patens cyp98a34* knock-out mutant might help to determine the biological role of *CYP98* in this non-lignin producing plant.

As observed previously for the lycopod *S. moellendorfii*, ferns also appear to use a distinct pathway or enzyme to produce monolignols. Ferns posses lignin composed of at least G lignin units. When the *P. vittata* CYP98 was incubated with potential substrates *in vitro,* it did not convert *p*-coumaroyl-shikimate, the substrate preferred by lignin biosynthesis-related CYP98s. Only little is known about the monolignol biosynthetic pathway in ferns, a group with known

extensive lignin content. The creation of a *CYP98* knock-out mutant in a fern species could give further information about an involvement of CYP98 in the monolignol biosynthetic pathway in ferns, and thereby test the hypothesis generated here that lignin is made independently of CYP98 mediated caffeoyl-shikimate biosynthesis in ferns. The gymnosperm CYP98 showed an *in vitro* substrate utilization profile which was intermediate between the bryophyte, lycopod and fern on one side and the angiosperms on the other side. Similar to fern species, a role of CYP98 in the monolignol biosynthetic pathway in gymnosperms has not yet been demonstrated. The *P. taeda CYP98A19* substrate utilization as well as the gene expression profile of the single-copy *P.* abies CYP98 is consistent with a dual role of CYP98s for both lignin and soluble phenolic biosynthesis in conifers. Thus, *p*-coumaroyl-shikimate/quinate specific isoforms dedicated largely to lignin biosynthesis only emerged within the angiosperms. It appears most likely that in angiosperms recruitment for lignin biosynthesis and *p*-coumaroyl-shikimate/quinate specificity evolved only once, and that multiple, independent duplications then led to isoform with relaxed substrate utilization profiles, as seen in both *A. trichopoda* and *P. trichocarpa*. In some species these isoforms may then have specialized towards alternative substrates as seen, for example, in rosmarinic acid or hydroxycinnamoyl-spermidine biosynthesis. However, our data do not exclude the possibility that, rather, broad-range substrate utilization was ancestral in angiosperms, and that recruitment for monolignol biosynthesis, alongside with *p*-coumaroyl-shikimate/quinate specificity, evolved convergently in distinct angiosperm lineages. In either case, the work of this thesis has provided further evidence that, despite the central role lignin plays in all vascular plants, its biosynthesis is plastic, distinct, and intersects with other metabolic and possibly regulatory pathways in different lineages within the vascular plants.

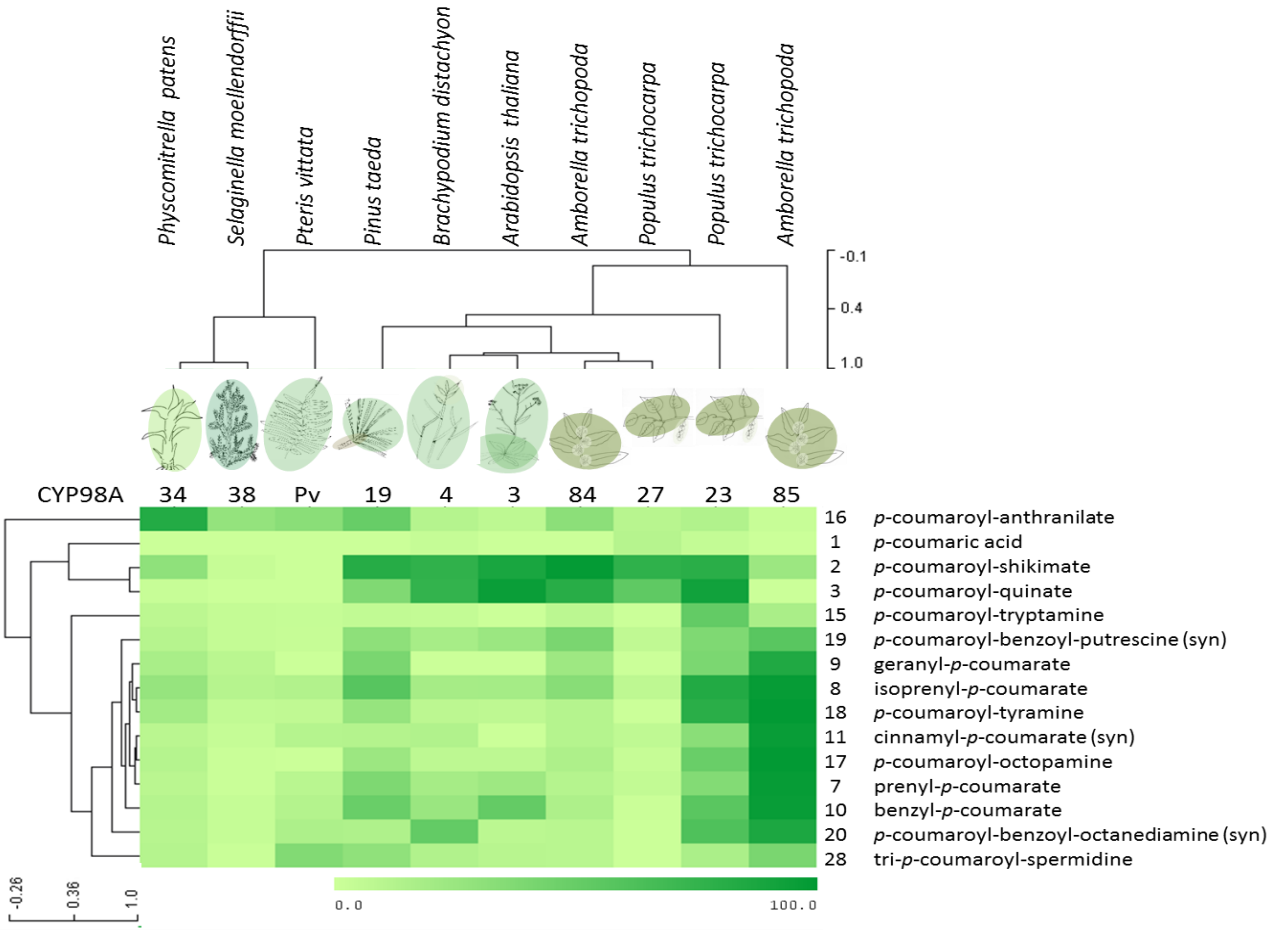


Figure 4.2 Hierarchical clustering analysis of the substrate conversion rates of all CYP98s investigated *in vitro* in this thesis.

Rates of substrate conversion in an enzymatic end point screening assay using 10 pmol of enzyme and 100µM of substrate (expected to be saturating) over 30 minutes at 28°C. The analysis of the incubation reactions was performed by HPLC/DAD. Hierarchical clustering was performed by Pearson Correlation, average linkage clustering.

5. Résumé français

Evolution de la famille *CYP98* de cytochromes P450 et de sa fonction chez les plantes terrestres

Ma recherche examine le métabolisme phénolique des plantes qui est à l'origine de la synthèse des précurseurs du bois et de molécules de défense contre les agressions de l'environnement. J'utilise des approches génomiques, biochimiques et évolutives pour comprendre comment les plantes produisent certains de ces composés qui ont permis aux plantes de conquérir le milieu terrestre, autorisant notamment la mise en place d'un système vasculaire performant et un port érigé et l'adaptation aux conditions hostiles de ce nouveau milieu.

Plus précisément, je travaille sur une famille de gènes appartenant à la famille *CYP98* de cytochromes P450, présente chez toutes les plantes terrestres. Sa présence s'étend de la mousse, qui a évolué il y a environ 450 millions d'années, jusqu'aux arbres, qui sont apparus il y a environ 300 millions d'années. Ces gènes remplissent une variété de fonctions qui semblent avoir changé au cours de l'évolution. Chez les végétaux vasculaires supérieurs la forme principale de l'enzyme catalyse l'étape d'hydroxylation en position 3 du noyau phénolique du coumaroyl-shikimate, conduisant à la production des principaux précurseurs de la lignine (monolignols). C'est une étape essentielle au bon développement de la plante. Le substrat de l'enzyme chez ces végétaux supérieurs est le *p*-coumarate conjugué à l'acide shikimique (i.e. *p*-coumaroyl-shikimate). Dans mon projet, je tente de couvrir toute l'évolution des plantes terrestres, travaillant sur les plantes représentatives des différents groupes phylogénétiques.

Hypothèses de travail

I) Les membres de la famille *CYP98* des plantes terrestres anciennes comme les Bryophytes sont impliqués dans la biosynthèse de conjugués d'acides hydroxycinnamiques, composés solubles impliqués dans la protection de la plante.

II) Le gène *CYP98* a évolué et s'est spécialisé chez les Euphyllophytes (fougères et les prêles/plantes à graines) pour la biosynthèse de la lignine, utilisant préférentiellement certaines formes d'esters phénoliques comme substrat.

III) Les gènes de la famille *CYP98* ont ensuite divergé chez les plantes terrestres supérieures pour produire des composés de défense contre les herbivores.

Objectifs

1. Caractérisation biochimique des CYP98 d'espèces représentatives de l'évolution des plantes terrestres (Bryophytes, Lycophytes, Monilophytes, Gymnospermes, Angiospermes) pour déterminer leur préférence de substrat. Les isoformes spécifiques de la biosynthèse de la lignine sont censées montrer une préférence pour le coumaroyl-shikimate, alors que les isoformes impliquées dans la formation des conjugués phénoliques solubles de la plante devraient avoir une spécificité de substrat moins restreinte.
2. Caractérisation fonctionnelle de certains CYP98 *in vivo* afin de vérifier les données obtenues lors des tests *in vitro*.
   a) La copie unique de *CYP98* chez la mousse *Physcomitrella patens* a été ciblée pour des approches de génétique inverse afin de tester l'impact du gène *in vivo* sur le développement de la plante.
   b) Le mutant *cyp98a3* d'*Arabidopsis thaliana* a été utilisé comme modèle pour tester si les membres de la famille CYP98 qui possèdent des spécificités de substrats similaires ou divergentes peuvent complémenter le déficit de fonction chez une plante vasculaire.
3. Analyse de l'évolution de la famille *CYP98* afin de comprendre les processus génétiques à l'origine de la diversification de la famille. En combinaison avec les données fonctionnelles obtenues pour les objectifs 1 et 2, un modèle d'évolution de la famille *CYP98* sera établi.

Chapitre 2 - Les CYP98 chez les plantes terrestres

Résumé

Quand les plantes ont envahi l'écosystème terrestre, il y a environ 480 millions d'années, de multiples mécanismes de protection étaient nécessaires pour faire face aux nouveaux défis environnementaux. Les plantes ont adapté leurs métabolismes, et leurs produits naturels sont devenus importants pour la survie. Une voie biosynthèse donnant naissance à de tels composés est la voie biosynthèse des phénylpropanoïdes. Nous décrivons ici le rôle d'une famille d'enzymes, les cytochromes P450 CYP98, impliquées dans le métabolisme des phénylpropanoïdes des plantes terrestres. Les CYP98s ont été décrits chez les Angiospermes comme étant impliqués dans la biosynthèse des monolignols dans la voie de biosynthèse de la lignine. Il a ensuite été démontré qu'ils participaient à la formation de produits naturels tels que l'acide chlorogénique et l'acide rosmarinique. Pour reconstruire l'évolution de la famille CYP98, nous avons étudié les CYP98s de la mousse *Physcomitrella patens*, du lycopode *Selaginella moellendorffii*, de la fougère *Pteris vittata*, du Gymnosperme *Pinus taeda*, et de deux Angiospermes, *Brachypodium distachyon* (monocotylédones) et *Arabidopsis thaliana* (eudicotylédones).

Nos reconstructions phylogénétiques suggèrent qu'une seule copie du gène *CYP98* a fondé chaque grande lignée de plantes terrestres et que la duplication de ce gène ne semblent avoir eu lieu que chez les angiospermes. Basé sur des essais biochimiques *in vitro*, nous avons montré que les CYP98s des Angiospermes testés préféraient le *p*-coumaroyl-shikimate comme substrat, tandis que les CYP98s de plantes ancestrales avaient d'autres préférences de substrat. Les CYP98s de *P. patens*, *S. moellendorffii* et *P. vittata,* ont montré une préférence pour le *p*-coumaroyl-anthranilate. Les enzymes ne produisent que peu voire pas du tout de caffeoyl-shikimate *in vitro.* Une implication de *CYP98* dans la biosynthèse de la lignine chez les fougères est discutée. Le profil métabolique du mutant knock-out dans *P. patens, cyp98a34*, indique que le *p*-coumaroyl-thréonate pourrait être le substrat de CYP98A34 *in vivo*. Le mutant knock-out de la mousse montre un phénotype de développement sévère. *CYP98A34* de *P. patens* ne complémente pas le phénotype du mutant *cyp98a3* de *A. thaliana*. Contrairement aux CYP98s des angiospermes testés, CYP98A19 du Gymnosperme *P. taeda* métabolise une gamme de

substrats large qui recouvre celles des enzymes d'Angiospermes et des plantes ancestrales, pouvant représenter une étape de transition entre les fonctions biochimiquement et physiologiquement distincts des CYP98s chez les Angiospermes et les plantes ancestrales.

Résultats

Une analyse de base des données informatiques a été effectuée pour collecter des séquences codant des CYP98 chez les Bryophytes, Lycophytes, fougères et Gymnospermes dans les bases à données des génomes et des transcriptomes. Au total, les 58 génomes actuellement disponibles sur « Phytozome v11 » (Goodstein et al., 2012) et les transcriptomes du projet 1000 plantes transcriptomes (1kp, www.onekp.com) ont été analysés. Cette recherche a montré la présence d'au moins un *CYP98* chez toutes les espèces appartenant à ces groupes. Des duplications du gène *CYP98* n'ont été identifiés que chez les Angiospermes, tant chez les monocotylédones que chez les dicotylédones. La taille de la famille *CYP98* peut varier d'une seule copie (chez par exemple *B. distachyon*, *Carica papaya*) à 12 (*Malus x domestica*) chez les Angiospermes, avec une médiane de 2 gènes, dans les 43 génomes Angiospermes analysés. Une recherche BLAST (Altschul et al., 1990) dans les génomes des algues vertes *Chlamydomonas reinhardtii*, *Volvox carteri*, *Coccomyxa subellipsoidea*, *Micromonas pusilla* et *Ostreococcus lucimarinus* n'a pas abouti à l'identification d'un homologue. De plus, une recherche dans Genbank, la base de séquence NCBI, en excluant les plantes terrestres (Embryophytes), n'a pas permi d'identifier de *CYP98*. Des espèces représentatives de chaque grande lignée de plantes terrestres (Bryophytes, Lycopodes, fougères, Gymnospermes et Angiospermes) ont été choisies pour une reconstruction phylogénétique. Des CYP98 de mousses, incluant le CYP98A34 de la mousse *P. patens*, ainsi que d'*Anthoceros* ont été inclus dans l'analyse. Les *CYP98s* de Lycopodes disponibles ont été inclus, parmi ceux-ci CYP98A38 de *S. moellendorffii*. En nous basant sur la classification des fougères établie par Smith *et al.* (Smith et al., 2006b), des CYP98s des quatre classes Polypodiopsida, Marattiopsida, Equisetopsida et Psilotopsida ont été identifiés pour l'analyse. Le CYP98 de *P. taeda*, CYP98A19, qui a été décrit précédemment (Anterola, 2002), a été inclu dans l'analyse, ainsi que d'autres CYP98s de conifères, cycas, gnétales et ginkgo. Des CYP98s d'Angiospermes monocotylédones et eudicotylédones ont été

ajoutés à l'analyse. Nous avons également inclus tous les CYP98s caractérisés biochimiquement. Ces CYP98s ont été connectés à la biosynthèse de la lignine, ainsi que la biosynthèse des composés phénoliques soluble et sont tous inclus dans la clade des Angiospermes. La topologie de la phylogénie des CYP98s suit celle des plantes terrestres, et toutes les lignées majeures forment des clades monophylétiques avec un fort soutien statistique (bootstrapping de 100 reproductions) (Figure 5.1).

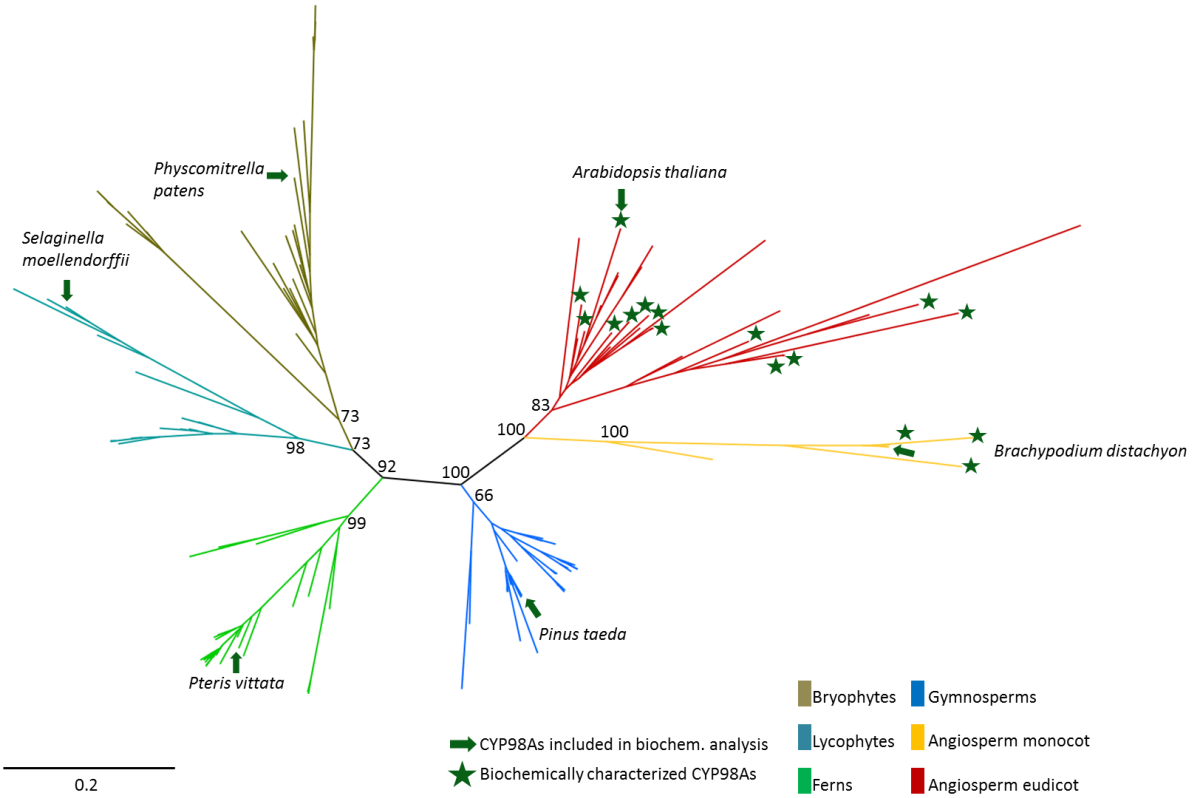


Figure 5.1 Reconstruction phylogénetique des CYP98s chez les plantes terrestres.

Arbre phylogénétique radial. Le bootstrap pour les points de ramification des grandes lignées de plantes terrestres est donné. L'arbre phylogénétique est basé sur un alignement des séquences d'acides aminés produit par DIALIGN (Morgenstern, 1999). L'arbre phylogénétique maximum likelihood a été reconstruit en utilisant PhyML (Guindon and Gascuel, 2003). Le soutien statistique bootstrap a été obtenu pour 100 répétitions. Tous les CYP98s caractérisés du point de vue biochimique appartiennent à la branche des Angiospermes et sont indiqués par des étoiles dans le cladogramme. Les membres de la famille de CYP98 caractérisés dans le cadre de mon travail de thèse ont été indiqués par une flèche verte et les noms d'espèces sont donnés dans le phylogramme.

Tous les gènes d'une lignée donnée sont issus d'un ancêtre commun. Il est clair que les deux fonctions connues des CYP98s, l'implication dans la biosynthèse de la lignine et dans la biosynthèse des composés phénoliques solubles, sont trouvés chez les Angiospermes, une clade monophylétique. Cela suggère que la diversification fonctionnelle au sein des Angiospermes ne s'est produite qu'après leur séparation des Gymnospermes. Une caractérisation détaillée de ces événements de duplication de gènes au sein des Angiospermes fera l'objet du chapitre 3 de cette thèse.

En supposant un recrutement des CYP98s pour la biosynthèse de la lignine seulement après la diversion Lycopodes / Euphyllophytes (les Lycopodes utilisent une voie indépendante de CYP98), il reste à savoir quand et à quelle fréquence ceci est arrivé. Tous les CYP98s décrits impliqués dans la biosynthèse de la lignine préfèrent le coumaroyl-shikimate comme substrat.

Des plantes représentatives de chaque lignée majeure ont été choisies. Le génome de la mousse *P. patens* a été le premier génome de Bryophyte séquencé, avec une bonne annotation du génome. *CYP98A34* est le seul *CYP98* chez *P. patens*. Le seul Lycopode avec un génome complet disponible, *S. moellendorffii*, a été étudié pour sa teneur et sa composition en lignine, mais son seul CYP98A38 n'a pas été caractérisé biochimiquement à ce jour. Des données transcriptomiques sont disponibles sur le projet de 1000 plante transcriptomes pour *P. vittata*, une espèce de fougères Leptosporangiate. En général, les fougères ont été peu étudiées. *P. vittata* est étudié pour la phytoremédiation, parce qu'il est un accumulateur d'arsenic et capable d'extraire des antibiotiques de l'eau (Danh et al., 2014; Li et al., 2015a). Une analyse de la lignine de *P. vittata* a montré une lignine constituée seulement d'unités G, similaire à la composition de la lignine des Gymnospermes (Weng et al., 2008b). Son seul CYP98, PvCYP98, a été inclus dans notre étude. CYP98A19 de *P. taeda* a été décrit dans une expérience de culture de cellules en suspension (Anterola, 2002). Quand les cellules ont été transférées dans un milieu contenant du saccharose et de l'iodure de potassium, la transcription des gènes codant les enzymes impliquées dans la biosynthèse de la lignine a été induite à l'exception de CYP98A19 et de la cinnamate 4-hydroxylase, qui n'ont montré qu'une très faible augmentation de transcription. CYP98A19 n'a pas été caractérisé biochimiquement jusqu'à présent. *B. distachyon* est une plante modèle des Angiospermes monocotylédones et fait l'objet de

recherches dans le domaine des parois cellulaires en utilisation pour produire des biocarburants (Coomey and Hazen, 2015). Nous avons sélectionné *B. distachyon* parce que son génome ne contient qu'un seul *CYP98*, *CYP98A4*, qui est sans doute impliqué également dans la biosynthèse de la lignine. Le CYP98A3 de l'Angiosperme eudicotylédone *A. thaliana* a été l'objet de recherches dans le passé et été à l'origine de la découverte de l'implication des CYP98s dans la biosynthèse des monolignols (Schoch et al., 2001; Franke and Hemm, 2002; Nair et al., 2002). Les séquences codantes de ces différents CYP98 ont été clonées dans un vecteur d'expression de levure et introduits dans la souche *S. cerevisiae* WAT11, qui contient la réductase ATR1 d'*A. thaliana*. Les microsomes de levures exprimant les enzymes ont été préparées pour effectuer des tests enzymatiques. Par spectrophotométrie, les CYPs correctement exprimés montrent une bande Soret à absorption maximale de 450 nm quand ils sont réduits et complexés avec du CO. La quantité d'enzyme fonctionnelle peut être déterminée par un spectre différentiel de P450 réduit et de P450 réduit et complexé avec du CO (Figure 5.2). Les spectres CO obtenus pour toutes les enzymes indiquent la présence d'enzyme fonctionnelle.

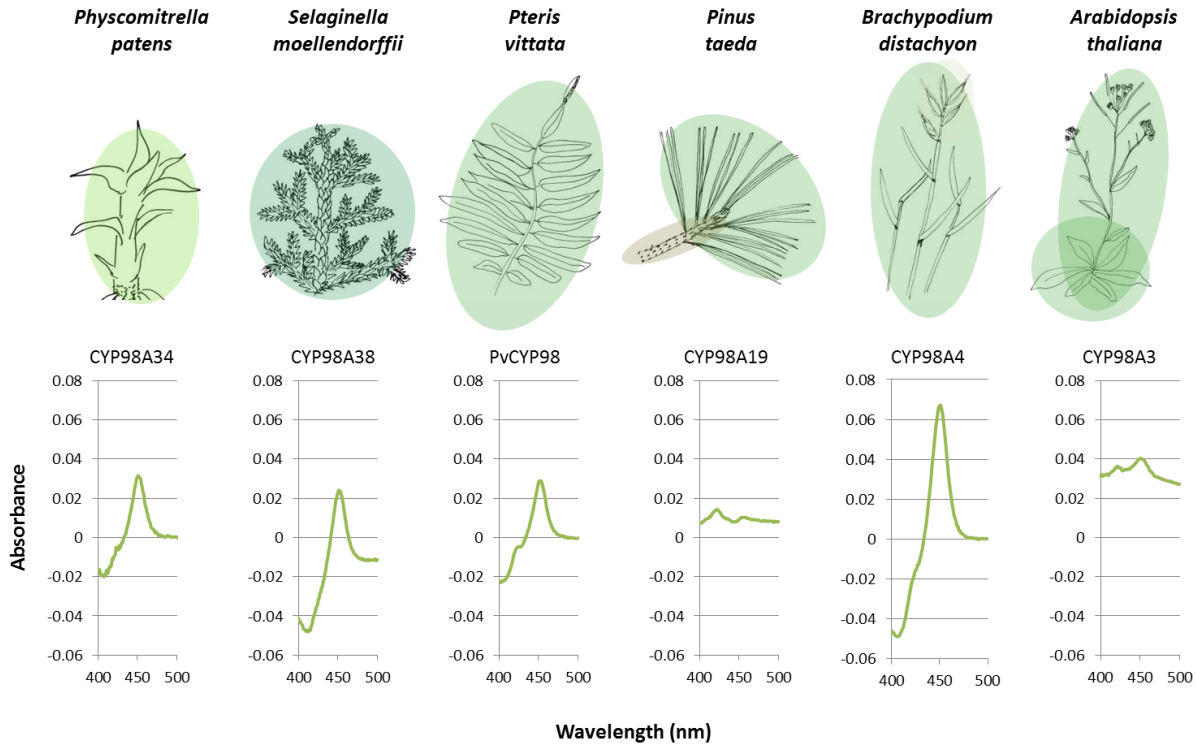


Figure 5.2 Spectres CO différentiel de CYP98s inclus dans l'analyse biochimique.

Les microsomes de levure ont été incubés avec divers substrats potentiels (Figure 2.4). La plupart de ces substrats ne sont pas disponibles dans le commerce et ont été synthétisés chimiquement ou enzymatiquement. La gamme de substrats comprenant des substrats connus des CYP98s, tels que le *p*-coumaroyl-shikimate (synthétisé par voie enzymatique) et le *p*-coumaroyl-quinate (synthétisé chimiquement). L'acide *p*-coumarique a été inclus dans l'expérience pour tester une hydroxylation sur l'acide libre. Plusieurs phénolamides ont été inclus car ils sont connus pour être métabolisés par certains CYP98s. Pour ne citer que deux exemples, *Triticum aestivum* CYP98A11 et CYP98A12 peuvent hydroxyler le *p*-coumaroyl-tyramine (Morant et al., 2007). En outre, plusieurs esters coumariques, précurseurs potentiels d'esters caféiques qui existent dans la nature, ont été synthétisés par voie chimique pour l'expérience. C'est le cas par exemple des prenyl-, isoprenyl- et benzyl-*p*-coumarate (Rubiolo et al., 2013). Pour élargir la gamme de substrats potentiels, certains substrats artificiels ont été synthétisés chimiquement et testés.
La formation du produit a été analysée par chromatographie en phase liquide (Figure 5.3).

Les résultats obtenus révèlent des préférences de substrat différentes pour les CYP98s d'Angiospermes et de Bryophytes, Lycopodes et fougères. Les CYP98s testés utilisent des substrats d'origine naturelle, mais aussi des substrats synthétiques comme des esters phénoliques et des phénolamides. L'acide *p*-coumarique libre et les esters et amides de cinnamate n'ont pas été convertit par les CYP98s testés. Le *p*-coumaroyl-shikimate et le *p*-coumaroyl-quinate sont les substrats préférés des CYP98s des Angiospermes, mais ne sont que faiblement utilisés par les enzymes de Bryophytes, Lycopodes et fougères. La gamme de substrats de CYP98A19 (Gymnosperme) est intermédiaire entre les deux groupes. Cette enzyme de *P. taeda* convertit de nombreux substrats. Les CYP98s de Bryophytes et Lycopodes ont une préférence *in vitro* pour *p*-coumaroyl-anthranilate. Une analyse de classification hiérarchique des données montre que les CYP98s de Bryophytes, Lycopodes et fougères d'une part et de Gymnospermes et d'Angiospermes d'autre part forment deux groupes distincts (Figure 5.4).

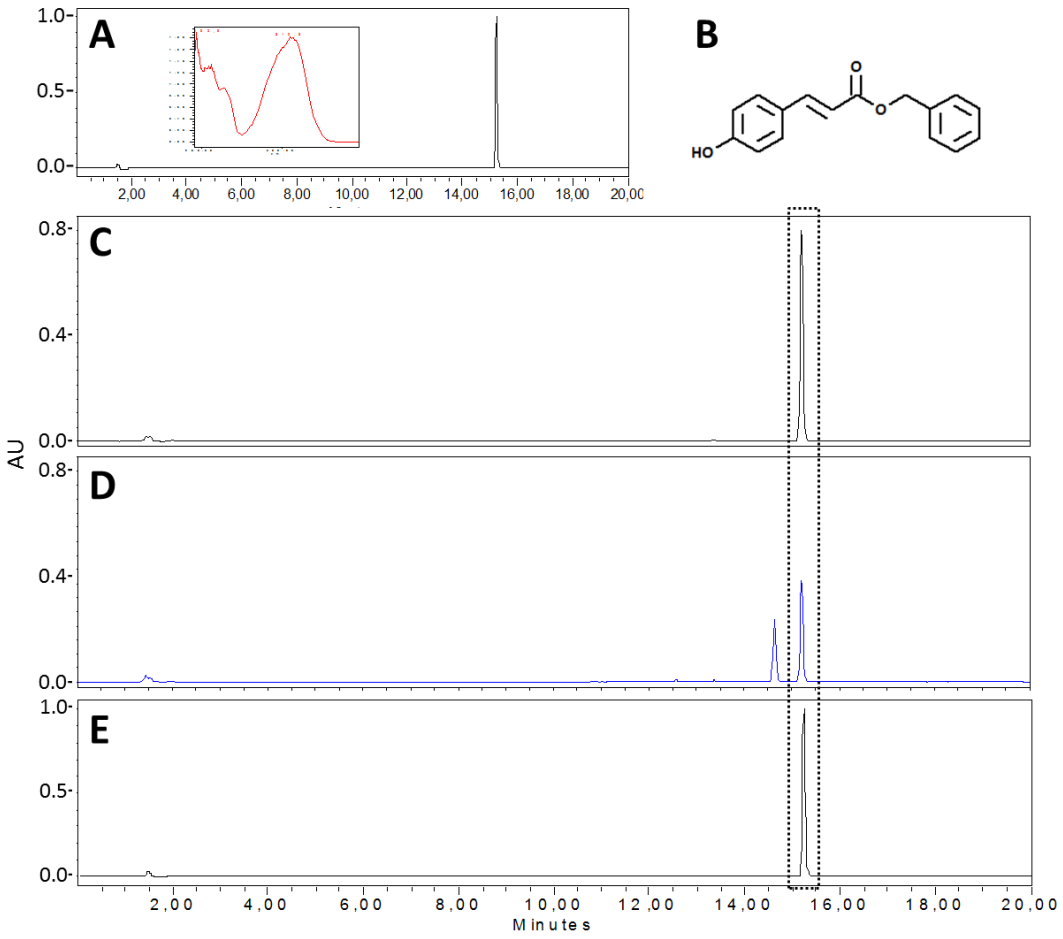


Figure 5.3 Incubation de microsomes de CYP98A19 de *P. taeda* avec le benzyl-*p*-coumarate. Analyse par HPLC / DAD.

A : Standard de benzyl-*p*-coumarate et spectre UV correspondant. **B :** structure du benzyl-*p*-coumarate **D et E :** Incubation de microsomes préparés à partir de levures exprimant CYP98A19 avec le benzyl-*p*-coumarate, **E :** Contrôle sans NADPH, **D :** Réaction avec NADPH. **C :** contrôle vecteur vide (microsomes préparés à partir de levures transformées avec pYeDP60).

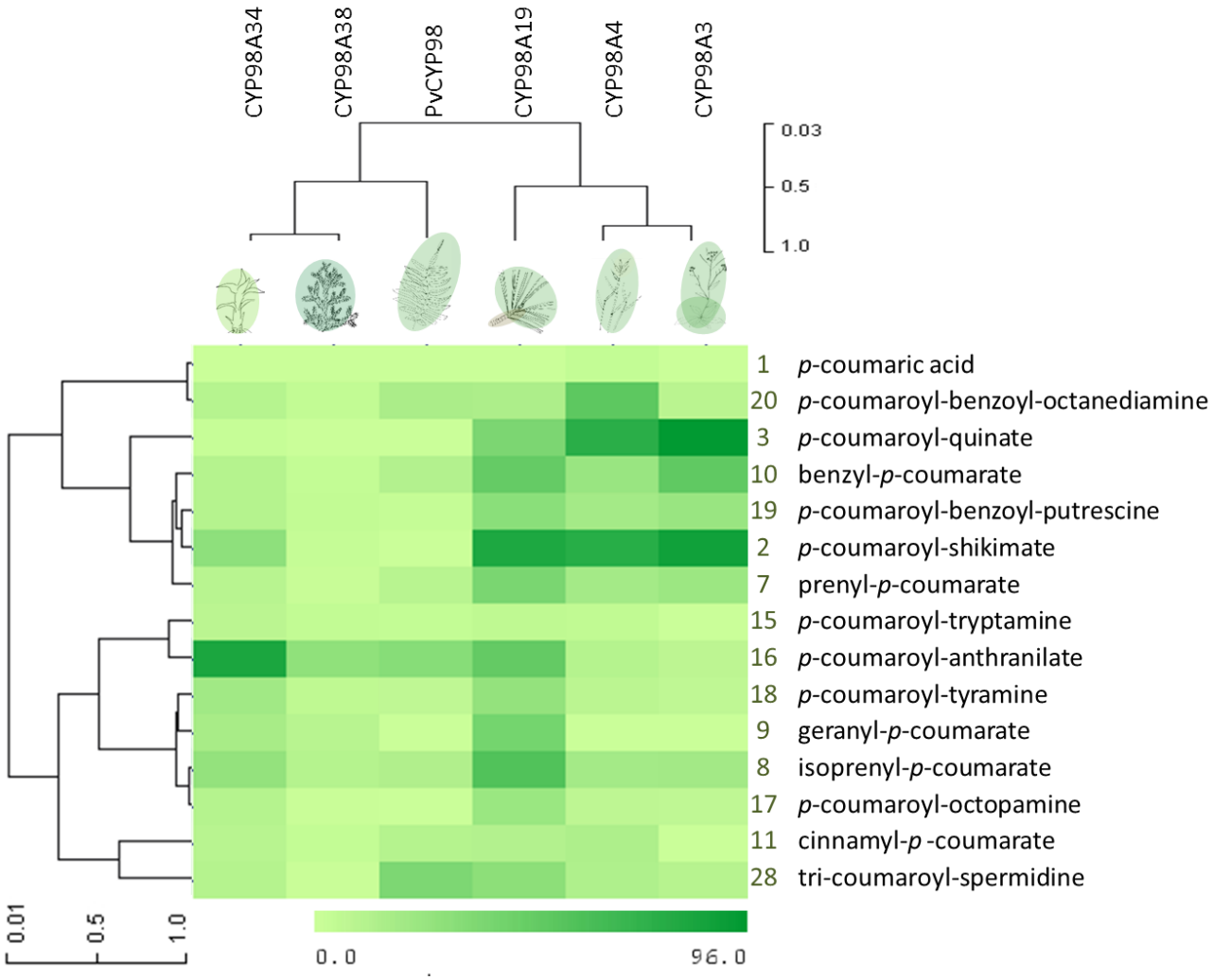


Figure 5.4 Classification hiérarchique des substrats et P450s testés biochimiquement.

Groupement par liaison moyenne utilisant la corrélation de Pearson. Les taux correspondants de conversion sont présentés en détail dans le Figure 2.6. Incubation de 10 pmole de P450 avec 100 µM de substrat pendant 30 minutes.

Comme le profil d'utilisation et de préférence de substrat des plantes ancestrales différait beaucoup de ceux des Angiospermes, le rôle biologique des CYP98s dans les plantes ancestrales a été étudié par une approche de génétique inverse.

Un mutant knock-out de *CYP98A34* de *P. patens* a été généré par G. Wiedemann et H. Renault, en profitant du taux élevé naturel de recombinaison homologue de la plante. La cassette de

sélection a été introduite vers l'extrémité 5' du gène, afin de supprimer les domaines P450 fonctionnels. Après vérification de l'intégration de la cassette, plusieurs lignées mutantes avec intégration unique ont été identifiées. Le phénotype de ces lignes est très fort, les plantes n'étant plus capables à pousser au stade gamétophore (Figure 5.5).

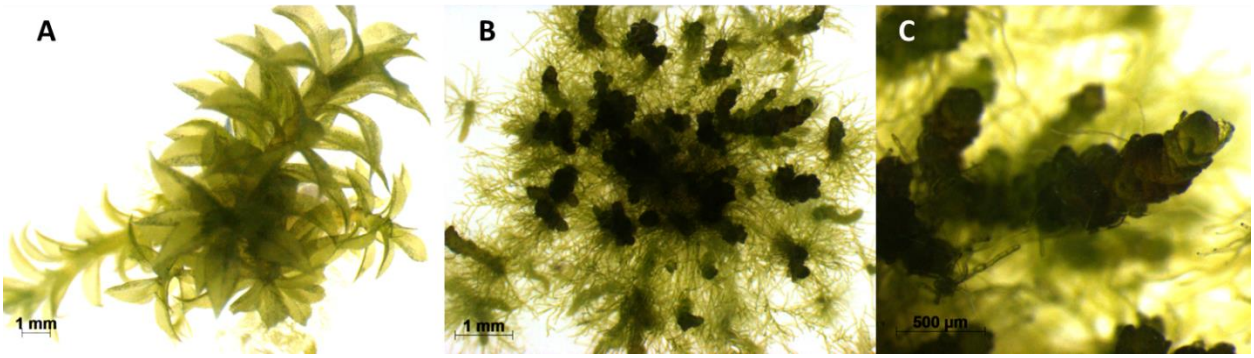


Figure 5.5 Phénotype du mutant knock-out *cyp99a34* de la mousse *P. patens*.

Généré par H. Renault. **A :** Gamétophores âgé de 8 semaines de type sauvage. **B:** Mutant knock-out *cyp98a34* âgé de 8 semaines montrant un phénotype sévère. **C:** Gros plan du phénotype du mutant knock-out.

Une analyse par HPLC des extraits méthanoliques des mutants par HPLC a révélé plusieurs différences entre la mousse sauvage et les mutants (Figure 5.6). Surtout deux pics éluant tôt dans les chromatogrammes des extraits de la mousse sauvage sont absents dans les chromatogrammes des mutants.

Pour déterminer les masses des composés d'intérêt, une analyse UPLC-MS / MS a été effectuée par H. Renault. Sur la base des spectres de masse et de la littérature, les pics ont été identifiés comme correspondant probablement de l'acide caféoyl-thréonique (Hahn and Nahrstedt, 1993; Kuczkowiak et al., 2014). La fonction de l'acide caféoyl-thréonique chez les plantes n'est pas décrite. Il a été détecté chez des variétés de plantes, comme *Fagus sylvatica*, *P. patens*, *Saniculiphyllum guangxiense*, *Miscanthus sacchariflorus*, *Miscanthus giganteus* et *Cornus controversa* (Lee et al., 1995; Richter et al., 2012; Parveen et al., 2013; Cadahía et al., 2014; Geng et al., 2014).

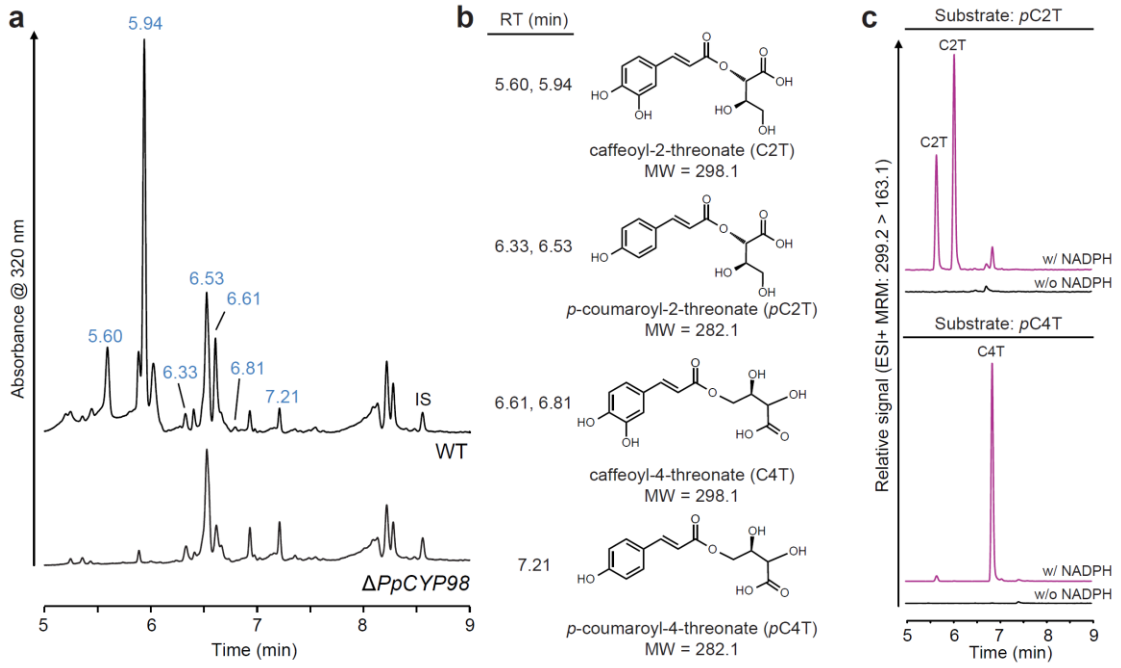


Figure 5.6 Spectres HPLC / DAD d'extraits de gamétophores de *P. patens* de type sauvage et de *cyp98a34* knock-out.

Analyse HPLC par H. Renault. **a:** UV chromatogram showing the absence of major peaks in the ΔPpCYP98 mutant gametophore crude extract. IS, internal standard (morin). **b:** Names and structures of molecules at the indicated retention times (RT). **c:** PpCYP98-dependent conversion of *p*-coumaroyl-2-threonate (pC2T) and *p*-coumaroyl-4-threonate (pC4T) esters into corresponding caffeoyl threonate esters (C2T and C4T). Control reactions without NADPH were concurrently analyzed. Molecules were detected using dedicated multiple reaction monitoring (MRM) methods.

L'acide caféoyl-thréonique a été identifié comme substrat d'une polyphénol oxydase chez *Dactylis glomerata* (Parveen et al., 2008). Les polyphénol oxydases sont considérées comme étant impliquées dans la défense des plantes (Constabel and Barbehenn, 2008). Treize polyphénols oxydases ont été identifiées dans le génome de *P. patens* (Tran et al., 2012). L'absence d'acide caféoyl-thréonique chez les lignées mutantes *cyp98a34* a suggère que l'acide *p*-coumaroyl-thréonique pourrait être le substrat de CYP98A34 *in vivo*. Le *p*-coumaroyl-thréonate a été synthétisé par voie chimique (M. Schmitt et coll., UMR CNRS 7200) et utilisé comme substrat pour les essais *in vitro*. Il a ainsi été possible de démontrer que CYP98A34

hydroxyle le *p*-coumaroyl-thréonate, quoique moins efficacement que le *p*-coumaroyl-anthranilate *in vitro*.

Comme CYP98A34 de la mousse montrait une - quoique plutôt faible - hydroxylation de *p*-coumaroyl-shikimate *in vitro*, une expérience de complémentation par le *CYP98A34* de *P. patens* du mutant *cyp98a3* d'*Arabidopsis thaliana* a été tentée. L'expression de CYP98A34 était sous contrôle du promoteur de la cinnamate-4-hydroxylase (C4H) d'*A. thaliana*. Des lignées transgéniques contenant la construction promoteur C4H et *CYP98A34*, et homozygotes pour le knock-out *cyp98a3,* ont été identifiées par criblage sur des milieux selectifs contenant du BASTA et de la kanamycine et par génotypage par PCR. La présence de transcrits de *CYP98A34* de *P. patens* a également été vérifiée par RT-PCR (Figure 5.7). Les résultats obtenus ont montré que CYP98A34 n'est pas capable de complémenter le phénotype sévère du mutant knock-out *cyp98a3* d'*A. thaliana*. En effet les plantes qui étaient homozygotes pour le *cyp98a3* knock-out et qui exprimaient *CYP98A34* présentant le même phénotype que les mutants de *cyp98a3* non-complémentés.

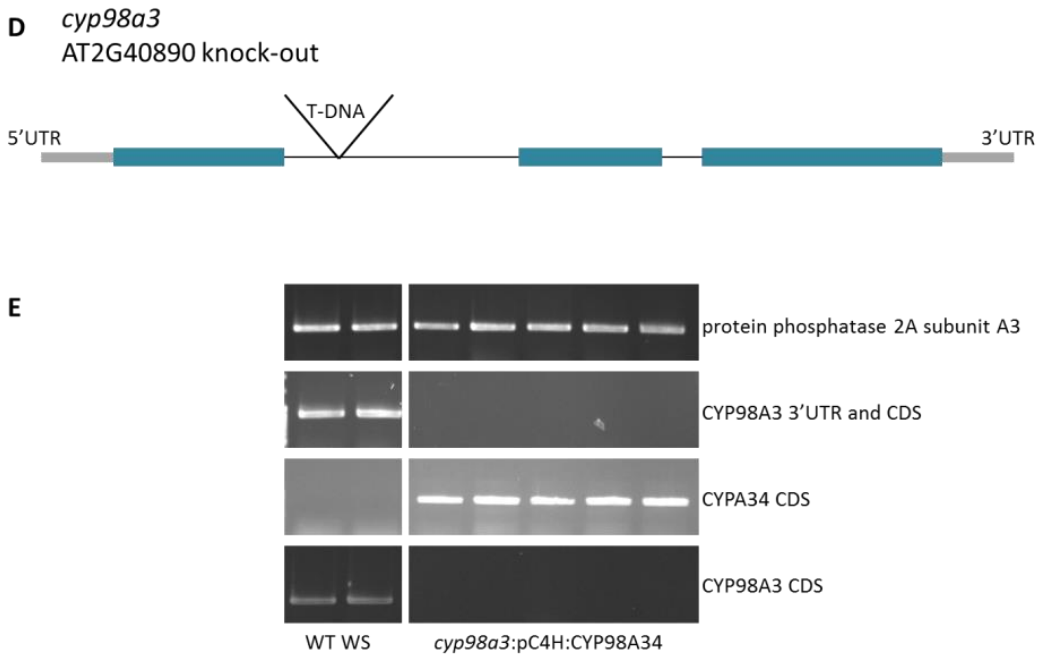


Figure 5.7 Complémentation du mutant *cyp98a3* d'*A. thaliana* par *CYP98A34* de *P. patens*.

A: Gros plan sur plantes d'*A. thaliana* âgées de 4 semaines. A gauche, une lignée homozygote pour le knock-out *cyp98a3* et exprimant *CYP98A34*; au centre, une lignée homozygote pour le knock-out cyp98a3 complémentée par *CYP98A3*; à droite *A. thaliana Wassilewskija*, type sauvage. **B:** Onze lignées de plantes homozygotes pour le knock-out *cyp98a3* et exprimant *CYP98A34*. **C:** Un gros plan du mutant knock-out homozygote *cyp98a3*. **D:** Schéma du locus *CYP98A3* chez *A. thaliana* et localisation de l'insertion d'ADN-T pour créer le *cyp98a3* knock-out. **E:** RT-PCR d'*A. thaliana* de type sauvage et des lignées homozygotes pour le knock-out *cyp98a3* exprimant *CYP98A34*.

Conclusion

Les résultats obtenus dans cette étude indiquent que le substrat préféré des CYP98s a changé au cours de l'évolution des plantes terrestres. Les formes qui présentent une préférence pour l'acide *p*-coumaroyl-shikimique ne sont présentes que chez les plantes à graines. Le *CYP98* unique du Bryophyte *P. patens*, mais aussi celui du Lycopode *S. moellendorffii* et de la fougère *P. vittata* n'ont quasiment pas d'activité de métabolisation de l'acide *p*-coumaroyl-shikimique mais produisent divers autres esters ou amides de caféoyle. *In vivo*, *CYP98A34* de *P. patens* ne peut pas complémenter le mutant *cyp98a3* d'*A. thaliana*. Néanmoins, la perte de fonction de CYP98 tant chez les Angiospermes que chez les Bryophytes, montre de sévères déficits du

développement qui vont au-delà des attentes associées à la perte d'activités dans le métabolisme secondaire. Chez *A. thaliana,* la formation de lignine et l'accumulation des flavonoïdes dans le mutant peuvent être découplés du phénotype nain (Li et al., 2010; Gallego-Giraldo et al., 2011; Kim et al., 2014). Ceci indique que ce n'est pas forcement le changement dans les principaux métabolites secondaires (lignine ou flavonoïdes) qui causent le phénotype nain. On retrouve une situation similaire chez *P. patens.* L'absence d'accumulation de caffeoyl-thréonate dans le mutant *cyp98a34* peut être soit la cause soit une coïncidence avec le phénotype de développement. Dans les deux cas, les données présentées ici démontrent un rôle crucial des *CYP98s* et des hydroxycinnamates 3,4-dihydroxylés dans le développement chez les Bryophytes et les Angiospermes. Nous montrons aussi que des esters distincts et non complémentaires sont produits chez les Bryophytes et chez les Angiospermes pour remplir ces rôles développementaux.

Les fougères produisent de larges quantités de lignine, mais le CYP98 de *P. vittata* ne montre pas un profil de substrat similaire à celui des CYP98s associés à la biosynthèse de la lignine. En particulier, aucune activité avec le *p*-coumaroyl-shikimate n'est détectable. Il apparaît ainsi que les fougères n'utilisent pas caféoyl-shikimate formé par un CYP98 pour la biosynthèse de la lignine. La même chose est constatée pour CYP98A38 de *S. moellendorffii*, qui est également incapable de produire du caféoyl-shikimate. A noter que *S. moellendorffii* possède une enzyme distincte, SmF5H / CYP788A1 (DN837863), qui est capable de contourner les étapes de 3'- et 5'-hydroxylation impliquées dans la biosynthèse des précurseurs de la lignine (Weng et al., 2008a). Cela est compatible avec une fonction non liée à la lignine du CYP98 de *S. moellendorffii*. Le profil de spécificité de substrat de CYP98A38 est également plus semblable à celui des CYP98s de *P. patens* et de *P. vittata* qu'à celui CYP98s des Angiospermes, *A. thaliana* et *B. distachyon*. Tout ceci semble indiquer que des CYP98s qui sont plus spécifiques pour le *p*-coumaroyl shikimate ont été recrutés pour la biosynthèse de la lignine chez les plantes à graines. Les Gymnospermes contiennent un seul *CYP98*. CYP98A19 de *P. taeda* montre une forte flexibilité d'acceptance de substrat, et est capable de métaboliser le *p*-coumaroyl shikimate *in vitro*. Cela pourrait signifier qu'il est impliqué à la fois dans la biosynthèse des précurseurs de la lignine, et dans la biosynthèse d'autres composés solubles. Cependant, il reste également possible que,

même chez les Gymnospermes, les CYP98s ne contribuent pas à la biosynthèse des monolignols. Dans une expérience sur des cultures de cellules en suspension, l'addition de Phe dans le milieu de culture seulement provoque une très faible hausse de l'expression de *CYP98A19* et cinnamate 4-hydroxylase, contrairement à d'autres gènes dans la voie de biosynthèse de la lignine (Anterola, 2002). Ceci avait été interprété comme un contrôle transcriptionnel des CYP indépendant du reste des gènes de la voie des phénylpropanoïdes. Cependant, les *CYP98s* chez les Angiospermes sont co-régulés avec la plupart des autres gènes de la voie biosynthèse des monolignols, comme indiqué par l'analyse génétique de co-expression chez *A. thaliana*, le peuplier et le riz (Ehlting et al., 2005; Hirano et al., 2012; Chen et al., 2014). En revanche, le seul *CYP98* du Gymnosperme *Picea glauca* est absent d'un réseau de co-expression de gène de la voie biosynthèse des monolignols (Porth et al., 2011). Il apparaît ainsi que les CYP98s des Gymnospermes ne sont pas seulement biochimiquement distincts de ceux des angiospermes associés à la lignine, mais qu'ils sont également sous un contrôle transcriptionnel distinct.

Seuls les Angiospermes possèdent plusieurs copies de *CYP98* ce qui pourrait indiquer que la diversification fonctionelle pouvait être en lien avec la duplication des gènes dans ce groupe.

Chapitre 3 Les CYP98 chez les Angiospermes

Résumé

Une grande diversité d'Angiospermes peut être trouvée dans presque tous les environnements. Dans ce chapitre nous décrivons l'évolution et la fonction des enzymes de la famille *CYP98* chez les Angiospermes. Bien que beaucoup de familles de gènes de plantes soient fortement conservées et retrouvées chez tous les Embryophytes, ce nombre de gène peut beaucoup varier au sein de différents groupes. Pour s'adapter aux défis d'un environnement spécifique, la famille peut s'élargir et se diversifier fonctionnellement. La duplication de gènes peut être à l'origine de ces événements adaptifs. L'implication des *CYP98* dans la voie biosynthèse des monolignols chez les Angiospermes a été décrite. Ces *CYP98s* sont également impliqués dans la voie de biosynthèse menant aux composées phénoliques solubles chez les plantes. Une duplication des *CYP98s* est observée seulement chez les Angiospermes. Toutes les Angiospermes étudiées possèdent au moins une copie de *CYP98*. Une reconstruction phylogénétique de la famille *CYP98*, prenant en compte les différents ordres d'Angiospermes, ne montre pas la formation de classes distinctes en corrélation avec la fonction biochimique ou physiologique des enzymes. Au contraire, les résultats observés suggèrent plusieurs duplications indépendantes dans la famille *CYP98*. Des événements de duplication indépendants chez *Populus trichocarpa* et *Amborella trichopoda* ont été caractérisés biochimiquement. Dans chacune des deux espèces, un CYP98 est spécialisé pour le *p*-coumaroyl-shikimate et semble donc potentiellement impliqué dans la biosynthèse des monolignols. Une deuxième isoforme métabolise une gamme de substrat très large. Une troisième isoforme chez *P. trichocarpa* ne montre aucune fonction *in vitro* et ne peut pas complémenter la déficience de CYP98A3 chez un mutant knock-out d'*A. thaliana*. Les données cinétiques des deux isoformes actives de *P. trichocarpa* renforcent, avec les données biochimiques et des données de co-expression de gène, l'hypothèse que l'une des isoformes est impliquée dans la formation des monolignols, pendant que l'autre isoforme est impliquée dans la formation d'esters phénoliques solubles. Les deux isoformes de *P. trichocarpa* complémentent avec la même efficacité le mutant knock-out *cyp98a3* d'*A. thaliana*. Chez les Salicaceae, la première duplication de *CYP98* a eu lieu avant la duplication complète du

génome. La duplication des gènes en tandem qui a donné naissance à *CYP98A23* et *CYP98A25* du peuplier a par contre eu lieu après la duplication complète du génome.

Résultats

Des espèces avec de multiples *CYP98* existent seulement chez les Angiospermes. Les résultats présentés au chapitre 2 montraient que tous les *CYP98* des Angiospermes résultaient d'un seul ancêtre. 43 génomes d'Angiospermes sont disponibles dans la base de données génomique phytozomev11. (Goodstein et al., 2012). 123 séquences *CYP98* ont été trouvées dans ces génomes d'Angiospermes. La taille de la famille CYP98 varie de 1 à 12 membres, avec une médiane de 2.

En suivant la dernière classification des Angiospermes de « l'Angiosperm Phylogeny Group » (APGIV) (Chase et al., 2016) pour la classification des ordres et des familles, un ensemble de données qui comprend deux espèces par ordre (si disponible) a été créé. Un schéma des ordres est présent en Figure_II_2. Ont été choisis préférentiellement des *CYP98* caractérisés, puis les espèces dont le génome était complètement séquencé, ensuite des données du projet 1000 transcriptomes de plantes (onekp.com). Dans une reconstruction phylogénétique globale avec toutes ces séquences, une distribution des *CYP98* corrélant à leur fonction n'est pas détectée. Au contraire, les *CYP98s* d'une même famille se trouvent le plus souvent très proches dans la même clade. Le soutien statistique pour les branches est fort pour quelques clades comme *A. trichopoda,* les monocotylédones et une clade contenaient *CYP98A8* et *CYP98A9* d'*A*. *thaliana*. Les branches des isoformes *CYP98* dans plusieurs familles montrent des soutiens statistiques forts. Mais le soutien reste faible pour les clades formées par plusieurs ordres. Pour affiner l'analyse, une reconstruction avec des *CYP98s* des espèces dont le génome a été séquencé a été effectuée.

Bien que la phylogénie ait été construite sur les séquences nucléotides sur la base d'une bonne qualité d'annotation quasiment complète, le soutien statistique reste faible entre plusieurs ordres d'Angiospermes. Il est possible que les différentes horloges moléculaires ne soient pas au même pas chez les différentes espèces. Les grandes différences en longueur de branches pourraient soutenir cette hypothèse. Pour beaucoup d'espèces, une isoforme de *CYP98* se

trouve dans un clade avec une branche courte, et une ou plusieurs autres se trouvent dans un clade avec une longueur de branche plus longue. Ils ont donc eu plus de changement au niveau moléculaire depuis la séparation de leur ancêtre commun (Figure 5.8).

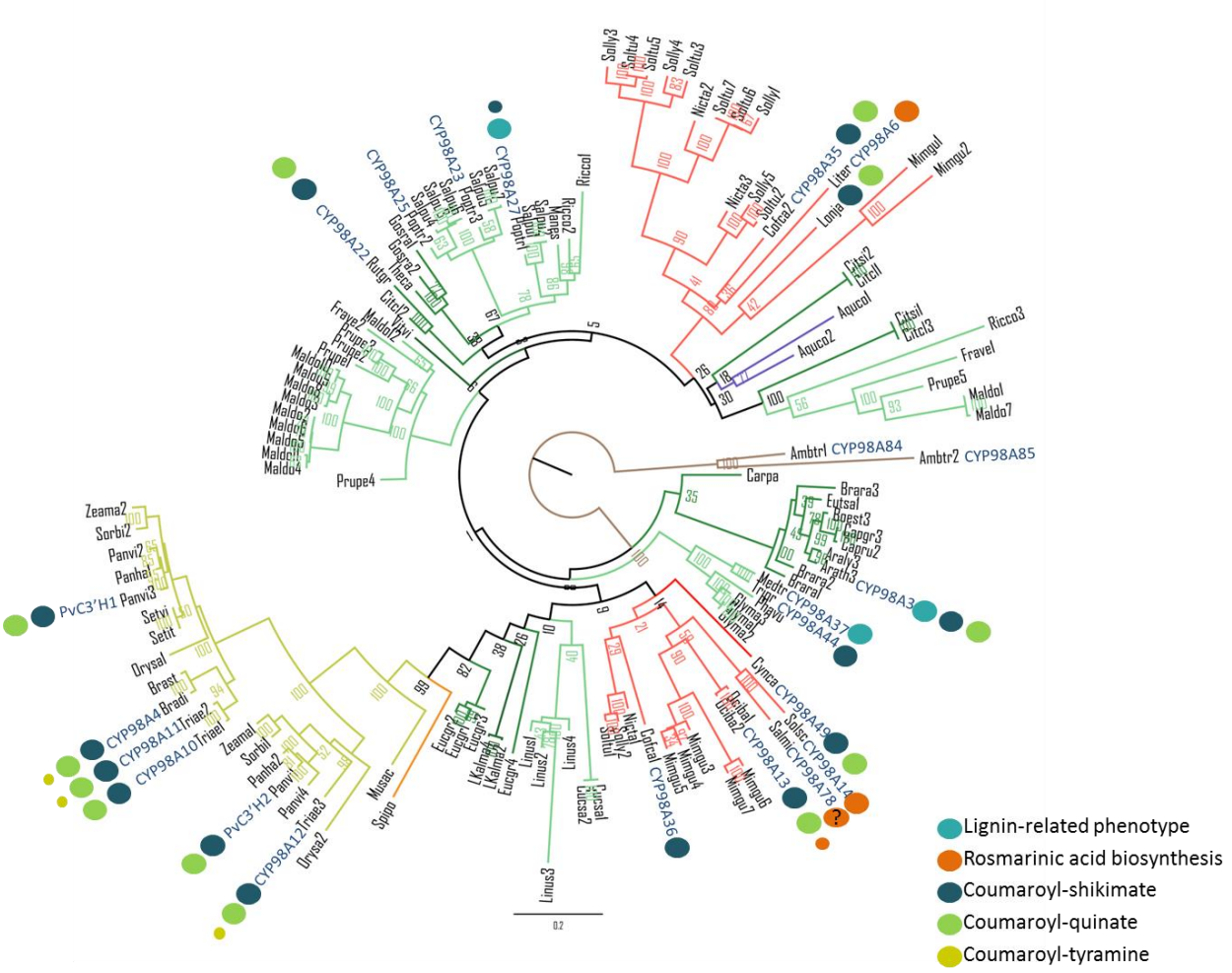


Figure 5.8 Reconstruction phylogénétique des gènes *CYP98* caractérisés et des gènes *CYP98* d'espèces avec des génomes séquencés.

Un alignement de séquences nucléotides de *CYP98* créé en utilisant DIALIGN, en gardant les positions d'alignement avec des similitudes diagonales au-dessus de zéro. La reconstruction phylogénétique maximum likelihood a été réalisée par PhyML, sur la base du modèle de HKY85. Le soutien statistique de la reconstruction phylogénétique a été obtenu par bootstrap avec 100 répétitions. Le codage couleur des branches suit la classification APG IV affichée de la Figure_III_ 2. Les noms des espèces sont donnés sur une liste d'espèces détaillées dans Table_III_ 5, en supplément. L'alignement est donné en annexe. Les numéros attribués aux *CYP98s* sont affichés en bleu foncé derrière les noms d'espèces. Les fonctions des CYP98s caractérisés trouvés dans la littérature sont indiquées par un point de couleur.

Une séparation majeure en deux classes liées à la fonction de l'enzyme n'est pas apparente. Une telle séparation aurait été attendue dans le cas de duplication et spécialisation anciennes. Au contraire, des duplications indépendantes sont observées dans chaque lignée, ce qui suggère une répétition des duplications, diversification et pertes de gènes dans l'histoire des Angiospermes. Pour les duplications indépendantes chez *P. trichocarpa* et *A. trichopoda,* on a obtenu un soutien statistique (par bootstrapping 100 reproductions) fort.

P. trichocarpa possède trois *CYP98s*. *A. trichopoda* possède de deux *CYP98s*. Les enzymes ont été exprimées dans des levures *S. cerevisiae* (Figure 5.9) et des microsomes ont été préparés.

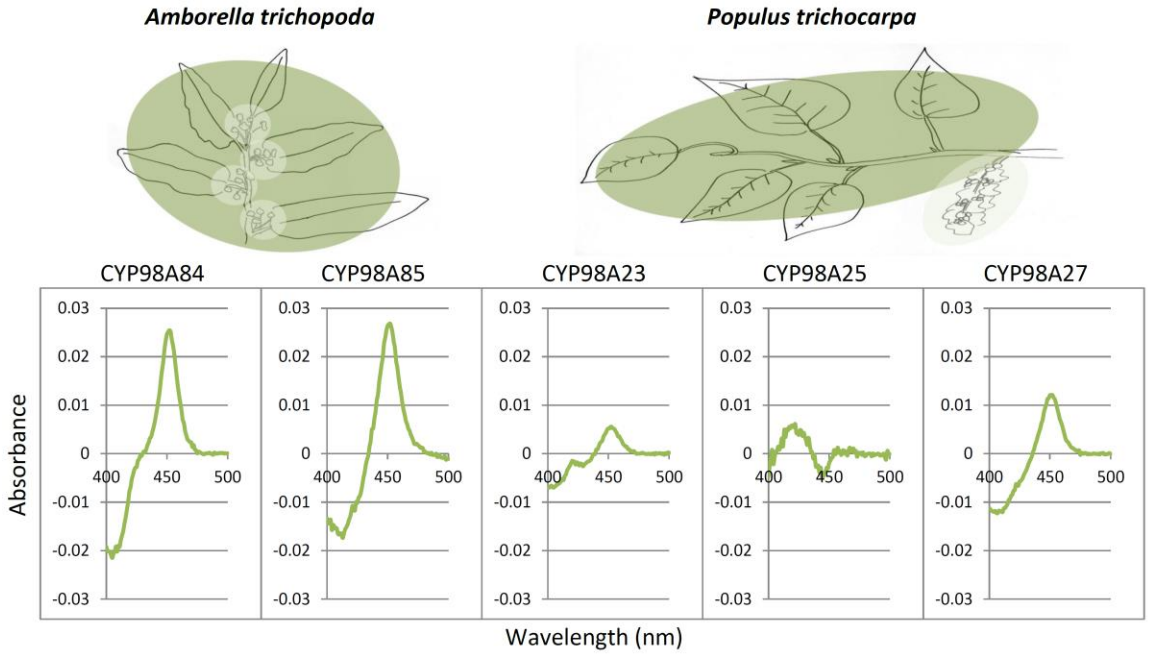


Figure 5.9 Spectres CO différentiels des CYP98s de *P. trichocarpa* et *A. trichopoda* réalisés sur des microsomes préparés à partir de levures.

Des tests enzymatiques ont été réalisés comme décrit dans le chapitre 2. CYP98A25 de *P. trichocarpa* n 'a montré aucune activité *in vitro* avec tous les substrats testés. Des substrats naturels et des substrats synthétiques ont été hydroxylés par les quatre autres CYP98s. Aucun des CYP98 testés n'a montré une activité avec l'acide *p*-coumarique libre dans nos expériences.

Dans l'ensemble, le CYP98A27 de *P. trichocarpa* et le CYP98A84 ont montré une préférence pour le *p*-coumaroyl-shikimate et le *p*-coumaroyl-quinate, alors que le CYP98A23 de *P. trichocarpa* et le CYP98A85 d'*A. trichopoda* ont converti une diversité de substrats (Figure 5.10).

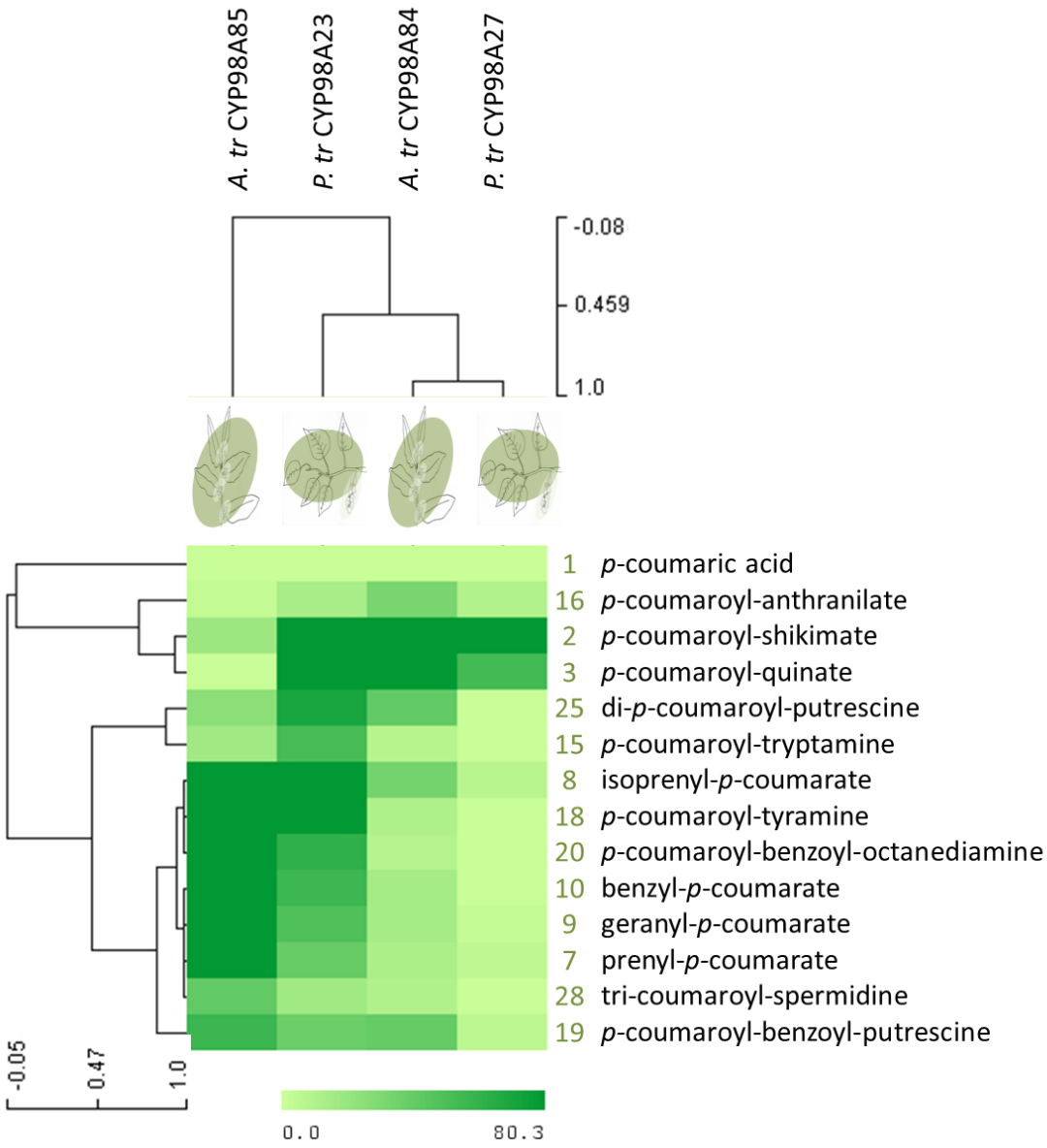


Figure 5.10 Classification hiérarchique des substrats et des P450 testés biochimiquement.

Groupement par liaison moyenne utilisant la corrélation de Pearson. Les taux correspondants de conversion sont présentés en détail dans Figure 3.8. Incubation de 10 pmole d'enzyme P450 avec 100 µM de substrat pendant 30 minutes à 28°C sous agitation.

Une corrélation des spectres des substrats du Gymnosperme testé en chapitre 2 montre une meilleure corrélation avec CYP98A84 d'*A. trichopoda* (Coefficient de Corrélation Pearson r=0.71). Si on considère les duplications indépendantes chez *A. trichopoda* et *P. trichocarpa* et le large spectre de substrat du CYP98 de Gymnosperme, on peut présumer que l'ancêtre des CYP98 des Angiospermes était soit spécifique pour le *p*-coumaroyl-shikimate, soit acceptait au moins le *p*-coumaroyl-shikimate comme un de ses substrats.

Les propriétés catalytiques des deux enzymes de peuplier ont ensuite été testées avec quatre substrats présents naturellement chez le peuplier. Les données obtenues montrent que CYP98A27 métabolise préférentiellement le *p*-coumaroyl-shikimate et que le spectre de substrat de CYP98A23 est plus large. Comme les deux enzymes *in vitro* montrent des fonctions redondantes, nous avons aussi étudié leur expression dans la plante *in silico* (Figure 5.11).

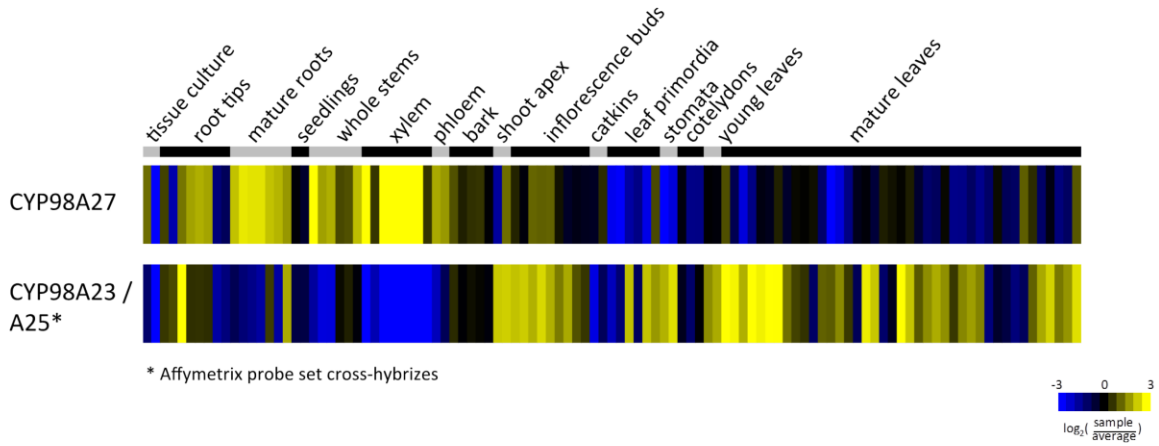


Figure 5.11 Expression des gènes *CYP98A23 / 25* (combinés) et de *CYP98A27* dans un ensemble de données de biopuces Affymetrix concernant organes et tissus.

Les valeurs d'expression de *CYP98A27* et *CYP98A23 / 25,* dans un ensemble de données Affymetrix, comprenant divers tissus et stades de développement de *P. trichocarpa,* sont montrées en médiane centrée. Pour une description du jeu des données et la normalisation voir Guo et al., 2014.

Alors que *CYP98A27* est exprimé dans des tissus fortement lignifiés, le *CYP98A23/25* (qui partagent une sonde commune dans les données Affymetrix) est plutôt exprimé dans les bourgeons d'inflorescences, les jeunes feuilles et les feuilles matures. Le patron d'expression de des deux gènes est donc bien en accord avec les activités détectées *in vitro*: CYP98A27 qui

montre une spécificité pour le *p*-coumaroyl-shikimate/*p*-coumaroyl-quinate est exprimé dans des tissues fortement lignifiés et *CYP98A23/25* dans des tissus riches en phénols solubles, potentiellement protecteurs (Greenaway and Whatley, 1990b; English et al., 1991).

Dans un jeu de données transcriptomiques du projet POPCAN, représentatif pour 371 individus de 197 accessions pour les jeunes feuilles et 390 individus de 194 accessions pour le xylème en développement, le résultat obtenu est cohérent avec les données Affymetrix. Alors que *CYP98A23* montre une forte expression dans les jeunes feuilles, *CYP98A25* ne montre presque aucune expression. Dans une analyse de co-expression, *CYP98A27* est co-exprimé avec de nombreux gènes de la voie biosynthèse de la lignine, impliqués dans la formation de la paroi cellulaire et des facteurs de transcription. *CYP98A23,* au contraire, n'est pas co-exprimé avec beaucoup de gènes (Coefficient de corrélation de Pearson r=0.75). En baissant le Coefficient de Corrélation de Pearson à r=0.65, on trouve plusieurs gènes co-exprimées, par exemple un homologue du facteur de transcription MYB4, impliqué dans la tolérance aux UV-B et l'accoutumance au froid, et un répresseur des gènes de la biosynthèse des monolignols chez *A. thaliana* et *Oryza sativa* (Jin et al., 2000; Vannini et al., 2004; Schenke et al., 2011). De plus, on trouve des gènes associés à la biosynthèse des flavonoïdes. Un de ces gènes, représente une isoforme de la 4CL associée à la biosynthèse des flavonoïdes et autres composés phénoliques solubles (Ehlting et al., 1999).

L'expression relative des trois gènes de *P. trichocarpa* a été suivie dans des feuilles de *P. trichocarpa*, suite à une attaque par des larves de *Lymantria dispar*. Une analyse statistique non paramétrique (Mann-Whitney U) montre une augmentation significative, mais très faible, de l'expression de *CYP98A23.* L'augmentation de l'expression relative de *CYP98A25* est plus forte, environ 10 fois en moyenne. Cependant, les niveaux absolus de transcriptions de *CYP98A25* restent cependant très faibles, à la fois dans le contrôle et dans les échantillons traités. Aucune différence significative n'a été observée pour les niveaux transcription de *CYP98A27*. Globalement, une induction de seulement deux fois de *CYP98A23* et *CYP98A27* ne suggère pas d'influence du traitement par *L. dispar* sur l'expression des gènes.

CYP98A23, *CYP98A25* et *CYP98A27* ayant montré des profils d'expression et co-expression distincts, ainsi que des propriétés biochimiques distinctes, nous avons ensuite cherché à

déterminer si la fonction de *CYP98A23* et *CYP98A27* était capable de complémenter le mutant knock-out *cyp98a3* d'*A. thaliana in vivo*. Nous avons aussi testé si *CYP98A25* pouvait complémenter le mutant *A. thaliana in vivo*, bien qu'aucune activité n'ait été détectée *in vitro*. Sous le contrôle du promoteur de la cinnamate-4-hydroxylase (C4H) (Bell-Lelong and Cusumano, 1997), les deux *CYP98s* du peuplier, *CYP98A23* et *CYP98A27*, ont permis de complémenter le phénotype de déficit de croissance du mutant *cyp98a3* (Figure 5.12). Ce résultat est en accord avec le profil d'utilisation de substrat des enzymes *in vitro*, les deux enzymes hydroxylant efficacement le *p*-coumaroyl-shikimate. Également en accord avec les données obtenues *in vitro* est l'absence de complémentation par *CYP98A25*. CYP98A25 qui ne semble guère exprimé en levure, dont la protéine ne montre aucune activité catalytique avec les substrats testés, qui est très peu exprimé dans la plante, et ne permet pas de complémenter le mutant *cyp98a3* d'*Arabidopsis* pourrait donc être en voie de pseudogènisation.

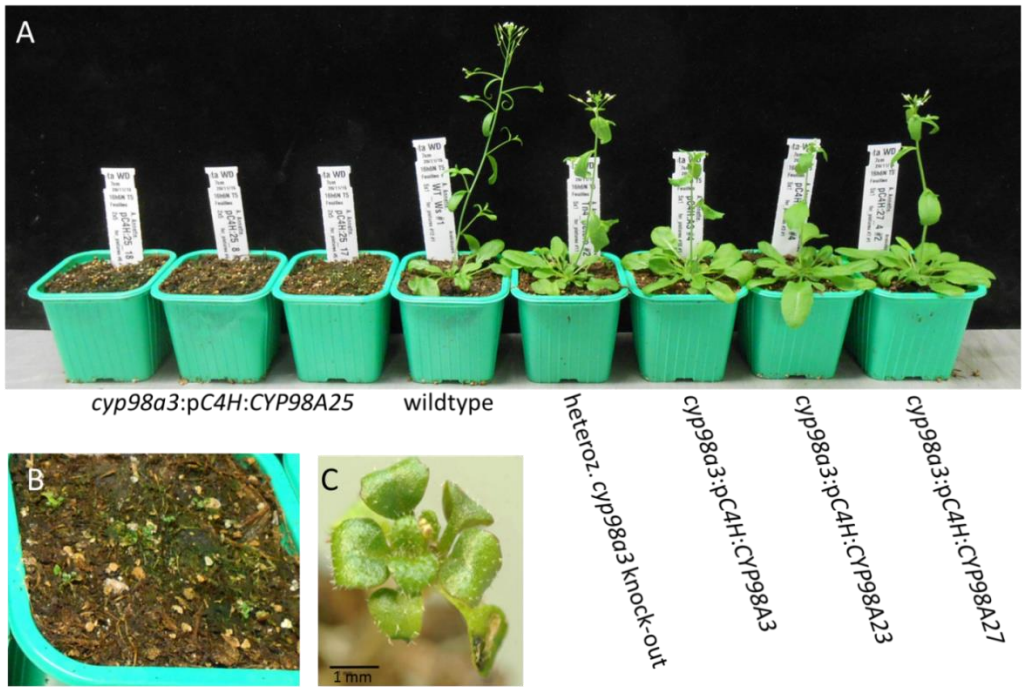


Figure 5.12 Complémentation du mutant knock-out *A. thaliana cyp98a3* avec les trois gènes *CYP98* de *P. trichocarpa.*

A) Arabidopsis sauvage et transformés. **B**) Gros plan sur les mutants d'*A. thaliana cyp98a3* portant la construction d'expression *CYP98A25* de *P. trichocarpa*. **C**) Mutant knock-out *A. thaliana cyp98a3* homozygote. Génotypage voir Figure 3.28.

Conclusion

Plusieurs membres de la famille *CYP98* n'existent que chez les Angiospermes. Tous les CYP98s caractérisés biochimiquement chez les Angiospermes (ici et dans la littérature) ont montré une préférence pour le *p*-coumaroyl-shikimate/*p*-coumaroyl-quinate lorsqu'ils étaient associés à la biosynthèse des monolignols. Notre hypothèse de départ, stipulant qu'une duplication ancienne précoce dans l'évolution des Angiospermes qui aurait généré un groupe lié à la biosynthèse de la lignine, et un autre groupe lié à la biosynthèse des phénols solubles, a été rejetée. Au lieu de cela, une série complexe de duplications caractérise la famille *CYP98* chez les Angiospermes suggérant le recrutement fréquent et indépendant des duplicats pour des fonctions spécifiques. Cela a été observé précédemment chez *Arabidopsis* (Matsuno, et al., 2009), et dans une certaine mesure pour le café (Mahesh et al., 2007). Cela semble aussi le cas pour le peuplier et Amborella. Il semble donc plus réaliste de supposer que les fonctions liées à la lignine communes à tous les Angiospermes ont été maintenues tout au long de l'évolution des Angiospermes et que des recrutements indépendants favorisant la formation de composés phénoliques solubles, parfois spécifiques d'une lignée, se sont produits à plusieurs reprises.

Conclusion Générale

Dans cette étude, nous avons analysé l'évolution fonctionnelle de la famille *CYP98* chez les plantes terrestres. Nous avons travaillé avec les familles *CYP98* de *Physcomitrella patens*, *Selaginella moellendorffii*, *Pteris vittata*, *Pinus taeda*, *Amborella trichopoda*, *Brachypodium distachyon*, *Populus trichocarpa* et *Arabidopsis thaliana*.

L'analyse phylogénétique a montré la présence de *CYP98* en copie unique depuis la mousse jusqu'aux Gymnospermes. La famille *CYP98* s'est élargie de paralogues chez de nombreux Angiospermes, chez lesquels de un à douze membres peuvent être observés. Bryophytes jusqu' aux Gymnospermes ne comptent qu'un seul membre. Un bon soutien statistique de l'arbre phylogénétique a été obtenu pour les clades formés par les *CYP98s* des Angiospermes n'ont généralement obtenu un bon soutien statistique qu'au niveau des espèces ou des familles. Au sein des Angiospermes, le soutien statistique des reconstructions phylogénétiques est resté faible. Des événements de duplication indépendants sont observés au sein des Angiospermes, ne résultant pas de la duplication complète d'un génome ancestral. Différentes plantes ont différents styles de vie et font face à diverses contraintes environnementales. On peut supposer que leurs horloges moléculaires peuvent avoir des rythmes très différents. Un exemple a été décrit par Tuskan et al., 2006, où l'horloge moléculaire de *A. thaliana* est environ six fois plus rapide que celle de *P. trichocarpa*.

Nous avons déterminé les substrats possibles des CYP98s de toutes les espèces étudiées, en testant leur activité avec une banque de substrats potentiels. Dans cette analyse biochimique, nous avons constaté que la spécificité de substrat s'est modifiée au cours de l'évolution des plantes terrestres. Un recrutement de CYP98s pour la biosynthèse des monolignols, tel que décrit dans la littérature et montrant une préférence pour le *p*-coumaroyl-shikimate, n'apparait qu'avec les Angiospermes. Les CYP98s de Bryophyte, de Lycopode et de fougère ne présentent pas cette caractéristique. La génération d'un mutant knock-out de *CYP98* de *P. patens* a par contre révélé que l'acide *p*-coumaroyl-thréonique était le substrat probable de l'enzyme *in vivo.* En accord avec cette observation *in vivo CYP98A34* de *P. patens* est incapable de complémenter le mutant *cyp98a3* d'*A. thaliana*. La perte de fonction des *CYP98* de mousse ou d'*Arabidopsis* entraine un déficit sévère de développement. Cela suggère un rôle crucial des *CYP98s* et des

phénols 3,4-dihydroxylés dans le développement aussi bien des Bryophytes que des Angiospermes. Il est probable que d'autres esters phénoliques distincts et non complémentaires seront produits par les Bryophytes et les Angiospermes pour le contrôle de leur développement. Une analyse plus approfondie des mutants *cyp98a34* de *P. patens* devra permettra déterminer le rôle des CYP98s dans les plantes qui ne produisent pas de lignine. Les fougères pourraient utiliser une voie distincte ou une autre enzyme pour produire des monolignols. La voie de biosynthèse des monolignols chez les fougères, un groupe à haute teneur en lignine, reste cependant très mal connue. La création d'un mutant knock-out dans les fougères permettrait de clarifier l'implication du *CYP98* dans leur voie de biosynthèse des monolignols. Le CYP98 de Gymnospermes a montré un spectre d'utilisation des substrat *in vitro* qui était intermédiaire entre celui des Bryophytes, Lycopodes et de la fougère d'une part et celui des Angiospermes d'autre part. Le rôle de ce *CYP98* dans la voie de biosynthèse des monolignols chez les gymnospermes reste à déterminer.

Notre première hypothèse était que l'ancêtre commun de tous les CYP98s chez les Angiospermes était spécifique pour le *p*-coumaroyl-shikimate et impliqué dans la biosynthèse des monolignols. La large gamme de substrats hydroxylés par le CYP98 de Gymnosperme inclut le *p*-coumaroyl-shikimate. Comme l'hydroxylation du *p*-coumaroyl-shikimate se retrouve pour tous les CYP98 d'Angiospermes testés *in vitro*, on peut supposer que l'ancêtre des CYP98*s* des Angiospermes préférait le *p*-coumaroyl-shikimate comme substrat. Comme les duplications des CYP98s trouvés chez les Angiospermes sont indépendantes, ces duplications ont offert à plusieurs reprises la possibilité d'une relaxation de la pression de sélection et donc de préférence de substrat pour arriver aux profils observés à ce jour.

6. Bibliography

Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. Bioinformatics **21**: 2104–2105

Abdulrazzak N, Pollet B, Ehlting J, Larsen K, Asnaghi C, Ronseau S, Proux C, Erhardt M, Seltzer V, Renou J-P, et al (2006) A coumaroyl-ester-3-hydroxylase insertion mutant reveals the existence of nonredundant meta-hydroxylation pathways and essential roles for phenolic precursors in cell expansion and plant growth. Plant Physiol **140**: 30–48

Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, et al (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol **52**: 399–451

Ahuja I, Kissen R, Bones AM (2012) Phytoalexins in defense against pathogens. Trends Plant Sci **17**: 73–90

Alber A, Ehlting J (2012) Chapter 4 - Cytochrome P450s in Lignin Biosynthesis. *In* J Lise, L Catherine, eds, Adv. Bot. Res. Academic Press, pp 113–143

Allen JM, Huang DI, Cronk QC, Johnson KP (2015) aTRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. BMC Bioinformatics **16**: 98

Allina SM (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

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol **215**: 403–10

Amborella Genome Project (2013) The Amborella genome and the evolution of flowering plants. Science **342**: 1241089

Anterola A (2002) Transcriptional Control of Monolignol Biosynthesis in Pinus taeda. FACTORS AFFECTING MONOLIGNOL RATIOS AND CARBON ALLOCATION IN PHENYLPROPANOID METABOLISM. J Biol Chem **277**: 18272–18280

Anterola A, Lewis NG (2002) Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. Phytochemistry **61**: 221–294

Baetz U, Martinoia E (2014) Root exudates: the hidden part of plant defense. Trends Plant Sci **19**: 90–98

Bak S, Beisson F, Bishop G, Hamberger B, Höfer R, Paquette S, Werck-Reichhart D (2011) Cytochromes P450. Arab B **9**: e0144

Barbehenn R, Dukatz C, Holt C (2010) Feeding on poplar leaves by caterpillars potentiates foliar peroxidase action in their guts and increases plant resistance. Oecologia **164**: 993–1004

Bartwal A, Mall R, Lohani P, Guru SK, Arora S (2013) Role of Secondary Metabolites and Brassinosteroids in Plant Defense Against Environmental Stresses. J Plant Growth Regul **32**: 216–232

Basile A, Giordano S, López-Sáez J, Cobianchi R (1999) Antibacterial activity of pure flavonoids isolated from mosses. Phytochemistry **52**: 1479–1482

Bassard J-E, Ullmann P, Bernier F, Werck-Reichhart D (2010) Phenolamides: Bridging polyamines to the phenolic metabolism. Phytochemistry **71**: 1808–1824

Basson AE, Dubery IA (2007) Identification of a cytochrome P450 cDNA (CYP98A5) from Phaseolus vulgaris, inducible by 3,5-dichlorosalicylic acid and 2,6-dichloro isonicotinic acid. J Plant Physiol **164**: 421–428

Bateman R, Crane P (1998) Early evolution of land plants: phylogeny, physiology, and ecology of the primary terrestrial radiation. Annu Rev Ecol Syst **29**: 263–292

Baucher M, Monties B (1998) Biosynthesis and genetic engineering of lignin. CRC Crit Rev Plant Sci **17**: 125–197

Bell-Lelong D, Cusumano J (1997) Cinnamate-4-hydroxylase expression in Arabidopsis (regulation in response to development and the environment). Plant Physiol **113**: 729–738

Besseau S, Hoffmann L, Geoffroy P (2007) Flavonoid accumulation in Arabidopsis repressed in lignin synthesis affects auxin transport and plant growth. Plant Cell **19**: 148–162

Bhardwaj R, Handa N, Sharma R (2014) Lignins and Abiotic Stress: An Overview, Physiologi. Springer New York

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

Bonawitz ND, Kim JI, Tobimatsu Y, Ciesielski PN, Anderson NA, Ximenes E, Maeda J, Ralph J, Donohoe BS, Ladisch M, et al (2014) Disruption of Mediator rescues the stunted growth of a lignin-deficient Arabidopsis mutant. Nature **509**: 376–380

Bowers JE, Chapman B, Rong J, Paterson AH (2003) Unraveling angiosperms genome evolution by phylogenitc analysis of chromosomal duplications events. Nature **422**: 433–438

Bowman JL, Floyd SK, Sakakibara K (2007) Green Genes—Comparative Genomics of the Green Branch of Life. Cell **129**: 229–234

Bremer K, Humphries CJ, Mishler BD, Churchill SP (1987) On Cladistic Relationships in Green Plants. Assoc Plant Taxon **36**: 339–349

Buchanan B, Gruissem W, Vickers K, Jones R (2015) Biochemistry & molecular biology of plants. John Wiley & Sons

Buchanan R, Beckett RD (2013) Green Coffee for Pharmacological Weight Loss. J Evid Based Complementary Altern Med **18**: 309–313

Buda GJ, Barnes WJ, Fich EA, Park S, Yeats TH, Zhao L, Domozych DS, Rose JKC (2013) An ATP Binding Cassette Transporter Is Required for Cuticular Wax Deposition and Desiccation Tolerance in the Moss Physcomitrella patens. Plant Cell **25**: 4000–4013

Cadahía E, Fernández de Simón B, Aranda I, Sanz M, Sánchez-Gómez D, Pinto E (2014) Non-targeted Metabolomic Profile of Fagus Sylvatica L. Leaves using Liquid Chromatography with Mass Spectrometry and Gas Chromatography with Mass Spectrometry. Phytochem Anal **26**: 171–182

Campbell M, Sederoff RR (1996) Variation in Lignin Content and Composition (Mechanisms of Control and Implications for the Genetic Improvement of Plants). Plant Physiol **110**: 3–13

Chabannes M, Ruel K (2001) In situ analysis of lignins in transgenic tobacco reveals a differential impact of individual transformations on the spatial patterns of lignin deposition at the cellular and subcellular level. Plant J **28**: 271–282

Chae L, Kim T, Nilo-Poyanco R, Rhee S (2014) Genomic signatures of specialized metabolism in plants. Science **344**: 510–513

Chao P, Hsu C, Yin M (2009) Anti-inflammatory and anti-coagulatory activities of caffeic acid and ellagic acid in cardiac tissue of diabetic mice. Nutr Metab (Lond) **6**: 1

Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, Sennikov AN, Soltis PS, Stevens PF, et al (2016) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc **181**: 1–20

Chaw S-M, Chang C-C, Chen H-L, Li W-H (2004) Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. J Mol Evol **58**: 424–41

Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. Nat Biotechnol **25**: 759–761

Chen F, Liu C-J, Tschaplinski TJ, Zhao N (2009) Genomics of Secondary Metabolism in Populus : Interactions with Biotic and Abiotic Environments. CRC Crit Rev Plant Sci **28**: 375–392

Chen H, Li Q, Shuford CM (2011) Membrane protein complexes catalyze both 4-and 3-hydroxylation of cinnamic acid derivatives in monolignol biosynthesis. Proc Natl Acad Sci **108**: 21253–21258

Chen H-C, Song J, Wang JP, Lin Y-C, Ducoste J, Shuford CM, Liu J, Li Q, Shi R, Nepomuceno A, et al (2014) Systems Biology of Lignin Biosynthesis in Populus trichocarpa: Heteromeric 4-Coumaric Acid:Coenzyme A Ligase Protein Complex Formation, Regulation, and Numerical Modeling. Plant Cell **26**: 876–93

Chen S, Krinsky BH, Long M (2013) New genes as drivers of phenotypic evolution. Nat Rev Genet **14**: 645–60

Christin P-A, Spriggs E, Osborne CP, Stromberg CAE, Salamin N, Edwards EJ (2013) Molecular Dating, Evolutionary Rates, and the Age of the Grasses. Syst Biol **63**: 153–165

Clifford M (2000) Chlorogenic acids and other cinnamates–nature, occurrence, dietary burden,

absorption and metabolism. J Sci Food Agric **80**: 1033–1043

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

Coleman H, Park J-Y, Nair R, Chapple C, Mansfield SD (2008a) RNAi-mediated suppression of p-coumaroyl-CoA 3'-hydroxylase in hybrid poplar impacts lignin deposition and soluble secondary metabolism. Proc Natl Acad Sci U S A **105**: 4501–4506

Coleman H, Samuels AL, Guy RD, Mansfield SD (2008b) Perturbed Lignification Impacts Tree Growth in Hybrid Poplar--A Function of Sink Strength, Vascular Integrity, and Photosynthetic Assimilation. Plant Physiol **148**: 1229–1237

Conant GC, Wolfe KH (2008) Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet **9**: 938–50

Constabel C, Barbehenn R (2008) Defensive roles of polyphenol oxidase in plants. Springer Netherlands

Coomey JH, Hazen SP (2015) Brachypodium distachyon as a Model Species to Understand Grass Cell Walls. 1–21

Corral-Lugo A, Daddaoua A, Ortega A, Espinosa-Urgel M, Krell T (2016) Rosmarinic acid is a homoserine lactone mimic produced by plants that activates a bacterial quorum-sensing regulator. Sci Signal **9**: ra1

Cronk QCB, Needham I, Rudall PJ (2015) Evolution of catkins: inflorescence morphology of selected Salicaceae in an evolutionary and developmental context. Front Plant Sci **6**: 1030

Danh LT, Truong P, Mammucari R, Foster N (2014) A critical review of the arsenic uptake mechanisms and phytoremediation potential of Pteris vittata. Int J Phytoremediation **16**: 429–53

Delaux P-M, Nanda AK, Mathé C, Sejalon-Delmas N, Dunand C (2012) Molecular and biochemical aspects of plant terrestrialization. Perspect Plant Ecol Evol Syst **14**: 49–59

Delwiche C, Cooper ED (2015) The Evolutionary Origin of a Terrestrial Flora. Curr Biol **25**: R899–R910

Delwiche C, Graham L, Thomson N (1989) Lignin-like compounds and sporopollenin in Coleochaete, an algal model for land plant ancestry. Science **245**: 399–401

Di P, Zhang L, Chen J, Tan H, Xiao Y (2013) 13C tracer reveals phenolic acids biosynthesis in hairy root cultures of Salvia miltiorrhiza. ACS Chem Biol **8**: 1537–1548

Dittmar K, Liberles D (2011) Evolution after gene duplication. John Wiley & Sons

Dixon RA (2001) Natural products and plant disease resistance. Nature **411**: 843–847

Dixon RA, Achnine L, Kota P, Liu C-J, Reddy MSS, Wang L (2002) The phenylpropanoid pathway and plant defence - A genomics perspective. Mol Plant Pathol **3**: 371–390

Dodsworth S, Chase MW, Leitch AR (2015) Is post-polyploidization diploidization the key to the

evolutionary success of angiosperms? Bot J Linn Soc **180**: 1–5

Douglas C (1996) Phenylpropanoid metabolism and lignin biosynthesis: from weeds to trees. Trends Plant Sci **1**: 171–178

Dudonné S, Poupard P (2011) Phenolic composition and antioxidant properties of poplar bud (Populus nigra) extract: individual antioxidant contribution of phenolics and transcriptional effect on skin. J Agric Food Chem **59**: 4527–4536

Eberle D, Ullmann P, Werck-Reichhart D, Petersen M (2009) cDNA cloning and functional characterisation of CYP98A14 and NADPH: cytochrome P450 reductase from Coleus blumei involved in rosmarinic acid biosynthesis. Plant Mol Biol **69**: 239–53

Edreva A, Velikova V, Tsonev T (2007) Phenylamides in plants. Russ J PLANT Physiol **54**: 287–301

Edwards K, Johnstone C, Thompson C (1991) A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Res **19**: 1349

Ehlting J, Büttner D, Wang Q (1999) Three 4-coumarate: coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms. Plant J **19**: 9–20

Ehlting J, Hamberger B, Million-Rousseau R, Werck-Reichhart D (2006) Cytochromes P450 in phenolic metabolism. Phytochem Rev **5**: 239–270

Ehlting J, Mattheus N, Aeschliman D (2005) Global transcript profiling of primary stems from Arabidopsis thaliana identifies candidate genes for missing links in lignin biosynthesis and transcriptional. Plant J **42**: 618–640

Ehlting J, Sauveplane V, Olry A (2008) An extensive (co-) expression analysis tool for the cytochrome P450 superfamily in Arabidopsis thaliana. BMC Plant Biol **8**: 1

El-Seedi HR, El-Said AM a, Khalifa S a M, Göransson U, Bohlin L, Borg-Karlson AK, Verpoorte R (2012) Biosynthesis, natural sources, dietary intake, pharmacokinetic properties, and biological activities of hydroxycinnamic acids. J Agric Food Chem **60**: 10877–10895

Ener ME, Lee Y-T, Winkler JR, Gray HB, Cheruzel L (2010) Photooxidation of cytochrome P450-BM3. Proc Natl Acad Sci U S A **107**: 18783–18786

English S, Greenaway W, Whatley F (1991) Analysis of phenolics of Populus trichocarpa bud exudate by GC-MS. Phytochemistry **30**: 531–533

English S, Greenaway W, Whatley F (1992) Analysis of phenolics in the bud exudates of Populus deltoides, P. fremontii, P. sargentii and P. wislizenii by GC-MS. Phytochemistry **31**: 1255–1260

Escamilla-Trevino L, Shen H, Hernandez T, Yin Y, Xu Y, Dixon RA (2014) Early lignin pathway enzymes and routes to chlorogenic acid in switchgrass (Panicum virgatum L.). Plant Mol Biol **84**: 565–576

Espiñeira JM, Novo Uzal E, Gómez Ros L V., Carrión JS, Merino F, Ros Barceló A, Pomar F (2011) Distribution of lignin monomers and the evolution of lignification among lower plants. Plant Biol

13: 59–68

Eudes A, Juminaga D, Baidoo EEK, Collins FW, Keasling JD, Loqué D (2013) Production of hydroxycinnamoyl anthranilates from glucose in Escherichia coli. Microb Cell Fact **12**: 62

Facchini P, Hagel J, Zulak K (2002) Hydroxycinnamic acid amide metabolism: physiology and biochemistry. Can J Bot **80**: 577–589

Field KJ, Pressel S, Duckett JG, Rimington WR, Bidartondo MI (2015) Symbiotic options for the conquest of land. Trends Ecol Evol **30**: 1–10

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151**: 1531–45

Fourquin C, Vinauger-Douard M, Fogliani B, Dumas C, Scutt CP (2005) Evidence that CRABS CLAW and TOUSLED have conserved their roles in carpel development since the ancestor of the extant angiosperms. Proc Natl Acad Sci U S A **102**: 4649–4654

Franke R, Hemm M (2002) Changes in secondary metabolism and deposition of an unusual lignin in the ref8 mutant of Arabidopsis. Plant J **30**: 47–59

Franke R, Humphreys JM, Hemm M, Denault JW, Ruegger MO, Cusumano J, Chapple C (2002) The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. Plant J **30**: 33–45

Friedman WE (2009) The meaning of Darwin's "abominable mystery." Am J Bot **96**: 5–21

Gaind K, Gupta R (1973) PHENOLIC COMPONENTS FROM THE LEAVES O F KALANCHOE PINNATA. Planta Med **23**: 149–153

Gallego-Giraldo L, Escamilla-Trevino L, Jackson L, Dixon RA (2011) Salicylic acid mediates the reduced growth of lignin down-regulated plants. Proc Natl Acad Sci **108**: 20814–20819

Gang DR, Beuerle T, Ullmann P, Werck-Reichhart D, Pichersky E (2002) Differential production of meta hydroxylated phenylpropanoids in sweet basil peltate glandular trichomes and leaves is controlled by the activities of specific acyltransferases and hydroxylases. Plant Physiol **130**: 1536–44

Gaquerel E, Gulati J, Baldwin I (2014) Revealing insect herbivory-induced phenolamide metabolism: from single genes to metabolic network plasticity analysis. Plant J **79**: 679–92

Gavira C, Höfer R, Lesot A, Lambert F, Zucca J, Werck-Reichhart D (2013) Challenges and pitfalls of P450-dependent (+)-valencene bioconversion by Saccharomyces cerevisiae. Metab Eng **18**: 25–35

Geng C-A, Chen H, Chen X-L, Zhang X-M, Lei L-G, Chen J-J (2014) Rapid characterization of chemical constituents in Saniculiphyllum guangxiense by ultra fast liquid chromatography with diode array detection and electrospray ionization tandem mass spectrometry. Int J Mass Spectrom **361**: 9–22

Gietz D, Jean AS (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res **20**: 1425

Goodstein D, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res **40**: D1178-86

Goremykin V V., Nikiforova S V., Biggs PJ, Zhong B, Delange P, Martin W, Woetzel S, Atherton RA, McLenachan PA, Lockhart P (2013) The evolutionary root of flowering plants. Syst Biol **62**: 50–61

Goremykin V V., Nikiforova S V., Cavalieri D, Pindo M, Lockhart P (2015) The Root of Flowering Plants and Total Evidence. Syst Biol **64**: 879–891

Greenaway W, English S (1992) Analysis of phenolics of bud exudates of Populus cathayana and Populus szechuanica by GC-MS. Zeitschrift für Naturforsch **47**: 308–312

Greenaway W, English S, May J, Whatley F (1991a) Chemotaxonomy of section Leuce poplars by GC-MS of bud exudate. Biochem Syst Ecol **19**: 507–518

Greenaway W, Gümüsdere I, Whatley F (1991b) Analysis of phenolics of bud exudate of Populus euphratica by GC-MS. Phytochemistry **30**: 1883–1885

Greenaway W, Scaysbrook T, Whatley F (1988) Phenolic analysis of bud exudate of Populus lasiocarpa by GC/MS. Phytochemistry **27**: 3513–3515

Greenaway W, Whatley F (1990a) Analysis of phenolics of bud exudate of Populus angustifolia by GC-MS. Phytochemistry **29**: 2551–2554

Greenaway W, Whatley F (1990b) Resolution of complex mixtures of phenolics in poplar bud exudate by analysis of gas chromatography—mass spectrometry data. J Chromatogr A **519**: 145–158

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

Gülçin İ (2006) Antioxidant activity of caffeic acid (3, 4-dihydroxycinnamic acid). Toxicology **217**: 213–220

Gunnison D, Alexander M (1975) Basis for the resistance of several algae to microbial decomposition. Appl Microbiol **29**: 729–738

Guo J, Carrington Y, Alber A, Ehlting J (2014) Molecular Characterization of Quinate and Shikimate Metabolism in Populus trichocarpa. J Biol Chem **289**: 23846–23858

Ha CM, Escamilla-Trevino L, Serrani Yarce JC, Kim H, Ralph J, Chen F, Dixon RA (2016) An essential role of caffeoyl shikimate esterase in monolignol biosynthesis in Medicago truncatula. Plant J **86**: 363–375

Hahn R, Nahrstedt A (1993) Hydroxycinnamic Acid Derivatives, Caffeoylmalic and New Caffeoylaldonic Acid Esters, from Chelidonium majus*,1. Planta Med **59**: 71–75

Hall TA (1999) BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 95–98

Hamada K, Nishida T, Yamauchi K (2004) 4-Coumarate: coenzyme A ligase in black locust (Robinia pseudoacacia) catalyses the conversion of sinapate to sinapoyl-CoA. J Plant Res **117**: 303–310

Hamberger B, 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

Hartmann T (2007) From waste products to ecochemicals: Fifty years research of plant secondary metabolism. Phytochemistry **68**: 2831–2846

Hasemann C a, Kurumbail RG, Boddupalli SS, Peterson J a, Deisenhofer J (1995) Structure and function of cytochromes P450:a comparative analysis of three crystal structures. Structure **3**: 41–62

Hayashi K, Kawaide H, Notomi M, Sakigi Y, Matsuo A, Nozaki H (2006) Identification and functional analysis of bifunctional *ent* -kaurene synthase from the moss *Physcomitrella patens*. FEBS Lett **580**: 6175–6181

Heller W, Kühnl T (1985) Elicitor induction of a microsomal 5-O-(4-coumaroyl) shikimate 3'-hydroxylase in parsley cell suspension cultures. Arch Biochem Biophys **241**: 453–460

Herrmann VK (1978) Hydroxyzimtsauren und Hydroxybenzoesauren enthaltende Naturstoffe in Pflanzen. Prog. Chem. Org. Nat. Prod. Springer Wien, pp 73–132

Hirano K, Aya K, Kondo M, Okuno A, Morinaka Y, Matsuoka M (2012) OsCAD2 is the major CAD gene responsible for monolignol biosynthesis in rice culm. Plant Cell Rep **31**: 91–101

Hoffmann L, Maury S, Martz F (2003) Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. J Biol Chem **278**: 95–103

Hohe A, Egener T, Lucht JM, Holtorf H, Reinhard C, Schween G, Reski R (2004) An improved and highly standardised transformation procedure allows efficient production of single and multiple targeted gene-knockouts in a moss, Physcomitrella patens. Curr Genet **44**: 339–347

Horst NA, Katz A, Pereman I, Decker EL, Ohad N, Reski R, Friedman WE, Lotan T, Ikeuchi M, Bowman JL, et al (2016) A single homeobox gene triggers phase transition, embryogenesis and asexual reproduction. Nat Plants **2**: 15209

Hu W, Kawaoka A, Tsai C-J (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate: CoA ligase genes in aspen (Populus tremuloides). Proc Natl Acad Sci **95**: 5407–5412

Hughes A (1994) The evolution of functionally novel proteins after gene duplication. Proc R Soc London B Biol Sci **256**: 119–24

Humphreys JM, Chapple C (2002) Rewriting the lignin roadmap. Curr Opin Plant Biol **5**: 224–229

Hurles M (2004) Gene duplication: The genomic trade in spare parts. PLoS Biol **2**: 0900–0904

Hutzler P (1998) Tissue localization of phenolic compounds in plants by confocal laser scanning microscopy. J Exp Bot **49**: 953–965

Imai Y, Sato R (1967) Conversion of P-450 to P-420 by neutral salts and some other reagents. Eur J Biochem 419–426

Inyushkina Y V., Kiselev K V., Bulgakov VP, Zhuravlev YN (2009) Specific genes of cytochrome P450 monooxygenases are implicated in biosynthesis of caffeic acid metabolites in rolC-transgenic culture of Eritrichium sericeum. Biochem **74**: 917–924

Isaji M, Miyata H, Ajisawa Y (1998) Tranilast: a new application in the cardiovascular field as an antiproliferative drug. Cardiovasc Drug Rev **16**: 288–299

Ishihara A, Ohtsu Y, Iwamura H (1999a) Biosynthesis of oat avenanthramide phytoalexins. Phytochemistry **50**: 237–242

Ishihara A, Ohtsu Y, Iwamura H (1999b) Induction of biosynthetic enzymes for avenanthramides in elicitor-treated oat leaves. Planta **208**: 512–518

Isidorov V a., Vinogorova VVT (2003) GC-MS analysis of compounds extracted from buds of Populus balsamifera and Populus nigra. Zeitschrift fur Naturforsch - Sect C J Biosci **58**: 355–360

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

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al (2011) Ancestral polyploidy in seed plants and angiosperms. Nature **473**: 97–100

Jin H, Cominelli E, Bailey P, Parr A, Mehrtens F, Jones J, Tonelli C, Weisshaar B, Martin C (2000) Transcriptional repression by AtMYB4 controls production of UV-protecting sunscreens in Arabidopsis. EMBO J **19**: 6150–6161

Jones L, Ennos A, Turner S (2001) Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of Arabidopsis. Plant J **26**: 205–216

Karamat F, Olry A, Doerper S, Vialart G, Bourgaud F, Hehn A, Ullmann P, Werck-Reichhart D (2012) CYP98A22, a phenolic ester 3'-hydroxylase specialized in the synthesis of chlorogenic acid, as a new tool for enhancing the furanocoumarin concentration in Ruta graveolens. BMC Plant Biol **12**: 152

Kenrick P, Wellman CH, Schneider H, Edgecombe GD (2012) A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. Philos Trans R Soc B Biol Sci **367**: 519–536

Kessler A, Baldwin IT (2004) Herbivore-induced plant vaccination. Part I. The orchestration of plant defenses in nature and their fitness consequences in the wild tobacco Nicotiana attenuata. Plant J **38**: 639–649

Kim JI, Ciesielski PN, Donohoe BS, Chapple C, Li X (2014) Chemically induced conditional rescue of the reduced epidermal fluorescence8 mutant of Arabidopsis reveals rapid restoration of growth and selective turnover of secondary metabolite pools. Plant Physiol **164**: 584–95

Kim M-S, Shin W-C, Kang D-K, Sohn H-Y (2015) Anti-thrombosis Activity of Sinapic Acid Isolated from the Lees of Bokbunja Wine. J Microbiol Biotechnol **26**: 61–65

Kolosova N, Miller B, Ralph S, Ellis BE, Douglas C, Ritland K, Bohlmann J (2004) Isolation of high-quality RNA from gymnosperm and angiosperm trees. Biotechniques **36**: 821–824

Korkina LG (2007) Phenylpropanoids as naturally occurring antioxidants: From plant defense to human health. Cell Mol Biol **53**: 15–25

Kranz H, Mikš D, Siegler M, Capesius I, Sensen C, Huss V (1995) The origin of land plants: Phylogenetic relationships among charophytes, bryophytes, and vascular plants inferred from complete small-subunit ribosomal RNA gene sequences. J Mol Evol **41**: 74–84

Kuczkowiak U, Petereit F, Nahrstedt A (2014) Hydroxycinnamic Acid Derivatives Obtained from a Commercial Crataegus Extract and from Authentic Crataegus spp. Sci Pharm **82**: 835–846

Kühnl T, Koch U, Heller W, Wellmann E (1987) Chlorogenic acid biosynthesis: characterization of a light-induced microsomal 5-O-(4-coumaroyl)-D-quinate/shikimate 3′-hydroxylase from carrot (Daucus carota L.). Arch Biochem Biophys **258**: 226–232

Labeeuw L, Martone PT, Boucher Y, Case RJ (2015) Ancient origin of the biosynthesis of lignin precursors. Biol Direct **10**: 23

Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. Mol Biol Evol **25**: 1307–1320

Lee D, Lee SH, Chung SR, Ro J, Lee K (1995) Phenolic components from the leaves of Cornus controversa H. Korean J Pharmacogn **26**: 327–336

Lee J, Chan BLS, Mitchell AE (2017) Identification/quantification of free and bound phenolic acids in peel and pulp of apples (Malus domestica) using high resolution mass spectrometry (HRMS). Food Chem **215**: 301–310

Li X, Bonawitz ND, Weng J-K, Chapple C (2010) The growth reduction associated with repressed lignin biosynthesis in Arabidopsis thaliana is independent of flavonoids. Plant Cell **22**: 1620–32

Li X, Yang X, Wang N, Xie Y (2015a) Potential of Pteris vittata to Remove Tetracycline Antibiotics from Aquatic Media. Int J Phytoremediation **17**: 895–9

Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, Barker MS (2015b) Early genome duplications in conifers and other seed plants. Sci Adv **1**: e1501084–e1501084

Li Z, Defoort J, Tasdighian S, Maere S, Van de Peer Y, De Smet R (2016) Gene duplicability of core genes is highly consistent across all angiosperms. Plant Cell **32**: TPC2015-00877-LSB

Lindermayr C, Möllers B (2002) Divergent members of a soybean (Glycine max L.) 4-coumarate: coenzyme A ligase gene family. Eur J Biochem **269**: 1304–1315

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative

PCR and the 2^(-ΔΔCT) method. Methods **25**: 402–408

Lowry B, Lee D, Hébant C (1980) The origin of land plants: a new look at an old problem. Taxon 183–197

Macoy DM, Kim W-Y, Lee SY, Kim MG (2015a) Biosynthesis, physiology, and functions of hydroxycinnamic acid amides in plants. Plant Biotechnol Rep **9**: 269–278

Macoy DM, Kim W-Y, Lee SY, Kim MG (2015b) Biotic stress related functions of hydroxycinnamic acid amide in plants. Plant Biotechnol Rep **58**: 156–163

Mahesh V, Million-Rousseau R, Ullmann P, Chabrillange N, Bustamante J, Mondolot L, Morant M, Noirot M, Hamon S, de Kochko A, et al (2007) Functional characterization of two p-coumaroyl ester 3'-hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis. Plant Mol Biol **64**: 145–59

Mansuy D (1998) The great diversity of reactions catalyzed by cytochromes P450. Comp Biochem Physiol Part C Pharmacol Toxicol Endocrinol **121**: 5–14

Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M, Toub O, Lund ST, Bohlmann J, Jaillon O, Aury J, et al (2010) Functional Annotation, Genome Organization and Phylogeny of the Grapevine (Vitis vinifera) Terpene Synthase Gene Family Based on Genome Assembly, FLcDNA Cloning, and Enzyme Assays. BMC Plant Biol **10**: 226

Martin-Tanguy J (1985) The occurrence and possible function of hydroxycinnamoyl acid amides in plants. Plant Growth Regul **3**: 381–399

Martone PT, Estevez JM, Lu F, Ruel K, Denny MW, Somerville C, Ralph J (2009) Discovery of Lignin in Seaweed Reveals Convergent Evolution of Cell-Wall Architecture. Curr Biol **19**: 169–175

Matsuno, M, Compagnon V, Schoch GA, Schmitt M, Debayle D, Bassard J-E, Pollet B, Hehn A, Heintz D, Ullmann P, et al (2009) Evolution of a Novel Phenolic Pathway for Pollen Development. Science **325**: 1688–1692

Matsuno M, Nagatsu A, Ogihara Y, Ellis BE, Mizukami H (2002) CYP98A6 from Lithospermum erythrorhizon encodes 4-coumaroyl-4'-hydroxyphenyllactic acid 3-hydroxylase involved in rosmarinic acid biosynthesis. FEBS Lett **514**: 219–224

Mayama S, Matsuura Y, Iida H, Tani T (1982) The role of avenalumin in the resistance of oat to crown rust, Puccinia coronata f. sp. avenae. Physiol Plant Pathol **20**: 189–199

Mayama S, Tani T, Matsuura Y (1981) The production of phytoalexins by oat in response to crown rust, Puccinia coronata f. sp. avenae. Physiol Plant Pathol **19**: 217–1N7

Mishler BD, Lewis LA, Buchheim MA, Renzaglia KS, Garbary DJ, Delwiche C, Zechman FW, Kantz TS, Chapman RL (1994) Phylogenetic Relationships of the "Green Algae" and "Bryophytes." Ann Missouri Bot Gard **81**: 451–483

Mizutani M, Ohta D (2010) Diversification of P450 Genes During Land Plant Evolution. Annu Rev Plant Biol **61**: 291–315

Moglia A, Comino C, Portis E, Acquadro A, De Vos RCH, Beekwilder J, Lanteri S (2009) Isolation and mapping of a C3'H gene (CYP98A49) from globe artichoke, and its expression upon UV-C stress. Plant Cell Rep **28**: 963–74

Mølgaard P, Ravn H (1988) Evolutionary Aspects of Caffeoyl Ester Distribution in Dicotyledons. Phytochemistry **27**: 2411–2421

Morant M, Schoch GA, Ullmann P, Ertunç T, Little D, Olsen CE, Petersen M, Negrel J, Werck-Reichhart D (2007) Catalytic activity, duplication and evolution of the CYP98 cytochrome P450 family in wheat. Plant Mol Biol **63**: 1–19

Morgenstern B (1999) DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics **15**: 211–218

Nagegowda DA, Gutensohn M, Wilkerson CG, Dudareva N (2008) Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. Plant J **55**: 224–239

Nair R, Xia Q, Kartha C, Kurylo E (2002) Arabidopsis CYP98A3 mediating aromatic 3-hydroxylation. Developmental regulation of the gene, and expression in yeast. Plant Physiol **130**: 210–220

Näsvall J, Sun L, Roth JR, Andersson DI (2012) Real-time evolution of new genes by innovation, amplification, and divergence. Science **338**: 384–7

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press

Nelson DR (1999) Cytochrome P450 and the individuality of species. Arch Biochem Biophys **369**: 1–10

Nelson DR (2006) Plant cytochrome P450s from moss to poplar. Phytochem Rev **5**: 193–204

Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW, et al (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. Pharmacogenet Genomics **6**: 1–42

Nour-Eldin H, Geu-Flores F, Halkier B (2010) USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories. Methods Mol Biol **643**: 185–200

Nour-Eldin H, Hansen B (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. Nucleic Acids Res **34**: e122

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, et al (2013) The Norway spruce genome sequence and conifer genome evolution. Nature **497**: 579–84

Okazaki Y, Isobe T, Iwata Y, Matsukawa T, Matsuda F, Miyagawa H, Ishihara A, Nishioka T, Iwamura H (2004) Metabolism of avenanthramide phytoalexins in oats. Plant J **39**: 560–572

Omura T, Sato R (1964) The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. J biol Chem **239**: 2370–2378

Palmer JD, Soltis DE, Chase MW (2004) the Plant Tree of Life: an Overview Oand Some Points of View. DNA Seq **91**: 1437–1445

Parveen I, Threadgill MD, Hauck B, Donnison I, Winters AL (2011) Isolation, identification and quantitation of hydroxycinnamic acid conjugates, potential platform chemicals, in the leaves and stems of Miscanthus × giganteus using LC-ESI-MSn. Phytochemistry **72**: 2376–84

Parveen I, Wilson T, Donnison I, Cookson AR, Hauck B, Threadgill MD (2013) Potential sources of high value chemicals from leaves, stems and flowers of Miscanthus sinensis "Goliath" and Miscanthus sacchariflorus. Phytochemistry **92**: 160–167

Parveen I, Winters AL, Threadgill MD, Hauck B, Morris P (2008) Extraction, structural characterisation and evaluation of hydroxycinnamate esters of orchard grass (Dactylis glomerata) as substrates for polyphenol oxidase. Phytochemistry **69**: 2799–806

Patten AM, Jourdes M, Cardenas C (2010) Probing native lignin macromolecular configuration in Arabidopsis thaliana in specific cell wall types: further insights into limited substrate degeneracy and assembly. Mol Biosyst **6**: 499–515

Pattewar S (2012) Kalanchoe pinnata: Phytochemical and pharmacological profile. Int J Phytopharm **2**: 1–8

Petersen M, Abdullah Y, Benner J, Eberle D (2009) Evolution of rosmarinic acid biosynthesis. Phytochemistry **70**: 1663–1679

Petersen M, Simmonds MMSJ (2003) Rosmarinic acid. Phytochemistry **62**: 121–125

Pichersky E, Lewinsohn E (2011) Convergent evolution in plant specialized metabolism. Annu Rev Plant Biol **62**: 549–566

Porth I, Hamberger B, White R, Ritland K (2011) Defense mechanisms against herbivory in Picea: sequence evolution and expression regulation of gene family members in the phenylpropanoid pathway. BMC Genomics **12**: 608

Powles SB, Yu Q (2010) Evolution in action: plants resistant to herbicides. Annu Rev Plant Biol **61**: 317–47

Proctor M (2014) The Diversification of Bryophytes and Vascular Plants in Evolving Terrestrial Environments. Photosynth. Bryophyt. Early L. Plants. Springer New York, pp 59–77

Pu G, Wang P, Zhou B, Liu Z, Xiang F (2013) Cloning and Characterization of Lonicera japonica p -Coumaroyl Ester 3-Hydroxylase Which Is Involved in the Biosynthesis of Chlorogenic Acid. Biosci Biotechnol Biochem **77**: 1403–1409

Pu Y, Chen F, Ziebell A, Davison B, Ragauskas AJ (2009) NMR Characterization of C3H and HCT Down-Regulated Alfalfa Lignin. BioEnergy Res **2**: 198–208

Radwan MA (1975) Genotype and Season Influence Chlorogenic Acid Content in Douglas-fir Foliage. **5**: 6–9

Ralph J, Akiyama T, Kim H, Lu F (2006) Effects of coumarate 3-hydroxylase down-regulation on lignin structure. J Biol Chem **281**: 8843–8853

Ramakers C, Ruijter JM, Deprez RHL, Moorman AF. (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett **339**: 62–66

Rana S, Bhat WW, Dhar N, Pandith SA, Razdan S, Vishwakarma R, Lattoo SK (2014) Molecular characterization of two A-type P450s, WsCYP98A and WsCYP76A from Withania somnifera (L.) Dunal: expression analysis and withanolide accumulation in response to exogenous elicitations. BMC Biotechnol **14**: 89

Rautengarten C, Ebert B, Ouellet M, Nafisi M, Baidoo EEK, Benke P, Stranne M, Mukhopadhyay A, Keasling JD, Sakuragi Y, et al (2012) Arabidopsis Deficient in Cutin Ferulate Encodes a Transferase Required for Feruloylation of -Hydroxy Fatty Acids in Cutin Polyester. Plant Physiol **158**: 654–665

Raven JA (1984) Physiological correlates of the morphology of early vascular plants. Bot J Linn Soc **88**: 105–126

Reddy MSS, Chen F (2005) Targeted down-regulation of cytochrome P450 enzymes for forage quality improvement in alfalfa (Medicago sativa L.). Proc Natl Acad Sci U S A **102**: 16573–16578

Rensing SA, Lang D, Zimmer ADA, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist EA, Kamisugi Y, et al (2008) The Physcomitrella Genome Reveals Evolutionary Insights into the Conquest of Land by Plants. Science **319**: 64–69

Reski R, Abel WO (1985) Induction of budding on chloronemata and caulonemata of the moss, Physcomitrella patens, using isopentenyladenine. Planta **165**: 354–8

Reyes-Prieto A, Weber APM, Bhattacharya D (2007) The origin and establishment of the plastid in algae and plants. Annu Rev Genet **41**: 147–168

Richter H, Lieberei R, Strnad M, Novák O, Gruz J, Rensing SA, von Schwartzenberg K (2012) Polyphenol oxidases in Physcomitrella: functional PPO1 knockout modulates cytokinin-dependent development in the moss Physcomitrella patens. J Exp Bot **63**: 5121–35

Rogosnitzky M, Danks R, Kardash E (2012) Therapeutic potential of tranilast, an anti-allergy drug, in proliferative disorders. Anticancer Res.

Rothfels CJ, Li F-W, Sigel EM, Huiet L, Larsson A, Burge DO, Ruhsam M, Deyholos M, Soltis DE, Stewart CN, et al (2015) The evolutionary history of ferns inferred from 25 low-copy nuclear genes. Am J Bot **102**: 1089–1107

Rubiolo P, Casetta C, Cagliero C, Brevard H, Sgorbini B, Bicchi C (2013) Populus nigra L. bud absolute: A case study for a strategy of analysis of natural complex substances. Anal Bioanal Chem **405**: 1223–1235

Sanderson MJ (2003) Molecular data from 27 proteins do not support a Precambrian origin of land plants. Am J Bot **90**: 954–956

Sarkanen K, Ludwig C (1971) Lignins: occurrence, formation, structure and reactions. Lignins Occur. Form. Struct. React.

Sarkar P, Bosneaga E, Auer M (2009) Plant cell walls throughout evolution: towards a molecular understanding of their design principles. J Exp Bot **60**: 3615–3635

Schenke D, Böttcher C, Scheel D (2011) Crosstalk between abiotic ultraviolet-B stress and biotic (flg22) stress signalling in Arabidopsis prevents flavonol accumulation in favor of pathogen defence compound production. Plant, Cell Environ **34**: 1849–1864

Schoch GA, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, Werck-Reichhart D (2001) CYP98A3 from Arabidopsis thaliana is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. J Biol Chem **276**: 36566–36574

Schoch GA, Morant M, Abdulrazzak N, Asnaghi C, Goepfert S, Petersen M, Ullmann P, Werck-Reichhart D (2006) The meta-hydroxylation step in the phenylpropanoid pathway: a new level of complexity in the pathway and its regulation. Environ Chem Lett **4**: 127–136

Schuler MA, Werck-Reichhart D (2003) Functional genomics of P450s. Annu Rev Plant Biol **54**: 629–67

Schween G, Fleig S, Reski R (2002) High-throughput-PCR screen of 15,000 transgenicPhyscomitrella plants. Plant Mol Biol Report **20**: 43–47

Shadle G, Chen F, Reddy MSS (2007) Down-regulation of hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and forage quality. Phytochemistry **68**: 1521–1529

Shi H, Dong L, Jiang J, Zhao J, Zhao G, Dang X (2013) Chlorogenic acid reduces liver inflammation and fibrosis through inhibition of toll-like receptor 4 signaling pathway. Toxicology **303**: 107–114

Smith AM, Coupland G, Dolan L, Harberd N, Jones J, Martin C, Sablowski R, Amey A (2006a) Plant biology. Garland Science

Smith AR, Pryer KM, Schuettpelz E, Korall P, Schneider H, Wolf PG (2006b) A classification for extant ferns. Taxon **55**: 705–731

Solecka D, Kacperska A (2003) Phenylpropanoid deficiency affects the course of plant acclimation to cold. Physiol Plant **119**: 253–262

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. Am J Bot **96**: 336–48

Soltis P, Soltis D (2016) Ancient WGD events as drivers of key innovations in angiosperms. Curr Opin Plant Biol **30**: 159–165

Soltis PS, Soltis DE (2014) Flower diversity and angiosperm diversification. Methods Mol Biol **1110**: 85–102

Sullivan ML, Zarnowski R (2010) Red clover coumarate 3'-hydroxylase (CYP98A44) is capable of

hydroxylating p-coumaroyl-shikimate but not p-coumaroyl-malate: implications for the biosynthesis of phaselic acid. Planta **231**: 319–28

Sykes RW, Gjersing EL, Foutz K, Rottmann WH, Kuhn S a., Foster CE, Ziebell A, Turner GB, Decker SR, Hinchee M a. W, et al (2015) Down-regulation of p-coumaroyl quinate/shikimate 3′-hydroxylase (C3′H) and cinnamate 4-hydroxylase (C4H) genes in the lignin biosynthetic pathway of Eucalyptus urophylla × E. grandis leads to improved sugar release. Biotechnol Biofuels **8**: 128

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. Science **320**: 486–8

Thompson J, Higgins D, Gibson T (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix. Nucleic Acids Res **22**: 4673–4680

Tran LT, Taylor JS, Constabel C (2012) The polyphenol oxidase gene family in land plants: Lineage-specific duplication and expansion. BMC Genomics **13**: 395

Tuskan G, DiFazio SP, Jansson S, Bohlmann J, Grigoriev I V., Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov a., et al (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science **313**: 1596–604

Tzin V, Galili G (2010) New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. Mol Plant **3**: 956–972

Umezawa T (2003) Diversity in lignan biosynthesis. Phytochem Rev **2**: 371–390

Urban P, Mignotte C, Kazmaier M (1997) Cloning, yeast expression, and characterization of the coupling of two distantly related Arabidopsis thalianaNADPH-cytochrome P450 reductases with P450 CYP73A5. J. Biol. …

Vandesompele J, Preter K De, Pattyn F, Poppe B, Roy N Van, Paepe A De, Speleman F Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol **3**: 1–12

Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, Araujo P, et al (2013) Caffeoyl Shikimate Esterase (CSE) Is an Enzyme in the Lignin Biosynthetic Pathway in Arabidopsis. Science **341**: 1103–1106

Vannini C, Locatelli F, Bracale M, Magnani E, Marsoni M, Osnato M, Mattana M, Baldoni E, Coraggio I (2004) Overexpression of the rice Osmyb4 gene increases chilling and freezing tolerance of Arabidopsis thaliana plants. Plant J **37**: 115–127

Vogt T (2010) Phenylpropanoid biosynthesis. Mol Plant **3**: 2–20

Wadenbäck J, von Arnold S, Egertsdotter U, Walter MH, Grima-Pettenati J, Goffner D, Gellerstedt G, Gullion T, Clapham D (2008) Lignin biosynthesis in transgenic Norway spruce plants harboring an antisense construct for cinnamoyl CoA reductase (CCR). Transgenic Res **17**: 379–92

Wagner A, Donaldson L, Kim H, Phillips L, Flint H, Steward D, Torr K, Koch G, Schmitt U, Ralph J (2009) Suppression of 4-coumarate-CoA ligase in the coniferous gymnosperm Pinus radiata. Plant Physiol **149**: 370–83

Wagner A, Tobimatsu Y, Phillips L, Flint H, Geddes B, Lu F, Ralph J (2015) Syringyl lignin production in conifers: Proof of concept in a Pine tracheary element system. Proc Natl Acad Sci U S A **112**: 6218–23

Wallace RJ (2004) Antimicrobial properties of plant secondary metabolites. Proc Nutr Soc **63**: 621–629

Wang B, Sun W, Li Q, Li Y, Luo H, Song J, Chen S (2015) Genome-wide identification of phenolic acid biosynthetic genes in Salvia miltiorrhiza. Planta **241**: 711–725

Weng JK (2014) The evolutionary paths towards complexity: A metabolic perspective. New Phytol **201**: 1141–1149

Weng J-K, Akiyama T, Bonawitz ND, Li X (2010a) Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte Selaginella and flowering plants. Plant Cell **22**: 1033–1045

Weng J-K, Banks JA, Chapple C (2008a) Parallels in lignin biosynthesis: A study in Selaginella moellendorffii reveals convergence across 400 million years of evolution. Commun Integr Biol **1**: 20–22

Weng J-K, Chapple C (2010) The origin and evolution of lignin biosynthesis. New Phytol **187**: 273–85

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

Weng J-K, Mo H, Chapple C (2010b) Over-expression of F5H in COMT-deficient Arabidopsis leads to enrichment of an unusual lignin and disruption of pollen wall formation. Plant J **64**: 898–911

Weng J-K, Philippe R, Noel JP (2012) The rise of chemodiversity in plants. Science **336**: 1667–1670

Werck-Reichhart D, Feyereisen R (2000) Cytochromes P450: a success story. Genome Biol **1**: 3003.1-3003.9

Werck-Reichhart D, Hehn A, Didierjean L (2000) Cytochromes P450 for engineering herbicide tolerance. Trends Plant Sci **5**: 116–123

Wink M (2003) Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry **64**: 3–19

Wolf PG, Sessa EB, Marchant DB, Li F-W, Rothfels CJ, Sigel EM, Gitzendanner MA, Visger CJ, Banks JA, Soltis DE, et al (2015) An Exploration into Fern Genome Space. Genome Biol Evol **7**: 2533–44

Xu J, Ding Z, Vizcay-Barrena G, Shi J, Liang W, Yuan Z, Werck-Reichhart D, Schreiber L, Wilson ZA, Zhang D (2014) ABORTED MICROSPORES Acts as a Master Regulator of Pollen Wall Formation in Arabidopsis. Plant Cell **26**: 1544–1556

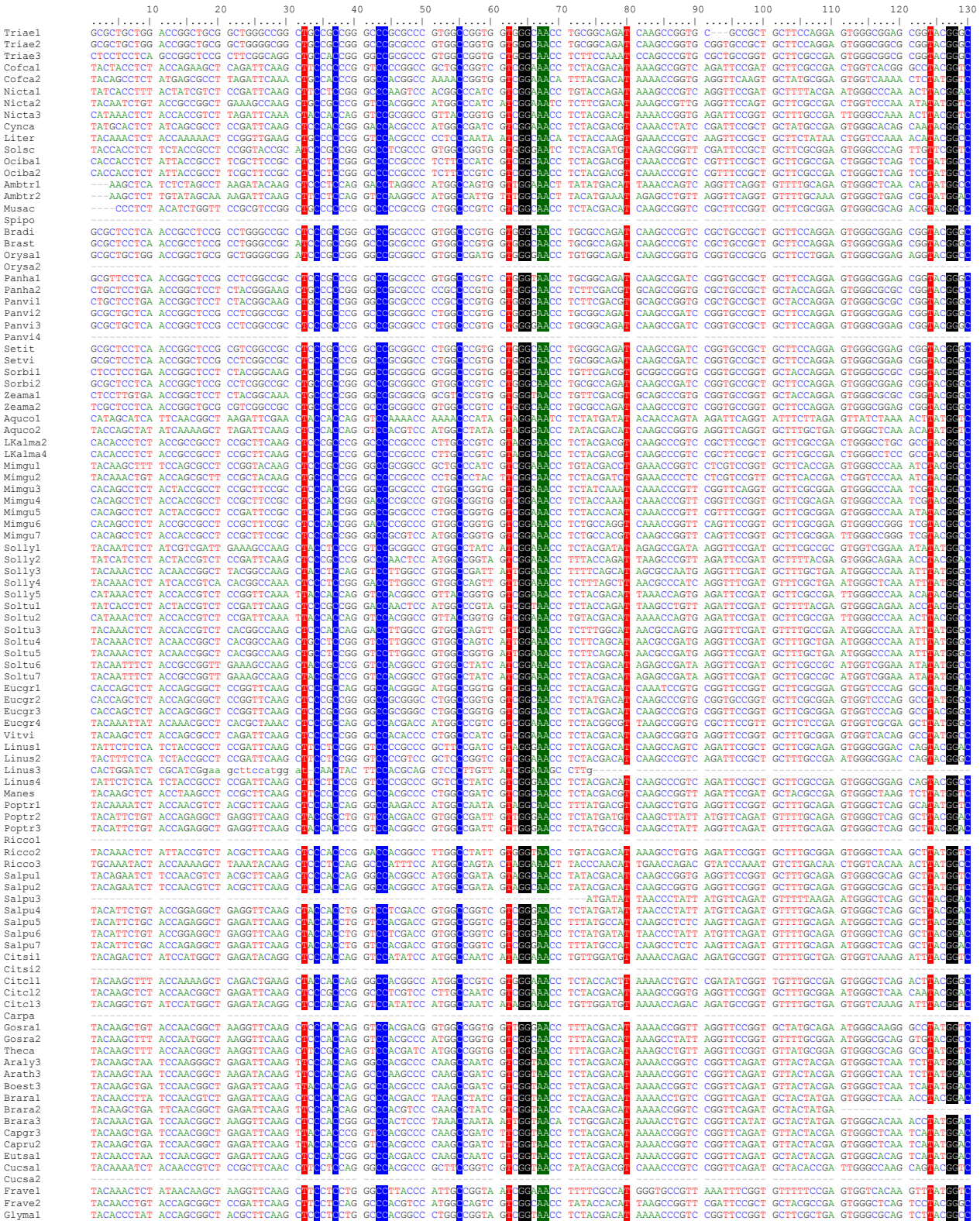
Xu Z, Zhang D, Hu J, Zhou X, Ye X, Reichel KL, Stewart NR, Syrenne RD, Yang X, Gao P, et al (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. BMC Bioinformatics **10**: 1

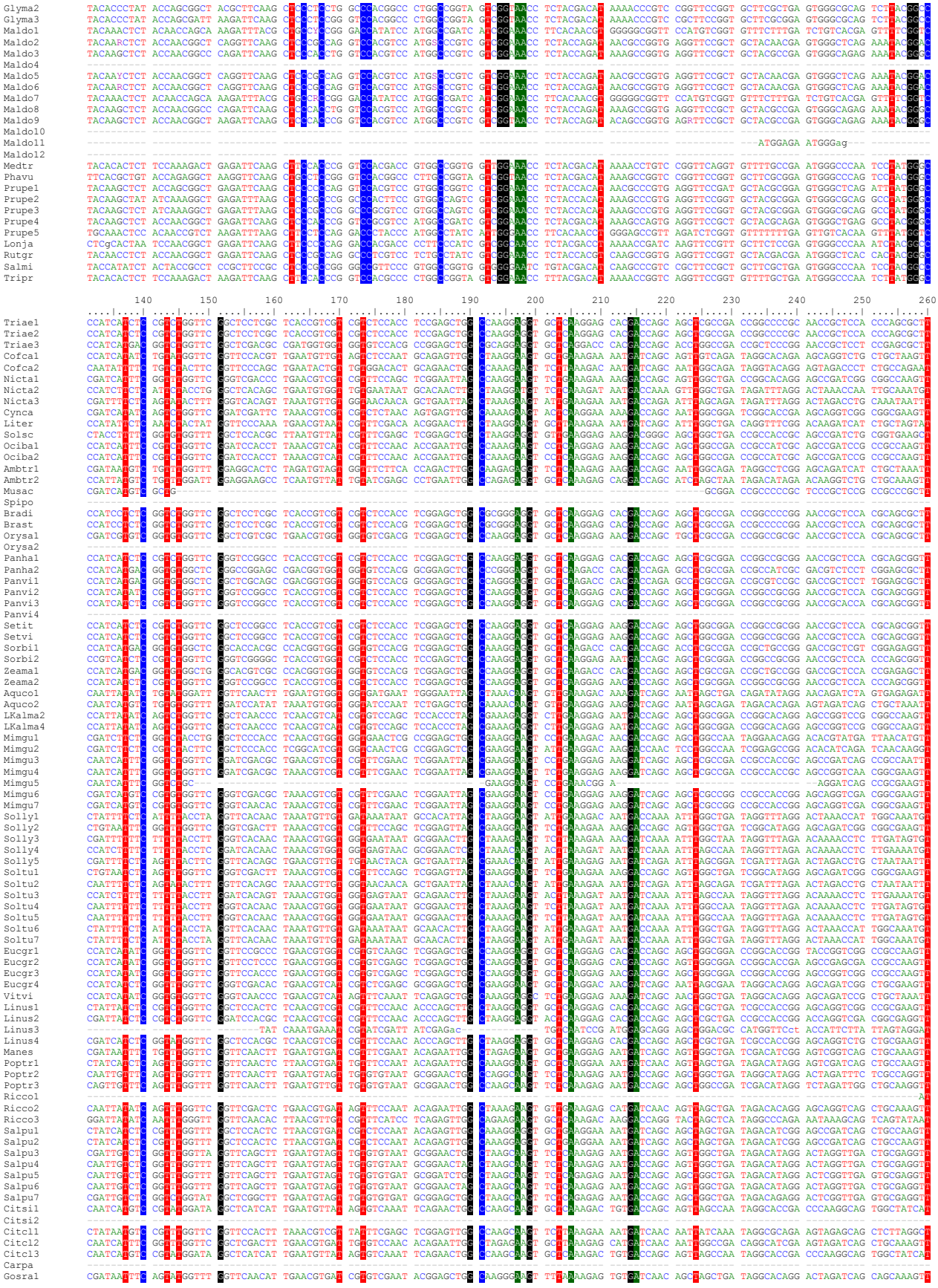
Yamasaki H, Sakihama Y, Ikehara N (1997) Flavonoid-peroxidase reaction as a detoxification mechanism of plant cells against H2O2. Plant Physiol **115**: 1405–1412

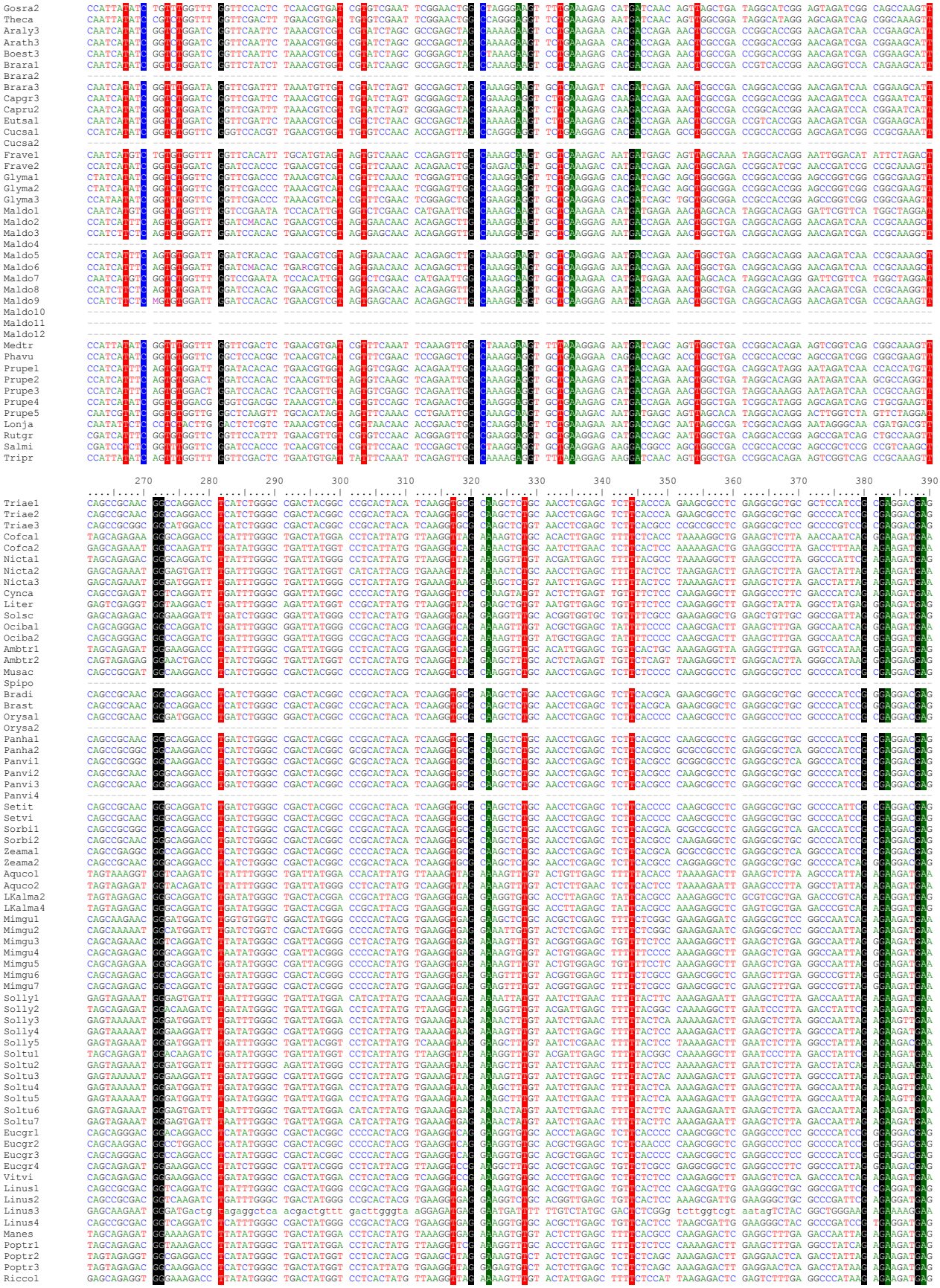
Zheng C, Santos Muñoz D, Albert VA, Sankoff D (2015) Syntenic block overlap multiplicities with a panel of reference genomes provide a signature of ancient polyploidization events. BMC Genomics **16**: S8

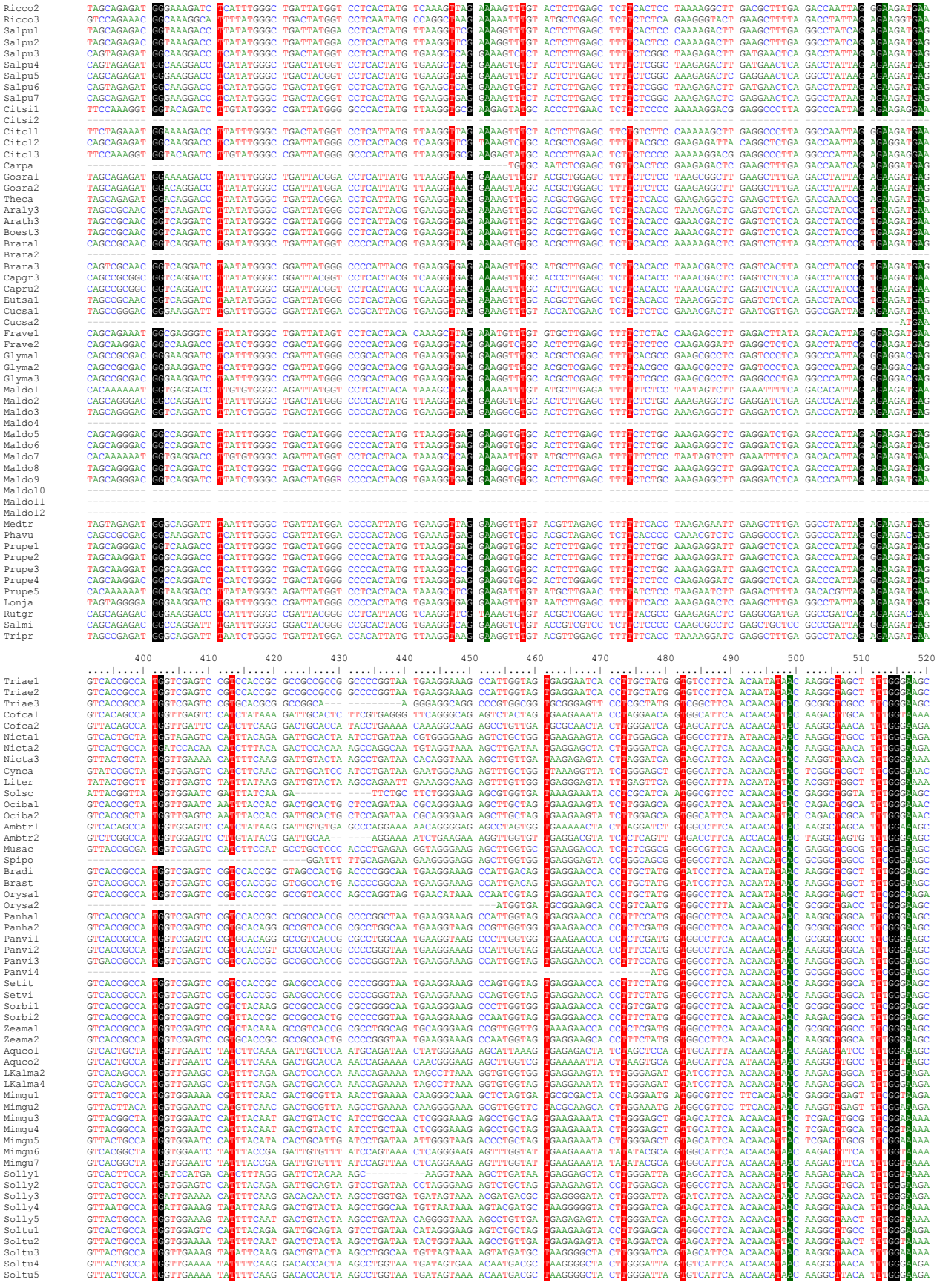
Zhong R, Lee C, Zhou J, McCarthy RL, Ye ZH (2008) A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. Plant Cell **20**: 2763–2782

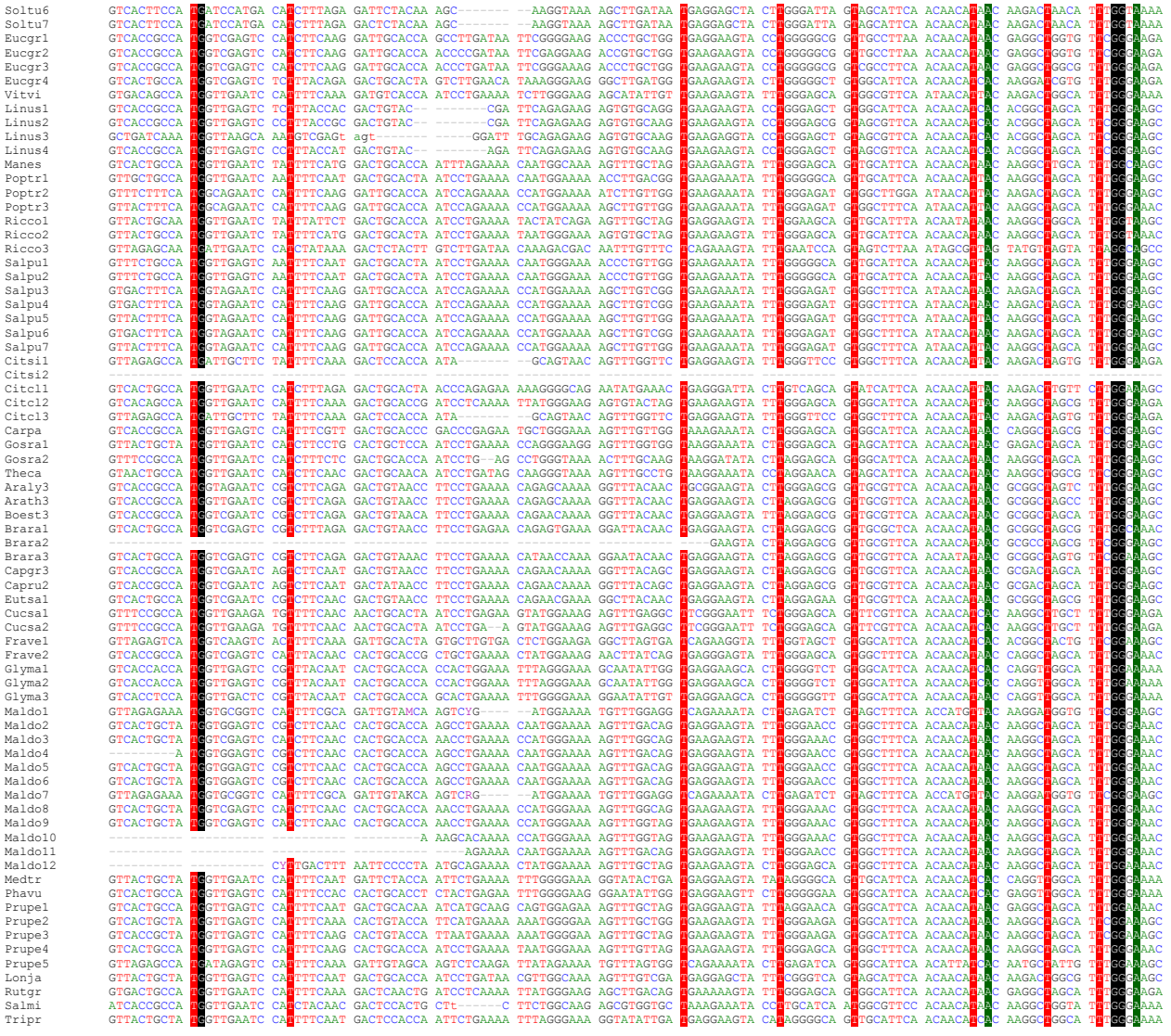
Appendix

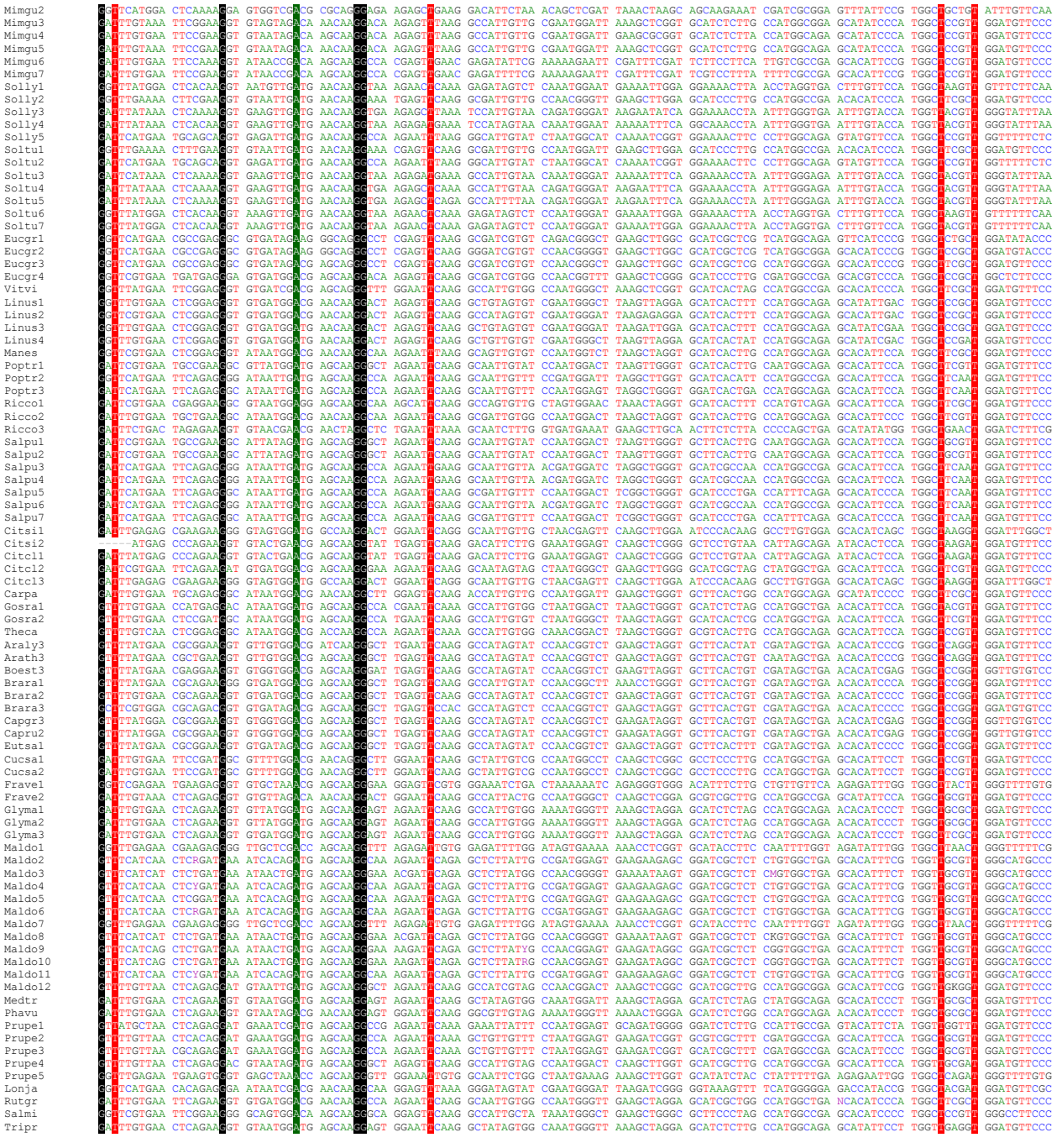


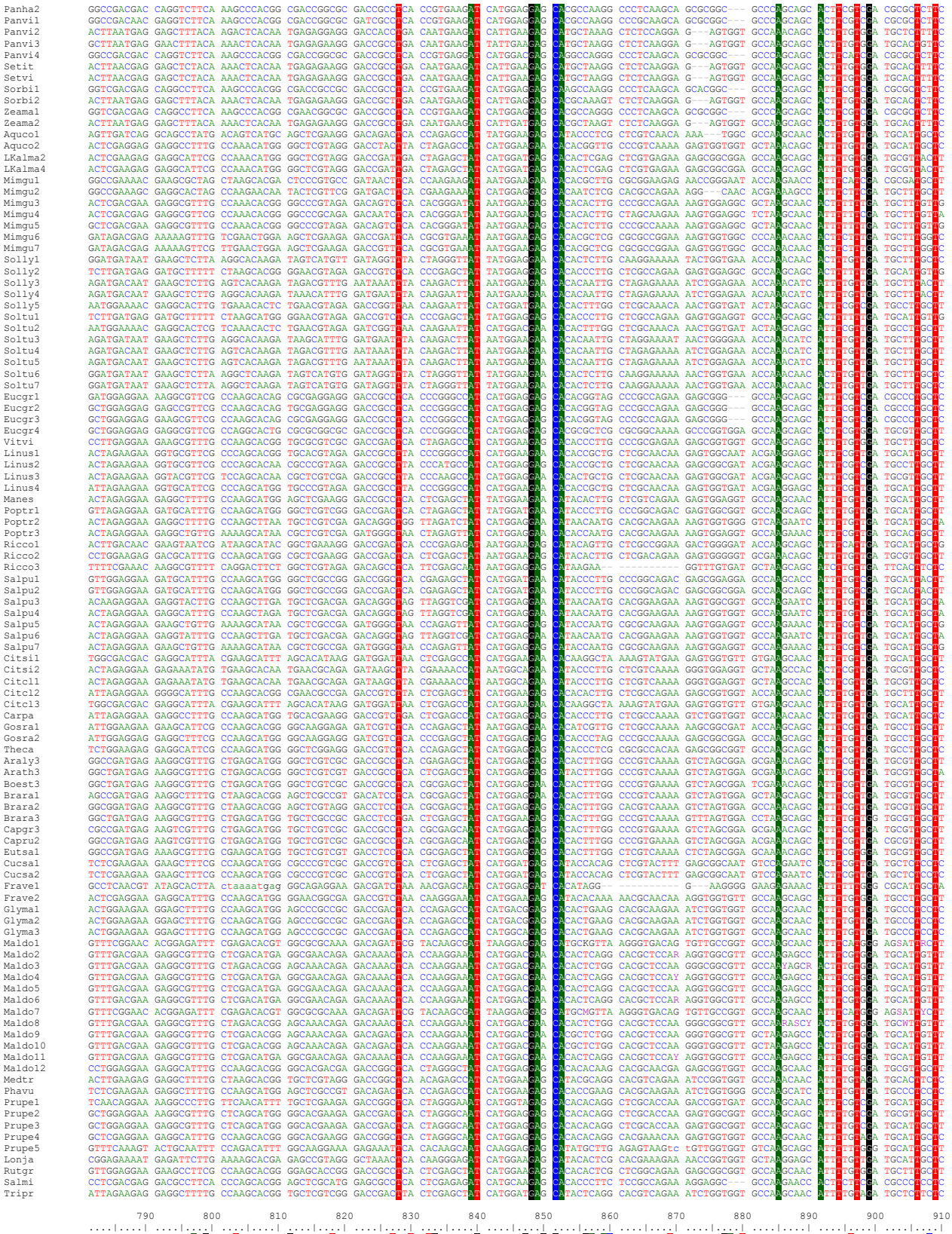



Gosra2
Theca
Araly3
Arath3
Boest3
Brara1
Brara2
Brara3
Capgr3
Capru2
Eutsa1
Cucsa1
Cucsa2
Frave1
Frave2
Glyma1
Glyma2
Glyma3
Maldo1
Maldo2
Maldo3
Maldo4
Maldo5
Maldo6
Maldo7
Maldo8
Maldo9
Maldo10
Maldo11
Maldo12
Medtr
Phavu
Prupe1
Prupe2
Prupe3
Prupe4
Prupe5
Lonja
Rutgr
Salmi
Tripr
270
280
290
300
310
320
330
340
350
360
370
380
390
Triae1
Triae2
Triae3
Cofca1
Cofca2
Nicta1
Nicta2
Nicta3
Cynca
Liter
Solsc
Ociba1
Ociba2
Ambtr1
Ambtr2
Musac
Spipo
Bradi
Brast
Orysa1
Orysa2
Panha1
Panha2
Panvi1
Panvi2
Panvi3
Panvi4
Setit
Setvi
Sorbi1
Sorbi2
Zeama1
Zeama2
Aquco1
Aquco2
LKalma2
LKalma4
Mimgu1
Mimgu2
Mimgu3
Mimgu4
Mimgu5
Mimgu6
Mimgu7
Solly1
Solly2
Solly3
Solly4
Solly5
Soltu1
Soltu2
Soltu3
Soltu4
Soltu5
Soltu6
Soltu7
Eucgr1
Eucgr2
Eucgr3
Eucgr4
Vitvi
Linus1
Linus2
Linus3
Linus4
Manes
Poptr1
Poptr2
Poptr3
Ricco1

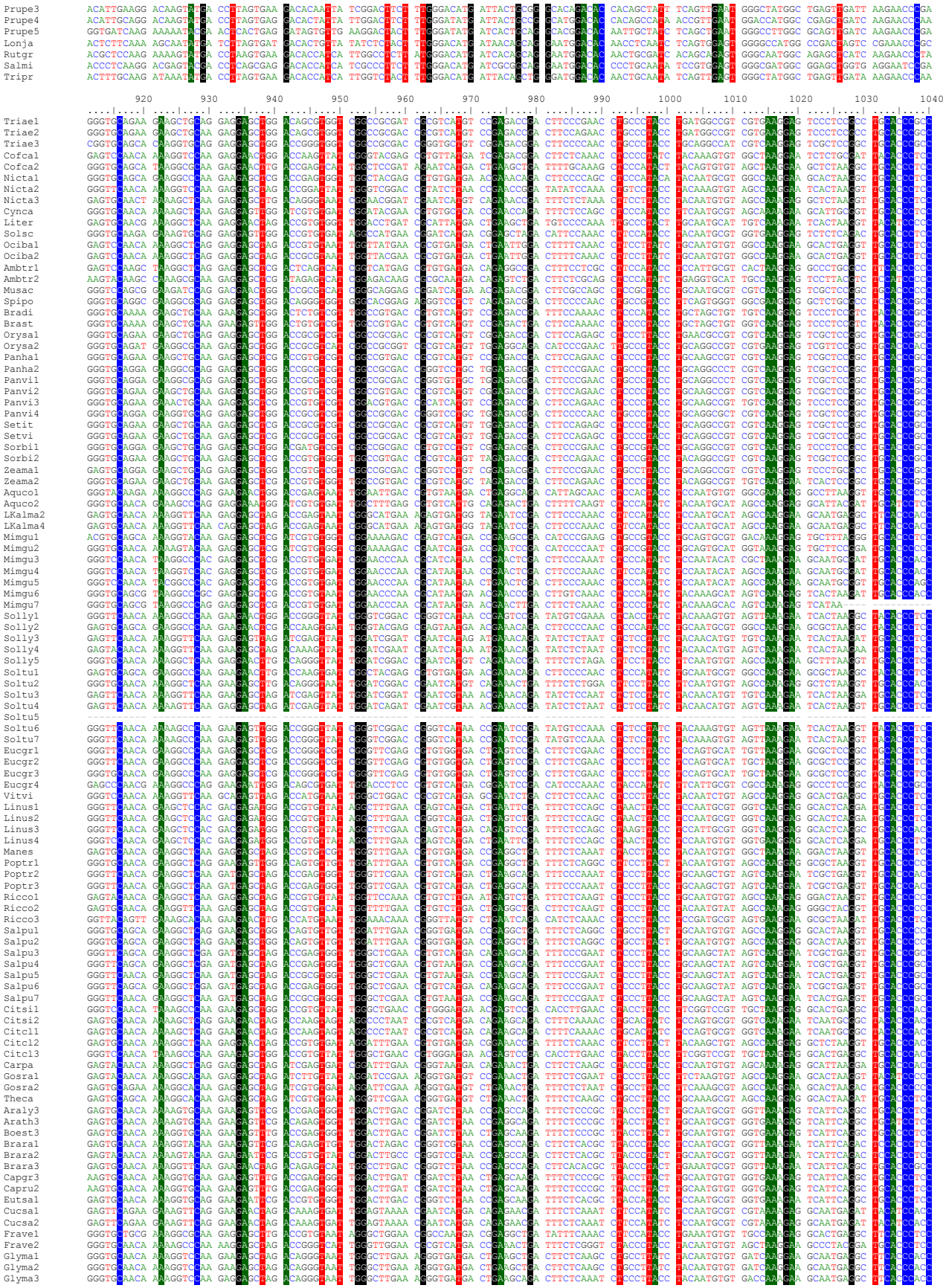

Ricco2
Ricco3
Salpu1
Salpu2
Salpu3
Salpu4
Salpu5
Salpu6
Salpu7
Citsi1
Citsi2
Citcl1
Citcl2
Citcl3
Carpa
Gosra1
Gosra2
Theca
Araly3
Arath3
Boest3
Brara1
Brara2
Brara3
Capgr3
Capru2
Eutsa1
Cucsa1
Cucsa2
Fravel
Frave2
Glyma1
Glyma2
Glyma3
Maldo1
Maldo2
Maldo3
Maldo4
Maldo5
Maldo6
Maldo7
Maldo8
Maldo9
Maldo10
Maldo11
Maldo12
Medtr
Phavu
Prupe1
Prupe2
Prupe3
Prupe4
Prupe5
Lonja
Rutgr
Salmi
Tripr
400
410
420
430
440
450
460
470
480
490
500
510
520
Triae1
Triae2
Triae3
Cofca1
Cofca2
Nicta1
Nicta2
Nicta3
Cynca
Liter
Solsc
Ociba1
Ociba2
Ambtr1
Ambtr2
Musac
Spipo
Bradi
Brast
Orysa1
Orysa2
Panha1
Panha2
Panvi1
Panvi2
Panvi3
Panvi4
Setit
Setvi
Sorbi1
Sorbi2
Zeama1
Zeama2
Aquco1
Aquco2
LKalma2
LKalma4
Mimgu1
Mimgu2
Mimgu3
Mimgu4
Mimgu5
Mimgu6
Mimgu7
Solly1
Solly2
Solly3
Solly4
Solly5
Soltu1
Soltu2
Soltu3
Soltu4
Soltu5

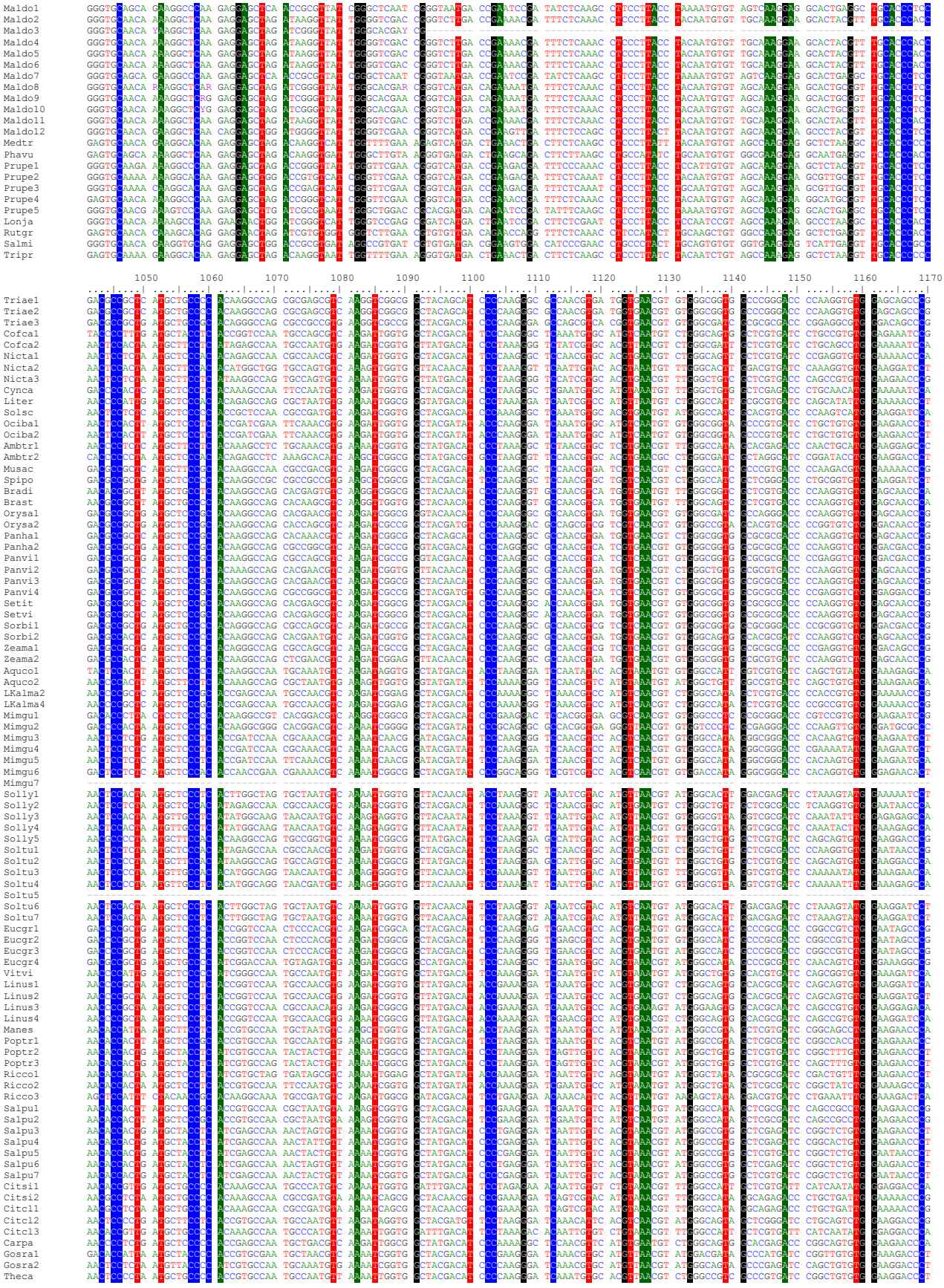

Soltu6
Soltu7
Eucgr1
Eucgr2
Eucgr3
Eucgr4
Vitvi
Linus1
Linus2
Linus3
Linus4
Manes
Poptr1
Poptr2
Poptr3
Ricco1
Ricco2
Ricco3
Salpu1
Salpu2
Salpu3
Salpu4
Salpu5
Salpu6
Salpu7
Citsi1
Citsi2
Citcl1
Citcl2
Citcl3
Carpa
Gosra1
Gosra2
Theca
Araly3
Arath3
Boest3
Brara1
Brara2
Brara3
Capgr3
Capru2
Eutsa1
Cucsa1
Cucsa2
Frave1
Frave2
Glyma1
Glyma2
Glyma3
Maldo1
Maldo2
Maldo3
Maldo4
Maldo5
Maldo6
Maldo7
Maldo8
Maldo9
Maldo10
Maldo11
Maldo12
Medtr
Phavu
Prupe1
Prupe2
Prupe3
Prupe4
Prupe5
Lonja
Rutgr
Salmi
Tripr
530
540
550
560
570
580
590
600
610
620
630
640
650
Triae1
Triae2
Triae3
Cofca1
Cofca2
Nicta1
Nicta2
Nicta3
Cynca
Liter
Solsc
Ociba1
Ociba2
Ambtr1
Ambtr2
Musac
Spipo
Bradi
Brast
Orysa1
Orysa2
Panha1
Panha2
Panvi1
Panvi2
Panvi3
Panvi4
Setit
Setvi
Sorbi1
Sorbi2
Zeama1
Zeama2
Aquco1
Aquco2
LKalma2
LKalma4
Mimgu1

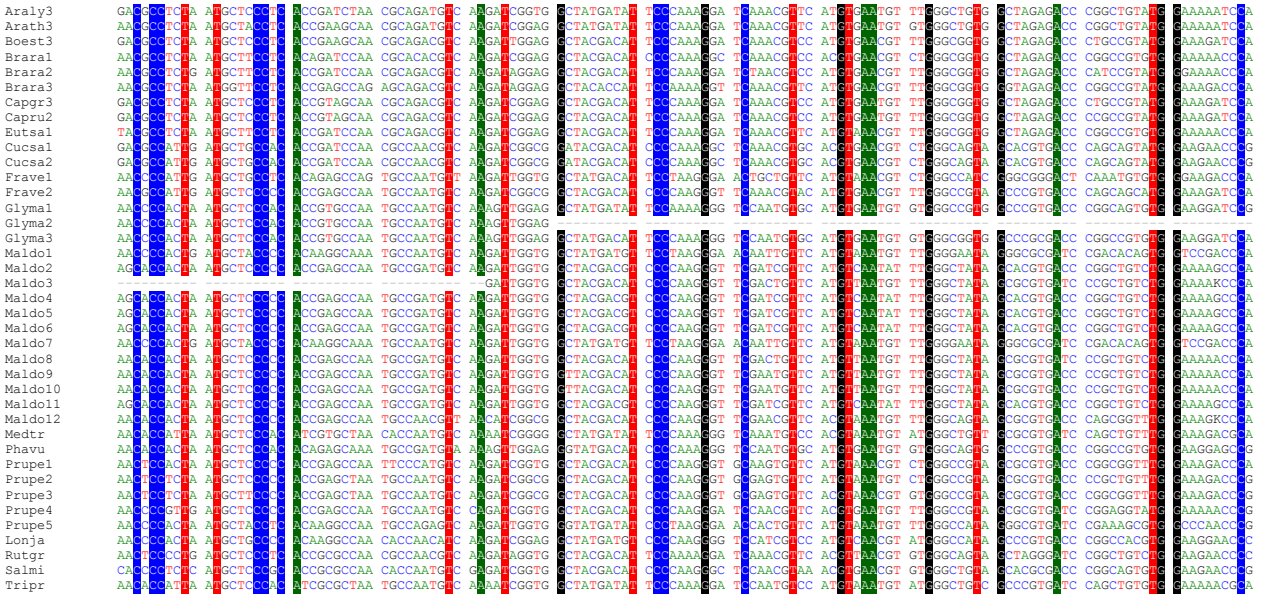

Mimgu2
Mimgu3
Mimgu4
Mimgu5
Mimgu6
Mimgu7
Solly1
Solly2
Solly3
Solly4
Solly5
Soltu1
Soltu2
Soltu3
Soltu4
Soltu5
Soltu6
Soltu7
Eucgr1
Eucgr2
Eucgr3
Eucgr4
Vitvi
Linus1
Linus2
Linus3
Linus4
Manes
Poptr1
Poptr2
Poptr3
Ricco1
Ricco2
Ricco3
Salpu1
Salpu2
Salpu3
Salpu4
Salpu5
Salpu6
Salpu7
Citsi1
Citsi2
Citcl1
Citcl2
Citcl3
Carpa
Gosra1
Gosra2
Theca
Araly3
Arath3
Boest3
Brara1
Brara2
Brara3
Capgr3
Capru2
Eutsa1
Cucsa1
Cucsa2
Frave1
Frave2
Glyma1
Glyma2
Glyma3
Maldo1
Maldo2
Maldo3
Maldo4
Maldo5
Maldo6
Maldo7
Maldo8
Maldo9
Maldo10
Maldo11
Maldo12
Medtr
Phavu
Prupe1
Prupe2
Prupe3
Prupe4
Prupe5
Lonja
Rutgr
Salmi
Tripr
660
670
680
690
700
710
720
730
740
750
760
770
780
Triae1
Triae2
Triae3
Cofca1
Cofca2
Nicta1
Nicta2
Nicta3
Cynca
Liter
Solsc
Ociba1
Ociba2
Ambtr1
Ambtr2
Musac
Spipo
Bradi
Brast
Orysa1
Orysa2
Panha1

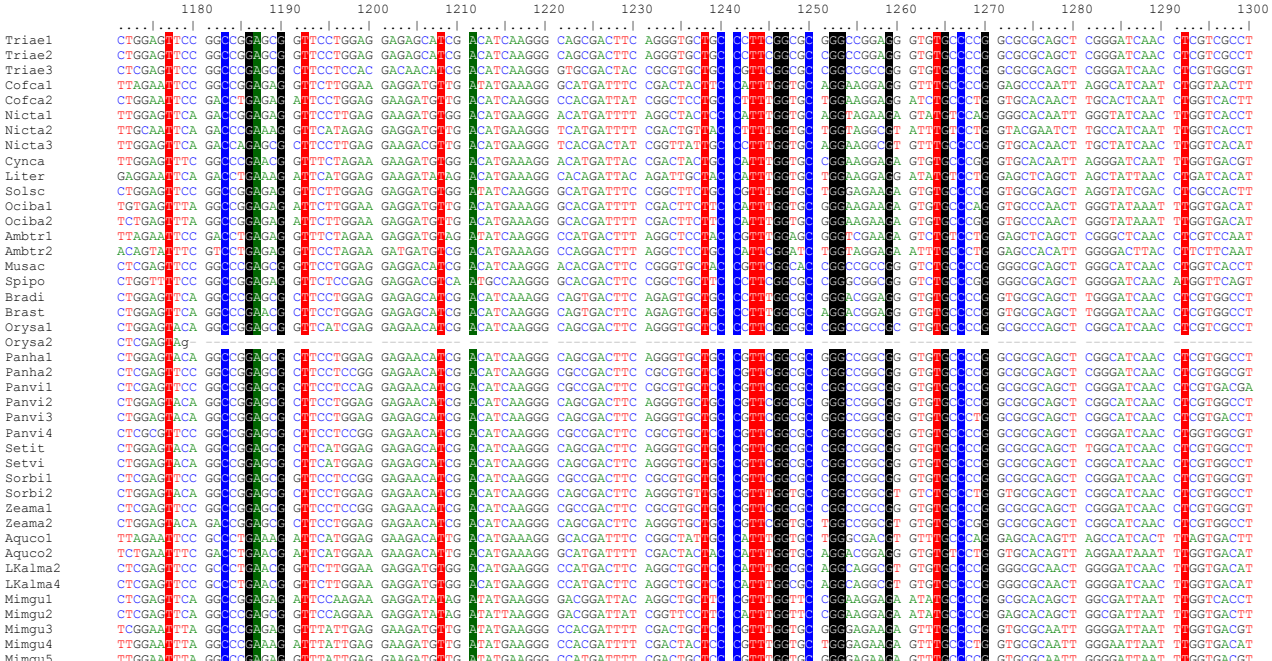

Panha2
Panvi1
Panvi2
Panvi3
Panvi4
Setit
Setvi
Sorbi1
Sorbi2
Zeama1
Zeama2
Aquco1
Aquco2
LKalma2
LKalma4
Mimgu1
Mimgu2
Mimgu3
Mimgu4
Mimgu5
Mimgu6
Mimgu7
Solly1
Solly2
Solly3
Solly4
Solly5
Soltu1
Soltu2
Soltu3
Soltu4
Soltu5
Soltu6
Soltu7
Eucgr1
Eucgr2
Eucgr3
Eucgr4
Vitvi
Linus1
Linus2
Linus3
Linus4
Manes
Poptr1
Poptr2
Poptr3
Ricco1
Ricco2
Ricco3
Salpu1
Salpu2
Salpu3
Salpu4
Salpu5
Salpu6
Salpu7
Citsi1
Citsi2
Citcl1
Citcl2
Citcl3
Carpa
Gosra1
Gosra2
Theca
Araly3
Arath3
Boest3
Brara1
Brara2
Brara3
Capgr3
Capru2
Eutsa1
Cucsa1
Cucsa2
Fravel
Frave2
Glyma1
Glyma2
Glyma3
Maldo1
Maldo2
Maldo3
Maldo4
Maldo5
Maldo6
Maldo7
Maldo8
Maldo9
Maldo10
Maldo11
Maldo12
Medtr
Phavu
Prupe1
Prupe2
Prupe3
Prupe4
Prupe5
Lonja
Rutgr
Salmi
Tripr
790
800
810
820
830
840
850
860
870
880
890
900
910
Triae1
Triae2
Triae3
Cofca1
Cofca2
Nicta1

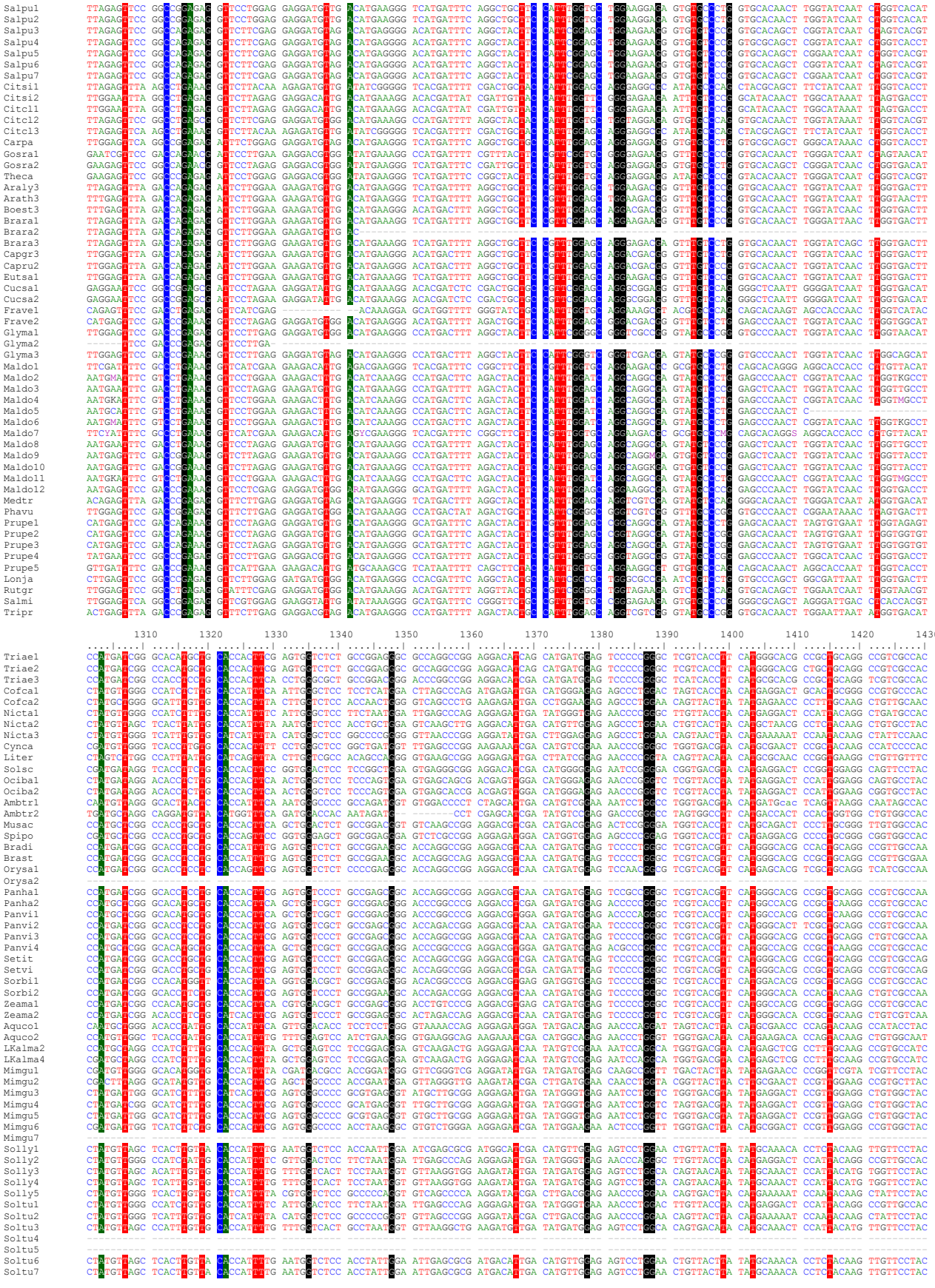

Nicta2
Nicta3
Cynca
Liter
Solsc
Ocibal
Ociba2
Ambtr1
Ambtr2
Musac
Spipo
Bradi
Brast
Orysa1
Orysa2
Panha1
Panha2
Panvi1
Panvi2
Panvi3
Panvi4
Setit
Setvi
Sorbi1
Sorbi2
Zeama1
Zeama2
Aquco1
Aquco2
LKalma2
LKalma4
Mimgu1
Mimgu2
Mimgu3
Mimgu4
Mimgu5
Mimgu6
Mimgu7
Solly1
Solly2
Solly3
Solly4
Solly5
Soltu1
Soltu2
Soltu3
Soltu4
Soltu5
Soltu6
Soltu7
Eucgr1
Eucgr2
Eucgr3
Eucgr4
Vitvi
Linus1
Linus2
Linus3
Linus4
Manes
Poptr1
Poptr2
Poptr3
Ricco1
Ricco2
Ricco3
Salpu1
Salpu2
Salpu3
Salpu4
Salpu5
Salpu6
Salpu7
Citsi1
Citsi2
Citcl1
Citcl2
Citcl3
Carpa
Gosra1
Gosra2
Theca
Araly3
Arath3
Boest3
Brara1
Brara2
Brara3
Capgr3
Capru2
Eutsa1
Cucsa1
Cucsa2
Fravel
Frave2
Glyma1
Glyma2
Glyma3
Maldo1
Maldo2
Maldo3
Maldo4
Maldo5
Maldo6
Maldo7
Maldo8
Maldo9
Maldo10
Maldo11
Maldo12
Medtr
Phavu
Prupe1
Prupe2

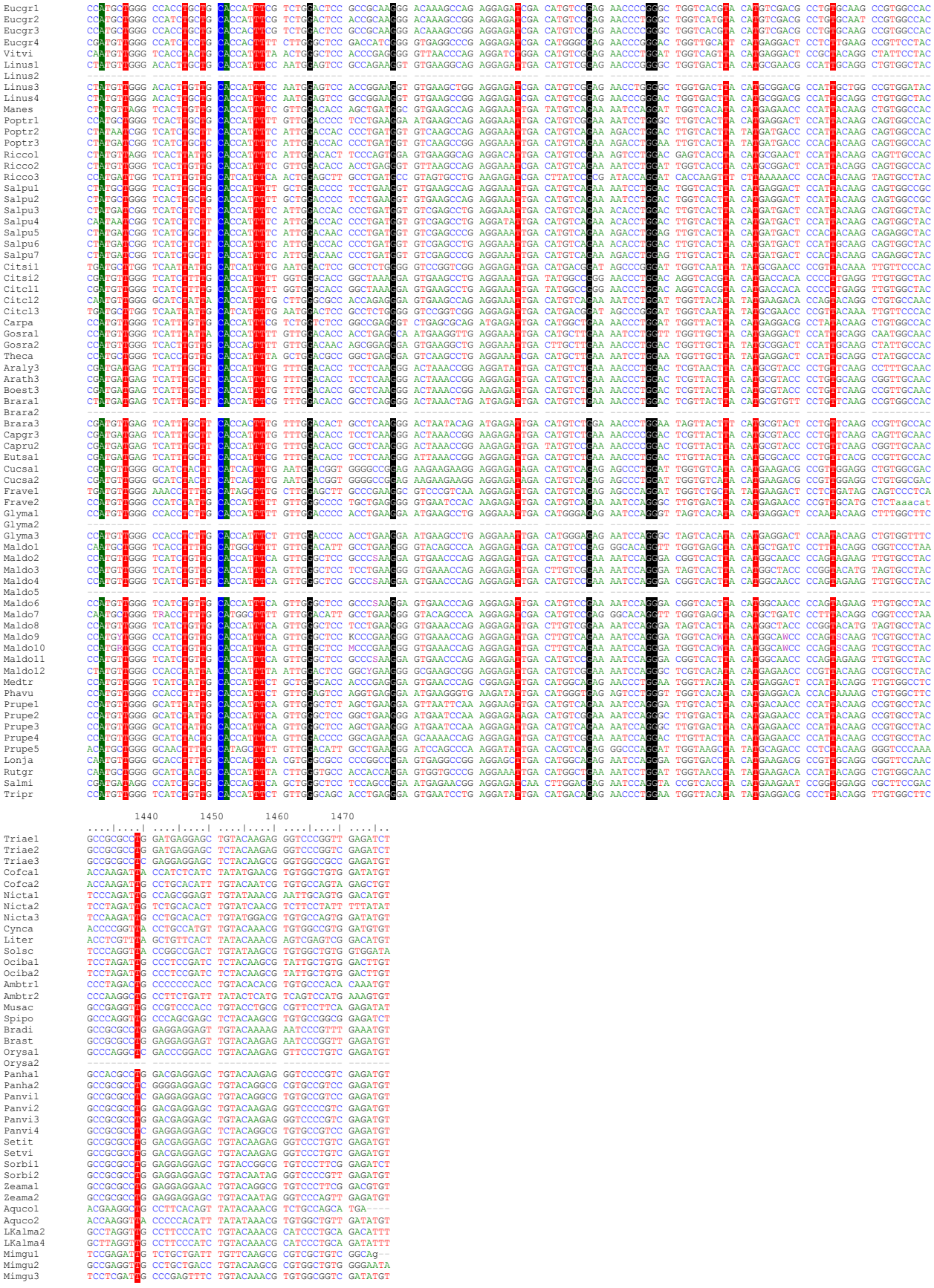

Prupe3
Prupe4
Prupe5
Lonja
Rutgr
Salmi
Tripr
920
930
940
950
960
970
980
990
1000
1010
1020
1030
1040
Triae1
Triae2
Triae3
Cofca1
Cofca2
Nicta1
Nicta2
Nicta3
Cynca
Liter
Solsc
Ociba1
Ociba2
Ambtr1
Ambtr2
Musac
Spipo
Bradi
Brast
Orysa1
Orysa2
Panha1
Panha2
Panvi1
Panvi2
Panvi3
Panvi4
Setit
Setvi
Sorbi1
Sorbi2
Zeama1
Zeama2
Aquco1
Aquco2
LKalma2
LKalma4
Mimgu1
Mimgu2
Mimgu3
Mimgu4
Mimgu5
Mimgu6
Mimgu7
Solly1
Solly2
Solly3
Solly4
Solly5
Soltu1
Soltu2
Soltu3
Soltu4
Soltu5
Soltu6
Soltu7
Eucgr1
Eucgr2
Eucgr3
Eucgr4
Vitvi
Linus1
Linus2
Linus3
Linus4
Manes
Poptr1
Poptr2
Poptr3
Ricco1
Ricco2
Ricco3
Salpu1
Salpu2
Salpu3
Salpu4
Salpu5
Salpu6
Salpu7
Citsi1
Citsi2
Citcl1
Citcl2
Citcl3
Carpa
Gosra1
Gosra2
Theca
Araly3
Arath3
Boest3
Brara1
Brara2
Brara3
Capgr3
Capru2
Eutsa1
Cucsa1
Cucsa2
Fravel
Frave2
Glyma1
Glyma2
Glyma3


Maldo1
Maldo2
Maldo3
Maldo4
Maldo5
Maldo6
Maldo7
Maldo8
Maldo9
Maldo10
Maldo11
Maldo12
Medtr
Phavu
Prupe1
Prupe2
Prupe3
Prupe4
Prupe5
Lonja
Rutgr
Salmi
Tripr
1050
1060
1070
1080
1090
1100
1110
1120
1130
1140
1150
1160
1170
Triae1
Triae2
Triae3
Cofca1
Cofca2
Nicta1
Nicta2
Nicta3
Cynca
Liter
Solsc
Ociba1
Ociba2
Ambtr1
Ambtr2
Musac
Spipo
Bradi
Brast
Orysa1
Orysa2
Panha1
Panha2
Panvi1
Panvi2
Panvi3
Panvi4
Setit
Setvi
Sorbi1
Sorbi2
Zeama1
Zeama2
Aquco1
Aquco2
LKalma2
LKalma4
Mimgu1
Mimgu2
Mimgu3
Mimgu4
Mimgu5
Mimgu6
Mimgu7
Solly1
Solly2
Solly3
Solly4
Solly5
Soltu1
Soltu2
Soltu3
Soltu4
Soltu5
Soltu6
Soltu7
Eucgr1
Eucgr2
Eucgr3
Eucgr4
Vitvi
Linus1
Linus2
Linus3
Linus4
Manes
Poptr1
Poptr2
Poptr3
Ricco1
Ricco2
Ricco3
Salpu1
Salpu2
Salpu3
Salpu4
Salpu5
Salpu6
Salpu7
Citsi1
Citsi2
Citcl1
Citcl2
Citcl3
Carpa
Gosra1
Gosra2
Theca


Araly3
Arath3
Boest3
Brara1
Brara2
Brara3
Capgr3
Capru2
Eutsa1
Cucsa1
Cucsa2
Frave1
Frave2
Glyma1
Glyma2
Glyma3
Maldo1
Maldo2
Maldo3
Maldo4
Maldo5
Maldo6
Maldo7
Maldo8
Maldo9
Maldo10
Maldo11
Maldo12
Medtr
Phavu
Prupe1
Prupe2
Prupe3
Prupe4
Prupe5
Lonja
Rutgr
Salmi
Tripr


1180
1190
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
1300
Triae1
Triae2
Triae3
Cofca1
Cofca2
Nicta1
Nicta2
Nicta3
Cynca
Liter
Solsc
Ociba1
Ociba2
Ambtr1
Ambtr2
Musac
Spipo
Bradi
Brast
Orysa1
Orysa2
Panha1
Panha2
Panvi1
Panvi2
Panvi3
Panvi4
Setit
Setvi
Sorbi1
Sorbi2
Zeama1
Zeama2
Aquco1
Aquco2
LKalma2
LKalma4
Mimgu1
Mimgu2
Mimgu3
Mimgu4
Mimgu5
Mimgu6
Mimgu7
Solly1
Solly2
Solly3
Solly4
Solly5
Soltu1
Soltu2
Soltu3
Soltu4
Soltu5
Soltu6
Soltu7
Eucgr1
Eucgr2
Eucgr3
Eucgr4
Vitvi
Linus1
Linus2
Linus3
Linus4
Manes
Poptr1
Poptr2
Poptr3
Ricco1
Ricco2
Ricco3


Salpu1
Salpu2
Salpu3
Salpu4
Salpu5
Salpu6
Salpu7
Citsi1
Citsi2
Citcl1
Citcl2
Citcl3
Carpa
Gosra1
Gosra2
Theca
Araly3
Arath3
Boest3
Brara1
Brara2
Brara3
Capgr3
Capru2
Eutsa1
Cucsa1
Cucsa2
Fravel
Frave2
Glyma1
Glyma2
Glyma3
Maldo1
Maldo2
Maldo3
Maldo4
Maldo5
Maldo6
Maldo7
Maldo8
Maldo9
Maldo10
Maldo11
Maldo12
Medtr
Phavu
Prupe1
Prupe2
Prupe3
Prupe4
Prupe5
Lonja
Rutgr
Salmi
Tripr
1310
1320
1330
1340
1350
1360
1370
1380
1390
1400
1410
1420
1430
Triae1
Triae2
Triae3
Cofca1
Cofca2
Nicta1
Nicta2
Nicta3
Cynca
Liter
Solsc
Ociba1
Ociba2
Ambtr1
Ambtr2
Musac
Spipo
Bradi
Brast
Orysa1
Orysa2
Panha1
Panha2
Panvi1
Panvi2
Panvi3
Panvi4
Setit
Setvi
Sorbi1
Sorbi2
Zeama1
Zeama2
Aquco1
Aquco2
LKalma2
LKalma4
Mimgu1
Mimgu2
Mimgu3
Mimgu4
Mimgu5
Mimgu6
Mimgu7
Solly1
Solly2
Solly3
Solly4
Solly5
Soltu1
Soltu2
Soltu3
Soltu4
Soltu5
Soltu6
Soltu7


Eucgr1
Eucgr2
Eucgr3
Eucgr4
Vitvi
Linus1
Linus2
Linus3
Linus4
Manes
Poptr1
Poptr2
Poptr3
Ricco1
Ricco2
Ricco3
Salpu1
Salpu2
Salpu3
Salpu4
Salpu5
Salpu6
Salpu7
Citsi1
Citsi2
Citcl1
Citcl2
Citcl3
Carpa
Gosra1
Gosra2
Theca
Araly3
Arath3
Boest3
Brara1
Brara2
Brara3
Capgr3
Capru2
Eutsa1
Cucsa1
Cucsa2
Frave1
Frave2
Glyma1
Glyma2
Glyma3
Maldo1
Maldo2
Maldo3
Maldo4
Maldo5
Maldo6
Maldo7
Maldo8
Maldo9
Maldo10
Maldo11
Maldo12
Medtr
Phavu
Prupe1
Prupe2
Prupe3
Prupe4
Prupe5
Lonja
Rutgr
Salmi
Tripr
1440
1450
1460
1470
Triae1
Triae2
Triae3
Cofca1
Cofca2
Nicta1
Nicta2
Nicta3
Cynca
Liter
Solsc
Ociba1
Ociba2
Ambtr1
Ambtr2
Musac
Spipo
Bradi
Brast
Orysa1
Orysa2
Panha1
Panha2
Panvi1
Panvi2
Panvi3
Panvi4
Setit
Setvi
Sorbi1
Sorbi2
Zeama1
Zeama2
Aquco1
Aquco2
LKalma2
LKalma4
Mimgu1
Mimgu2
Mimgu3


Mimgu4 TCCTCGATTG CCCGAGTTTC TATACAAGCG TGTGGCGGTC GATATGT
Mimgu5 TCCTCGATTA CCCAAGTTTC TGTACAAGCG TGTGGCGGTC GATATGT
Mimgu6 TCCTCGATTG CCCGAGTTTC TATACAAGCG CACGGCTGCG GACTTGT
Mimgu7
Solly1 TCCTAGGTTG CCTCAACACT TGTATGAACG CAGTCCTATC GTTATAT
Solly2 TCCCAGATTA CCAGCTGAGT TGTATAAACG AATTGCAGTC GACATGT
Solly3 TCCTAGGTTA CCTAGAAACC TATATACACA TGTTGAGTga
Solly4 TCCTAGGTTA CCTAGAAACT TATACAACAC ATGTTCAATG AACATGT
Solly5 TCCAAGATTG CCTGCACACT TGTATAAACG TGTGCCGATG GATATGT
Soltu1 TCCCAGATTG CCAGCTGAGT TGTATAAACG AATTGCAGTC GACATGT
Soltu2 TCCAAGATTG CCTGCACACT TGTATAAACG TGTGCCAATG GATATGT
Soltu3 TCCTAGGTTA CCTATACACT TATACAACAC ATGTTCAATG AACATGT
Soltu4
Soltu5
Soltu6 TCCTAGATTG CCTCTACACT TATATGAACG CGGTCCTATC GTTATAT
Soltu7 TCCTAGATTG CCTCTACACT TATATGAACG CGGTCCTATC GTTATAT
Eucgr1 GCCTAGGCTA CCCCCCGAAC TGTACAAACG CATGCCGTAT GAAATGT
Eucgr2 GCCTAGGCTA CCCTCCGAAC TGTACAAACG CGTGCCGTAT GAAATGT
Eucgr3 GCCTAGGCTA CCCTCCGAAC TGTACAAACG CGTGCCGTAT GAAATGT
Eucgr4 TCCAAGGTTG GCCGCAGAGT TGTACAAGCG TGTACCTGTG GACATAT
Vitvi TTCAAGGTTA CCTGCAAGCT TATACAAAAG GATGGCTGTG GATATTT
Linus1 TCCTCGGCTG CCGTCTCACT TGTACAAACG TGTAGCCGTC GACATGT
Linus2
Linus3 TCCTCGGCTA CCGTCTCACT TGTACAAACG TGTAGCCGTC GACATGT
Linus4 TCCCCGGCTG CCGTCTCATT TGTACAAACG TGTAGCCGTC GACATGT
Manes ACCTCGGCTG CCTTCAGATT TGTACAAACG AGTGGCTGTA GATATGT
Poptr1 TCCTCGGTTG CCTTCACATT TGTACAAACG TGTTGCTGTT GATATTT
Poptr2 TCCTCGCCTG CCTTCACACT TGTACAAACG GATGGCTTCA GATATGT
Poptr3 TCCTCGCCTG CCTCCACACT TGTACAAGCG GGTGGCTTCA GATATGT
Ricco1 CCCTCGTCTT CCTTGA
Ricco2 TCCTCGACTC CCTTCAGAAT TGTACAAACG CGTGGCTGTG GATATGT
Ricco3 ATTGAGATTG CCAGCCCACT TATACAAACA TGAGAGTTAT GATGc
Salpu1 TCCTCGGTTG CCTTCACATT TGTACAAACG TGTTGCTGTT GATATTT
Salpu2 TCCTCGGTTG CCTTCACATT TGTACAAACG TGTTGCTGTT GATATTT
Salpu3 TCCACGCCTG CCTCCACACT TGCACAAACG GATGGCTTCA GATATGT
Salpu4 TCCACGCCTG CCTCCACACT TGCACAAACG GATGGCTTCA GATATGT
Salpu5 TCCTCGCCTG CCTCCACACT TGCACAAACG GATGGCATCA GATATGT
Salpu6 TCCACGCCTG CCTCCACACT TGCACAAACG GATGGCTTCA GATATGT
Salpu7 TCCTCGCCTG CCTCCACACT TGCACAAACG GATGGCATCA GATATGT
Citsi1 ACCGAGGTTA CCTGCACAAG TGTATGAACG TGTGTGA
Citsi2 CCCGCGGTTG GCTGATGGCC TGTACAAACG TGTCCTTGTG GACATGT
Citcl1 CCCGCGGTTG GCTGATGGCC TGTACAAACG TGTCCTTGTG GACATGT
Citcl2 TCCTAGGCTG CCTTCGCACT TGTATAAACG TGTGGCCGCT GATATGT
Citcl3 ACCGAGGTTA CCTGCACAAG TGTATGAACG TGTGTGA
Carpa TCCACGGTTG CCTTCTCATT TGTACAAACG GGTGCCTGCG GATATCT
Gosra1 TCCTAGGTTG CCTTATCATC TTTACAAACG CATTGCTGTT GATATGT
Gosra2 TCCTAGGCTG CCTTCTCATC TCTACAAACG TGTGGCCGCG GATATGT
Theca TCCTAGGTTG CCTTCCCATC TGTACAAACG TGTGGCTGTG GATGTCT
Araly3 GCCTCGGTTG CCTTCAGATC TGTACAAAAG GGTGCCTTTT GATATGT
Arath3 GCCTCGGTTG CCTTCGGATC TGTACAAACG CGTGCCTTAC GATATGT
Boest3 GCCTCGGTTG CCTTCGGATC TGTACAAACG CGTGCCTTTC GATATGT
Brara1 GCCTCGGTTG CCTTCTGATC TGTATAAACG TGTGCCTTAT GAAATGT
Brara2
Brara3 GCCTCGGTTG GCTTCGGATC TGTATAAACG TGTGCCTTTT GATATGT
Capgr3 ACCACGGTTG CCTTCGGACC TGTACAAACG CGTGCCTTAC TATATGT
Capru2 GCCACGGTTG CCTTCGGACC TGTACAAACG CGTGCCTTAC TATATGT
Eutsa1 GCCTCGGTTG CCTTCGGATC TGTGTAAACG CGTGCCTTTC GATATGT
Cucsa1 TCCAAGGCTG ACGAAATTGT TGTACAAACG AGTGGCTGTG GACATGT
Cucsa2 TCCAAGGCTG ACGAAATTGT TGTACAAACG AGTGGCTGTG GACATGT
Frave1 ATTGAGATTG CCAGAGCACT CTTACAAGCA TATGCTTTGA
Frave2 ttgTCTGCTG CCTCTTGGTT TCAACAAATt aagtctt
Glyma1 TCCTAGGCTC CCCTCACATT TGTACAAACG TGTGCCTGCT GAGATCT
Glyma2
Glyma3 TCCTAGGCTC CCCTCACATT TATACAAACG TGTGCCTGCT GAGATCT
Maldo1 AGAGAGATTG CCAGCACACT TGTACATGAA TTAG
Maldo2 TCCAAGGCTG CCATCACACT TGTACAAGCG TGTGGAGGCG ACTATAA
Maldo3 TCCGAGGKTC CCTTCACACT TGTACAAGCG TGTGGCCGCG AATATAT
Maldo4 TCCAAGGCTG CCATCACACT TGTACAAGCG TGTGGAGGCG ACTATAA
Maldo5
Maldo6 TCCAAGGCTG CCATCACACT TGTACAAGCG TGTGGAGGCG ACTATAA
Maldo7 AGAGAGATTG CCAGCACACT TGTACATGAA TTAG
Maldo8 TCCGAGGATC CCTTCACACT TGTACAAGCG TTcttgcaaa attgg
Maldo9 TCCGAGGCTC CCTTCACACT TGTACAAGCG AGTGGCSGCG AATATGT
Maldo10 TCCGAGGCTC CCTTCACACT TGTACAAGCG AGTGGCSGCG AATATGT
Maldo11 TCCAAGGCTG CCATCACACT TGTACAAGCG TGTGGAGGCG ACTATGT
Maldo12 TCCTAGGCTG CCATCACAGT TATACAAACG TGTCGCTGCA GATATGT
Medtr TCCTAGGCTT CCGTCGGAGT TGTACAAACG TGTGACAGCT GATATCT
Phavu TCCTAGGCTC CCTTCACACT TGTACAAACG TCTACCTGCT GAGATTT
Prupe1 ACCAAGGCTT CCTTCGCACT TGTACAAACG CGTCGCCGCA GAAATCT
Prupe2 TCCAAGGCTC CCCTCGCATT TGTACAAACG TGTTGCCGCA GATATGT
Prupe3 TCCAAGGCTC CCCTCGCATT TGTACAAACG TGTTGCCGCA GAAATGT
Prupe4 TCCAAGGCTT CCATCGCACT TGTACAAACG TGTGGCTGCA GATATGT
Prupe5 TTTGAGATTG CCACCACACT TGTACAAGCA GTCCCAAGAT TGA
Lonja GCCGAGGTTG CCGTCAAGGT TGTACGCGCG TGTCCCAGTG GATtaCA
Rutgr TCCTAGGCTT CCTTCACAAT TGTACAAACG TGTTGCTGCT GATTTGT
Salmi GCCAAGACTA GCGCCTCACT TGTACAAGCG CGTTGCTGTC GACACCA
Tripr TCCTAGGCTT CCCTCAGAAT TGTACAAACG TGTGCCAGCT GATATCT

UNIVERSITÉ DE STRASBOURG

Annette Veronika ALBER

École Doctorale des Sciences de la Vie et de la Santé STRASBOURG

Phenolic 3-hydroxylases in land plants: biochemical diversity and molecular evolution

Les plantes produisent une grande variété de produits naturels pour faire face aux conditions environnementales. Les enzymes de la famille CYP98 des cytochromes P450 sont des enzymes clés dans la production des composés dérivés de la voie des phénylpropanoïdes. Ces enzymes sont impliquées dans l'hydroxylation des esters phénoliques pour la biosynthèse des monolignols chez les angiospermes, mais elles sont également impliquées dans la production de divers autres composés phénoliques solubles. Nous avons caractérisé des CYP98 représentatifs des mousses, Lycopodes, fougères, Gymnospermes, Angiospermes basales, Monocotylédones et Eudicotylédones et démontré que leur préférence de substrat a changé au cours de l'évolution. Un mutant knock-out de *CYP98* de mousse a révélé un phénotype sévère et que le *p*-coumaroyl-thréonate est substrat de l'enzyme *in vivo.* Une duplication des *CYP98s* ne peut être observée que dans le génome des Angiospermes, qui présentent généralement une isoforme potentiellement impliquée dans la biosynthèse de la lignine et autres isoformes, résultant de duplications indépendantes, dont le spectre de substrats est plus large *in vitro*.

Plants produce a rich variety of natural products to face environmental constraints. Enzymes of the cytochrome P450 CYP98 family are key actors in the production of phenolic bioactive compounds. They hydroxylate phenolic esters for lignin biosynthesis in angiosperms, but also produce various other bioactive phenolics. We characterized CYP98s from a moss, a lycopod, a fern, a conifer, a basal angiosperm, a monocot and from two eudicots. We found that substrate preference of the enzymes has changed during evolution of land plants with typical lignin-related activities only appearing in angiosperms, suggesting that ferns, similar to lycopods, produce lignin through an alternative route. A moss *CYP98* knock-out mutant revealed coumaroyl-threonate as CYP98 substrate *in vivo* and showed a severe phenotype. Multiple CYP98s per species exist only in the angiosperms, where we generally found one isoform presumably involved in the biosynthesis of monolignols, and additional isoforms, resulting from independent duplications, with a broad range of functions *in vitro.*