Molecular Characterization of Shikimate and Quinate Biosynthesis in *Populus trichocarpa*: Functional Diversification of the Dehydroquinate Dehydratase/Shikimate (Quinate) Dehydrogenase (DQD/SDH/QDH) Superfamily via Gene Duplication

by

Jia Guo
B.Sc. (Honours), Trinity Western University, 2008

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biology

Supervisory Committee

Molecular Characterization of Shikimate and Quinate Biosynthesis in *Populus trichocarpa*: Functional Diversification of the Dehydroquinate Dehydratase/Shikimate (Quinate) Dehydrogenase (DQD/SDH/QDH) Superfamily via Gene Duplication

by

Jia Guo
B.Sc. (Honours), Trinity Western University, 2008

Supervisory Committee

Dr. Jürgen Ehlting, (Department of Biology)
Supervisor

Dr. C. Peter Constabel, (Department of Biology)
Departmental Member

Dr. John Taylor, (Department of Biology)
Departmental Member

Dr. Caren Helbing, (Department of Biochemistry and Microbiology)
Outside Member

Abstract

Supervisory Committee
Dr. Jürgen Ehlting, (Department of Biology)
Supervisor
Dr. C. Peter Constabel, (Department of Biology)
Departmental Member
Dr. John Taylor, (Department of Biology)
Departmental Member
Dr. Caren Helbing, (Department of Biochemistry and Microbiology)
Outside Member

The shikimate pathway connects primary metabolism with the biosynthesis of the three aromatic amino acids (phenylalanine, tyrosine and tryptophan), which are essential protein building blocks. This pathway also provides precursors for a wide array of plant secondary metabolites with adaptive functions in plant adaptation and defense. The third and fourth steps of the shikimate pathway (the conversion of shikimate from 3-dehydroquinate via 3-dehydroshikimate) are catalyzed by a bi-functional enzyme called 3-dehydroquinate dehydratase/shikimate dehydrogenase (DQD/SDH). DQD/SDHs have been biochemically characterized in a few plant species including *Arabidopsis thaliana*, *Solanum lycopersicum* and *Nicotiana tabacum*. The embryo-lethal phenotype of Arabidopsis null mutants lacking *DQD/SDH* highlights a critical role of shikimate in primary metabolism. Quinate shares high structural similarity with shikimate and is an important secondary metabolite present in many plant species. Quinate and its derivatives (e.g. chlorogenic acid) serve important functions in plant defense due to their astringent (i.e. bitterness) and antimicrobial properties. Quinate can be derived from 3-dehydroquinate, and this reaction is catalyzed by quinate dehydrogense (QDH), the reaction mechanism of which resembles that of SDH. With a functional genomics

approach, I demonstrated that two of the five poplar putative DQD/SDHs (Poptr1 and Poptr5, poplar DQD/SDH1 and 2) have exclusive specificity for shikimate, while the other three (Poptr2 to Poptr4, poplar QDH1 to 3) are involved in quinate biosynthesis. Phylogenetic reconstruction of the DQD/SDH/QDH superfamily has identified two distinct clades in seed plants that may act preferentially on either shikimate or quinate, whereas lineages that have diverged prior to the angiosperm/gymnosperm split, only have a single copy DQD/SDH. An evolutionary analysis was carried out, and the sequence of the immediate pre-duplication ancestral DQD/SDH (>300MYA) was estimated and reconstructed. Protein structure modelling and *in vitro* biochemical characterization of the ancestral recombinant protein was performed along with some extant members of this family (pre-duplication representatives: *Rhodopirellula baltica* (Rhoba), *Chlamydomonas reinhardtii* (Chlre), *Physcomitrella patens* (Phypa) and *Selaginella moellendorffii* (Selmo); post-duplication species: *Pinus taeda* (Pinta1 & Pinta2) and *Populus trichocarpa* (Poptr1 & Poptr3). Together, the results indicate that quinate biosynthetic activity was gained prior to duplication and remained low until it became beneficial and favored by selection. The optimization of quinate biosynthetic activity was at the expense of losing some primary shikimate biosynthetic function creating a pleiotropic conflict. This was then resolved by gene duplication and further specialization leading to genes encoding specialized enzymes (either SDH or QDH). Diversification of the DQD/SDH/QDH superfamily likely occurred through sub-functionalization via a mechanism described as “Escape from Adaptive Conflict.”

Table of Contents

List of Tables

List of Figures

List of Abbreviations

4CL	4-coumarate: CoA ligase
ADH	Arogenate dehydrogenase
ADP	Adenosine diphosphate
ADT	Arogenate dehydratase
AMP	Adenosine monophosphate
APT	Anthranilate phosphoribosyltransferase
Arg	Arginine
AS	Anthranilate synthase
Asn	Asparagine
Asp	Aspartic acid
ATP	Adenosine triphosphate
BLAST	Basical Local Alignment Search Tool
C3'H	*p*-Coumaroyl-CoA 3'-hydroxylase
C3H	4-hydroxycinnamate 3-hydroxylase
C4H	Cinnamate 4-hydroxylase
CAD	Cinnamyl alcohol dehydrogenase
CCoA-3H	Coumaroyl-coenzyme A 3-hydroxylase
CCoA-OMT	Caffeoyl-coenzyme A O-methyltransferase
CCR	Cinnamoyl-CoA reductase
CdRP	1-(*o*-carboxyphenylamino)-1-deoxy-ribulose 5-phosphate

CM	Chorismate mutase
COMT	Caffeic acid 3-O-methyltransferase
CS	Chorismate synthase
cTP	Chloroplast targeting peptide
DAHP	3-deoxy-d-*arabino*-heptulosonate 7-phosphate
DDC	Duplication, Degeneration, Complementation
dN	Non-synonymous substitution rate
DQA	3-dehydroquinate
DQD	3-dehydroquinate dehydratase
DQS	3-dehydroquinate synthase
dS	Synonymous substitution rate
DSA	3-dehydroshikimate
E4P	Erythrose-4- phosphate
EAC	Escape from Adaptive Conflict
EB	Empirical Bayes
EPSP	5-enolpyruvylshikimate 3-phosphate
F5H	Ferulate 5-hydroxylase
G3P	Glyceraldehyde 3-phosphate
GEO	Gene Expression Omnibus
Gln	Glutamine
Glu	Glutamate
Gly	Glycine

GT	Glycosyltransferase
HCT	Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase
His	Histidine
HPLC	High Pressure Liquid Chromatography
HPP-AT	4-hydroxyphenylpyruvate aminotransferase
IAD	Innovation, Amplification, Divergence
IGPS	Indole-3-glycerol phosphate synthase
Ile	Isoleucine
IPTG	Isopropyl-1-thio-B-D-galactoside
JGI	Joint Genome Institute
Km	Michaelis-Menten constants
Lys	Lysine
MAS5	MicroArray Suite 5.0
MEME	Mixed Effects Model of Evolution
MYA	Million Years Ago
NAD+	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide, reduced form
NADP+	Nicotinamide adenine dinucleotide phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate, reduced form
NCBI	National Centre for Biotechnology Information
OMT	*o*-Methyltransferase

PAI	Phosphoribosylanthranilate isomerase
PAL	Phenylalanine ammonia-lyase
PAML	Phylogenetic Analysis by Maximum Likelihood
PAT	Phosphoribosylanthranilate transferase
PCC	Pearson correlation coefficients
PCR	Polymerase chain reaction
PDH	Prephenate dehydrogenase
PDT	Prephenate dehydratase
PEP	Phosphoenolpyruvate
Phe	Phenylalanine
Pi	Inorganic phosphate
PPA-AT	Prephenate aminotransferase
PPi	Inorganic diphosphate
PPO	Polyphenol oxidase
PPY-AT	Phenylpyruvate aminotransferase
PRPP	Phosphoribosyl pyrophosphate
PVDF	Polyvinylidene fluoride
QA	Quinate
QD	Quinate dehydratse/hydrolyase
QDH	Quinate dehydrogenase
SA	Shikimate
SDH	Shikimate dehydrogenase

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	Serine
SK	Shikimate kinase
SLAC	Single Likelihood Ancestor Counting
TAIR	The Arabidopsis Information Resource
Thr	Threonine
TIC	Translocon at the inner envelop membrane of chloroplasts
TOC	Translocon at the outer envelop membrane of chloroplasts
Trp	Tryptophan
TS	Tryptophan synthase
Tyr	Tyrosine
USER	The Uracil-Specific Excision Reagent
UV	Ultraviolet
Vmax	Maximal velocity value
YFP	Yellow Fluorescent Protein
α-KG	α-ketoglutarate

Acknowledgments

With this PhD dissertation, I am approaching the end of my challenging doctorial journey. It has been a truly life-changing experience. It would not be possible for me to come this far without help and support from many people.

First of all, I would like to say a big thank you to my supervisor Dr. Jürgen Ehlting for giving me the honor of being his very first student as well as allowing me the opportunity to carry out this project. He patiently provided advice and encouragement which are essential for me to proceed through my PhD program. He has also been very supportive and has given me the freedom to pursue independent work. Without his guidance, it would be almost impossible for me achieve this much.

Special thanks to my committee, Dr. Peter Constabel, Dr. John Taylor, Dr. Caren Helbing for their support, guidance and insightful suggestions. I would also like to thank Dr. Martin Boulanger for his help with protein structure modeling and interpretation, and Dr Raad Nashmi for allowing me to use his confocal laser scanning microscope and providing help whenever it was needed.

My deep appreciation goes out to the former and current members of the Ehlting lab and Constabel lab. Their friendship and support has meant a lot to me. I really think that I could not finish my work without support from all my dear friends. Very special thanks to Yuriko Carrington for being willing to continue working on this project and finish what I started.

I would also like to thank my parents, who have been working really hard and have sacrificed their lives for mine. Their unconditional love, understanding and support are

invaluable to me throughout my entire life. They always have faith in me even when I don't have faith in myself. I know that I can always count on them when things don't go well. Without them, it would not be possible for me to finish this journey.

Special thanks to my husband, Ning, and his wonderful family who have always been supportive and considerate. Ning is a great partner, who has loved me during my good times and bad times. The past few years have not been easy to me, both personally and academically. I truly appreciate that Ning has always been by my side no matter what happened.

Dedication

I would like to dedicate this dissertation to my family and my husband, Ning, for their love and support. I hope that this work makes you proud.

Chapter 1 Introduction

1.1 Land plant evolution and adaptation

The history of plant evolution is marked by innovations that have led to increasing levels of complexity. Colonization of the land by plants is one of the most significant evolutionary events in plant history. From there, fast radiations took place that resulted in the diversity of photosynthetic organisms (plants in particular) we see today. These innovations are of great importance in shaping the modern terrestrial ecosystem as well as allowing adaptive traits to be gained by plants living in a specific habitat. The evolution of these traits help plants thrive in their environments and become more successful compared with their ancestors.

1.1.1 A brief overview of land plant evolution

The *Viridiplantae* (green lineages) contains two major evolutionary groups: *chlorophyta* (green algae) and *streptophyta* (charophytes and embryophytes) (Figure 1.1). With different molecular clock methods, the separation of these two evolutionary lineages has been dated back to 1,200-725 MYA (Delaux *et al.*, 2012; Floyd and Bowman, 2007). It is suggested that land plants (embryophytes) have evolved from multi-cellular freshwater green algae, which are closely related to the extant charophyte group (Pires and Dolan, 2012; Lewis and McCourt, 2004, Lemieux *et al.*, 2007; Delaux *et al.*, 2012). The charophytic ancestry of land plants is supported by phylogenetic studies and some biochemical and structural characters shared by charophytes and land plants (e.g. cell division via a phragmoplast, lignin-like material in charophytic placenta, the

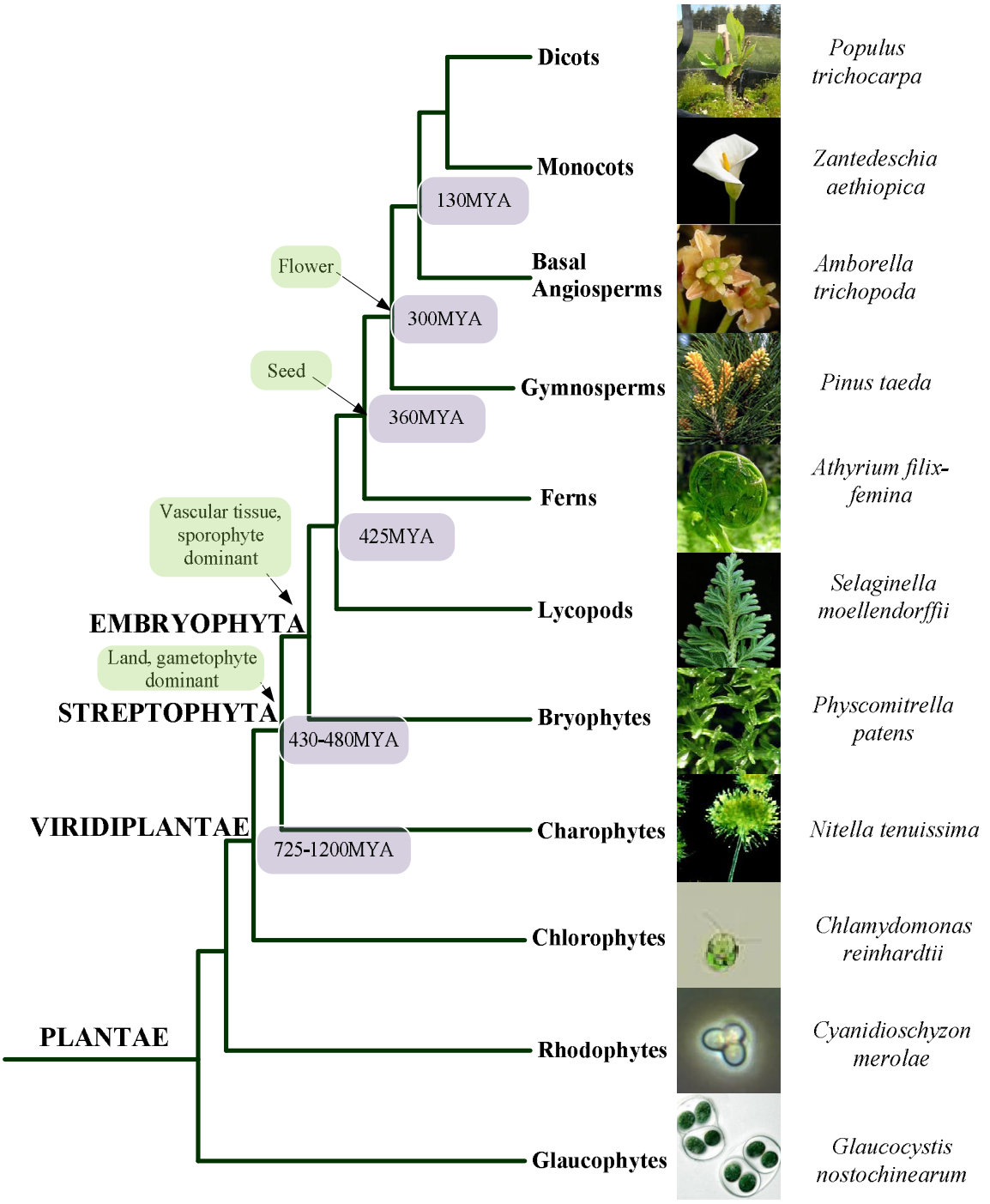


Figure 1.1: Evolution of the plant kingdom with significant innovations (green)**.** The plant kingdom can be subdivided into three groups: glaucophytes (fresh water algae), rhodophytes (red algae) and the green lineage (*Viridiplantae*) including chlorophytes (green algae), charophytes and embyrophytes (land plants). Pictures of sample species from each group are shown on the right. The estimated dates of some nodes are in purple.

Sources of the pictures in Figure 1.1:

Zantedeschia aethiopica http://www.prlog.org/10175218-less-stress-planning-weddings-with-fabulous-personalized-uniquely-designed-wedding-accessories.html
Amborella trichopoda
http://blogkarinaj.blogspot.ca/2011/01/la-clave-para-desentranar-el-origen-de.html
Pinus taeda
http://www.wildflower.org/collections/printable_QR_main.php?collection=NC
Athyrium filix femina
http://winemaking.jackkeller.net/request213.asp
Selaginella moellendorffi
http://genome.jgi-psf.org/Selmo1/Selmo1.home.html
Physcomitrella patens
http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html
Nitella tenuissima
http://www.kranswieren.nl/N.%20tenuissima.foto.html
Chlamydomonas reinhardtii
https://ncma.bigelow.org/files/strains/CCMP222_Chlamydomonas_CCMP222_d.jpg
Cyanidioschyzon merolae
http://en.wikipedia.org/wiki/Cyanidioschyzon_merolae
Glaucocystis nostochinearum
http://fmp.conncoll.edu/Silicasecchidisk/Pics/Other%20Algae/Blue_Green%20jpegs/Glaucocystis1.jpg

presence of sporopollenin, etc.) (Delaux *et al.*, 2012; Becker and Marin, 2009). Based on fossil records, the emergence of land plants, a key evolutionary event, occurred about 480-430 MYA (Delaux *et al.*, 2012; Strother *et al.*, 1996; Wellman *et al.*, 2003; Tomescu *et al.*, 2009). From this point on, land plant community further diversified to form the current complex terrestrial ecosystem. The discovery of fossil spore tetrads, which are also the oldest paleoecological evidence from land plants, suggests that the first land plant was closely related to extant liverworts (Wellman *et al.*, 2003; Pires and Dolan, 2012). The basal position of liverworts in land plants is further supported by phylogenomic analyses with sequences from bryophytes (liverworts, mosses and hornworts) and vascular plants (Qiu, 2008). Liverworts, along with mosses and hornworts, form the non-vascular group, which is characterized by the lack of vascular tissues, true roots and true leaves. Both gametophyte and sporophyte generations are present in bryophytes' life cycles with gametophyte phase being dominant (Delaux *et al.*, 2012). The sporophyte is largely dependent on the gametophyte. Due to the lack of vascular (water conducting) tissues and an indispensible role of water in spore production, the presence of most bryophytes is restricted to damper places (Pires and Dolan, 2012). Vascular plants only came to appear about 425 MYA (Gensel, 2008). During the Devonian period, land plant diversity was increased dramatically. Specializations and adaptations (e.g. the presence of highly specialized organs, sporophyte-dominant life cycle, water-independent reproduction, etc.) have been made to allow successful colonization of the terrestrial (drier) ecosystem by land plants. Around 300 MYA, the earth was covered by ferns, lycophytes and horsetails (Pires and Dolan, 2012). Gymnosperms also appeared during that period and then became dominant

between 260-70 MYA (Taylor *et al.*, 2009; Pires and Dolan, 2012). Major groups of angiosperms appeared during the early Cretaceous period and became dominant in terms of species diversity since the late Cretaceous period (100-65 MYA) (Pires and Dolan, 2012).

1.1.2 Plant terrestrialization and adaptation

Colonization of the terrestrial environment by land plants was an essential step towards the formation of modern terrestrial ecosystem. However, it was not an easy process. The early land environment was rather harsh. The ancestors of land plants had to face many challenges during their transition from the aquatic environment to land (Kenrick and Crane, 1997; Delaux *et al.*, 2012; Emiliani *et al.*, 2009; Lowry *et al.*, 1980), which included desiccation, extreme temperatures and temperature variations, UV irradiation, loss of support, attacks by microbes from the soil and air, etc. These challenges drove the accumulation of adaptive traits, which favored plant terrestrialization and contributed to successful conquest of the terrestrial environment by land plants.

1.1.2.1 Morphological aspects

Life cycles of land plants are characterized by two distinct generations: a sexual haploid gametophyte phase and an asexual diploid sporophyte phase. It is suggested by phylogenetic analyses that the gametophyte phase was inherited by land plants from their algal ancestor while the sporophyte evolved from a dependant of the gametophyte to a physiologically independent organism during the evolution of land plants (Kenrick and Crane, 1997; Pires and Dolan, 2012). Gametophytes were reduced during this process,

and this was coupled by organ differentiation in sporophytes. The appearance of highly specialized organs enabled plants to better adapt to the complex terrestrial environments. Humidity is relatively low in the earth atmosphere, which has led to increased chance of desiccation for colonizing plants. In order to adapt to the dry environment, land plants have evolved a distinct structure, a thin waxy covering (cuticle), which serves as a physical barrier against water loss. This layer also protects plants against xenobiotics and microbial infection. A cuticle is present in most extant land plants but is absent from green algae (Delaux *et al.*, 2012), which further illustrates its importance to terrestrial plants. In response to the desiccating terrestrial environment, land plants including many bryophytes have evolved some water conducting structures (e.g. hydrome in mosses). Increase in organismal complexity and size coincided with the evolution of vascular tissue with thickened cells walls. Deposition of lignin within the vascular tissue leads to extra strength, which allows long-distance water transport (Popper *et al.*, 2011). The presence of strengthened vascular tissue allows plants to grow upward against gravity as well as high-pressure water transport from roots to the top of a plant. Land plants have also developed specialized epidermal structures (stomata), which play a key role in preventing water loss via evapotranspiration and regulating carbon dioxide uptake for photosynthesis (Haworth *et al.*, 2011). By doing this, stomata help maintain a constant internal environment within a leaf. Stomatal structures were found to be highly conserved across extant plant species, which suggested a critical role of stomata in plant terrestrialization and adaptation (Edwards *et al.*, 1998). Colonization of the terrestrial environments by land plants required specialized organs involved in plant anchorage to a substrate as well as absorption of water and mineral nutrients from the surrounding

environments (Pires and Dolan, 2012; Delaux *et al.*, 2012). Root-like structures (rhizoids) were already present in early land plants (bryophytes). However, true rooting system only came to existence later in land plant evolution. In addition, symbiotic relationships were established between plant roots and the underground microbial community. These interactions facilitate the uptakes of water and nutrients, which are essential to plant overall fitness and successful adaptation to the terrestrial environments.

1.1.2.2 Genetic aspects

Plant genomes were subject to numerous local gene duplications as well as many whole-genome duplications. Genomic studies on a diverse set of plants suggests that most, if not all, seed plants have a polyploid ancestry (Beike and Rensing, 2010; Vandepoele *et al.*, 2003; Blanc and Wolfe, 2004; Cui *et al.*, 2006). It is interesting that there has been an increased number of genome doubling events occurring independently during the Cretaceous-Tertiary period (65 MYA), which was marked by massive species extinction followed by radiation (Fawcett *et al.*, 2009). Frequent fixation of polyploidization events among land plants during that period suggests that whole-genome duplications may provide an advantage during the extinction event to allow plants to better and faster adapt to the changing environments (Pires and Dolan, 2012; Delaux *et al.*, 2012; Rensing *et al.*, 2008). Whole-genome duplication, or gene duplications in general, is found to be responsible for the expansion of some gene families with regulatory roles in gene transcription, signal transduction as well as development (Sterck *et al.*, 2007; Pires and Dolan, 2012). It is estimated that 90 percent of transcription factors in Arabidopsis were created within the last 150 million years via gene duplication events (Maere *et al.*, 2005). Almost all the transcription factor classes found in angiosperms are

already present in early land plants (i.e. extant bryophytes) (Pires and Dolan, 2012), but only a small number of these are found in green algae (Riano-Pachon *et al.*, 2008; Richardt *et al.*, 2007). This suggests that the number of transcription factor gene families was increased in the early stage of land plant evolution. Although basal land plants and angiosperms share similar number of transcription factor gene families, the sizes of these families are quite different. Bryophytes have much smaller gene family sizes compared with angiosperms (Richardt *et al.*, 2007). Again, the evolution of these gene families may be driven by frequent gene duplication events, whole-genome duplications in particular. This has led to the regulatory diversity observed in land plants, which may allow them to adapt and acclimate to environmental changes during and after land colonization.

1.1.2.3 Biochemical aspects (plant secondary metabolism)

The ancestors of land plants faced many challenges during the water-to-land transition. Being sessile, plants are not able to escape from stresses the same way as animals can. Instead, plants have evolved alternative strategies to deal with challenges imposed by the surrounding environments and to allow them to thrive even under extreme conditions. One such strategy is to produce and release various chemical compounds (plant secondary metabolites) with potentially protective and defensive roles. It appears that biochemical innovations and metabolic modifications have played crucial roles in plant adaptation to the stressful terrestrial environments.

1.2 Plant secondary metabolism

Across all species, plants are able to produce over 200,000 secondary metabolites with relatively low molecular weights, which fall into three major classes:

phenylpropanoids, terpenoids, and nitrogen containing compounds. These compounds were considered to be metabolic wastes or detoxification products with no direct effect on plant growth or development for a long period of time (Hartmann, 2007). The functional aspects of plant secondary metabolites were largely neglected due to their non-essential roles to plant cellular survival. However, the significance of having these compounds in plants has been revisited, and the adaptive role of these compounds in chemical ecology has been put front and centre.

1.2.1 Roles of plant secondary metabolites in coping with abiotic stresses

Since moving onto land, plants are subject to high UV radiation, which could potentially lead to DNA damage, and subsequently reduced primary productivity and low survival rates (Popper *et al.*, 2011). In order to cope with UV stress, land plants have evolved sunscreening mechanisms by producing phenolic compounds (such as flavonols and sinapoyl esters), which absorb light within the UV-B region. These compounds serve as filters that protect land plants from being damaged by high level UV radiation in the terrestrial environment (Agati *et al.*, 2011; Emiliani *et al.*, 2013). Flavonols also function as free radical scavengers with a potential role in excess photoenergy dissipation and UV response (Agati *et al.*, 2009).

In the aquatic habitat, thalli of macroalgae are supported by buoyancy and there is no need for additional strength. Successful colonization of land by plants was constrained by challenges imposed by the terrestrial environments including gravity and water limitation. The ability of vascular plants to produce and deposit the phenolic polymer lignin within secondary cell walls provides adaptive advantages during plant

terrestrialization by increasing plant strength and water transport efficiency (Popper *et al.*, 2011). These traits also allow plants to grow taller, which reduces the possibility of being shaded and subsequently leads to increased primary productivity.

1.2.2 Plant secondary metabolites as defensive compounds

Plant secondary metabolites are suggested to have important ecological functions, which allow plants to compete and to survive in the changing environments. Plants are constantly challenged by pathogens. In response, plants have evolved several survival strategies during evolution, which include reinforcement of cell walls, activation of defense genes, the production of reactive oxygen species as well as the biosynthesis of antimicrobial chemicals (phytoalexins) (Ahuja *et al.*, 2011). The concept of phytoalexin was introduced about 70 years ago. It was based on the experimental evidence that potato tuber tissues with previous exposure to an incompatible *P. infestans* stain demonstrated induced resistance to a compatible strain (Ahuja *et al.*, 2011 refering toMuller and Borger, 1940). The induced production of a phytoalexin with anti-fungal activity was thought to be responsible for triggering the defense response, which protected plants against subsequent infection by the compatible stain. This long-term resistance (or systemic acquired resistance) involves communications between infected tissues and uninfected tissues facilitated by signaling molecules such as salicylic acid and jasmonic acid (Durrant and Dong, 2004). The reception of those signaling molecules causes global responses in a plant including the induction of genes encoding phytoalexin biosynthesis. Much recent efforts have been made to identify potential phytoalexins in different plant species and to elucidate the production and regulation of these compounds. A large number of phytoalexins have been characterized including some alkaloids, terpenoids,

indolic glucosinolates, indoles, and polyphenols (Ahuja *et al.*, 2011; Astani *et al.*, 2010, Bednarek *et al.*, 2009; Tsuji *et al.*, 1992; Timperio *et al.*, 2012). It is worth mentioning that the production of these compounds is induced upon microbial infection and serves as one fundamental defense mechanism against viruses, bacteria and fungi.

Interactions between plants and herbivores have also been intensively studied to determine the role of plant secondary metabolites in both constitutive and induced plant defenses against herbivores (Wimp *et al*., 2005; Newton *et al*., 2009). In response to herbivory, induced plant defense via plant secondary metabolites may have negative effects on the survival rate of herbivores (Agrawal, 2000; Poelman et al., 2010) either directly by killing/repelling (De Moraes *et al*., 2001), or indirectly by attracting predators and parasitoids of the herbivores (Schnee *et al*., 2006). Some plant secondary metabolites have toxic effects on herbivores including tissue damages and cyst induction, which can potentially lead to death (Iason, 2005). For example, pyrrolizidine alkaloids (PAs) are constitutive plant defense compounds against mammalian herbivores. PAs can be degraded in the guts of cattle to form pyrrols, which are highly toxic and can cause severe liver damage (Mattocks, 1986; Macel, 2011). There has been ample evidence supporting the role of volatile compounds in attracting herbivore enemies. For example, several terpenes (including linalool) can be produced by birch trees (*Betula pubescens*) upon infestation by caterpillars of *Epirrita. autumnata* (autumnal moth). The emission of these compounds can attract passerine birds that feed on the caterpillars (Mäntylä *et al*., 2008).

1.2.3 Plant secondary metabolites as signaling molecules in chemical ecology

Plant secondary metabolites also serve as signaling molecules with a role in mediating plant-animal interactions. Most plants are able to produce flowers and/or fruits of multiple colors, which are created by pigments. Pigments are molecules with the ability to reflect light of specific wavelengths. By doing so, pigments create colors (Grotewold, 2006), which function as visual signals in attracting animal pollinators and seed dispersers (Miller *et al*., 2001). Plant color pigments fall into four major groups: tetrapyrroles (chlorophylls), carotenoids, flavonoids (anthocyanins, flavones, flavonols) and betalains, which are responsible for color variations observed in nature (Miller *et al*., 2001). In addition to visual cues, many pollinators are able to locate a flower by its scent. The scent signal emitted by a host plant contains a complex blend of volatile compounds including terpenoids, aliphatic compounds and benzenoids, and variations in chemical composition can be used by plants to attract their specific mutualistic insects for pollination (Hossaert-McKey *et al*., 2010).

A role of plant secondary metabolites in mediating symbiotic interaction has also been identified. Experimental evidence suggests that legumes can produce and secrete specific flavonoids that function as signaling molecules in promoting the production of Nod factors in bacteria, which is critical to nodule formation during rhizobial symbiosis (Fisher and Long, 1992).

1.2.4 The role of plant secondary metabolites in mediating plant-plant interactions

Some plant secondary metabolites are also involved in mediating plant-plant interactions. Positive interactions between plants involve chemicals exuded from the

roots of one plant that can increase herbivore resistance in its neighboring plants (Bais *et al*., 2006). Aboveground, volatile compounds (terpenes and green leaf volatiles) released by one damaged plant are involved in the priming of defense response in undamaged leaves of the same plant or neighboring plants, which have not experienced pathogen attacks (Dudareva *et al*., 2007; Unsicker *et al*., 2009). Resource competition leads to negative interactions between plants, and allelopathy is one mechanism through which a plant is able to defeat its competitors. Phytotoxins such as coumarin can be produced and released by one plant in order to reduce the growth and survival of susceptible neighboring plants, which can lead to increased resource availability (Bais *et al*., 2006). The ability to produce allelopathic chemicals has become a major contributing factor to the remarkable invasive success of some exotic species, which have been causing severe economic problems and threatening ecological and biological diversity and integrity in North America (Callaway and Aschehoug, 2000, Grotkopp *et al*., 2002, Bais *et al*., 2003).

1.2.5 Medicinal use of plant secondary metabolites

There has been an increasing interest in using plant secondary metabolites as novel medical drugs, flavors and industrial materials. Historically, people were highly dependent on plants and animal parts for their survival and disease healing. Some of these traditional medicines have been tested and indeed proven to have medicinal functions. Plants can produce a diverse range of chemicals and have always been recognized as a rich source of bioactive compounds, which can be used either directly as drugs or indirectly as a basis for synthetic drugs (Hanson, 2003). For example, quinones (e.g. aloe-emodin, juglone, β-lapachol, plumbagin, shikonin, and thymoquinone), which

are plant-derived secondary metabolites, demonstrate pharmacological activities with anti-cancer properties (Lu *et al.*, 2013). Taxol (Paclitaxel), another plant-derived anti-cancer drug, is one of the most effective drugs used to treat breast cancer (Malik *et al.*, 2011). Much research has been dedicated to discovering drugs from plant natural products. According to a recent assessment, 60% of anti-cancer drugs are shown to be of a natural origin, and 75% of the drugs used to cure infectious diseases are structurally derived from plant natural products (Newman *et al.*, 2003).

1.2.6 Characteristics of plant secondary metabolites and evolution of plant secondary metabolism

Plants are able to produce a large array of plant secondary metabolites, which demonstrate an amazing chemical and structural diversity (Pichersky and Gang, 2000; De Luca and St Pierre, 2000). Notably, such diversity is derived from a much smaller set of compounds containing just a few core sets of structures, which are often intermediates or derivatives of primary metabolism. Although some plant secondary metabolites are wide-spread across the plant kingdom, most secondary metabolites appear to be restricted to a certain taxonomic group. For example, alkaloids are only metabolized in 10-15% of vascular plant species (Haslam, 1994). Unlike primary metabolism, the production of plant secondary metabolites is not a constitutive process but is limited to specific organs or developmental stages and may be induced by environmental cues (Hartmann, 1996; Haslam, 1994). Regarding the origin of plant secondary metabolic genes, which create such diversity, multiple mechanisms have been put forward. For example, an extensive phylogeny of phenylalanine ammonia lyase (PAL), the enzyme catalyzing the first step of the phenylpropanoid pathway, suggests that the ancestor of land plants acquired a PAL

via horizontal gene transfer (HGT) from symbiotic soil bacteria (Emiliani *et al*., 2009). More widespread, there has also been evidence suggesting that genes recruited from primary metabolic pathways may be involved in generating such diversity. A comparative genomic study of aquatic algae, *Physcomitrella patens* and vascular plants suggests that most plant gene families are highly conserved. The major groups of land plants have not invented many new gene families, but largely maintained the basic genetic toolbox inherited from their ancestors. However, lineage-specific variations in gene family size are obvious across the plant kingdom (Rensing *et al*., 2008, Flagel and Wendel, 2009). In addition, some gene families contain both primary and secondary metabolic genes (e.g. *O*-methyltransferases, cytochrome-P450-dependent monooxygenases, etc) (Lupien *et al*., 1999; Wang and Pichersky, 1999). Along with the fact that many plant secondary metabolites share core structural units derived from primary metabolism, this indicates that the metabolic diversity does not arise from the creation or acquisition of a new set of genes but instead by duplication of existing metabolic genes followed by innovation or specialization events (Flagel and Wendel, 2009). Gene duplication, which has been proposed as a central mechanism for gene functional diversification and adaptation, thus plays a crucial role in shaping the diversity in plant secondary metabolism.

1.3 Plant secondary metabolites derived from aromatic amino acids

The aromatic amino acids phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp) are essential to animal diets, which serve as basic protein building blocks. Besides their role in protein biosynthesis, these three aromatic amino acids also function as precursors of a large variety of secondary metabolites in plants, which play critical roles in plant

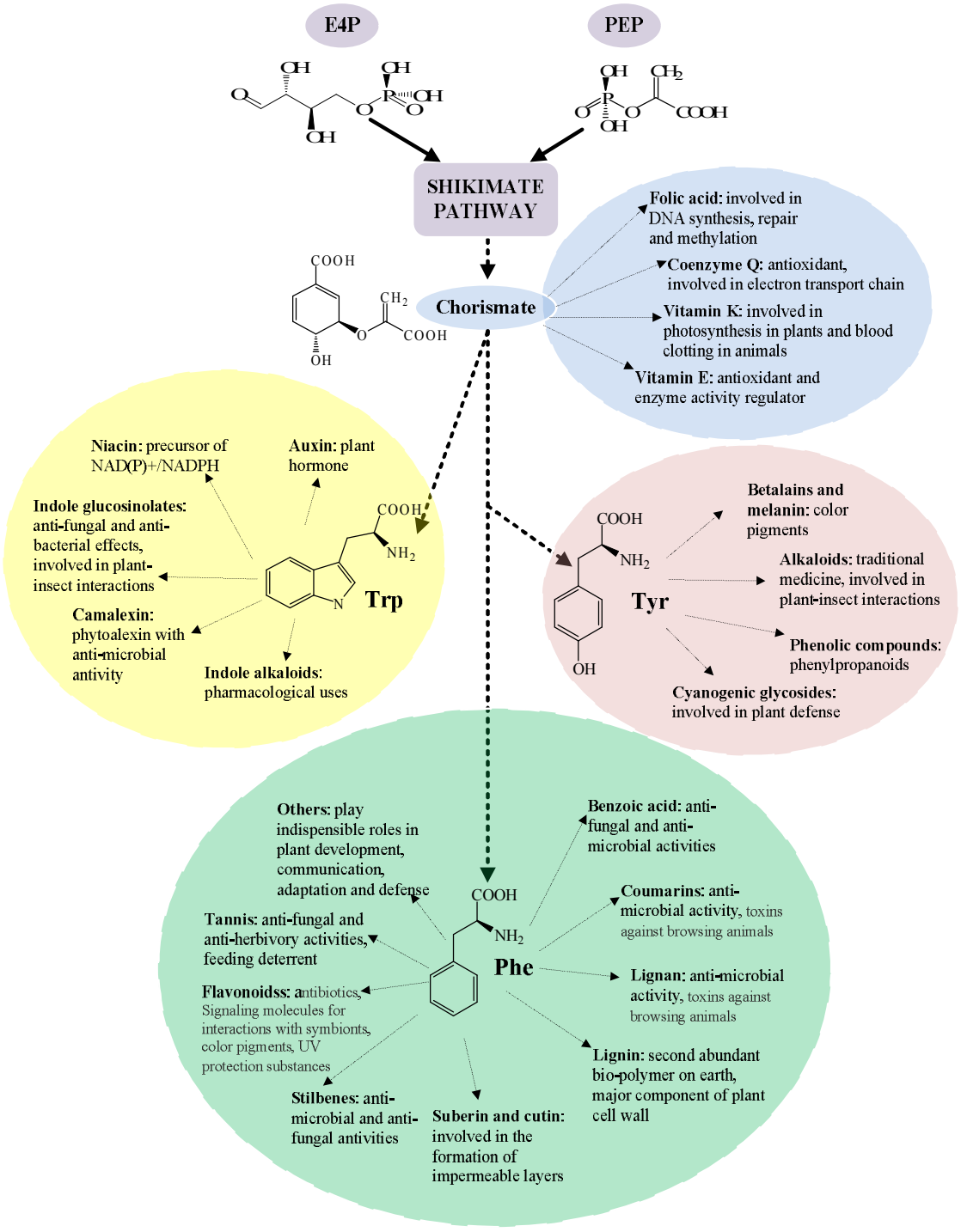

E4P
PEP
SHIKIMATE PATHWAY
Chorismate
Folic acid: involved in DNA synthesis, repair and methylation
Coenzyme Q: antioxidant, involved in electron transport chain
Vitamin K: involved in photosynthesis in plants and blood clotting in animals
Vitamin E: antioxidant and enzyme activity regulator
Niacin: precursor of NAD(P)+/NADPH
Auxin: plant hormone
Indole glucosinolates: anti-fungal and anti-bacterial effects, involved in plant-insect interactions
Camalexin: phytoalexin with anti-microbial antivity
Indole alkaloids: pharmacological uses
Trp
Betalains and melanin: color pigments
Alkaloids: traditional medicine, involved in plant-insect interactions
Phenolic compounds: phenylpropanoids
Cyanogenic glycosides: involved in plant defense
Tyr
Others: play indispensible roles in plant development, communication, adaptation and defense
Benzoic acid: anti-fungal and anti-microbial activities
Coumarins: anti-microbial activity, toxins against browsing animals
Tannis: anti-fungal and anti-herbivory activities, feeding deterrent
Lignan: anti-microbial activity, toxins against browsing animals
Flavonoidss: antibiotics, Signaling molecules for interactions with symbionts, color pigments, UV protection substances
Lignin: second abundant bio-polymer on earth, major component of plant cell wall
Stilbenes: anti-microbial and anti-fungal antivities
Suberin and cutin: involved in the formation of impermeable layers
Phe

Figure 1.2: Plant natural products derived from chorismate (the end product of the shikimate pathway) and the three aromatic amino acids (phenylalanine Phe, tyrosine Tyr and tryptophan Trp).
The shikimate pathway converts E4P (erythrose-4-phosphate) and PEP (phosphoenolpyruvate) into chorismate via 7 reactions, and chorismate is a common precursor for the biosynthesis of Phe, Try, Trp and some plant natural products (blue). Phe (green), Try (pink) and Trp (yellow) are further converted to a wide array of pivotal plant natural products which play important adaptive roles in plant development and defense. Some of these plant-derived chemicals are also essential to human diet and health (e.g. vitamins).

development, reproduction and defense (Figure 1.2) (Herrmann and Weaver, 1999; Weaver and Herrmann, 1997; Maeda and Dudareva, 2012; Dewick, 1995). Trp serves as a precursor of a large array of plant natural products including alkaloids, some phytoalexin, indoles, glucosinolates as well as the phytohormone auxin (Gunatilaka, 2008; Kroymann, 2011; Gibson *et al*., 1972; Radwanski and Last, 1995; Sanchez-Vallet, 2010). Tyr is involved in metabolizing alkaloids, cyanogenic glycosides, and the color pigments betalains (Kutchan, 1995, Koch *et al*., 1995; Dewick, 1995). Phe, being the largest carbon sink among the three aromatic amino acids, is a common precursor of a large group of chemicals collectively called phenolic compounds, which include the C6-C1 benzenoids and the C6-C3 phenylpropanoids, as well as their derivatives (Maeda and Dudareva, 2012; Vogt, 2010).

1.3.1 Phenylpropanoids

Phenylpropanoids, and more generally most phenolic compounds, are derived from phenylalanine. These compounds contain at least one benzenol ring in their chemical structures as a distinguishing characteristic. Phenylpropanoids have been investigated extensively for their important roles as protective and defensive compounds against both abiotic and biotic stresses, as signaling molecules in mediating plant defense, and as major structural components in plants (Vogt, 2010). These compounds can be involved in plant defense against pathogen (Dixon and Paiva, 1995) and herbivores (Felton *et al*., 1999). Plants can produce a large array of phenylpropanoids in response to abiotic stresses, such as drought, wounding (Bernards and Lewis, 1992), UV irradiation (Lois, 1994) and low temperature (Christie *et al*., 1994). Some phenylpropanoids also have important functions in chemical ecology including roles in attracting pollinators and seed

dispersers (Nishida *et al*., 2004; Weisshaar and Jenkins, 1998). Phenylpropanoid-based polymers, such as lignin and tannin, contribute substantially to the strength and robustness of tree species towards challenges imposed by their enemies and changing environments.

Phenylpropanoids are synthesized via the phenylpropanoid pathway. Enzymatic steps of this pathway are comparably well understood and documented in numerous species. The phenylpropanoid pathway converts phenylalanine, an end product of the shikimate pathway, to cinnamic acid by the action of phenylalanine ammonia lyase (PAL) (Figure 1.3) (Dixon and Paiva, 1995; Vogt, 2010). *p*-Coumaric acid is derived from cinnamic acid through a hydroxylation reaction catalyzed by cinnamate 4-hydroxylase (C4H), and is further converted into *p*-coumaroyl-CoA by 4-coumarate coenzyme A ligase (4CL). *p*-Coumaroyl-CoA is considered to be the central intermediate of the phenylpropanoid pathway. Further derivatization and modification of this intermediate contributes to the great diversity observed in this group of compounds.

1.3.1.1 Lignin

Among all phenolic compounds produced in plants, lignin, a complex polymer formed largely from *p*-coumaryl (H), coniferyl (G), and sinapyl (S) monolignols (Boerjan *et al*., 2003), is quantitatively the major carbon sink from the shikimate pathway. It accounts for 30% of the world's biomass, and it is the second most abundant biopolymer on earth (Weng and Chapple, 2010; Vogt, 2010). Lignin is deposited most abundantly in secondary cell walls in vascular plants, and it makes up about one third of the wood dry mass. It provides structural rigidity to the plant body which allows erect growth and high-

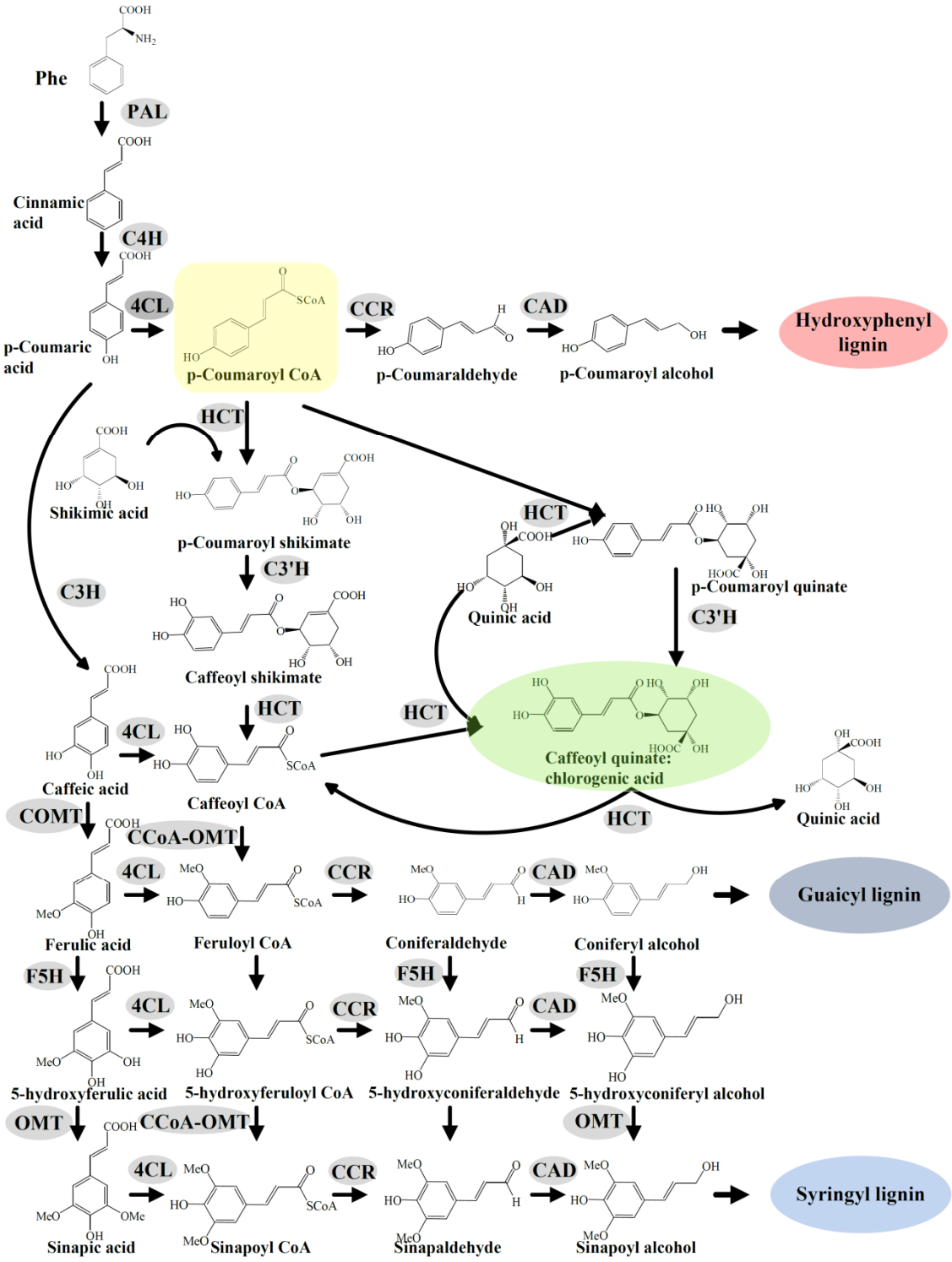

Phe
PAL
Cinnamic acid
C4H
p-Coumaric acid
4CL
p-Coumaroyl CoA
CCR
p-Coumaraldehyde
CAD
p-Coumaroyl alcohol
Hydroxyphenyl lignin
HCT
Shikimic acid
p-Coumaroyl shikimate
C3'H
C3H
Caffeoyl shikimate
HCT
Quinic acid
p-Coumaroyl quinate
C3'H
Caffeic acid
4CL
Caffeoyl CoA
HCT
Caffeoyl quinate: chlorogenic acid
HCT
Quinic acid
COMT
CCoA-OMT
Ferulic acid
4CL
Feruloyl CoA
CCR
Coniferaldehyde
CAD
Coniferyl alcohol
Guaicyl lignin
F5H
F5H
F5H
5-hydroxyferulic acid
4CL
5-hydroxyferuloyl CoA
CCR
5-hydroxyconiferaldehyde
CAD
5-hydroxyconiferyl alcohol
OMT
CCoA-OMT
OMT
Sinapic acid
4CL
Sinapoyl CoA
CCR
Sinapaldehyde
CAD
Sinapoyl alcohol
Syringyl lignin

Figure 1.3: The lignin and chlorogenic acid biosynthetic pathway.
This pathway starts with phenylalanine and converts it into p-Coumaroyl CoA (an important intermediate of this pathway) via three steps. From there, three types of monolignols and chlorogenic acid (green) are being produced. Monolignols further polymerize to form the three subunits of lignins: hydroxyphenyl lignin (pink), guaicyl lignin (purple) and syringyl lignin (blue).

Abbreviations:
4CL: 4-coumarate-CoA ligase; **C3'H**: p-Coumaroyl-CoA 3'-hydroxylase; **C3H**: 4-hydroxycinnamate 3-hydroxylase; **C4H**: Cinnamate 4-hydroxylase; **CAD**: Cinnamyl alcohol dehydrogenase; **CCoA-OMT**: Caffeoyl-coenzyme A O-methyltransferase; **CCR**: Cinnamoyl-CoA reductase; **COMT**: Caffeic acid 3-O-methyltransferase; **F5H**: Ferulate 5-hydroxylase; **HCT**: Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase; **OMT**: O-methyltransferase; **PAL**: Phenylalanine ammonia-lyase

tension water transport (Weng and Chapple, 2010). Lignin also plays a very important role in plant defense against herbivores and pathogens (Campbell and Sederoff, 1996). The biosynthesis of lignin monomers involves several hydroxylation and methylation steps of the aromatic ring and reduction of the propanoic acid moiety to the respective alcohols. It is suggested that rate-limiting steps regulating carbon flow into the lignin pathway are catalyzed by a series of hydroxylase encoded by genes from the cytochrome P450 monooxygenase family (C4H, C'3H, and F5H; Figure 1.3) (Vogt, 2010; Weng and Chapple, 2010). After being synthesized, monolignols are translocated to the cell walls of vascular tissue, where they are polymerized to form lignin (Vanholme *et al*., 2008). Lignin biosynthesis is a very complex process. Most of the knowledge has been derived from the study of enzymes involved in monolignol biosynthesis. However, some of the later steps (translocation and polymerization) have not been well demonstrated.

Global warming and pollution resulting from the burning of fossil fuels have been of great public concerns and have raised a particular interest in replacing the current fuel system with renewable and cleaner biofuels from plants (Chapple *et al*., 2007). It has been suggested that lignin quantity and composition are two factors that largely determine wood properties and have major impacts on pulping efficiency as well as biofuel production (Li *at el*., 2003). High lignin content hinders the access to and the degradation of polysaccharides (cellulose) in cell walls, which subsequently leads to lower fermentation rate of sugars to ethanol, a major product of fermentation based biofuel production. In order to overcome these difficulties, numerous of studies have been dedicated to gain better understanding of the lignin biosynthesis and its regulatory

processes in the hope of reducing energy costs as well as increasing yield in pulping and biofuel production (Boudet *et al.*; 2003; Ragauskas *et al*., 2006; Chen and Dixon, 2007).

1.3.1.2 Chlorogenic acid

Plants are able to produce tremendous amounts of non-lignin phenylpropanoids, among which chlorogenic acid (Figure 1.3) is of particular interest in the context of this study. Chlorogenic acids refer to a group of esters of hydroxycinnamic acids and quinate while the term chlorogenic acid is used specifically for an ester of caffeic acid and quinate (3-O-caffeoylquinic acid). The phenylpropanoid moiety of chlorogenic acid is an intermediate of the phenylpropanoid metabolism and quinate is a derivative of shikimate or 3-dehydroquinate. Thus, these two moieties derived from the end product and the intermediates of the shikimate pathway. Chlorogenic acid has been widely used as a nutraceutical with effects on preventing cancer development and lowering risk of acquiring cardiovascular diseases and type II diabetes (Morton *et al.,* 2000; Laranjinha *et al.,* 1994, Sawa *et al.,* 1999; Paynter *et al*., 2006). In addition, due to its antioxidant properties, chlorogenic acid is also used in many other fields, e.g. as food additives and in the cosmetic industry.

Chlorogenic acid is produced in many lineages across the plant kingdom, but as expected for secondary metabolites, is not detectable in all species (Petersen *et al.*, 2009). Notably, there appears to be no evidence for the existence of chlorogenic acid in the plant model system *Arabidopsis thaliana*. In contrast, chlorogenic acid accumulates to high levels in some plant species such as coffee, tobacco and poplar (Farah *et al*., 2008). As a naturally occurring phenolic compound in plants, chlorogenic acid also plays important

roles in plant adaptation and defense. Furthermore, its constituents can be re-mobilized to form other phenolic compounds with more complex structures including lignin, tannin, etc. (Boerjan *et al*., 2008, Tsai *et al*., 2006). Chlorogenic acid also plays roles in protecting plants against oxidative stresses (excess photoenergy) by acting as a direct scavenger of free radicals and a reducing agent for guaiacol peroxidase (Grace and Logan, 2000). Early research demonstrated that transgenic tobacco plants with suppressed levels of phenylalanine ammonia-lyase (PAL) had low levels of chlorogenic acid production and subsequently showed more rapid lesion development compared with wild type plants after being exposed to virulent fungal pathogen (Maher *et al*., 1994). Negative effects of chlorogenic acid on insect herbivores are also apparent (Beninger *et al*., 2004, Jassib, 2003, Dowd and Vega, 1996). The anti-herbivory property of chlorogenic acid relies on the ability of its derivative (chlorogenoquinone: an electrophilic molecule with high binding affinity towards amino groups of amino acids or proteins) to bind free amino acids and small peptides, which can potentially reduce the bioavailability of amino acids and small proteins. This can subsequently lead to starvation and reduction in herbivore fitness (Leiss *et al*., 2009). Finally, chlorogenic acid is claimed to have negative effects on the growth of bacteria, fungi and virus (Sung and Lee, 2010; Lou *et al*., 2005; Hoover *et al*., 1998; Ma *et al*., 2007; De Sotillo *et al*., 1998).

1.4 The shikimate pathway

All phenylpropanoids contain a core structure derived from the upstream pathway, which is commonly referred to as the shikimate pathway. The shikimate pathway connects primary metabolism with the biosynthesis of the three aromatic amino acids (Phe, Try and Trp) (Herrmann and Weaver, 1999; Maeda and Dudareva, 2012; Dewick,

1995). This pathway has only been found in microorganisms and plants, but is absent in animals (Roberts *et al*., 1998; Herrmann and Weaver, 1999). Bacteria and fungi primarily synthesize the three aromatic amino acids as essential protein building blocks. However, in plants, these three amino acids also serve as precursors of a large number of secondary metabolites with critical roles in plant survival and adaptation (see above; Hermann, 1995; Weaver and Herrmann, 1997, Herrmann and Weaver, 1999). It is estimated that over 30% of photosynthetically fixed carbon is directed through the shikimate pathway largely to plant secondary metabolism. Due to the absence of the shikimate pathway from animals, the three aromatic amino acids, which are the end products of this pathway, are essential to animal diets. It also makes this pathway a target for herbicides, live vaccines, antibiotics and anti-infectious drugs (Steinrucken and Amrhein, 1980; O'Callaghan *et al*., 1988; Zhang *et al*., 2005).

1.4.1 The seven enzymatic steps of the shikimate pathway

The shikimate pathway (Figure 1.4) begins with phosphoenolpyruvate (PEP) and erythrose-4-phosphate (E4P), and consists of seven enzymatic steps leading to the production of chorismate, which is a common precursor of the three aromatic amino acids. The first committed step of the shikimate pathway is the aldol condensation of PEP and E4P giving rise to 3-deoxy-D-*Arabino*-heptulosonate 7-phosphate (DAHP) and inorganic phosphate. This reaction is catalyzed by DAHP synthase (Herrmann and Weaver, 1999; Maeda and Dudareva, 2012; Dewick, 1995). Based on sequence and structural analysis, DAHP synthases fall into two distinct groups sharing only 10% protein sequence similarity (Dewick, 1995; Maeda and Dudareva, 2012). Type I

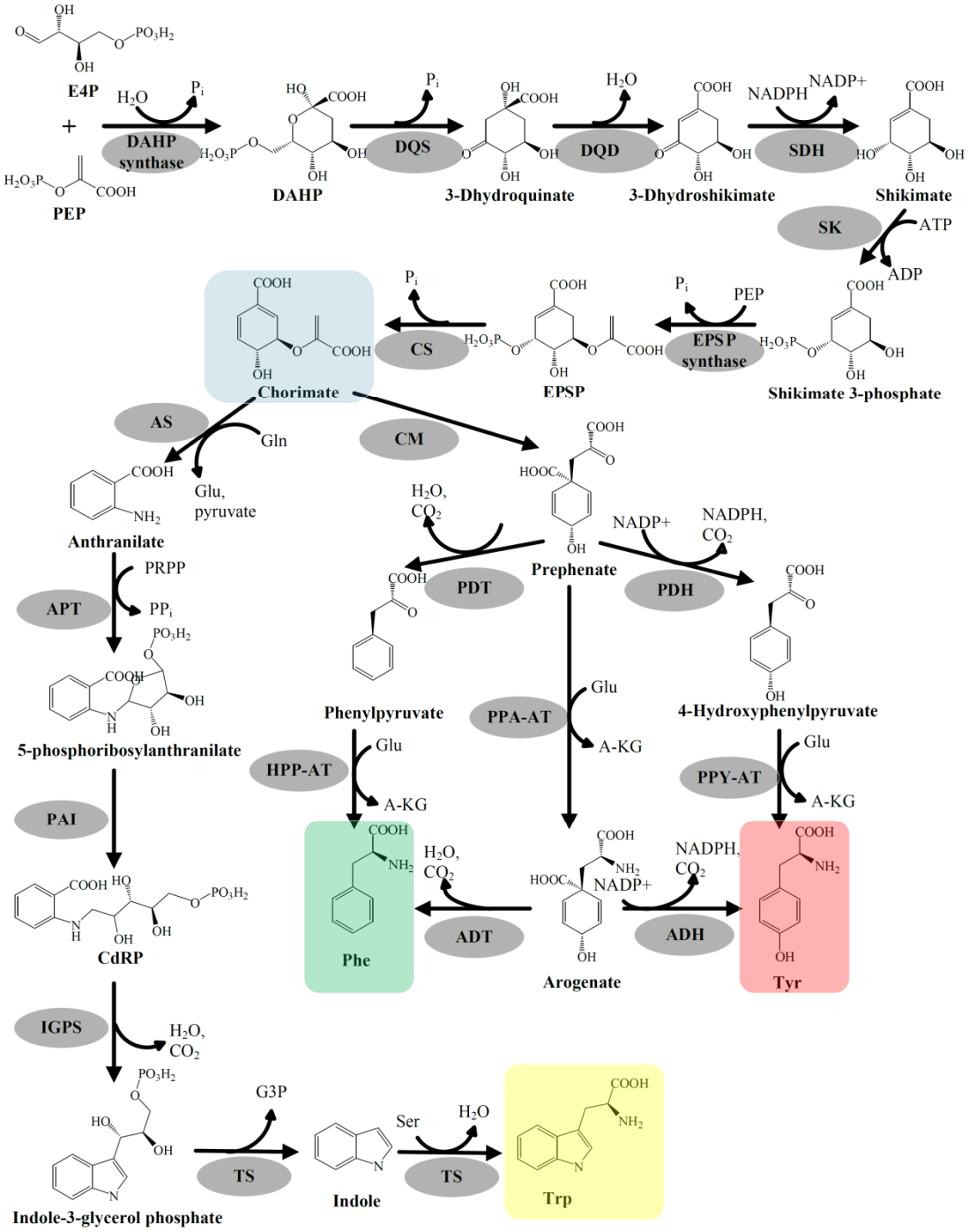


Figure 1.4: The shikimate pathway and its downstream aromatic amino acid pathways.
The shikimate pathway starts with PEP (from the glycolysis pathway) and E4P (from the pentose phosphate pathway) and converts these into chorimate (blue) via 7-step reactions. Chorismate is further converted into either anthranilate, a precursor of the tryptophan biosynthetic pathway, or prephenate, which can be then converted into phenylalanine and tyrosine (Maeda and Dudareva, 2013).

Abbreviations:
ADT: Arogenate dehydratase; **A-KG**: α-ketoglutarate; **APT**: Anthranilate phosphoribosyltransferase; **AS**: Anthranilate synthase; **CdRP**: 1-(o-carboxyphenylamino)-1-deoxy-ribulose 5-phosphate; **CM**: Chorismate mutase; **CS**: Chorismate synthase; **DAHP**: 3-deoxy-d-arabino-heptulosonate 7-phosphate; **DQD**: 3-dehydroquinate dehydratase; **DQS**: 3-dehydroquinate synthase; **EPSP**: 5-enolpyruvylshikimate 3-phosphate; **G3P**: Glyceraldehyde 3-phosphate; **Gln**: Glutamine; **Glu**: Glutamate; **HPP-AT**: 4-hydroxyphenylpyruvate aminotransferase; **IGPS:** Indole-3-glycerol phosphate synthase; **PAI**: Phosphoribosylanthranilate isomerase; **PDH:** Prephenate dehydrogenase; **PDT**: Prephenate dehydratase; **Pi**: Inorganic phosphate; **PPA-AT**: Prephenate aminotransferase; **PPi**: Inorganic diphosphate; **PPY-AT**: Phenylpyruvate aminotransferase; **PRPP**: Phosphoribosyl pyrophosphate; **SDH**: Shikimate dehydrogenase; **Ser**: Serine; **SK**: Shikimate kinase; **TS**: Tryptophan synthase

isozymes contain enzymes from bacteria (*Escherichia coli*) and fungi (*Saccharomyces cerevisiae; Neurospora crassa*) (Herrmann and Weaver, 1999; Dewick, 1995; Kunzler *et al*., 1992; Paravicini *et al*., 1988; Paravicini *et al*., 1989) while Type II DAHP synthases are found mainly in plants and some bacterial species including *Streptomyces rimosus* and *Mycobacterium tuberculosis* (Stuart and Hunter, 1993; Webby *et al*., 2005). 3-dehydroquinate synthase (DQS) catalyzes the second step of the shikimate pathway, which is the conversion of 3-dehydroquinate from DAHP via a sequence of reactions including alcohol oxidization, phosphate elimination, carbonyl reduction and aldol condensation (Dewick, 1995; Maeda and Dudareva, 2012; Bender *et al*., 1989). The third step of the shikimate pathway (dehydration of 3-dehydroquinate to 3-dehydroshikimate) is catalyzed by 3-dehydroquinate dehydratase (DQD). DQDs are differentiated by reaction mechanisms into two types: type I and II, which catalyze the syn- and anti-elimination of water, respectively (Shneier *et al*., 1993; Maeda and Dudareva, 2012). The fourth step of the shikimate pathway (reduction of 3-dehydroshikimate to shikimate) is catalyzed by an NADP-dependent shikimate dehydrogenase (SDH, AroE). Shikimate kinase (SK) catalyzes the fifth reaction of the shikimate pathway, which is the phosphorylation of shikimate to produce shikimate 3-phosphate using ATP (Whipp and Pittard, 1995; Maeda and Dudareva, 2012; Fucile *et al*., 2008; Fucile *et al*., 2011; Kasai *et al*., 2005). 5-Enolpyruvylshikimate 3-phosphate (EPSP) synthase catalyzes the sixth step of the shikimate pathway, which is the condensation of shikimate 3-phosphate and PEP to yield EPSP and inorganic phosphate. EPSP synthase is the primary target of the broad spectrum herbicide glyphosate, which acts as a competitive inhibitor with respect to phosphoenolpyruvate (Herrmann and Weaver, 1999). Glyphosate-tolerant crop plants

carrying genes encoding glyphosate-insensitive EPSP synthase (from either *Agrobacterium tumefaciens* or *Petunia hybrida*) have been successfully generated and are commercially used extensively (i.e. 'RoundUp Ready' crops). Chorismate synthase (CS) catalyzes the last step of the shikimate pathway, which is the elimination of phosphate from EPSP to produce chorismate. Although the overall reaction is redox neutral, reduced flavin is required, which functions as a transient electron donor in the phosphate elimination reaction. There are indeed two types of CS enzymes separated by their ability to reduce flavin (Maclean and Ali, 2003; Quevillon-Cheruel *et al*., 2004; Schaller *et al*., 1991).

The seven enzymes of the shikimate pathway are organized differently in different lineages. In bacteria, the seven steps are catalyzed by separate monofunctional enzymes. In contrast, steps two to six are catalyzed by a pentafunctional enzyme complex (AROM) in fungi. Similar to bacteria, most of the shikimate pathway enzymes in plants are monofunctional, but the third and fourth steps (conversion of 3-dehydroquinate to shikimate via 3-dehydroshikimate) are catalyzed by a bifunctional enzyme named dehydroquinate dehydratase/shikimate dehydrogenase (DQD/SDH).

1.4.2 Regulation of the shikimate pathway

Although the shikimate pathway is present in different domains of life, enzymes involved in this pathway are quite distinct in terms of modes of regulatory mechanism. As the enzyme catalyzing the first step of the shikimate pathway, DAHP synthase is likely to be a target for regulation of carbon flow into this pathway. In *E. coli* (or bacteria in general), there are three DAHP synthase isoforms, which demonstrate different

sensitivities towards feedback inhibition by the three aromatic amino acids. However, most plant DAHP synthases are not inhibited by aromatic amino acids with only a few exceptions (e.g. maize and pea) (Maeda and Dudareva, 2012). In Arabidopsis, there are two DAHP synthases, which are differently expressed and regulated. One of the two isoforms is constitutively expressed while the other is highly induced by wounding and pathogen infection (Keith *et al*., 1991). It was also found that the application of jasmonic acid can lead to an induction of genes encoding DAHP synthase in Arabidopsis plants (Devoto *et al*., 2005; Maeda and Dudareva, 2012). Tomato QDS is preferentially expressed in roots and is induced by the treatment with an elicitor in suspension culture (Bischoff *et al*., 1996). Spinach SK is regulated by energy charge (ATP, ADP and AMP concentrations) (Schmidt *et al*., 1990), which suggests that SK may be involved in maintaining a balance between cellular energy production and the energetic cost of the shikimate pathway. The Arabidopsis genome encodes two SKs. AtSK1 is extremely thermally stable while AtSK2 can be inactivated at 37 ºC. The appearance of a thermal-stable SK in Arabidopsis is thought to be an adaptive trait that allows maintaining flux through the shikimate pathway under heat stress (Fucile *et al*., 2011). Plant cDNAs encoding EPSP synthases have been isolated from a few plant species (Klee *et al*., 1987; Gasser and Klee, 1990; Gasser *et al*., 1988; Shah *et al*., 1986). It seems that EPSP synthases are expressed constitutively at low levels, and their expressions are tissue specific and developmentally regulated. Plant CSs were found to be differentially expressed in various parts of plants and in response to fungal elicitation (Gorlach *et al*., 1995; Gorlach *et al*., 1994; Gorlach *et al*., 1993; Braun *et al*., 1996). In microorganisms, regulations of the shikimate pathway enzymes are most likely to occur post-

transcriptionally via feedback inhibition. However, the shikimate pathway in plants is most likely to be regulated at the transcriptional level in response to both biotic (wounding, herbivory and pathogen infection) and abiotic (light intensities, nitrogen and amino acid starvation, etc.) stresses.

1.4.3 Subcellular localization of the shikimate pathway

It is suggested that plastids contain a complete set of enzymes involved in the shikimate pathway (Bickel *et al*., 1978; Bagge and Larsson, 1986; Benesova and Bode, 1992; Homeyer and Schultz, 1988; Mousdale and Coggins, 1985; Mousdale and Coggins, 1986; Klee *et al*., 1987; Schmidt *et al*., 1992). Based on sequence analysis, many shikimate pathway enzymes have extensions at their N-terminuses, which encode potential chloroplast targeting peptides (cTPs). The plastidial localization of the complete shikimate pathway is further supported by experimental evidence including detection of enzymatic activities in the plastidial fractions of plant extracts, detection of radio-labeled aromatic amino acids in isolated chloroplasts fed with labeled precursors (e.g. PEP), GFP (green fluorescent protein) fusion protein experiments and protein import assays (Maede and Dudareva, 2012; Ding *et al*., 2007; Della-Cioppa *et al.,* 1986; Kasai *et al.,* 2005; Zhao *et al.,* 2002). Identification of cytosolic isoforms for some shikimate pathway enzymes (DAHP synthase, DQD/SDH and EPSP synthase) has led to the debate on whether or not another set or at least a subset of the shikimate pathway enzymes are present outside of plastids. The shikimate pathway plays critical roles in both protein biosynthesis (primary metabolism) and providing precursors for the biosynthesis of downstream secondary metabolites. This leads to the proposal of a dual pathway hypothesis that the plastidial shikimate pathway is responsible for the production of

amino acids for protein production, while the shikimate pathway located in the cytosol is responsible for the production of precursors of secondary metabolites, which may be involved in plant defense (Herrmann, 1995; Weaver and Herrmann, 1997; Schmid and Amrhein, 1995). This hypothesis was further supported by the discovery that genes encoding Arabidopsis cytosolic DAHP synthase are induced by wounding and pathogen infection (Keith *et al.*, 1991; Maede and Dudareva, 2012) and that all arogenate dehydrogenase and arogenate dehydratase isoforms, which catalyze the ultimate steps of tyrosine and phenylalanine biosynthesis, respectively, are located solely in plastids (Rippert *et al.,* 2009). Although cytosolic isoforms of some shikimate pathway have been identified, it still remains unclear if a whole set of shikimate pathway enzymes are present in the cytosol.

1.5 Shikimate biosynthesis

Shikimate is a central intermediate of the shikimate pathway. Shikimate biosynthesis, as a part of primary metabolism, is essential. This is highlighted by the embryo lethal phenotype of an Arabidopsis mutant (*emb3004*) lacking the shikimate biosynthetic gene 3-dehydroquinate dehydratase/shikimate dehydrogenase (*DQD/SDH*) (Meinke *et al*., 2008). This implies that shikimate biosynthetic activity has been maintained by strong purifying selection pressure during plant evolution. Besides an absolute need of shikimate in aromatic amino acid biosynthesis for protein production, shikimate is required both as a precursor and as a cofactor in caffeoyl-CoA synthesis, which is the first committed step of G- and S-monolignol biosynthesis in the phenylpropanoid pathway (Boerjan *et al*., 2008). For this, shikimate may function as a negative regulator to ensure sufficient

carbon flow into protein biosynthesis (primary metabolism) by limiting the biosynthesis of G- and S-lignin when cellular shikimate concentration is relatively low.

The shikimate pathway contains many branches, and almost all the intermediates of this pathway are potential branch points leading to other metabolic pathways (secondary metabolism) (Bentley and Haslam, 1990). Shikimate may function as a branch point leading to the biosynthesis of cyclohexane carboxylates (e.g. ansatrienin and asukamycin) in bacteria, which are antibiotics with potential anticancer properties (Bentley and Haslam, 1990). The CoA-ester of cyclohexane carboxylate also functions as a potential starting material for the elongation of polyketides, which can be further derivatized and modified into bioactive secondary metabolites (Bentley and Haslam, 1990). In addition, shikimate can also be esterified with hydroxycinnamic acids to form hydroxycinnamoyl shikimate, one of which is an intermediate of the lignin biosynthetic pathway while others may serve roles in plant chemical ecology.

1.5.1 3-Dehydroquinate dehydratase/shikimate dehydrogenase (DQD/SDH)

In plants, shikimate biosynthesis from 3-dehydroquinate via 3-dehydroshikimate (Fig. 1.5) is catalyzed by the bifunctional DQD/SDH structurally resembling a fusion protein of the two distinct enzymes found in bacteria (AroD: DQD and AroE: SDH). DQD/SDHs have been functionally characterized in a few plant species including *Arabidopsis thaliana* (Arabidopsis; Singh and Christendat, 2006), *Solanum lycopersicum* (tomato; Bischoff *et al*., 2001) and *Nicotiana tabacum* (tobacco; Bonner and Jenson, 1994; Ding *et al*., 2007). The crystal structure of Arabidopsis DQD/SDH has been established, which allows identification of key amino acid residues involved in catalysis

and substrate binding of the DQD and SDH domains (Singh and Christendat, 2006). The two active sites of Arabidopsis DQD/SDH are localized in close proximity, allowing efficient channeling of 3-dehydroshikimate (the product of DQD and the substrate for SDH) from the DQD domain to the SDH domain, which is essential to efficient SDH activity (Singh and Christendat, 2006). It was also found that SDH activity is much higher than DQD activity (Ding *et al*., 2007), which ensures directional carbon flow through the shikimate pathway. Arabidopsis only contains one DQD/SDH, while most other plants contain multiple family members. For example, tobacco contains at least two isoforms (NtDQD/SDH 1 and 2). NtDQD/SDH1 is localized to chloroplasts. In contrast, NtDQD/SDH2 is localized to the cytosol, and its SDH activity is largely reduced in comparison with NtDQD/SDH1. The role of the cytosolic DQD/SDH in tobacco remains unknown, but it was suggested that this enzyme might be involved in a similar reaction with a different substrate (Ding *et al*., 2007).

1.6 Quinate biosynthesis and the shikimate/quinate cycle

Quinate, which shares high structural similarity with shikimate, is widely distributed across the plant kingdom, particularly in gymnosperms as well as herbaceous and woody angiosperms (Yoshida *et al*., 1995; Boudet, 1973). Quinate concentration can reach up to 14% of leaf dry mass in the developing tissues of some plant species (Leuschner *et al*., 1995; Ossipov *et al.,* 2000). In addition, quinate esters with phenylpropanoid acids (i.e. chlorogenic acids) are also common and widespread (see above). The distribution of quinate exhibits tissue specificity and seasonal fluctuation (Osipov and Shein 1986; Marsh *et al.,* 2009). Quinate accumulates to high levels (10-14%) in actively growing needles of some coniferous species such as *Pinus sylvestris* and

Larix sibirica but levels drop off as needles mature in later season (Ossipov *et al.,* 1995). In contrast, the concentration of quinate in vascular tissues (developing xylem) is extremely low (0.5%), and this is the place where phenolic compounds, lignin in particular, are synthesized (Ossipov *et al.,* 1995). It was found that quinate can be transported in phloem (Gora *et al*., 1994). Together, this suggests that quinate may function as a mobile carbon source for the biosynthesis of a large number of phenolic compounds including lignin: Quinate accumulates in early growing season in leaves as a result of high photosynthetic activity, and it is later mobilized to increase or maintain phenylalanine production to satisfy the demands for phenolic compounds (Ossipov and Shein 1986). The role of quinate in the production of phenolic compounds is further supported by the discovery of an increased level of quinate in the leaf phloem sap of beech seedlings in response to herbivory. It is speculated to act as a transported precursor of procyanidins, which are defense-related phenolic compounds in beech (Gora *et al*., 1994). Quinate may also have a role in modulating osmotic potential during the development of tolerance against drought (Gebre *et al.,* 1997). Finally, as described above, quinate can be esterified with caffeic acid and other hydroxycinnamates to form chlorogenic acids, which serve defensive and adaptive roles in plants (Grace and Logan, 2000). Quinate, together with its derivatives (e.g. chlorogenic acids), is also responsible for bitterness and astringency in plants and plant extracts such as coffee (Vaast *et al*., 2006).

1.6.1 Quinate biosynthesis

Quinate catabolism has been well studied in bacteria and fungi, where quinate is utilized as a carbon source. In bacteria, quinate can be channeled back to the shikimate

pathway and used for aromatic amino acid production. This reaction is catalyzed by shikimate/quinate dehydrogenase (YdiB). However, quinate metabolism in plants remains largely unknown. There has been some older evidence supporting the idea that quinate functions as a reserve compound of the shikimate pathway that can be derived from the intermediates of the shikimate pathway (Minamikawa *et al*., 1969; Yoshida, 1969; Boudet, 1980; Gamborg, 1967, Weinstein *et al*., 1961; Goldschmidt and Quimby, 1964). Based on a ^{14}C-labeled substrate feeding assay, an alternative pathway responsible for quinate biosynthesis, independent of the shikimate pathway, was proposed. It is based on the observation that more ^{14}C-labeled quinate was detected when using ^{14}C-labeled glucose 6-phosphate as a substrate compared with using ^{14}C-labeled 3-dehydroquinate (Boudet, 1980). However, this pathway has not been further investigated. Quinate can also be derived from the intermediates of the shikimate pathway by three enzymes. Quinate may be converted from 3-dehydroshikimate via 3-dehydroquinate by a bifunctional enzyme 3-dehydroquinate dehydratase/quinate dehydrogenase (DQD/QDH). This is supported by the purification of a DQD/QDH from corn seedlings (Graziana *et al*., 1980). Quinate can also be formed in a single-step reaction from either 3-dehydroquinate or shikimate catalyzed by quinate dehydrogenase (QDH) and quinate dehydratase (QD; also referred to as quinate hydrolyase, QH), respectively (Figure 1.5) (Herrmann, 1995). Both reactions have only been characterized biochemically. Genes encoding these two enzymes have not been identified to date.

1.6.1.1 Quinate dehydrogenase (QDH)

Quinate dehydrogenase (QDH) catalyzes the NAD(P)-dependent oxidation of quinate to 3-dehydroquinate, and its reaction mechanism resembles that of SDH

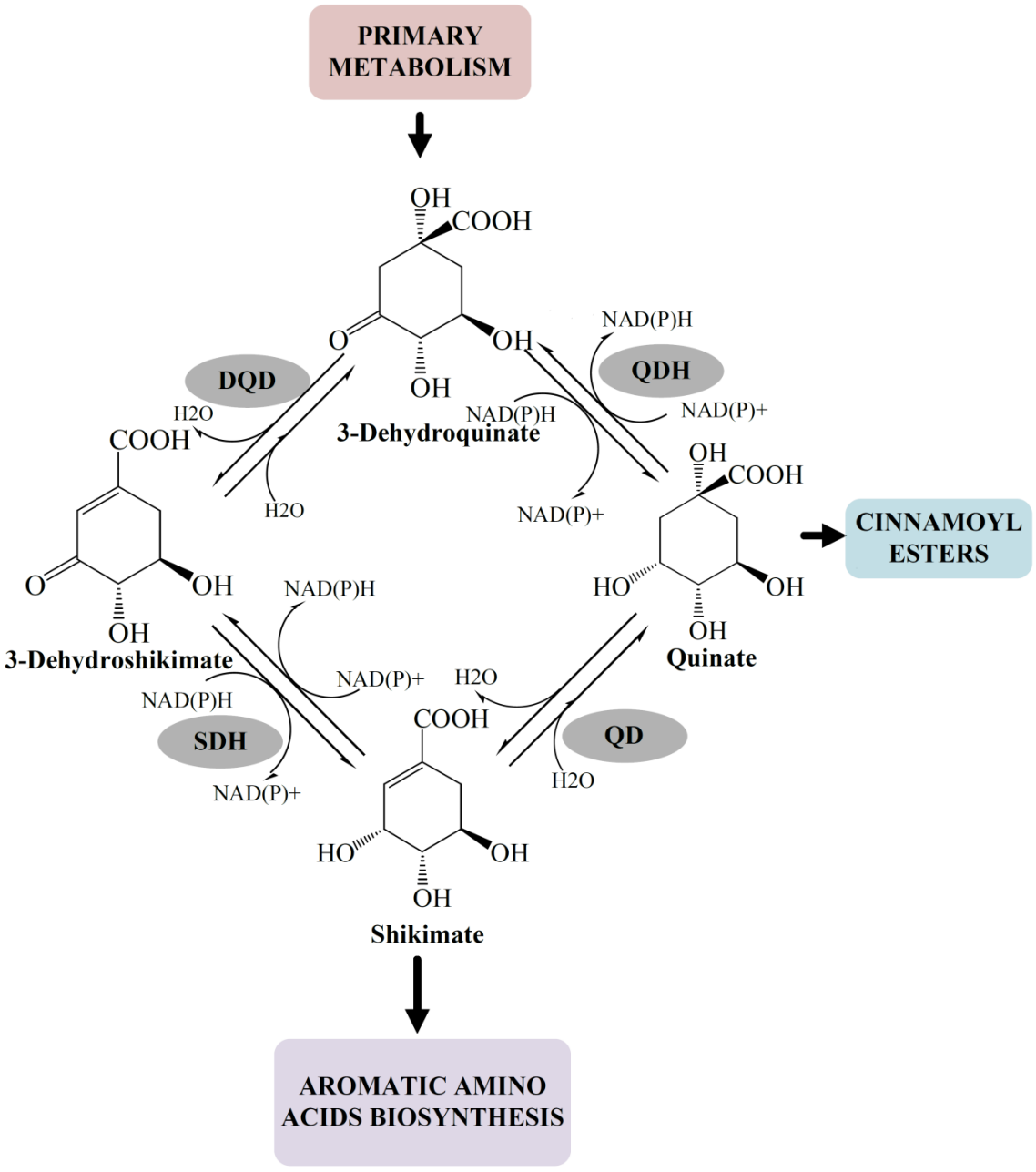


Figure 1.5: Schematic view of the shikimate/quinate cycle.
The conversion of shikimate from 3-dehydroquinate via 3-dehydroshikimate is catalyzed by a bi-functional enzyme: 3-deydroquinate dehydratase/shikimate dehydrogenase (DQD/SDH). Quinate can be synthesized from both 3-dehydroquinate and shikimate, and these two reactions are catalyzed by (quinate dehydrogenase) QDH and (quinate dehydratase) QD, respectively.

catalyzing the fourth step of the shikimate pathway (Figure 1.5). QDH enzymes have been partially purified and functionally characterized in a few plant species including corn, carrot, mung bean, pea and some coniferous species (Gamborg, 1966; Boudet, 1980, Kang and Scheibe, 1993; Ossipov *et al.,* 1995; Ossipov *et al.,* 2000; Minamikawa, 1977; Refeno *et al.,* 1982; Graziana and Boudet, 1983). It was suggested that QDH from carrot suspension culture is activated by phosphorylation (Refeno *et al.,* 1982). Later, a Ca^{2+} and calmodulin-dependent kinase was identified to be responsible for the phosphorylation (activation) of QDH from carrot (Ranjeva *et al.,* 1983). However, these enzymatic properties appear to be specific to carrot since QDH from mung bean sprouts does not respond to NaF (a phosphatase inhibitor, which promotes phosphorylation) or Ca^{2+} (Kang and Scheibe, 1993). Variations in cofactor preference are also observed across different plant species. QDHs from angiosperms such as bean, pea, carrot and corn prefer NAD as a cofactor (Kang and Scheibe, 1993; Minamikawa, 1977, Refeno *et al.,* 1982; Graziana and Boudet, 1983), while QDHs from gymnosperms (*Larix sibirica, Pinus taeda*) are more specific for NADP (Ossipov *et al.,* 1995; Ossipov *et al.,* 2000). A plastidial localization is suggested for QDH from mung bean (Kang and Scheibe, 1993). However, subcellular localizations of QDHs from other plant species remain largely enigmatic.

1.6.1.2 Quinate dehydratase (QD)

Quinate dehydratase (QD) can also be involved in quinate biosynthesis. It catalyzes the hydration of shikimate to form quinate. The reaction mechanism of QD resembles that of DQD, the third enzyme of the shikimate pathway (Figure 1.5). This reaction has been observed in mung bean plants (Minamikawa, 1977), and it has only been partially

purified from pea roots (Leuschner *et al*., 1995). A localization study suggested that QD from pea is localized to plastids.

1.7 Molecular evolutionary models

Some evidence suggests that the 3-dehydroquinate dehydratase/shikimate dehydrogenase (DQD/SDH) superfamily contains members involved in both shikimate and quinate biosynthesis. Based on biochemical data from partly purified enzyme fractions, some members of this family have exclusive specificity towards shikimate (Ding *et al*., 2007, Singh and Chritendat, 2006), while some have been experimentally proven to be able to take both shikimate and quinate as substrates (suggesting gene sharing of both functions) (Ossipov *et al.,* 2000; Lindner *et al.,* 2005). These suggest that genes involved in quinate metabolism (secondary metabolism) may have evolved from shikimate biosynthetic (primary metabolism) genes via gene duplication. Gene duplication has been proposed as a central mechanism for gene functional diversification and adaptation, which plays important roles in shaping the diversity in plant secondary metabolism. After gene duplication, functional variations between duplicated genes can arise in different ways, which include accumulation of mutations leading to the gain of a new function, separation and optimization of ancestral functions and changes in gene dosages (Conant and Wolfe, 2008).

1.7.1 Neofunctionalization

The process of neofunctionalization starts with a gene duplication event, which gives rise to two functionally redundant copies performing exactly the same function as their ancestor. Functional redundancy might lead to relaxed selection constraint on one of

the two daughter duplicates, which allows a new function to be gained. In the meanwhile, purifying selection acts on the other copy which conserves the original function. Assuming the novel function is advantageous, duplicate fixation can occur and the new function is then fixed in a population (Conant and Wolfe, 2008; Innan and Kondrashov, 2010; Hughes, 1994). An example of functional innovation after gene duplication is provided by the study of two Arabidopsis *CYP98* genes (*CYP98A8* and *CYP98A9*) (Matsuno *et al.*, 2009). Protein sequence analysis demonstrates that CYP98A8 and CYP98A9 share about 50% similarity with CYP98A3 (C3'H), which was previously shown to catalyze a meta-hydroxylation reaction in lignin biosynthesis. It was then suggested by an evolutionary study that CYP98A3, CYP98A8 and CYP98A9 may have evolved from a common ancestor, and that CYP98A8 and CYP98A9 have evolved recently under positive selection (Matsuno *et al.*, 2009). Both CYP98A8 and CYP98A9 are expressed predominantly in inflorescence tips, young flower buds, and stamen whereas CYP98A3 is highly expressed in roots, stems and flowers. UPLC-MS/MS analysis of the *CYP98A8/9* overexpression and knock-out mutant plants indicates that these two genes may be involved in metabolizing *N*1, *N*5-di(hydroxyferuloyl)-*N*10-sinapoylspermidine, which is known to exist only in pollen and may play a role in pollen formation. The gain of the new function and new expression profiles is likely to be a recent evolutionary event, which has occurred after gene duplication.

1.7.2 Subfunctionalization

Subfunctionalization, in contrast, begins with the presence of a multi-functional ancestor. Upon gene duplication, these multiple functions are subdivided and fixed separately among daughter duplicates (Conant and Wolfe, 2008; Innan and Kondrashov,

2010; Force *et al*., 1999). It has to be kept in mind that subfunctionalization is a rather general idea that covers a diverse range of mechanisms.

1.7.2.1 Duplication, Degeneration, Complementation (DDC) model

In DDC, a special case of subfunctionalization, neutral mutations accumulate after gene duplication independently among the duplicated copies. This results in the removal of different subsets of original functions in different daughter duplicates (degeneration). The loss of a subset of functions in one daughter duplicate is however complemented by the presence of this subset of functions in another duplicated copy (complementation). By working together, the ancestral task can still be accomplished by several enzymes encoded by functionally specialized genes instead of one multi-functional enzyme (Hahn, 2009; Conant and Wolfe, 2008). The DDC model has been found to be responsible for functional diversification in a few cases (Force *et al*., 1999, Huminieki and Wolfe, 2004, Semon and Wolfe, 2008). It is noteworthy that DDC does not involve optimization of the original functions, but describes solely a redistribution of functions present in one gene originally onto two daughter copies.

1.7.2.2 Escape from Adaptive Conflict (EAC) model

EAC has been proposed to be an alternative model, in which the ancestor gained a new function while maintaining the old function at the same time; thereby a multi-functional ancestor emerges. If the two functions compete with each other (e.g. they both employ the same active site of the encoded protein), any optimization of one function will be at the expense of the other. This creates opposing selection pressure (adaptive conflict) and results in both functions being sub-optimal. This pleiotropic conflict can

only be resolved after gene duplication. Each of the daughter duplicates evolves independently and becomes specialized in one of the two original functions (Sikosek *et al*., 2012; Des Marais and Rausher, 2008; Huang *et al*., 2012). Being different from DDC, the daughter duplicates undergo adaptive changes (non-neutral mutation) after gene duplication to allow functional specialization and optimization (Conant and Wolfe, 2008). An example of functional diversification via EAC was proposed by Des Marais and Rausher (2012), who worked with dihydrofavonol reductase enzymes in morning glories. They suggested in their paper that two consecutive gene duplication events have occurred giving rise to three genes. Strong positive selection was detected before the second duplication, and enzymes encoded by genes derived from this duplication have largely lost the ancestral functions. In contrast, the enzyme encoded by the third gene, which has diverged before the second duplication, demonstrated optimized ancestral functions. They argued in their paper that the three genes have diverged into two distinct functional groups with one being specialized for the ancestral functions and the other two being specialized for other function, which was not defined or tested in their study. Since the second function has not been identified, it is impossible to demonstrate that this function is also present in the ancestor. As a result, whether or not this gene family has evolved via EAC is still questionable.

1.7.3 Dosage effects

In this model, gene duplication occurs after a secondary function is gained by a gene and becomes beneficial. Gene duplication can provide an immediate advantage by increasing the protein level (dosage), which may compensate for the low efficiency of the novel function (Soskine and Tawfik, 2010). As a result, duplicated copies are preserved

by selection in a population. In some cases, mutations accumulate later among duplicates, which may lead to an optimization of the second function or the gain of another function in addition to the two original functions. The dosage selection model served as a basis for a new model named the Innovation, Amplification, Divergence (IAD) model (Soskine and Tawfik, 2010; Conant and Wolfe, 2008).

1.7.3.1 Innovation, Amplification, Divergence (IAD) model

Before duplication, the ancestral gene has a promiscuous side activity in addition to its original function (innovation). This secondary function may later become beneficial to the host and is then favoured by natural selection to increase its copy numbers by (multiple) gene duplication(s) (amplification). Besides the dosage effect resulted from increased amplification frequency driven by selection, this may also lead to relaxed constraints to maintain the original function in the additional copies. These copies are free to accumulate mutations that could improve the novel function. Once one copy is able to show significant improvement of the new function, selection pressure on the other extra copies to optimize the new function is relaxed leaving one copy responsible for all of the new function in the end (Bergthorsson *et al*., 2007; Nasvall *et al*., 2012). Due to the requirement of frequent amplification by IAD, this model was originally proposed for microbes, and most of the supporting evidence is from bacteria. It still remains open if this model can be used to describe functional innovations in more complex organisms like plants.

1.8 *Populus trichocarpa* as a model tree species

Populus trichocarpa was the first tree species that had its genome completely sequenced (Tuskan *et al*., 2006), and it was chosen as a model tree species for several reasons: it has a relatively small genome size (423Mb); it has a rapid growth rate and can reach reproductive maturity in four to six years; it is relatively easy to propagate; there is a large collection of genomic resources and well-established molecular toolbox available (Jansson and Douglas, 2007). Poplar trees (including aspens and cottonwoods) are found widespread across the northern hemisphere and are a major component of the Canadian boreal forest. They are of great importance, both ecologically and economically. They provide a shelter and food sources for a large number of wildlife (Stettler *et al*., 2006). In addition, poplar trees have been identified as a promising feedstock for pulping and biofuel production owing to their high growth rates and high cellulose/lignin ratios (Sannigrahi and Ragauskas, 2010). Poplars are also known to produce and accumulate a wide array of plant secondary metabolites (especially phenolic compounds) including phenolic glycosides, condensed tannins and hydroxycinnamoyl esters (e.g. chlorogenic acids), which makes poplar an excellent system to study plant secondary metabolism.

1.9 Objectives

Our knowledge of the shikimate pathway and its regulation has been inferred largely from microbial studies. However, this pathway has not been as comprehensively studied and understood in plants. The third and fourth steps of the shikimate pathway (shikimate biosynthesis) are catalyzed by the bifunctional enzyme DQD/SDH, which has only been molecularly characterized in a few plant species. Quinate shares high structural

similarity with shikimate and is a typical defense-related secondary metabolite present in many plant species. Quinate can be derived from 3-dehydroquinate and shikimate by quinate dehydrogense (QDH) and quinate dehydratase (QD). Reaction mechanisms of QDH and QD resemble that of SDH and DQD, respectively, but no genes have been identified encoding these functions. The *P. trichocarpa* genome encodes five putative DQD/SDHs named Poptr1 to Poptr5 here (Hamberger *et al.*, 2006). The annotation of these five genes is based on sequence analysis, and the functionalities of enzymes encoded by these five genes have not been experimentally proven. **The first objective of this project was to functionally characterize the five annotated DQD/SDHs from poplar and to determine if any of these are potentially involved in quinate biosynthesis (QDH and/or QD).** Due to similarities in the reaction mechanisms, I hypothesized that some of the five putative poplar DQD/SDH genes indeed encode QDH and/or QD activities. To test this hypothesis, a reverse biochemistry approach was followed, where the cDNAs of the five putative DQD/SDHs from poplar were expressed in *E.coli* allowing biochemical characterization of their substrate preferences (Chapter 2). Shikimate plays biologically divergent roles in both protein production and phenylpropanoid / lignin biosynthesis. I therefore hypothesized that genes encoding poplar DQD/SDH are differentially expressed. Some isoforms may be constitutively expressed in nonlignified tissues (specific to amino acid production for protein biosynthesis) while others may be expressed principally in lignified tissues and specific to lignin biosynthesis. QDH and QD encoding genes are predicted to be highly expressed in plant leaf and bark tissues due to their roles in quinate and chlorogenic acid synthesis (i.e. tissues where chlorogenic acids accumulates). The expression patterns of poplar

DQD/SDHs and QDHs as well as the co-expressed genes were characterized by compiling and analyzing public large-scale Affymetrix microarray data (Chapter 2). It has been shown that shikimate biosynthesis is localized to plastids, while the subcellular localization of quinate biosynthesis remains largely unknown. Subcellular localizations of poplar DQD/SDHs and QDHs were tested by transient transformation assays using fluorescent protein-fusion (Chapter 3).

It has been widely accepted that gene duplications and divergence play important roles in generating evolutionary novelty and adaptation. Functional variations (shikimate and quinate biosynthetic activities) were observed in the DQD/SDH/QDH superfamily, and this suggests that genes involved in quinate metabolism (secondary metabolism) may have evolved from shikimate biosynthetic (primary metabolism) genes via gene duplication. **The second objective of this work was to study diversification of the DQD/SDH/QDH gene family in the plant lineages and to pinpoint the evolutionary mechanism responsible for the functional diversity observed in this gene family.** Towards this goal, the phylogenetic relationship of the DQD/SDH/QDH superfamily was reconstructed and used to identify signatures of positive selection and to reconstruct the protein sequence of the immediate pre-duplication ancestor. Biochemical properties of the reconstructed ancestral protein were determined together with enzymes from select extant members representing major systematic groups (i.e. pre-duplication enzymes from a bacterium, a green alga, a bryophyte, and a lycopod; as well as post-duplication isoforms from a gymnosperm and an angiosperm). Combining the biochemical data with the evolutionary studies allowed pinpointing the evolution model through which functional diversity was created in the DQD/SDH/QDH gene superfamily (Chapter 4).

Chapter 2 Molecular characterization of quinate and shikimate biosynthesis in *Populus trichocarpa*

2.1 Introduction

The shikimate pathway connects primary with secondary metabolism through the biosynthesis of the three aromatic amino acids: phenylalanine (Phe), tyrosine (Tyr) and tryptophan (Trp) (Floss, 1979; Herrmann, 1995a; Schmid and Amrhein, 1995). This pathway is present only in bacteria, fungi, apicomplexan parasites (Roberts *et al*., 1998) and plants (Tohge *et al*., 2013), but is not present in animals. In plants, more than 30% of all photosynthetically fixed carbon may be directed through the shikimate pathway (Maeda and Dudareva, 2012; Tohge *et al*., 2013). The absence of the shikimate pathway in animals makes the three aromatic amino acids (the end products of this pathway) essential to animal diets, and it also makes this pathway a primary target for herbicides (Steinrucken and Amrhein, 1980), live vaccines (O'Callaghan et al., 1988), antibiotics and anti-infectious drugs (Zhang *et al*., 2005). Microorganisms primarily synthesize the three aromatic amino acids as protein building blocks. However, in plants, these three amino acids and many intermediates of the shikimate pathway also serve as the precursors of a large variety of natural products (e.g. UV protectant, pigments, antioxidant, phytoalexins, lignin, etc.), which play important physiological and ecological roles in plants (Herrmann, 1995b; Herrmann and Weaver, 1999; Weaver and Herrmann, 1997; Maeda and Dudareva, 2012; Dewick, 1995). Among all compounds derived from the shikimate pathway, lignin is quantitatively the largest carbon sink of this pathway. Lignin is a major component of plant secondary cell walls supporting erect growth and

long-distance water transport, and constitutes up to 35% of wood dry mass (Byrne and Nagle, 1997). It also functions in plant defense against herbivores and pathogens (Campbell and Sederoff, 1996). Global warming and pollution resulting from the burning of fossil fuels have been of great public concern and have raised a particular interest in replacing the current fossil fuel based system with new biofuels, many of which are derived from cellulose (Chapple *et al*., 2007). However, high lignin content and lignin composition (low S/G ratios) can hinder biofuel production and necessitate the pretreatment of lignocellulosic feedstock (Li *at el*., 2003; Ragauskas *et al*., 2006; Chen and Dixon, 2007). Poplar trees have been identified as a promising feedstock owing to its high growth rates and high cellulose/lignin ratio (Sannigrahi and Ragauskas, 2010).

The use of plant secondary metabolites by humans has a long history. Some have been used as dyes, flavours, fragrances, stimulants, poisons, as well as therapeutic drugs ever since early human history (Wink, 2009). Also in modern medicine, plant secondary metabolites derived from the shikimate pathway have been attracting lots of public and scientific attention for their roles in improving general health condition (e.g. the use of chlorogenic acid as an antioxidant nutraceutical).

The shikimate pathway consists of seven enzymatic steps starting with the condensation of phosphoenolpyruvate (PEP) and D-erythrose 4 phosphate (E4P) and ending with the production of chorismate, which is the last common precursor of Trp, Tyr and Phe (Herrmann and Weaver, 1999). Despite the presence of this pathway in different domains of life, enzymes involved in this pathway are quite distinct in terms of their protein sequences, molecular organizations and modes of regulatory mechanisms. In bacteria, the first enzyme 3-deoxy-D-*arabino*-heptulosonate-7-phosphate (DAHP)

synthase is regulated either through feedback inhibition or transcriptional regulation (Herrmann and Weaver, 1999). In plants, the whole pathway seems to be regulated primarily at the transcriptional level. It is responsive to light, developmental signals, abiotic stresses (e.g. nitrogen and amino acid starvation) and regulated by biotic stresses including wounding, herbivory and pathogen infection (Herrmann and Weaver, 1999). Enzymes of the shikimate pathway are also organized differently in different lineages. The seven steps are catalyzed by separate monofunctional enzymes in bacteria. In contrast, enzymes involved in the catalysis of steps two to six in fungi are fused into one single polypeptide encoded by a single pentafunctional *AROM* gene. In plants, most enzymes are encoded by distinct genes, but the conversion of 3-dehydroquinate to shikimate via 3-dehydroshikimate, representing the third and fourth steps of the shikimate pathway, is catalyzed by a bifunctional enzyme designated as 3-dehydroquinate dehydratase/shikimate dehydrogenase (DQD/SDH). Structurally, plant DQD/SDHs resemble fusion proteins with the N-terminal half housing the DQD domain and the C-terminal half containing the SDH domain, each of which possesses a distinct active site. DQD/SDHs have been functionally characterized in a few plant species including *Arabidopsis thaliana* (Arabidopsis; Singh and Chritendat, 2006), *Solanum lycopersicum* (tomato; Bischoff *et al*., 2001) and *Nicotiana tabacum* (tobacco; Bonner and Jenson, 1994; Ding *et al*., 2007). The DQD/SDH crystal structure from Arabidopsis allowed identification of some key amino acid residues involved in the catalytic activities of DQD/SDH (Singh and Christendat, 2006). In *Populus trichocarpa*, five putative DQD/SDH genes (here referred to as Poptr1 through Poptr5) have been identified based

on sequence similarity, but none of these has been characterized to date (Hamberger *et al.*, 2006).

Not only the end products, but all intermediates of the shikimate pathway can be branch points leading to secondary metabolic processes (Bentley and Haslam, 1990). Among these, quinate, without being a direct intermediate, is formed in a single step reaction from either shikimate or 3-dehydroquinate. Quinate is widely distributed across the plant kingdom, and can accumulate to high levels in some plant species. For example, in the developing needles of larch and pine, its concentration can reach up to 14% of leaf dry weight (Bentley and Haslam, 1990; Ossipov *et al.,* 1995). Chlorogenic acid, an ester of caffeic acid and quinate, is also known to accumulate to high levels in coffee and poplar trees, and it is gaining popularity due to its potential role in weight control and association with lower risk for type 2 diabetes and cardiovascular diseases (Morton *et al.,* 2000; Laranjinha *et al.,* 1994, Sawa *et al.,* 1999; Paynter *et al*., 2006). Quinate catabolism has been well studied in bacteria and fungi. In bacteria, quinate can be utilized as a carbon source and channeled back to the shikimate pathway by shikimate/quinate dehydrogenase (YdiB), an enzyme with affinities towards both shikimate and quinate (Lindner *et al.,* 2005). In contrast, quinate metabolism in plant remains largely enigmatic. The interconversion of shikimate and quinate has been demonstrated biochemically, and an accessory role of quinate as a reserve compound that can be channeled back into the shikimate pathway has been proposed early on (Minamikawa *et al*., 1969; Yoshida, 1969; Boudet, 1980; Gamborg, 1967, Weinstein *et al*., 1961; Goldschmidt and Quimby, 1964). Later, quinate was shown to be the precursor of chlorogenic acid in plants, which may also serve as an intermediate and storage reserve of lignin biosynthesis (Maher *et al*.,

1994; Hoffmann *et al.,* 2004). Biochemical evidence suggests that quinate can be synthesized either from shikimate or from 3-dehydroquinate, which is the product and the substrate of DQD/SDH, respectively. Enzymes involved in those two processes are quinate dehydratase/hydrolyase (QD) and quinate dehydrogenase (QDH) (Figure 1.5) (Herrmann, 1995). These two enzymes have been (partially) purified and characterized in a few plant species including corn, carrot, mung bean, pea and some coniferous species (Gamborg, 1966; Boudet, 1980, Kang and Scheibe, 1993; Ossipov *et al.,* 1995; Ossipov *et al.,* 2000; Minamikawa, 1977, Refeno *et al.,* 1982; Leuschner *et al*., 1995; Graziana and Boudet, 1983). However, genes encoding those two enzymes have not been identified to date. The reaction mechanisms and substrate structures of QDH and QH resemble that of SDH and DQD, respectively (Figure 1.5). SDH and QDH catalyze NAD(P)H-dependent reductions of a keto-group to an alcohol acting on 3-dehydroshikimate and 3-dehydroquinate, respectively. Likewise, DQD and QD catalyze reversible water elimination reactions with 3-dehydroquinate and quinate, respectively. Due to similarities in reactions catalyzed and substrates utilized; and knowing that enzymes having both SDH and QDH activities exist in conifers (Lindner *et al.,* 2005; Ossipov *et al.,* 2000), I hypothesize that QDH and/or QD activities may be encoded by genes similar to DQD/SDH.

2.2 Methods

2.2.1 DQD/SDH protein sequence analyses and structure modeling

The five poplar putative DQD/SDH protein sequences (Poptr 1 to Poptr 5; Table 2.1; Hamberger et al., 2006) were retrieved from the *Populus trichocarpa* genome database

Table 2.1: DQD/SDH family members (plant DQD/SDHs and bacterial AroD, AroE and YdiB) included in amino acid sequence analysis.

Name	Species	Database	Gene models
EsccoAroE	*Escherichia coli W*	NCBI	YP_006126142.1
EsccoAroD	*Escherichia coli W*	NCBI	YP_006124567.1
EsccoYdiB	*Escherichia coli W*	NCBI	YP_006124566.1
Arath	*Arabidopsis thaliana*	NCBI	NP_187286.1
Solly	*Solanum lycopersicum*	NCBI	NP_001234051.1
Nicta1	*Nicotiana tabacum*	NCBI	gb|AAS90325.1
Nicta2	*Nicotiana tabacum*	NCBI	gb|AAS90324.2
Poptr1	*Populus trichocarpa*	Phytozome	Potri.010G019000.2
Poptr2	*Populus trichocarpa*	Phytozome	Potri.013G029900.2
Poptr3	*Populus trichocarpa*	Phytozome	Potri.005G043400.1
Poptr4	*Populus trichocarpa*	Phytozome	Potri.014G135500.3
Poptr5	*Populus trichocarpa*	Phytozome	Potri.013G029800.1

version 1.1 hosted at the Joint Genome Institute (Tuskan *et al.,* 2006; JGI: http://genome.jgi.doe.gov/Poptr1_1/Poptr1_1.info.html). Sequence accuracy and gene annotations were assessed by comparison among different genome versions (1.1, 2.0 and 3.0). Completeness of coding sequences was assessed by BLASTP (Basical Local Alignment Search Tool) searches against the Plant GDB transcript assembly database (http://www.plantgdb.org/prj/ESTCluster/progress.php?orderby=Version). Amino acid sequences of functionally characterized DQD/SDHs from *Arabidopsis thaliana* (Arath; Singh and Chritendat, 2006)*, Solanum lycopersicum* (Solly; Bischoff *et al*., 2001) *and Nicotiana tabacum* (Nicta 1 and 2; Ding *et al*., 2007) were retrieved from GenBank at the NCBI (National Centre for Biotechnology Information: http://www.ncbi.nlm.nih.gov/). *Escherichia coli* (Escco) AroD (DQD), AroE (SDH) and YdiB (SDH/QDH) protein sequences were also retrieved from NCBI.

Multiple sequence alignment of the collected DQD/SDH amino acid sequences was performed using ClustalW implemented in BioEdit (Hall, 1999) (http://www.mbio.ncsu.edu/bioedit/bioedit.html). This alignment was split into two fragments to allow separate examination of DQD and SDH domains. Pairwise sequence similarity scores for both SDH and DQD domains were calculated separately and used to generate sequence similarity tables in BioEdit. The two alignments (containing either DQD or SDH) were also used to build Neighbor Joining (NJ) trees with Phylip3.69 (Felsenstein, 1989).

The full-length protein alignment containing both DQD and SDH domains was manually trimmed to retain only amino acid residues involved in reaction catalysis, cofactor and substrate binding. The modified alignment was then subjected to creation of

sequence logos using the 'WebLogo' tool found at http://weblogo.berkeley.edu/logo.cgi (Crooks et al, 2004).

3D protein structures of the five putative poplar DQD/SDHs (Poptr1 to Poptr5) were predicted using Phyre (http://www.sbg.bio.ic.ac.uk/~phyre/; Kelley and Sternberg, 2009). Model coordinates were generated using Arabidopsis DQD/SDH (Protein DataBase ID: c2o7qA; http://www.rcsb.org/pdb/home/home.do) coupled with either 3-dehydroshikimate and tartrate or shikimate as a template, and aligned to the known Arabidopsis DQD/SDH structure. Pymol (http://www.pymol.org) was used to further examine and visualize the generated models.

2.2.2 Cloning of the five DQD/SDH genes from *Populus trichocarpa*

In order to obtain full-length poplar DQD/SDH clones, total RNA was isolated from *Populus trichocarpa* Nisqually-1 plant tissues (leaf, bark, xylem and phloem tissues collected at University of Victoria) using the cetyltrimethylammonium bromide (CTAB) method as described (Haruta *et al.*, 2001). Total RNA was then reverse transcribed into cDNAs using 5 μM oligo dT primer and 1 U reverse transcriptase (Invitrogen). The five putative poplar DQD/SDHs were amplified from the cDNA pool via PCR with USER (Uracil-Specific Excision Reagent) modified gene specific primers (Table 2.2). Proofreading Pfu Turbo Cx Hotstart DNA polymerase (Agilent) was used in PCR amplification (30 cycles, denaturation at 94 ºC for 30 s, annealing at 56 ºC for 30 s, extension at 72 ºC for 120 s). PCR amplicons were treated with USER^{TM} enzyme mixture (New England Biolabs) and cloned into pQE30-USER-6xHis over-expression vectors (with His tag at the N terminus) separately following the standardized protocol (Geu-

Table 2.2: USER-modified (uracil-containing) PCR primer sets used to generate pQE30 over-expression constructs carrying the full-length coding sequences of Poptr1 to Poptr5.
Start and stop codons are underlined, and extensions used for USER cloning are indicated in bold.

	Forward primer	Reverse primer
Poptr1	5'-**GGCTTAAU**ATGGATTCTGCAAGCAACGTCC-3'	5'-**GGTTTAAU**CTAGTACTTTGACATGATCTTCTGAAAGAGTTC-3'
Poptr2	5'-**GGCTTAAU**ATGGGGCGTGCTGGGATC-3'	5'-**GGTTTAAU**TCAGAATTTGGCTAGAACAATCTCCC-3'
Poptr3	5'-**GGCTTAAU**ATGGGGAGTGTTGGAGTCCTGAC-3'	5'-**GGTTTAAU**TCAGAATTTGGCTAAAACAATCTCCC-3'
Poptr4	5'-**GGCTTAAU**ATGGCATTCAAGAACAACCTCTTAG-3'	5'-**GGTTTAAU**TCAAAATTGCTCCAAGACAAGC-3'
Poptr5	5'-**GGCTTAAU**ATGGATCTCCAAAGCGCTG-3'	5'-**GGTTTAAU**TTATGTATTCCTTGCTAACACATCTCTAAT-3'

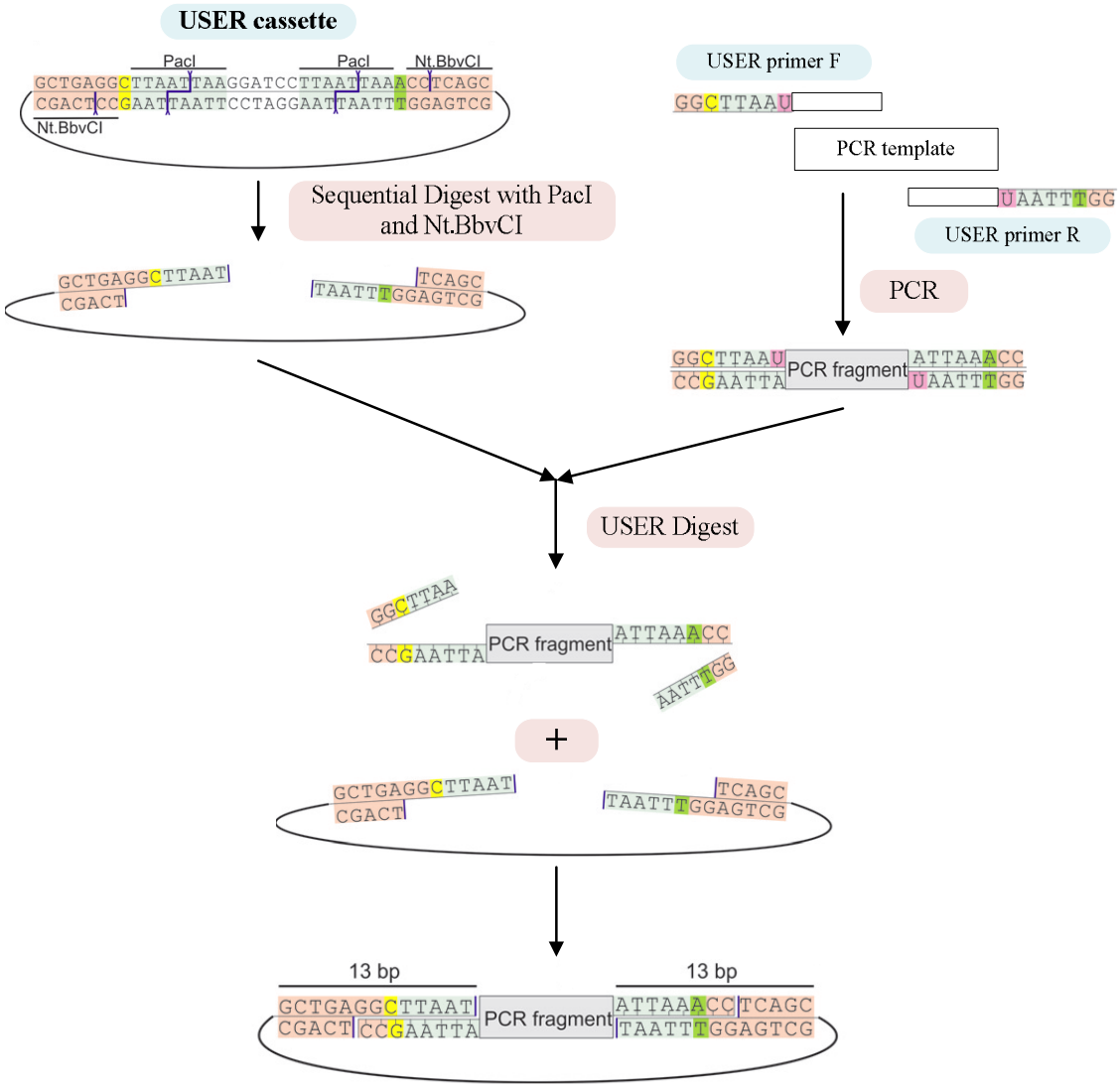


Figure 2.1: Overview of the USER cloning technique.
A vector containing the USER cassette (upper left corner) was digested with PacI and Nt.BbvCI to generate linearized vector with 8 nt single-stranded extensions at both ends. A PCR reaction was carried out with USER-modified primers by the PfuTurbo Cx Hotstart DNA polymerase. PCR amplicons were treated with USER enzyme mixture and mixed with the linearized vector. Reaction mixture was incubated at 37°C for 20min followed by incubation at25°C for another 20 min. The products were used for bacterial transformation without ligation. Nt.BbvCI recognition sites are marked in pink, PacI recognition sites are marked in light blue. Yellow and green colors mark single base differences between the two extensions, which are important in directional cloning (modified from Halkier *et al.,* 2006).

Flores *et al*., 2007; Figure 2.1). The final constructs were verified by DNA sequencing. A pQE30 construct carrying cDNA of a glycosyltransferase (GT) (kindly provided by Dr. Vasko Veljanovski; Veljanovski and Constabel, 2013) was used for the subsequent experiments as a negative control.

2.2.3 Production and purification of poplar putative DQD/SDH recombinant proteins

Over-expression constructs carrying the five putative poplar DQD/SDH coding sequences were transferred into the M15 *E. coli* competent cells by electroporation to allow the expression of recombinant proteins in a bacterial system. Experimental conditions including temperature, isopropylthio-β-galactoside (IPTG) concentration and induction time were optimized to yield high levels of soluble recombinant protein expression. For each set of conditions tested, samples were retained and the abundance of recombinant proteins in each sample was assessed by SDS-PAGE and western blot with the SuperSignal® West HisProbe^{TM} Kit (Thermo Scientific) following the manufacturer's instruction. The set of conditions yielding the highest amounts of soluble recombinant proteins were used for subsequent enzyme experiments: 5 mL of LB broth containing selective antibiotics (100 mg/L ampicillin and 25 mg/L kanamycin) was inoculated with a bacterial colony containing the right construct and grown at 37 °C overnight with shaking. The main culture (50 mL) was inoculated with the pre-culture next morning and grown until OD_{600} = 0.6, and then IPTG was added to a final concentration of 0.06 mM. Protein production was carried out at 19 °C with shaking for 16 hours. *E. coli* cells were harvested by centrifugation at 4000 x g for 20 min and the cell pellet was stored at -80°C for later use.

6xHis-tagged recombinant proteins were extracted and purified under native condition using Nickel-NTA agarose (Qiagen) (Figure 2.2). A frozen cell pellet was resuspended in 4 mL of lysis buffer (50 mM NaH_2PO_3, 300 mM NaCl, and 10 mM imidazole, NaOH, pH 8, 1 mg/ml lysozyme) and incubated on ice for 1 hr. The mixture was further lysed using a sonicator followed by centrifugation at 10000 x g for 30 min at 4°C. The supernatant was incubated with 1 mL of 50% Nickel-NTA agarose (Qiagen) on ice for 1 hr with shaking. The lysate-NTA slurry was loaded onto a column and washed twice with wash buffer without imidazole (50 mM NaH_2PO_4, 300 mM NaCl, NaOH, pH 8) and three times with buffer containing imidazole (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, NaOH, pH 8) before elution (with 50 mM NaH_2PO3, 300 mM NaCl, and 250 mM imidazole, NaOH, pH 8). Protein concentration was determined by Bradford assay (Bradford, 1976) using BSA as a standard.

Denaturing SDS-PAGE was performed to assess the purity of recombinant proteins. Boiled protein samples were loaded on a 12% polyacrylamide gel and separated at 20 mA for 45 min. Proteins were transferred from the gel to PVDF membrane (Millipore) by electroblotting at 30 volts overnight at 4 °C. 6xHis-tagged proteins were detected by chemiluminescence using the SuperSignal® West HisProbe™ Kit (Thermo Scientific) following the manufacturer's instruction.

2.2.4 Measurement of dehydrogenase activities using spectrophotometry

Dehydrogenase activities of the recombinant proteins with shikimate (SDH) and quinate (QDH) were determined photometrically by measuring NADPH or NADH production at 340 nm (Ding *et al*., 2007). An appropriate amount of purified 6xHis-

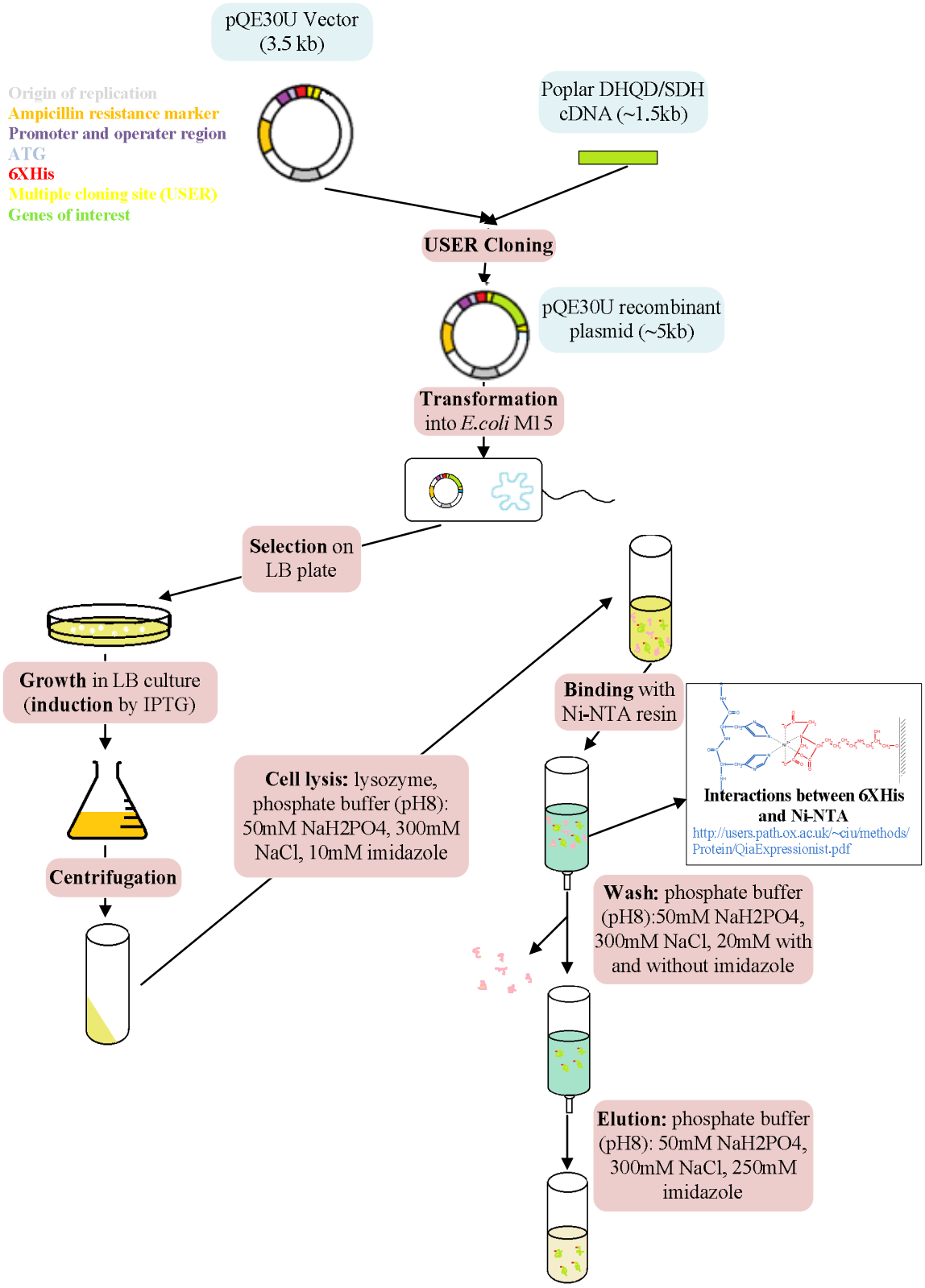

pQE30U Vector (3.5 kb)
Origin of replication
Ampicillin resistance marker
Promoter and operater region
ATG
6XHis
Multiple cloning site (USER)
Genes of interest
Poplar DHQD/SDH cDNA (~1.5kb)
USER Cloning
pQE30U recombinant plasmid (~5kb)
Transformation into E.coli M15
Selection on LB plate
Growth in LB culture (induction by IPTG)
Centrifugation
Cell lysis: lysozyme, phosphate buffer (pH8): 50mM NaH2PO4, 300mM NaCl, 10mM imidazole
Binding with Ni-NTA resin
Interactions between 6XHis and Ni-NTA
http://users.path.ox.ac.uk/~ciu/methods/Protein/QiaExpressionist.pdf
Wash: phosphate buffer (pH8):50mM NaH2PO4, 300mM NaCl, 20mM with and without imidazole
Elution: phosphate buffer (pH8): 50mM NaH2PO4, 300mM NaCl, 250mM imidazole

Figure 2.2: Schematic overview of recombinant protein (Poptr1 to Poptr5) production, extraction and purification.

Over-expression constructs carrying the five putative poplar DQD/SDH (Poptr1 to Poptr5) coding sequences (derived from USER cloning) were introduced into *E.coli* (strain M15) through electroporation. Transformed bacteria were plated on LB plates with selective antibiotic (100 mg/L ampicillin and 25 mg/L kanamycin) and incubated at 37 °C overnight. Colonies containing desired construct are verified by colony PCR and used to inoculate 5 mL LB broth containing selective antibiotics. This pre-culture was grown at 37 °C overnight with vigorous shaking. A 50 mL culture was then inoculated with the pre-culture and grown at 37 °C until OD_{600} = 0.6, and isopropylthio-β-galactoside (IPTG) was added to a final concentration of 0.06 mM. Protein production was carried out at 19 °C with shaking for 16 hours. Cells were harvested by centrifugation at 4000 x G for 20 min and the cell pellet was used for protein purification. 6xHis-tagged recombinant proteins were extracted and affinity-purified under native condition using Nickel-NTA agarose (Qiagen). The cell pellet was resuspended in 4 mL of lysis buffer (50 mM NaH_2PO_3, 300 mM NaCl, 10mM imidazole, and 1 mg/ml lysozyme) and incubated on ice for 1 hr. The mixture was further lysed using a sonicator followed by centrifugation at 10000 x G for 30 min at 4 °C. The supernatant was incubated with 1 ml of 50% Nickel-NTA agarose suspension on ice for 1hr with moderate shaking. The lysate-NTA slurry was loaded onto a polypropylene column and washed twice with wash buffer without imidazole (50 mM NaH_2PO_3, 300 mM NaCl, and 250 mM) and subsequently three time with buffer containing imidazole (with 50 mM NaH_2PO_3, 300 mM NaCl, and 250 mM imidazole). Purified proteins were eluted with elution buffer (with 50 mM NaH_2PO_3, 300 mM NaCl, and 250 mM imidazole).

tagged protein (0.01-0.05 mg) was mixed with 750 μl of reaction buffer (100 mM Trizma base, HCl, pH 8) and 100 μl of 2 mM $NADP^+$ or 50 μl of 5 mM NAD^+ at room temperature. Reaction was initiated by adding 100 μl of 10 mM shikimate or 100 mM quinate. Absorbance changes at 340 nm were measured and recorded every 10s for the first one minute and every minute for the following 4 min. Both shikimate and quinate dehydroganse activities as a function of pH were determined using different buffer systems: $NaCH_3CO_2$-CH_3CO_2H buffer solutions (pH 4 and 5), K_3PO_4-H_3PO_4 buffer (pH 6, 7 and 7.5), Trizma base-HCl buffer (pH 8 and 8.5), and $NaHCO_3$-NaOH buffer (pH 9, 10 and 11) (Diaz and Merino, 1996). The experimental procedure was the same as described above.

Kinetic properties (Vmax and Km values) of each enzyme with its preferred substrate (shikimate for Poptr1 and Poptr5, quinate for Poptr2 and Poptr3) were determined at the optimal pH (pH8.5) using at least 12 concentrations of shikimate (ranging from 0.5-1000 μM) and quinate (50-2000 μM). Changes in absorbance were converted to enzyme activities ($\mu M \cdot s^{-1} \cdot mg^{-1}$) using the equation $A=c\epsilon l$, where A is absorbance, c is cofactor concentration, ϵ is NADPH extinction coefficient (6.2×10^3 $L \cdot mol^{-1} \cdot cm^{-1}$) and l is the length of light path (1cm). Three replicates (independent protein purifications) were carried out for each reaction to assess the accuracy of this experiment, and means were calculated using Excel (Microsoft). Kinetic data of each enzyme were plotted and fitted to the Michaelis-Menten model using plotting and curve-fitting tools implemented in MATLAB. Maximal velocity values (Vmax) and Michaelis-Menten constants (Km) were estimated. Confidence of curve fitting (i.e. 95% confidence bounds,

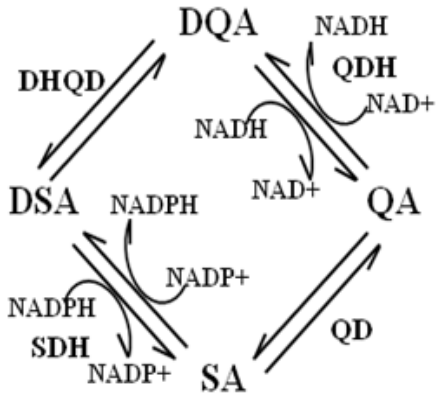
R-square values and root mean square error) was also assessed with MATLAB. Enzyme substrate specificities were calculates by dividing Vmax by Km.

2.2.5 Measurement of dehydratase activities and confirmation of dehydrogenase activities using high performance liquid chromatography (HPLC)

Due to the reversibility of the shikimate/quinate cycle (Figure 1.5), possible multifunctionality of the enzymes analyzed, and variations in co-factor requirements, a systematic experimental design was developed to cover all four presumed functions of the five putative poplar DQD/SDHs. Recombinant proteins were incubated in reaction buffer at pH 8.5 with different combinations of substrates and cofactors as shown in Table 2.3. To test for reversible dehydratase activities (DQD and QD), the standard reaction mixture consists of purified recombinant protein, pH 8.5 buffer and 3-dehydroquinate, 3-dehydroshikimate, shikimate or quinate without addition of a cofactor. The standard reaction mixture of a dehydrogenase enzyme assay contains enzyme, buffer, cofactor and substrate as above. Thirty min incubation time at room temperature was allowed. Upon incubation, reaction mixtures were filtered through a 0.22 μm filter and loaded onto an Aminex HPX-87H organic acid analysis column (300X7.8 mm I.D.; 9 μm particle diameter) (Mousdale and Coggins, 1985, Givry *et al*., 2007) and separated on an Ultimate 3000 HPLC system with a diode array detector (Dionex). Products were separated using an isocratic elution system (10 mM H_2SO_4) over a 40 min period with a flow rate of 0.4 ml/min. A control set of assays were prepared with boiled enzymes, which were used as negative controls. Reduction in peak area representing substrates or increase in peak area representing products was indicative of enzymatic activities.

Table 2.3: Experimental design used to test for dehydratase and dehydrogenase activities with HPLC.
Reactions tested for are given in the first column, and the substrates and cofactors included are indicated by an 'X'. NAD^+ or NADH were used for Poptr 2 and Poptr3, $NADP^+$ and NADPH for Poptr1 and Poptr5. SA: shikimate; QA: quinate; DSA: 3-dehydroshikimate; DQA: 3-dehydroquinate.

Reaction tested	SA	QA	DSA	DQA	NAD+ or NADP+	NADH or NADPH
QA->DQA->(DSA)		X			X	
QA->SA		X				
DQA->QA->(SA)				X		X
DQA->DSA				X		
SA->DSA->(DQA)	X				X	
SA->QA	X					
DSA->SA->(QA)			X			X
DSA->DQA			X			



2.2.6 Poplar DQD/SDH and QDH expression profiling and co-expression analyses based on combined Affymetrix microarray expression data

All publicly available poplar Affymetrix microarray data (*GeneChip® Poplar Genome Array*) were retrieved from ArrayExpress and GEO *(Gene Expression Omnibus)* at NCBI on July 4th-5th, 2012. This collection of microarray data consists of 43 diverse experiment series with a total of 748 array hybridizations covering 17 poplar species. Microarray raw expression data were normalized using the MAS5 algorithm (sc=500) implemented in R (Gautier *et al*., 2003; Toufighi *et al*., 2005). Normalized expression data were then divided into three groups (organs and tissues, treatments, and transgenic) based on their sample annotations. Log_2-transformed expression values were used for the 'organ' dataset while log_2 ratios compared to controls were calculated for the 'treatment' and 'transgenic' groups. From this normalized expression data matrix, expression data were extracted to generate expression profiles of the five poplar genes of interest across different tissue types and treatments. E-northerns were created using HeatMapper (an online tool provided by BAR; Toufighi *et al*., 2005).

Normalized data were also used for a co-expression analysis using the "Expression Angler" algorithm also implemented at BAR (Toufighi *et al*., 2005). Pairwise Pearson correlation coefficients (PCC) between genes of interest (poplar DQD/SDHs and QDHs) and all other genes were calculated across each dataset, and only the top 25 co-expressed genes were retained.

2.3 Results

2.3.1 Sequence diversity in the DQD/SDH superfamily

The poplar genome contains five genes, which share sequence similarity with functionally characterized DQD/SDHs, and these five genes were annotated as poplar DHQD/SDH1 through DHQD/SDH5 (Poptr1 to Poptr5) (Hamberger *et al*., 2006). Protein sequences encoded by the five genes were aligned with functionally characterized DQD/SDHs from Arabidopsis, tobacco and tomato (Figure 2.3). There was a high degree of sequence conservation throughout the entire sequences, and several key amino acids (Singh and Christendat, 2006) essential for substrate binding and catalysis were highly conserved as well. Sequence similarity comparisons of these proteins with characterized plant enzymes indicated that Poptr1 and Poptr5 share high sequence identity with characterized proteins ranging from 56% to 78% for Poptr1 and from 47% to 66% for Poptr5. Poptr1 and 5 share relatively high sequence identity (61% for the N-terminal DQD domain and 66% for the C-terminal SDH domain) with each other. In contrast, Poptr2, Poptr3, and Poptr4 show lower sequence similarities with functionally characterized proteins ranging from 24% to 44% (Table 2.4). Protein sequence NJ (neighbor joining) trees further highlighted that poplar Poptr1 and Poptr5 are more similar to characterized DQD/SDH proteins than Poptr2, Poptr3, and Poptr4. This suggests that poplar Poptr1 and Poptr5 are likely to encode true DQD/SDH activities while Poptr2, Poptr3, and Poptr4 define a DQD/SDH like group and are potentially involved in catalyzing reactions that are similar to DQD/SDH, perhaps using a different substrate. All five poplar proteins have comparable low similarities to bacterial SDH (AroE) and SDH/QDH (YdiB), but all plant

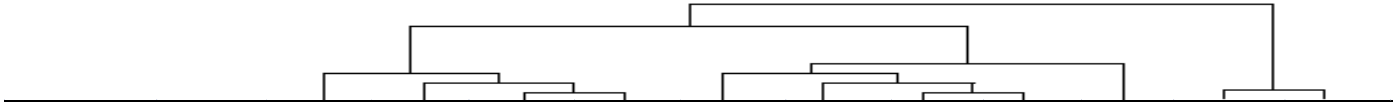

G-motif
Arath
Solly
Nicta1
Poptr1
Poptr5
Nicta2
Poptr2
Poptr3
Poptr4
EsccoAroD
EsccoAroE
EsccoYdiB

Figure 2.3: Alignment of *Populus trichocarpa* DQD/SDHs (Poptr1 to Poptr5) with functionally characterized plant DQD/SDHs (*Arabidopsis thaliana*: Arath, *Nicotiana tabacum*: Nicta1-2, *Solanum lycopersicum*: Solly), and *Escherichia coli* AroD (DQD), AroE (SDH) and YdiB (S/QDH).

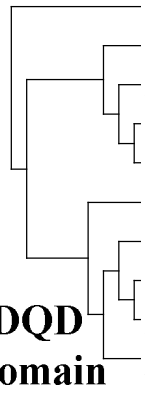
This alignment was generated using ClustalW (Bioedit). The two functional domains (DQD and SDH) are highlighted in pink and light blue, respectively. Residues with >70% similarity are shaded. Amino acid residues potentially involved in catalysis or substrate binding are indicated by arrows.

Table 2.4: DQD/SDH protein sequence similarity tables generated with Bioedit.
The top right triangle gives similarities in the SDH domain of the proteins while the lower left triangle gives similarities in the DQD portion. Neighbor Joining trees were produced using the same alignments with Phylip3.69 and placed on the left side (DQD) and top (SDH) of their corresponding tables. Sequences included are DQD/SDHs from Arabidopsis (Arath), tomato (Solly), tobacco (Nicta1-2), poplar (poptr1-5), *E. coli* (Escco) AroD, AroE and YdiB.

SDH Domain



DQD Domain



Seq->	Escco AroD	Poptr4	Nicta2	Poptr2	Poptr3	Poptr5	Poptr1	Solly	Nicta1	Arath	Escco AroE	Escco YdiB
EsccoAroD	ID	--	--	--	--	--	--	--	--	--	--	--
Poptr4	0.182	ID	0.684	0.647	0.658	0.528	0.582	0.575	0.5	0.578	0.258	0.301
Nicta2	0.202	0.408	ID	0.776	0.78	0.501	0.539	0.521	0.435	0.556	0.224	0.298
Poptr2	0.212	0.395	0.612	ID	0.903	0.501	0.521	0.517	0.45	0.535	0.229	0.291
Poptr3	0.2	0.404	0.612	0.848	ID	0.501	0.528	0.524	0.453	0.535	0.229	0.278
Poptr5	0.231	0.351	0.384	0.453	0.465	ID	0.664	0.655	0.557	0.606	0.234	0.295
Poptr1	0.238	0.328	0.384	0.45	0.453	0.613	ID	0.784	0.639	0.736	0.25	0.288
Solly	0.226	0.308	0.363	0.431	0.42	0.591	0.669	ID	0.736	0.708	0.242	0.297
Nicta1	0.242	0.332	0.385	0.458	0.446	0.612	0.722	0.832	ID	0.552	0.212	0.248
Arath	0.197	0.246	0.322	0.362	0.359	0.468	0.552	0.475	0.509	ID	0.276	0.296
EsccoAroE	--	--	--	--	--	--	--	--	--	--	ID	0.25
EsccoYdiB	--	--	--	--	--	--	--	--	--	--	--	ID

SDHs are more similar to bacterial YdiB (25-30% similarity) than to AroE (21-28%).

2.3.2 Protein structure modeling of poplar DQD/SDHs

Protein function is strongly associated with its structure, and changes in structure can lead to diversification in protein function (e.g. enzymatic activity). 3D protein structures of the five putative poplar DQD/SDHs were modeled based on the known structural data of the Arabidopsis DQD/SDH protein (Singh and Christendat, 2006; Singh and Christendat, 2007) using the structure prediction tool Phyre. Reliable prediction of a protein structure can typically be achieved if it shares sufficient sequence similarity with the known structure. As seen in the primary sequence similarity comparison, Poptr 1 and Poptr5 share higher sequence similarities with the Arabidopsis DQD/SDH (than members from the like group. While Poptr2, Poptr3 and Poptr4 (members from the DQD/SDH like group) are more divergent, but were expected to be similar enough to allow determination of reliable structure models. Indeed, for all five sequences, 489 residues (over 91% of the sequence) were modeled with 100% confidence using the Arabidopsis template.

2.3.2.1 3-dehydroquinate dehydratase (DQD)

The N-terminal DQD domain encompasses the first 316 amino acids of the Arabidopsis protein. Analysis of the active sites of the Arabidopsis DQD domain by Singh and Christendat (2006) have revealed that Lys241 and His214 are the major catalytic groups in this domain: Lys241 forms a Schiff base intermediate with the carbonyl group of 3-dehydroquinate during reaction while His214 helps modulate the formation and breakdown of this Schiff base intermediate. In addition, Arg279 serves as the key binding site for DQA, and some other residues (Glu124, Met271, Phe291 and

Gln304) are involved in properly orienting and positioning the substrate (Singh and Christendat, 2006). The corresponding amino acids were found to be divergent between the DQD and DQD-like groups, but were highly conserved within each group. Again, Poptr1 and 5 share most of these amino acid residues with the Arabidopsis DQD template, further supporting DQD function being conserved in Poptr1 and 5. In contrast, major differences were observed between the Arabidopsis DQD and the DQD domains of Poptr2, Poptr3 and Poptr4. For example, the binding group Arg279 in Arabidopsis is replaced by a Gln in Poptr2, while Lys241, which is involved in the formation of Schiff base intermediate, is still present, His214, which functions as a modulator of the Schiff base formation step, is replaced by a Tyr in Poptr2, Poptr3, and Poptr4 (Figure 2.4 and Figure 2.5, Appendix A). The replacement of His by Tyr in Poptr2, Poptr3, and Poptr4 suggests a loss of DQD function in these enzymes. However, the conservation of Tyr within the DQD-like group may be an indication of a distinct function performed by Poptr2, Poptr3, and Poptr4.

2.3.2.2 Shikimate dehydrogenase (SDH)

The SDH domain spans amino acids 328 to 588 at the C-terminus of the Arabidopsis protein. Key amino acid residues required for SDH function have been previously identified in Arabidopsis: Lys385 and Asp423 are likely to play an important role in catalysis, and Ser336, Ser338 and Try550 are involved in interacting with the C1 carboxylate (Singh and Christendat, 2006). The two catalytic sites were found to be highly conserved across all the examined models. This may be an indication that the same basic catalytic capacity is probably maintained across all five enzymes. However, variations at the substrate binding sites were observed between members from the SDH

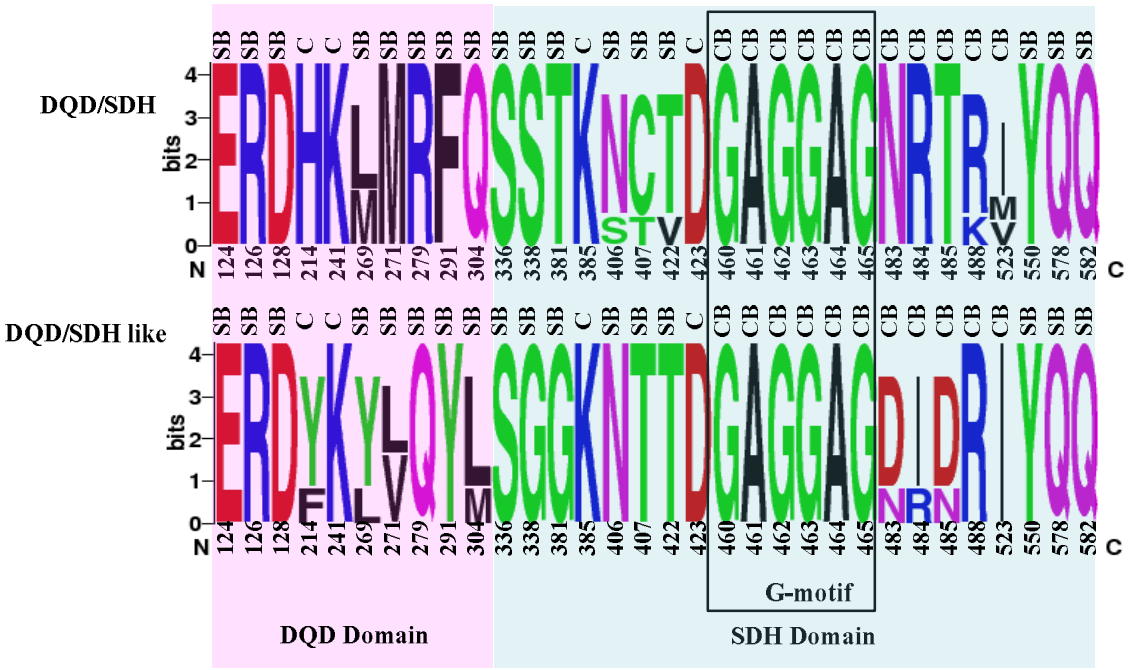


Figure 2.4: Weblogos generated with key amino acid residues of previously characterized DQD/SDHs (from Arabidopsis, tobacco, tomato) and the five putative poplar DQD/SDHs (Poptr1 to Poptr5), which are potentially involved in reaction catalysis (C), substrate binding (SB) or cofactor binding (CB).
Sites are numbered based on their relative positions in the Arabidopsis DQD/SDH structure. Images were produced using the 'WebLogo' tool found at http:// weblogo.berkeley.edu/ logo.cgi.

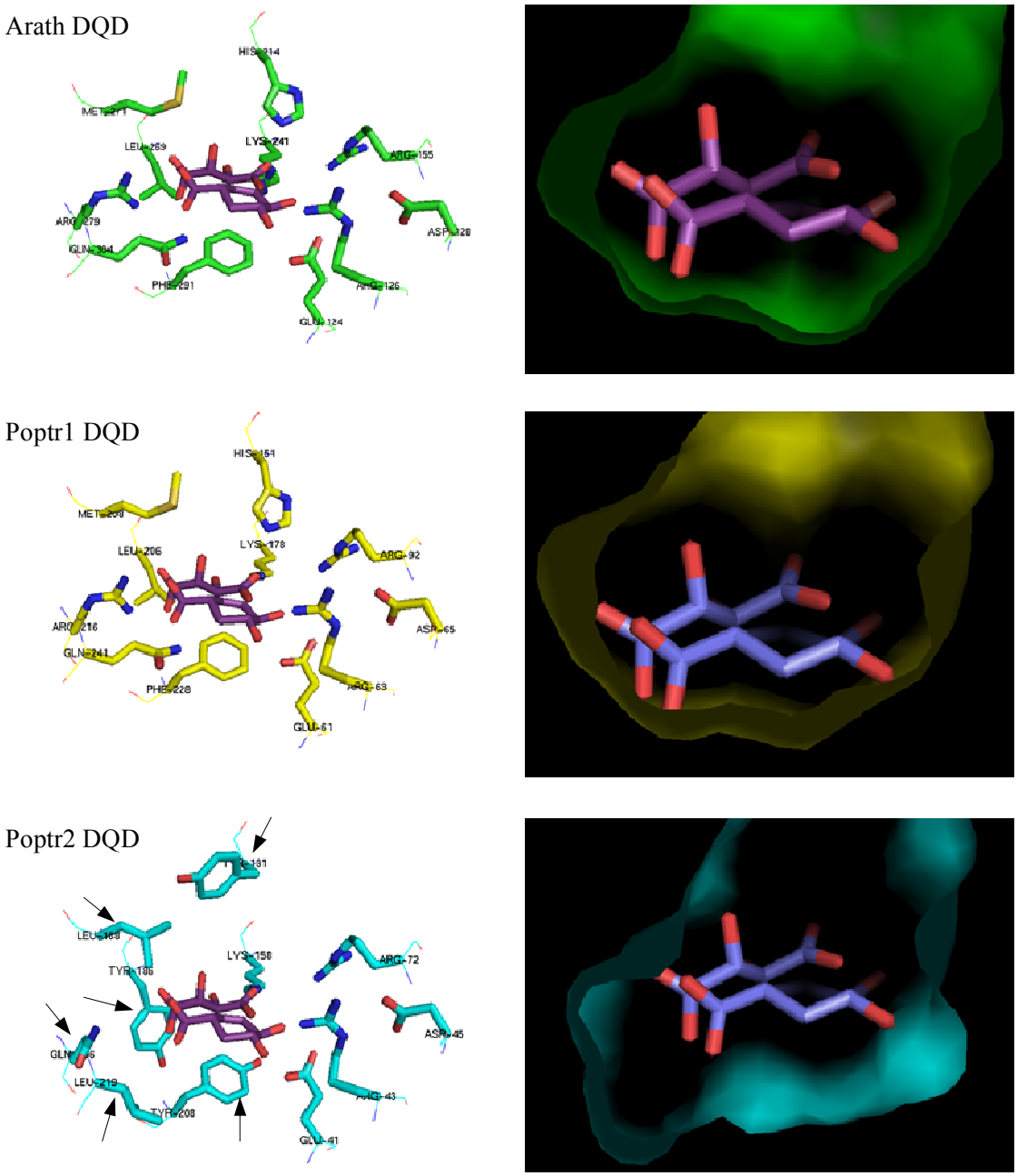


Figure 2.5: Active site structure models of the DQD domains of Poptr1 and Poptr2 in comparison to the known DQD/SDH structure from Arabidopsis.
DQD active sites are each shown as stick models (left) and as surface models (right). The actual structure from Arabidopsis (Arath) is shown in green. The model for Poptr1, belonging to the DQD/SDH group, is shown in yellow, and the model for Poptr2, belonging to the DQD/SDH-like group, is shown in blue. Amino acids differing from the Arabidopsis structure are indicated by arrows. Structure models of Poptr3 to Poptr5 are shown in Appendix A.

and SDH-like groups. A key amino acid residue (Ser338 in Arabidopsis) involved in interacting with the C1 carboxylate group of the substrate was found to be replaced by Gly in Poptr2, Poptr3 and Poptr4 (Figure 2.4 and Figure 2.6, Appendix A). The shortening of the side chain results in the weakening of hydrogen bonds, which leads to the appearance of an extra pocket in the active site. This allows more bulky substrates to fit in. Quinate, which has an additional hydroxyl group at the C1 position when compared with the conventional substrate shikimate, would be a good fit. In addition, a Gly residue exchanged for Thr381 in Arabidopsis was found to be highly conserved in the SDH-like group, and this Thr is replaced by Ser in the bacterial YdiB enzymes with affinities towards both shikimate and quinate (Singh and Christendat, 2006). Taken together, the two changes to Gly from Ser338 and Thr381 may define recognition of quinate as an alternative substrate for Poptr2, Poptr3 and Poptr4.

2.3.3 Enzymatic characterization of poplar DQD/SDH family members

Full-length cDNA fragments encoding the five putative poplar DQD/SDHs were PCR amplified and cloned into pQE30 expression vectors. Constructs were validated by DNA sequencing, and no difference was observed compared to the poplar genome v3.0 gene models. The five clones were introduced into *E. coli* to allow the expression of recombinant proteins. Experimental procedures were optimized to allow sufficient soluble proteins to be produced. His-tagged versions of the native proteins were purified using affinity chromatography. For four out of the five members, sufficient amounts of soluble recombinant proteins were purified. Only for Poptr4, no soluble recombinant protein was yielded despite further optimization attempts. It was therefore not included in further examination. Recombinant proteins (Poptr1, Poptr2, Poptr3 and Poptr5) were

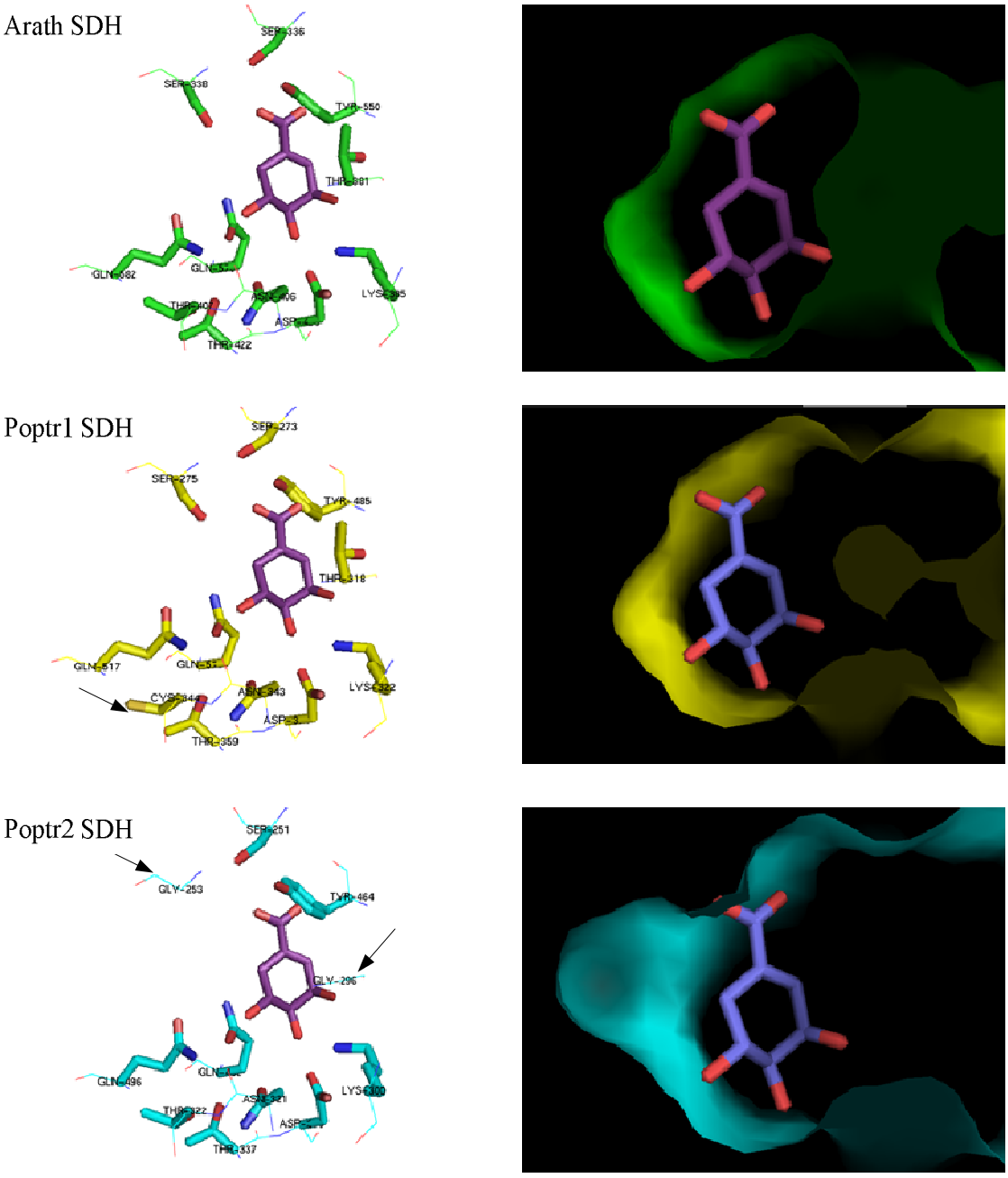


Figure 2.6: Active site structure models of the SDH domains of Poptr1 and Poptr2 in comparison to the known DQD/SDH structure from Arabidopsis.
SDH active sites are each shown as stick models (left) and as surface models (right). The actual structure from Arabidopsis (Arath) is shown in green. The model for Poptr1, belonging to the DQD/SDH group, is shown in yellow, and the model for Poptr2, belonging to the DQD/SDH-like group, is shown in blue. Amino acids differing from the Arabidopsis structure are indicated by arrows. Structure models of Poptr3 to Poptr5 are shown in Appendix A.

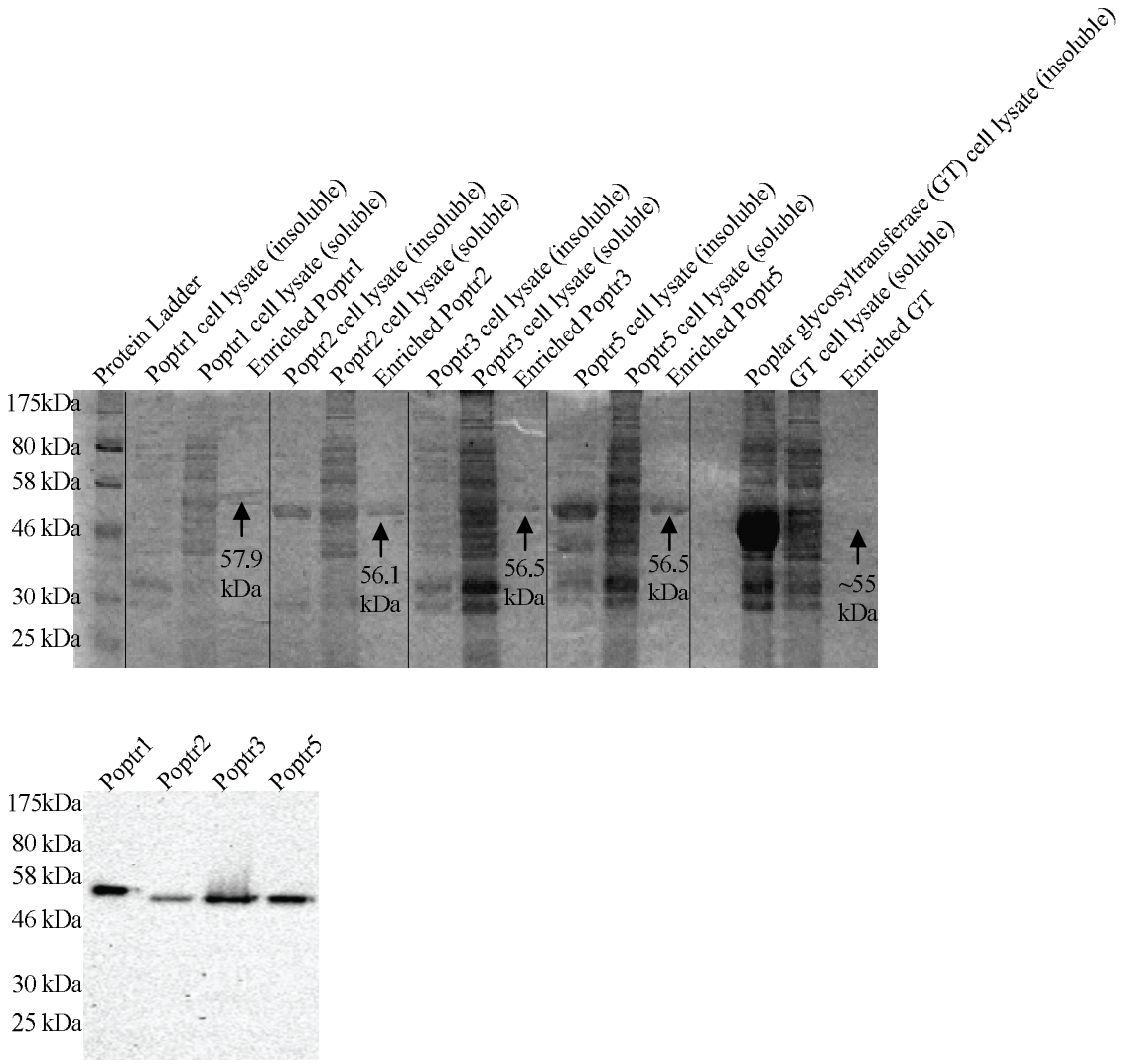


Figure 2.7: SDS-PAGE gel and Western blot analyses of purified proteins (Poptr1-3, Poptr5 and GT).
Recombinant DQD/SDHs (putative) from poplar (Poptr1 to Poptr5) were purified from *E. coli* through Ni-NTA affinity chromatography and separated on a polyacrylamide gel (top, stained with coomassie) followed by Western blotting and detection of His-tagged proteins using a HisProbe fused to horseradish peroxidase (bottom).

assessed with SDS-PAGE and only a single band was apparent after Coomassie staining for each poplar putative DQD/SDH (Figure 2.7). Molecular weights of the expressed proteins were as expected (ranging from 56 to 58 kDa). Western blotting was performed to assess the quality of protein purification, and again, only a single band of the expected size was observed for each enzyme. This indicates that no degradation has happened during purification steps. The obtained full-length proteins were used for subsequent enzyme assays.

2.3.3.1 Photometric measurement of dehydrogenase activities

Dehydrogenase activities of the recombinant proteins with shikimate (SDH) and quinate (QDH) were determined photometrically by measuring NADH or NADPH production at 340 nm. Under saturating conditions, Poptr1 and Poptr5 displayed strong activities with shikimate but no detectable activity with quinate, thus confirming SDH activity (Figure 2.8). In terms of co-factor requirements, maximal enzyme activities were observed when $NADP^+$ was used as a cofactor, and activities dropped dramatically when replacing $NADP^+$ with NAD^+. In contrast, Poptr2 and Poptr3 could act on both substrates but displayed a much higher preference for quinate compared to shikimate. Only residual activities were observed when shikimate was used as a substrate. This shows that Poptr2 and Poptr3 may encode QDH. Additionally, Poptr2 and Poptr3 appeared to be specific for NAD^+ since activities for both shikimate and quinate were reduced when $NADP^+$ was used as a cofactor. To validate that no endogenous *E. coli* SDH or QDH were co-purified, an unrelated protein of the same size (a glycosyl transferase family protein from poplar) was purified in parallel, which resulted in no SDH or QDH activity (Figure 2.7 and Figure 2.8).

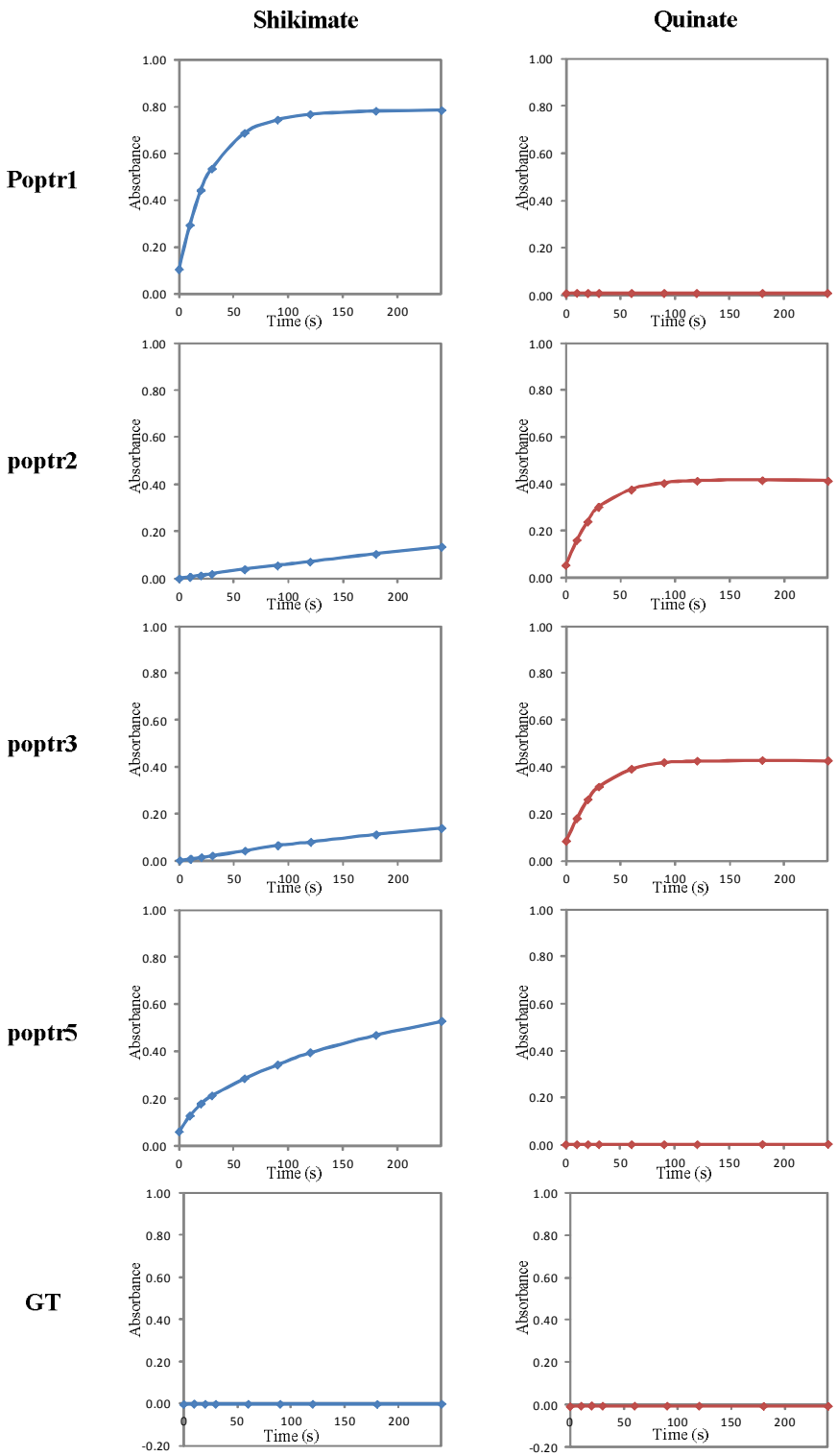


Figure 2.8: Enzyme activities of recombinant Poptr1, Poptr2, Poptr3 and Poptr5 at saturating substrate concentrations (1 mM shikimate or 10 mM quinate). Cofactor (NADPH or NADH) formation was measured spectrophotometrically at 340 nm for 4 min. A purified glycosyl transferase family protein (GT) was used as a negative control showing that no endogenous *E. coli* SDH or QDH were co-purified.

2.3.3.2 Determination of dehydratase activities and confirmation of dehydrogenase activities using HPLC

Traditionally, SDH (QDH) activities are determined by measuring NADPH or NADH formation during the dehydrogenation of shikimate (or quinate) to form 3-dehydroshikimate (or 3-dehydroquinate). This is justified when only one type of reaction is present (SDH or QDH). However, the situation is more complex if an enzyme has both SDH and QDH activity and/or additional dehydratase activity (either with quinate, i.e. QD or with 3-dehydroquinate, i.e. DQD). For example, co-factor consumption when using quinate as a substrate can be due to QDH activity (as assumed), but it can also be due to a combined QD and SDH activity. Vice versa, apparent SDH activity with shikimate could also be due to combined QD and QDH activities (Figure 1.5). Thus, theoretically Poptr2 and Poptr3 may not have QDH activity, but QD and SDH activities instead. Inversely, it could theoretically also be possible that an apparent SDH activity observed with shikimate is instead due to combined QD and QDH activities. To distinguish between these possibilities and to validate the photometric assay results, recombinant enzymes were mixed with different combinations of substrates and cofactors, and reaction products were assessed by HPLC (Figure 2.9, Table 2.5, Appendix B).

HPLC analysis confirmed the bi-functionality of Poptr1 and Poptr5: they were able to catalyze the production of shikimate from 3-dehydroquinate via 3-dehydroshikimate. Also the reverse reactions could be catalyzed by the same enzyme. Neither enzyme was able to use quinate as a substrate in the presence or absence of a cofactor (i.e. no apparent QD or QDH activity). This confirmed the specificity of Poptr1 and 5 towards shikimate

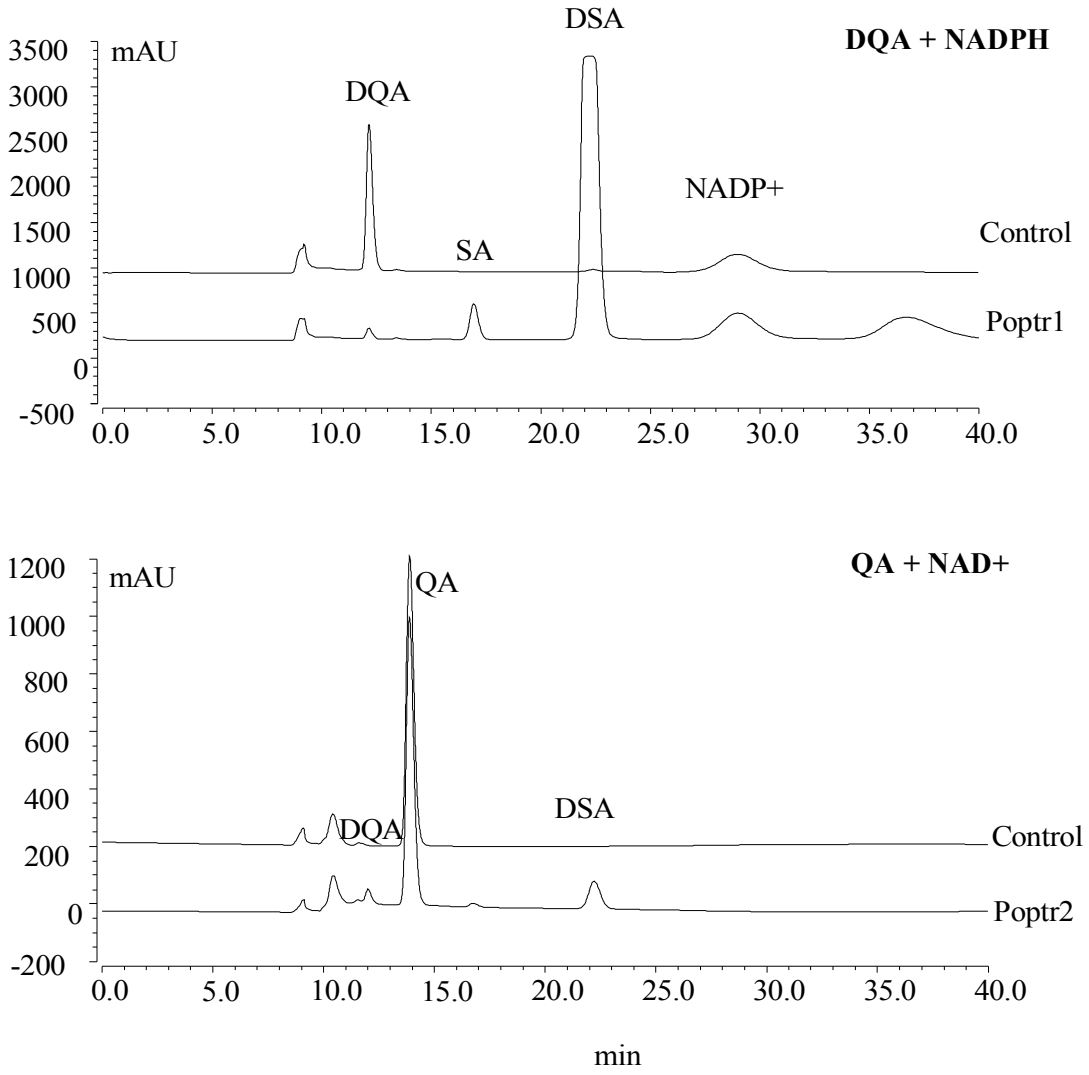


Figure 2.9: HPLC elution profiles of enzyme assays catalyzed by recombinant poplar DQD/SDH1 (Poptr1, top) and QDH1 (Poptr2, bottom).
Substrates, products, and cofactors were separated using HPLC equipped with a diode array detector. Poptr1 enzyme incubated with 3-dehydroquinate (DQA) and NADPH resulted in the formation of shikimate (SA) and 3-dehydroshikimate (DSA), confirming DQD/SDH activity. Poptr2 was incubated with quinate (QA) and NAD^+ resulting in the formation of 3-dehydroquinate (DQA) and 3-dehydroshikimate (DSA), confirming DQD/QDH activity. Boiled enzymes were used in control experiments. Additional HPLC results from a systematic evaluation of all four anticipated activities (Table 2.5) are shown in Appendix B.

Table 2.5: Enzymatic activities of poplar recombinant DQD/SDHs (Poptr1 and Poptr5) and QDHs (Poptr2 and Poptr3) determined with HPLC.
Substrates and cofactor were used at saturating concentrations and are given in the first two columns, which are followed by the possible reactions and enzymatic activities hypothesized. Products (SA: shikimate; QA: quinate; DSA: 3- dehydroshikimate; DQA: 3-dehydroquinate) identified based on authentic standards through HPLC analysis from the given substrate (first column) are indicated as '+', compounds not detectable are indicated as '-'.

Substrate used	Cofactor used	Possible reactions (enzyme activity)	Products formed				Enzyme activity detectable			
			DQA	DSA	SA	QA				
Poptr1 and Poptr5										
DQA	NADPH	DQA → QA --> SA(QDH, QD); DQA → DSA → SA (DQD, SDH)		+	+	-	DQD	SDH	No QD	No QDH
DQA	-	DQA → DSA (DQD)		+	-	-	DQD			
DSA	NADPH	DSA → SA → QA (SDH, QD); DSA → DQA → QA (DQD, QDH)	+		+	-	DQD	SDH	No QD	No QDH
DSA	-	DSA → DQA (DQD)	+		-	-	DQD			
SA	$NADP^+$	SA → DSA → DQA (SDH, QDH); SA → QA → DQA(QD, QDH)	+	+		-	DQD	SDH	No QD	No QDH
SA	-	SA → QA (QD)	-	-		-			No QD	
QA	$NADP^+$	QA → DQA → DSA (QDH, DQD) QA → SA → DSA (QD, SDH)	-	-	-				No QD	No QDH
QA	-	QA → SA (QD)	-	-	-				No QD	
Poptr2 and Poptr3										
DQA	NADH	DQA → QA → SA (QDH, QD); DQA → DSA → SA (DQD, SDH)		+	-	+	DQD		No QD	QDH
DQA	-	DQA → DSA (DQD)		+	-	-	DQD			
DSA	NADH	DSA → SA → QA (SDH, QD); DSA → DQA → QA (DQD, QDH)	+		+	-	DQD	SDH	No QD	
DSA	-	DSA → DQA (DQD)	+		-	-	DQD			
SA	NAD^+	SA → DSA → DQA (SDH, DQD); SA → QA → DQA (QD, QDH)	+	+		-	DQD	SDH	No QD	
SA	-	SA → QA (QD)	-	-		-			No QD	
QA	NAD^+	QA →DQA → DSA (QDH, DQD) QA →SA → DSA (QD, SDH)	+	+	-		DQD		No QD	QDH
QA	-	QA → SA (QD)	-	-	-				No QD	

Note: based on photometric assay results, Poptr1 and Poptr5 prefer NADP as a cofactor while Poptr2 and 3 demonstrated higher activities with NAD. Enzymatic activity of each enzyme was tested in the presence of its preferred cofactor with HPLC.

biosynthesis via 3-dehydroshikimate only. HPLC results also confirmed that Poptr2 and Poptr3 were able to act on quinate to produce 3-dehydroquinate using NAD^+ as a cofactor, confirming that the photometric assays indeed measured QDH activity. Also SDH activity was detectable, but it was significantly lower when compared to QDH activity. In the absence of a cofactor, no conversion of quinate to shikimate was observed for either of the two enzymes (i.e. Poptr2 and Poptr3 do not have QD activity). In contrast, both Poptr2 and 3 maintained DQD activity, but at a very low rate. Thus it appeared that rather than gaining QD activity, the DQD activity was maintained, but the low levels of activities suggests that they might be in the process of losing this activity.

2.3.3.3 Determination of optimal pH condition

The apparent absence of QD activity allowed me to distinguish SDH and QDH activities quantitatively using spectrophotometry. pH optima were determined for both SDH and QDH reactions using different buffer systems ranging from pH 5 to 11. Both types of reactions had similar pH optima with all enzymes displaying high activity under basic conditions. The optimal pH for all reactions was determined to be between 8.5 and 9.5, and activities dropped quickly when the pH was shifted from the optimal pH (Figure 2.10).

2.3.3.4 Kinetic analyses of Poplar SDHs and QDHs

Saturation kinetics of each enzyme were determined at various substrate concentrations in triplicates using saturating amounts of cofactor (NADP+ for SDHs and NAD+ for QDHs) under optimal assay conditions (Figure 2.11, Table 2.6). Enzymes exhibited typical Michealis-Menten kinetics, and Km and Vmax values were determined

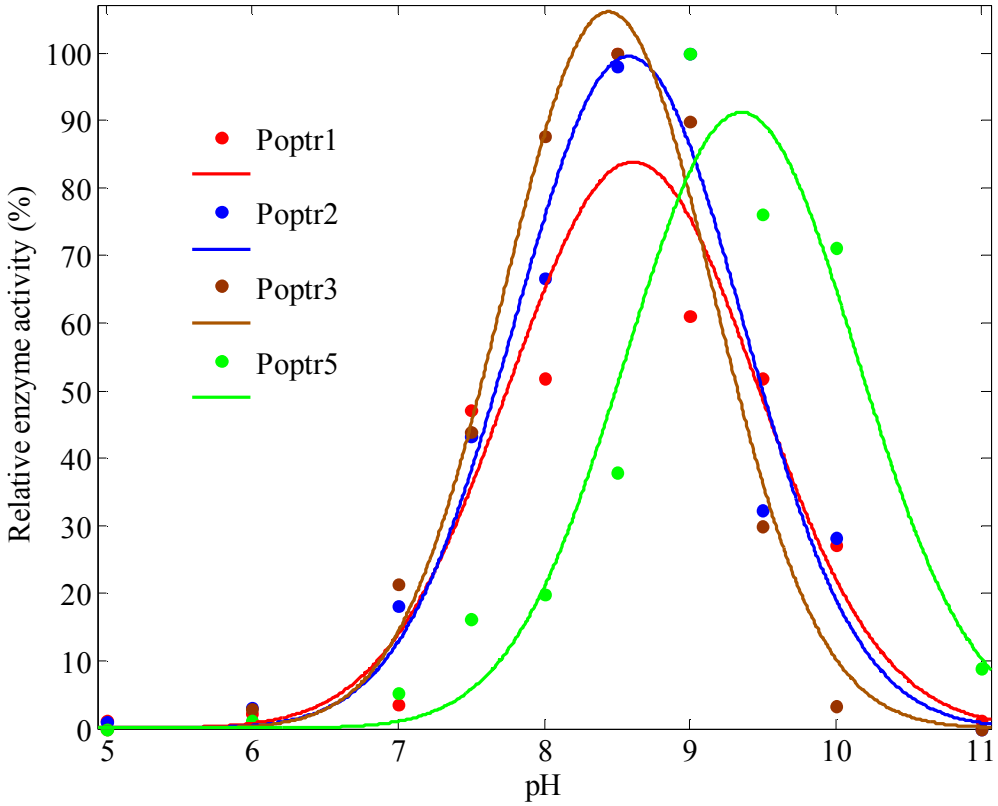


Figure 2.10: Enzymatic activities of poplar recombinant DQD/SDHs (Poptr1 and Poptr5) and QDHs (Poptr2 and Poptr3) as a function of pH.
Enzyme activities were determined spectrophotometrically with substrate concentration of 100 μM. Shikimate was used as a substrate for Poptr1 and Poptr5 while quinate was used for Poptr2 and Poptr3. Relative Activities (the maximal activity of a given enzyme was set to 100%) at pH ranging from 5 to 11 were determined and used to generate pH curves using Curve Fitting Toolbox™ implemented in MATLAB assuming Gaussian distribution.

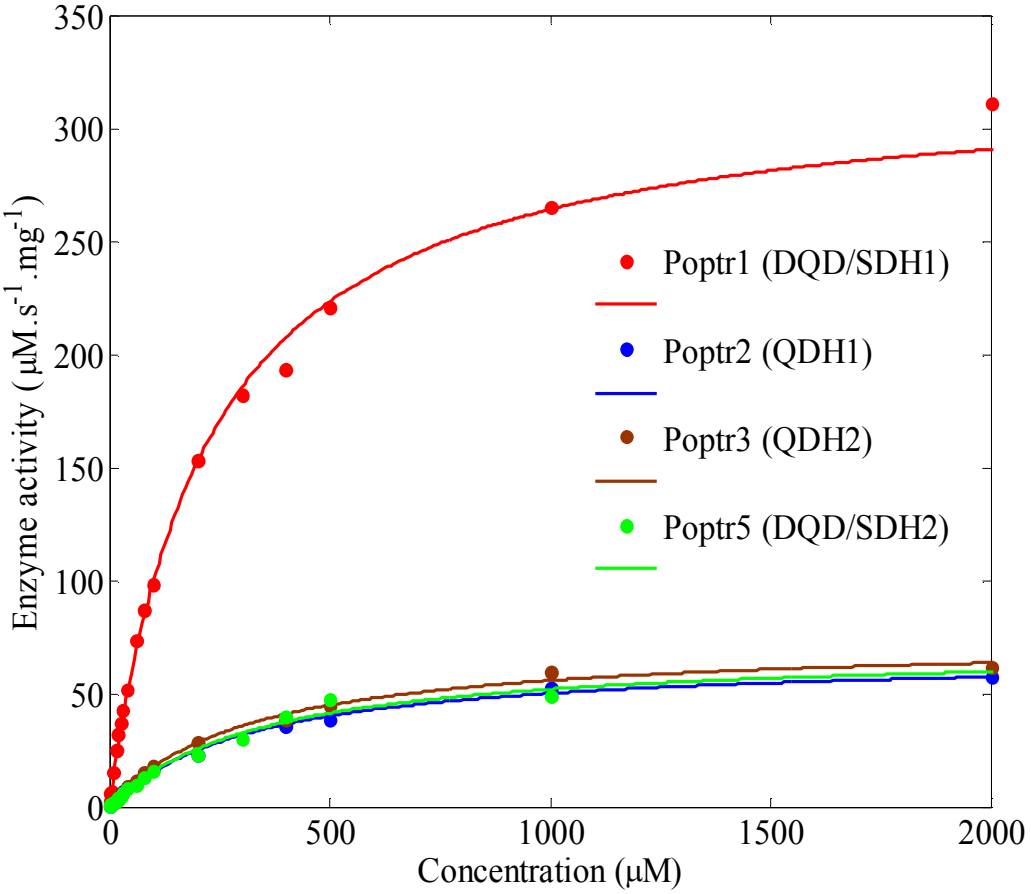


Figure 2.11: Kinetic analysis of poplar recombinant DQD/SDHs (Poptr1 and Poptr5) and QDHs (Poptr2 and Poptr3).
Either shikimate (DQD/SDHs) or quinate (QDHs) was used as a substrate with at least 12 concentrations of substrate (ranging from 0.5 to 2000 μM). The preferred cofactor was kept at saturating concentrations. Kinetic data were averaged across three replicates, plotted and curve-fitted to the Michaelis-Menten (non-linear) model with Curve Fitting Toolbox™ implemented in MATLAB.

Table 2.6: Kinetic properties of poplar recombinant DQD/SDHs (Poptr1 and Poptr5) and QDHs (Poptr2 and Poptr3).
Either shikimate (DQD/SDHs) or quinate (QDHs) was used as a substrate with at least 12 concentrations of substrate (ranging from 0.5 to 2000 μM). The preferred cofactor was kept at saturating concentrations. Vmax and Km values were estimated based on non-linear fitting to the Michaelis-Menten model using Curve Fitting Toolbox™ implemented in MATLAB. Vmax and Km values represent averages of 3 independent replicates. Confidence of curve fitting (i.e. 95% confidence bounds, R^2 values and root mean square error (RMSE)) was assessed with MATLAB. Enzyme substrate specificities were calculates by dividing Vmax by Km.

	Vmax	**Vmax 95%**		**Km**	**Km 95%**		**R^2**	**RMSE**	**Vmax/Km**
	($\mu M.s^{-1}.mg^{-1}$)	**confidence bounds**		**(μM)**	**confidence bounds**				
Poptr1 (DQD/SDH1)	**323.1**	311.9	334.3	**223.1**	198.6	247.5	1.00	6.95	**1.45**
Poptr5 (DQD/SDH2)	**70.0**	60.9	79.0	**345.8**	249.7	442.0	0.99	1.99	**0.20**
Poptr2 (QDH1)	**67.0**	63.4	70.5	**334.4**	287.4	381.5	1.00	1.24	**0.20**
Poptr3 (QDH2)	**73.8**	69.0	78.5	**321.3**	265.9	376.6	0.99	1.67	**0.23**

using a non-linear regression model (Michealis-Menten model). For Poptr1, recombinant enzyme showed a relatively high affinity (Km=223.1 μM) and high enzyme activity (Vmax=323.1 $\mu M \cdot s^{-1} \cdot mg^{-1}$) towards shikimate. In comparison, Poptr5 showed lower shikimate affinity (Km=345.8 μM) and lower maximal activity (Vmax=67.0 $\mu M \cdot s^{-1} \cdot mg^{-1}$) than Poptr1. Using quinate as a substrate, Poptr2 and Poptr3 showed similar kinetic properties (Km=334.4 μM and 321.3 μM, Vmax=67.0 $\mu M \cdot s^{-1} \cdot mg^{-1}$ and 73.8 $\mu M \cdot s^{-1} \cdot mg^{-1}$, respectively). Determination of Km and Vmax for Poptr2 and Poptr3 with shikimate was not feasible because activities with shikimate were undetectable at substrate concentration of less than 100 μM. Based on these enzymatic properties, I assigned the names DQD/SDH1 and DQD/SDH2 to Poptr1 and Poptr5, and QDH1 and QDH2 to Poptr2 and Poptr3, respectively. Poptr4 was assigned the name QDH3 owing to its higher similarity to QDHs than to SDHs.

2.3.4 Expression profiling of Poplar DQD/SDHs and QDHs using combined Affymetrix microarray data

To assess expression profiles of Poplar DQD/SDHs and QDHs across different organs as well as their induction in response to environmental stimuli and genetic modifications, publicly available poplar Affymetrix microarray data (748 hybridizations) were retrieved, normalized, and grouped into organ-specific samples, treatments (compared to the respective controls), and mutant (transgenics) vs. wildtype comparisons. Processed data were extracted from each group to generate expression profiles of DQD/SDHs and QDHs. DQD/SDH1 (Poptr1) is constitutively expressed (at a moderate level) across a wide range of tissues, and it is highly expressed in reproductive organs and

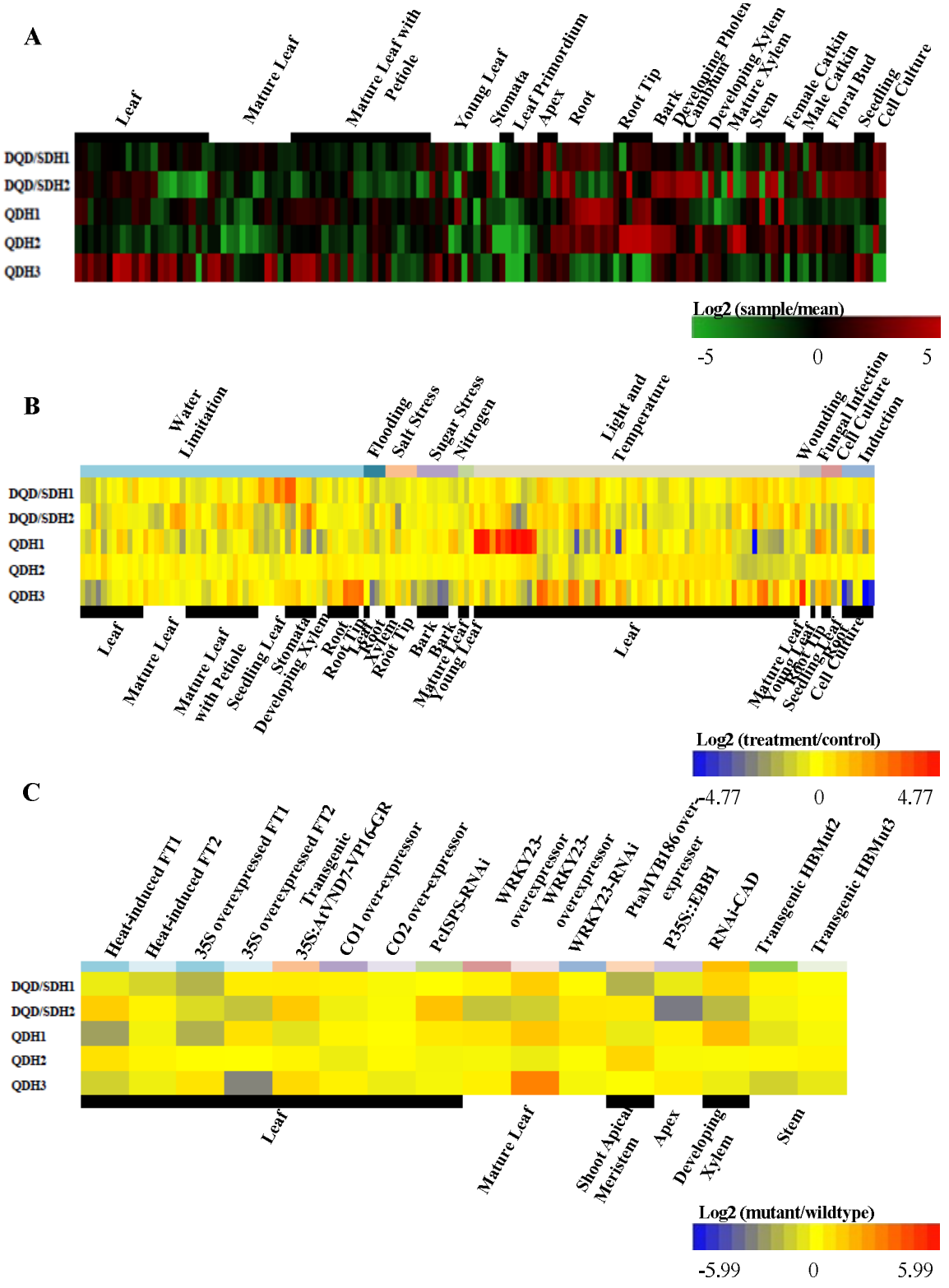

A
Leaf
Mature Leaf
Mature Leaf with Petiole
Young Leaf
Stomata
Leaf Primordium
Apex
Root
Root Tip
Bark
Developing Phoem
Cambium
Developing Xylem
Mature Xylem
Stem
Female Catkin
Male Catkin
Floral Bud
Seedling
Cell Culture
DQD/SDH1
DQD/SDH2
QDH1
QDH2
QDH3
Log2 (sample/mean)
-5
0
5
B
Water Limitation
Flooding
Salt Stress
Sugar Stress
Nitrogen
Light and Temperature
Wounding
Fungal Infection
Cell Culture Induction
DQD/SDH1
DQD/SDH2
QDH1
QDH2
QDH3
Leaf
Mature Leaf
Mature Leaf with Petiole
Seedling Leaf
Stomata
Developing Xylem
Root
Root Tip
Root
Xylem
Root Tip
Bark
Bark
Mature Leaf
Young Leaf
Leaf
Mature Leaf
Young Leaf
Seedling Leaf
Root
Cell Culture
Log2 (treatment/control)
-4.77
0
4.77
C
Heat-induced FT1
Heat-induced FT2
35S overexpressed FT1
35S overexpressed FT2
Transgenic 35S:AtVND7-VP16-GR
CO1 over-expressor
CO2 over-expressor
PcISPS-RNAi
WRKY23-overexpressor
WRKY23-overexpressor
WRKY23-RNAi
PtaMYB186 over-expresser
P35S::EBB1
RNAi-CAD
Transgenic HBMut2
Transgenic HBMut3
DQD/SDH1
DQD/SDH2
QDH1
QDH2
QDH3
Leaf
Mature Leaf
Shoot Apical Meristem
Apex
Developing Xylem
Stem
Log2 (mutant/wildtype)
-5.99
0
5.99

Figure 2.12: Organ-specific expression of poplar DQD/SDHs and QDHs (A), and their responses to changing environmental factors (B) and to genetic modifications (C).

Publicly available large-scale Affymetrix microarray expression data for any *Populus* species were compiled, normalized, and grouped based on experiment annotations. Log-transformed expression ratios (A: Organ-specific expression data are shown relative to the mean expression of the respective gene across all experiments; B and C: treatment and mutant expression data relative to the respective control experiments) were visualized as heatmaps.

actively growing tissues, such as shoot apices and leaf primordia (Figure 2.12A). In contrast, DQD/SDH2 (Poptr5) was found to be expressed to high levels in lignified tissues, while its expression in most other tissue types are negligible. Both QDH 1 and 2 (Poptr2 and Poptr3) were found to be highly expressed in roots. QDH2 is in addition highly expressed in vascular tissues and reproductive organs. QDH3 (Poptr4) showed a distinct expression pattern. Its expression was found predominantly in leaves, bark and seedlings. Quinate and its derivatives are known to accumulate to high levels in leaves and bark. Thus, its expression pattern suggests a potential role of QDH3 in the production of quinate and quinate derivatives in leaves even though this could not be proven biochemically due to difficulties in producing recombinant protein. QDQ/SDHs and QDHs respond differently to environmental stresses and genetic modifications (Figure 2.12B and C). The expression of DQD/SDH1 is induced by water limitation especially in cotyledons and stomata, while the other genes do not seem to be influenced by water shortage. QDH1 showed high sensitivity towards light condition and temperature changes, and its expression was induced by cold temperature at night. Most interestingly, QDH3 showed a marked response to wounding and fungal infection, and its expression level was also increased in plants over-expressing a transcription factor involved in plant defense (WRKY23). This suggests an important defensive role of QDH3.

2.3.5 Co-expression analysis using combined Affymetrix microarray data

Genes encoding the same biochemical function may have distinct physiological functions, and this can be inferred by differential expressions of different isoforms. On the other hand, genes encoding different enzymatic functions, but acting in the same metabolic or physiological pathway may show similar gene expression patterns and

response profiles. In other words, co-expressed genes are likely to be involved in the same physiological process. To further correlate distinct gene functions of DQD/SDHs and QDHs with their physiological roles, a co-expression analysis was carried out using Pearson correlation analysis with Affymetrix microarray expression data. The top 25 co-expressed genes were retained (Figure 2.13, Appendix C). DQD/SDH1 was found to be co-expressed with three shikimate pathway genes (DHAP synthase (annotated as 2-dehydro-3-deoxyphosphoheptonate aldolase in Figure 2.13), EPSP synthase and chorismate mutase), three tRNA synthetases including a phenylalanyl-tRNA synthetase, and a putative chaperonin involved in protein folding, which suggests its role in aromatic amino acid biosynthesis destined for protein production (primary metabolism). In contrast, results for the other four genes were less informative. A large number of the co-expressed genes encode hypothetical proteins with unknown function, preventing immediate interpretations.

2.4 Discussion

The present work demonstrated that the poplar DQD/SDH family has diverged into two groups with one being specific for shikimate (SDHs) and the other being specific for quinate (QDHs). This is one of the rare examples of having both primary (SDH) and secondary (QDH) functions present in a comparably small family. The three poplar QDHs were originally annotated as DQD/SDHs due to fact that they share significant sequence similarity with previously characterized DQD/SDHs. More careful sequence comparison with functionally characterized DQD/SDHs revealed that QDHs are indeed distinct, and this was validated by molecular modeling and *in vitro* enzymatic analysis. Moreover, DQD/SDHs and QDHs showed different expression patterns and stress

A

B

C

Figure 2.13: Identification of genes that are co-expressed with DQD/SDH1 by doing co-expression analysis with large-scale microarray expression data (A: organ-specific; B: treatment; C: transgenic).
Normalized microarray expression data were subject to co-expression analysis using the 'ExpressionAngler' algorithm. Pairwise Pearson correlation coefficients between a gene of interest and all other genes were calculated across each dataset, and the top 25 genes ($r > 0.5$) were retained. See Appendix C for lists of genes which are co-expressed with DQD/SDH2 and QDH1-3.

responses, which suggests different physiological and ecological roles of these two enzymes in plants.

2.4.1 Determination of key amino acid residues involved in substrate discrimination

Conservation of key amino acid residues between poplar QDHs and Arabidopsis SDH domain was assessed by modeling the active sites from these enzymes. Comparison revealed two distinct sites which were specific for QDHs. Ser338, which has been previously identified to be essential to proper function of SDH in Arabidopsis (Singh and Christendat, 2006), is replaced by Gly in all Poplar QDHs. It has been experimentally proven that mutations at this position in Arabidopsis SDH domain could lead to reduction in reaction rate (Singh and Christendat, 2006). The same Ser-to-Gly replacement has also been found in one of the two tobacco DQD/SDH isoforms (NtDQD/SDH2), and NtDQD/SDH2 has been found to have reduced SDH activity compared to NtDQD/SDH1 (Ding *et al*., 2007). The ability of NtDQD/SDH2 to use quinate as a substrate was not tested in their study, but it was proposed that "the preferred substrate of the enzyme *in vivo* is in fact a derivative of shikimate exhibiting a larger functional group at the C1 position" (Ding *et al*., 2007), essentially describing quinate. Protein modeling also highlighted the importance of this replacement to QDH activity. This Ser-to-Gly conversion has led to the appearance of extra space around the C1 position, which allows quinate to fit in. All these together suggest an indispensable role of this Gly residue to QDH activity. Beside Ser338, one more site was found to be different between poplar QDHs and Arabidopsis SDH: Thr381 (Arabidopsis) is changed to Gly in all QDHs. The role of this residue in catalysis has not been well documented. However, it is noted that this Thr residue is replaced by Ser in YdiB and SDH-like enzymes (Singh and

Christendat, 2006), which are able to take both shikimate and quinate as substrates. Both Gly and Ser have relatively smaller side groups than Thr, and this allows more space to accommodate quinate. It seems likely that Thr has to be changed to a smaller amino acid in order for QDH activity to be gained, possibly by accommodating quinate's less planar cyclohen**ane** ring chair or twist-boat conformation compared to the more planar half-chair cyclohex**ene** conformation of shikimate. The replacement of Thr by Gly may not be absolutely necessary for the gain of this function. However, it is apparently important for QDH activity to be optimized since poplar QDHs with Gly prefer quinate greatly over shikimate as a substrate whereas YdiB having a Ser at this position accepts both quinate and shikimate. It will be interesting to see the effect if Thr is converted into Ala, which is also a small amino acid lacking a hydroxyl group in the side chain (similar to Gly).

2.4.2 Functional characterization of DQD/SDHs and QDHs in poplar

Enzymatic properties of recombinant poplar DQD/SDHs and QDHs were assessed. The observed $Km_{(shikimate)}$ values for poplar SDHs were 223.1 μM (DQD/SDH1) and 345.8 μM (DQD/SDH2). These values are within the $Km_{(shikimate)}$ range of functionally characterized SDHs (Tobacco1: 130$\pm$15μM; Arabidopsis: 685$\pm$36μM; English walnut: 860μM; *Pinus taeda*: 700 μM and 800 μM) (Singh and Chritendat, 2006; Muir *et al*., 2001; Ding *et al*., 2007; Ossipov *et al.,* 2000). Poptr2 and Poptr3 (QDH1 and QDH2) showed smaller $Km_{(quinate)}$ values (334.4 μM and 321.3 μM) when comparing with the Km values of previously characterized (i.e partially purified) QDHs ranging from 1.84 mM to 3.6 mM (Ossipov *et al.,* 1995; Kang and Scheibe, 1993; Ossipov *et al.,* 2000). In other words, Poplar QDHs showed much higher affinities towards quinate. This may be

an indication that they have been more optimized for quinate biosynthesis. Both SDHs and QDHs shared similar optimal pH (between 8.5 and 9.5), which was consistent with the pH range seen in tobacco (9.0-9.4) (Ding *et al*., 2007). $NADP^+$ is preferred by tobacco SDHs as a cofactor. The same preference was observed for poplar SDHs. However, variations in cofactor preference have been reported for QDHs. QDHs from bacteria and plant such as beans and corns (angiosperms) use NAD preferentially as a cofactor (Minamikawa, 1970; Graziana *et al.,* 1980; Kang and Scheibe, 1993). In contrast, QDHs from gymnosperms prefer NADP (Ossipov and Shein, 1986). Thus, it was not surprising that poplar QDHs preferred NAD over NADP as a cofactor. SDH activity has been primarily found in chloroplast, where photosynthesis takes place. NADPH is generated during photosynthetic steps and ready to be used by SDH as a cofactor. In contrast, the location of QDH activity remains unclear. There have been some speculations on where QDHs are localized. However, inconsistency exists and experimental evidences are scarce. The possibility of having different isozymes localized to different subcellular compartments remains high. This may help explain the variations in cofactor preference observed for plant QDHs, which is largely dependent on the availability of a certain cofactor in a specific subcellular compartment.

2.4.3 Association of SDH and QDH activities

In bacteria, there are two types of SDHs: one can react with only shikimate (AroE) while the other is able to take both shikimate and quinate as substrates at similar rates (YdiB) (Michel *et al*., 2003). Based on sequence similarity analysis, both plant SDHs and QDHs are more closely related to bacterial YdiB (sequence identity scores ranging from 0.25-0.30) than AroE (scores ranging from 0.21-0.28) (Table 2.4) although scores are low

in general. Thus, it is tempting to speculate that plants may have inherited both activities initially, at least to some extent. However, one of the two functions may not be experimentally apparent or biologically significant. SDH and QDH activities have been measured in a few plant species. In some plant species, the activities of SDH and QDH are encoded by two separate enzymes (Marsh *et al.,* 2009; Graziana *et al.,* 1980) while these two activities can be present in a single enzyme in some coniferous species (Ossipov *et al.,* 2000). However, in the latter case, it remains unclear if the two activities are encoded by one enzyme with two functional domains or one protein with dual substrate specificities. In this work, two poplar QDHs were presented with residual SDH activity, and sequence analysis and protein modeling supported these two as enzymes with dual substrate specificities. Enzymes which can act on both shikimate and quinate equally have only been found in coniferous species, which appear to have relatively slow evolution rate (Buschiazzo *et al.,* 2012). In contrast, most angiosperms have specialized SDHs or QDHs. Since gymnosperms have evolved at a lower rate than angiosperms, they would probably represent the ancestral situation better. Taken together, this suggests that the common ancestor of seed plants probably had both activities in one enzyme. After gene duplication, angiosperms have evolved fast enough to allow gene duplicates to be specialized in one of the two original functions while slowly-evolving gymnosperms may still be in the process of optimizing either function.

2.4.4 Absence of QD activity

Previous work has demonstrated that quinate can be taken up by intact leucoplasts from pea roots and shuffled to the shikimate pathway for aromatic amino acid biosynthesis (Leuschner and Schultz, 1991). An enzyme (quinate dehydratase/hydrolyase

QD) catalyzing the direct conversion of shikimate from quinate has been isolated and characterized from pea root by the same group (Leuschner *et al.,* 1991; Schmidt *et al.,*1991). The reaction mechanism of QD resembles that of DQD. Poplar QDHs (Poptr2, Poptr3 and Poptr4) contain domains that are similar to DQD. As a result, a hypothesis was made that QD activity may be encoded by the DQD-like domain in poplar QDHs. *In vitro* enzyme assays rejected this hypothesis. DQD activity was maintained in QDHs at extremely low levels. However, QD activity has not been gained since shikimate cannot be converted into quinate or vice versa. It remains possible that QD activity is encoded by a different gene, which shares no similarity with DQD. More work needs to be done to biochemically characterize QD in poplar if it is present. The presence of residual DQD activity in poplar QDHs suggests that it may be in the process of losing it. However, this is not accompanied by the gain of QD activity. Instead, it may have gained an alternative activity, which was not tested here. However, it still remains possible that the presence of the DQD-like domains in poplar QDHs is only for structural reasons.

2.4.5 Organ specific expression of poplar DQD/SDHs and QDHs

The divergent biological roles of shikimate as a precursor in both primary metabolism (e.g. protein biosynthesis) and secondary metabolism (e.g. lignin biosynthesis) suggest that genes encoding the same biochemical function may have different physiological roles. This can be indicated by differential expression of the isoforms: one isoforms may be constitutively expressed in all tissues (specific to aromatic amino acid production for protein biosynthesis) while others may be predominantly expressed in tissues undergoing secondary cell wall biogenesis such as developing xylem (specific to lignin biosynthesis). Likewise, QDH isoforms may be preferentially expressed in leaf or bark tissue due to

their roles in quinate and chlorogenic acid synthesis (chlorogenic acid accumulates in poplar leaves and barks). Expression profiling based on large-scale microarray data revealed that DQD/SDH1 was constitutively expressed across different tissues with highest expressed in actively growing tissues. Along with co-expression analysis result, it suggests that DQD/SDH1 is involved in housekeeping primary function (protein production). DQD/SDH2 was found to be highly expressed in vascular tissues, which suggests a role in lignin biosynthesis. However, this was not further supported by co-expression analysis. DQD/SDH2 was not found to be co-expressed with gene involved in monolignol biosynthesis. QDH1 and 2 were preferentially expressed in roots. Some shikimate pathway enzymes are expressed at high levels in roots (Weaver and Herrmann, 1997), and a role of shikimate-derived compounds was suggested in creating and maintaining the rhizospheric microbial community (Phillips *et al.*,1994). In contrast, the presence of quinate and its derivatives in roots has not been well studied and documented. However, the high expression of quinate specific genes in roots does indicate a role of quinate-derived compounds in underground tissues. QDH3 was found to be expressed predominantly in leaves, bark and seedlings suggesting that QDH3 may be responsible for the production of quinate and its derivatives in plants. Unfortunately, the expression of recombinant QDH3 in bacterial system proved to be unsuccessful, which made *in vitro* determination of the biochemical properties of this enzyme impossible. This may be attributed to the fact that plants and bacteria have different preferences on codons encoding the same amino acid; plant proteins may not be properly expressed due to the lack of a specific codon in the bacterial system. Future work may include trying to

express this protein in a eukaryotic system, and study of recombinant QDH3 will provide more insight into both biological and physiological roles of this enzyme.

2.4.6 Response of poplar DQD/SDHs and QDHs to environmental stresses

The shikimate pathway serves as a bridge connecting primary carbohydrate metabolism with the production of the three aromatic amino acids, which are precursors of plant secondary metabolites playing adaptive and protective roles in plants. As plants thrive through challenges posed by changing environment, their demands for shikimate-derived compounds also change (Weaver and Herrmann, 1997). It is evident that enzymes catalyzing the shikimate pathway are strongly influenced by environmental factors including light, temperature, pathogens, wounding and nitrogen availability. Although the mechanism has not been well understood, the production of DAHP synthase (the first enzyme of the shikimate pathway) has been found to be induced in response to nitrogen starvation. In addition, light is an important factor that influences the expression of shikimate pathway genes. Expression levels of the shikimate pathway enzymes are also affected by biotic stresses. In response to wounding, potato is able to produce more active DAHP synthase proteins which results in an increase in total enzyme activity (Dyer *et al*., 1989). DAHP synthase was also shown to have a remarked response to fungal elicitors in Arabidopsis (Keith *et al*., 1991). In this study, stress profiles of poplar DQD/SDHs and QDHs in response to wounding and fungal elicitors were assessed using extensive microarray expression data. Both DQD/SDHs and QDHs showed response to some levels, but QDH3 was the only one which can respond markedly to both wounding and fungal infection. It was demonstrated previously in this work that QDH3 is preferentially expressed in leaves, supporting its role in the production of quinate and its

derivatives (e.g. chlorogenic acids) in leaves. Chlorogenic acids have been claimed to have both anti-microbial (Desotillo *et al.,* 1998) and anti-fungal activities (Ma *et al*., 2007). Early work has shown that transgenic tobacco plants with suppressed levels of phenylalanine ammonia-lyase (PAL) show low levels of chlorogenate production and subsequently more rapid lesion development than wildtype plants after being exposed to virulent fungal pathogen (Maher *et al*., 1994). Additionally, quinate is also found to accumulate to a high level in immature fruits (of *Amelanchier alnifolia*), and its concentration drops gradually during the ripening process. This suggests a role of quinate in maintaining the astringency in immature fruits, which is necessary in protecting young fruits against herbivory (Rogiers and Knowles, 1997). QDH3, which is potentially involved in quinate biosynthesis, may therefore be important in plant defense. This is supported by the induction of this gene by wounding and fungal infection. Moreover, the expression of QDH3 was found to be induced in poplar plants over-expressing WRKY23, a transcription factor involved in the regulation of plant defense to pathogens and abiotic stresses (Rushton *et al*., 2010). Together this suggests an important role of QDH3 in plant defense.

In conclusion, it was demonstrated in this study that some of the previously annotated DQD/SDHs in poplar are indeed QDHs. It is also the first time that plant QDHs have been cloned and functionally characterized. Despite similarities in reaction mechanism and substrate structure, individual DQD/SDH and QDH isoforms are distinct in terms of their expression patterns and stress profiles, which may be a reflection of different physiological functions. DQD/SDH1and 2 are likely involved in functions (protein production and lignin biosynthesis) that are essential to plant survival, while

QDHs are more likely related to plant adaptation and defense (secondary metabolism). There has been evidence supporting that genes of plant secondary metabolism are derived from primary metabolic genes (Pichersky and Gang, 2000), and here we are presenting an example of it with QDHs (secondary metabolism) being derived from DQD/SDHs (primary metabolism). However, additional investigation is necessary to provide insight into the evolutionary history of both shikimate and quinate biosynthesis.

Chapter 3 Subcellular Localization of Poplar DQD/SDHs and QDHs

3.1 Introduction

Plant cells contain distinct subcellular compartments (organelles) which are responsible for specialized functions including photosynthesis, energy generation, amino acid biosynthesis, etc. The majority of proteins present in different organelles are encoded by the nuclear genome and synthesized as precursor proteins in the cytoplasm. Protein precursors contain targeting signals, which are crucial for correct sorting of these proteins to their appropriate subcellular destinations (Della-Cioppa *et al.,* 1987). There are two types of signaling peptides depending on their positions in a protein: terminal signaling peptides and internal signaling peptides (Lytovchenko *et al.,* 2013; Nilsson *et al.,* 2001). Signaling sequences are recognized by receptor proteins localized on the surface membrane of the destination organelles, and specific protein complexes enable these proteins to travel through membranes. Delivery of proteins may be coupled with protein translation (co-translational translocation; e.g. proteins localized to rough endoplasmic reticulum) or occur after translation (post-translational translocation; proteins with other localizations) (Della-Cioppa *et al.,* 1987).

The chloroplast is an organelle that is found in plants and some other eukaryotic organisms, and it plays indispensible roles in many plant metabolic processes including photosynthesis and biosynthesis of many important metabolites such as carbohydrates, lipids and amino acids (Smeekens *et al.,* 1990). It has been well accepted that chloroplasts have evolved from cyanobateria through endosynbiosis. Chloroplasts are

semi-autonomous organelles that possess their own genome. However, the number of genes encoded in the plastid genome has been reduced dramatically in the course of plant evolution. Many genes have been transferred from the plastid genome to their host nuclear genome (Bhushan *et al.,* 2006; Li and Chiu, 2010; Bruce, 2000). Over 90% of the proteins present in chloroplasts are encoded in the host nucleus. The nuclear-encoded plastidial proteins are synthesized in the cytoplasm as precursor proteins containing N-terminal transit peptides (chloroplast targeting peptides: cTPs), which facilitate protein transport into chloroplasts (von Heijne *et al.,* 1989; Emanuelsson and von Heijne, 2001; Zhang and Glaser, 2002; Shi and Theg, 2013). After being imported, cTPs are cleaved off by proteases inside chloroplasts to produce mature proteins. Import of proteins into chloroplasts is a highly regulated process, which includes several steps: phosphorylation of a specific serine or threonine residue in the cTP region, the formation of a hetero-oligmeric guidance complex, the recognition of cTPs by the translocase complexes located in the outer and inner membranes of plastids (TOC and TIC) (Glaser and Soll, 2004; Jarvis and Robinson, 2004) and translocation through the TOC and TIC complexes (Soll and Schleiff, 2004; Keegstra and Froehlich, 1999; Jarvis and Soll, 2001; Waegemann and Soll, 1996; May and Soll, 2000). Early work has been devoted to determine conserved “homology blocks” in cTPs. However, unlike signaling peptides targeting proteins to other organelles, cTPs do not have a highly conserved sequence motif. One study suggests that Arabidopsis cTPs can be subdivided into at least seven groups with distinct sequence motifs (Lee *et al.,* 2008). Although attempts to identify conserved motifs in cTPs have proven to be unsuccessful, studies of amino acid composition of cTPs revealed that they contain a large number of hydrophobic and

positively charged amino acids but low abundance of acidic amino acids (Zhang and Glaser, 2002; von Heijne *et al.,* 1989; Gavel and von Heijne, 1990). In addition, cTPs are found to share a three-domain structure composed of an N-terminal segment with uncharged amino acids, a central segment lacking acidic amino acids, and a C-terminal part involved in the formation of amphiphilic β strand (Gavel and von Heijne, 1990). Also, a loosely defined cTP sequence motif ((Val/Ile)-X-(Ala/Cys)lAla) was reported by the same group with a relatively low matching rate (9/32) (Gavel and von Heijne, 1990).

A complete set of enzymes involved in the shikimate pathway have been demonstrated to be located in plastids (Bickel *et al*., 1978; Bagge and Larsson, 1986; Benesova and Bode, 1992; Homeyer and Schultz, 1988; Mousdale and Coggins, 1985; Mousdale and Coggins, 1986). However, most of the available evidence is based on detection of enzymatic activities in plastidial fractions of plant extracts or detection of radio-labeled aromatic amino acids in isolated chloroplasts fed with labeled precursors (e.g. PEP). There have only been a few shikimate pathway enzymes with their plastidial localizations confirmed with GFP (green fluorescent protein) fusion protein experiments or protein import assays (Maede and Dudareva, 2012; Ding *et al*., 2007; Della-Cioppa *et al.,* 1986; Kasai *et al.,* 2005; Zhao *et al.,* 2002).

There has been a debate on whether or not there is a second set of shikimate (aromatic amino acid biosynthesis) pathway enzymes present outside of plastids. This is supported by the identification of cytosolic isoforms for some shikimate pathway enzymes including DAHP synthase, DQD/SDH and EPSP synthase (Schmid and Amrhein, 1995, Ding *et al*., 2007; Feierabend and Brassel, 1977). Recent research has demonstrated that chorismate mutase (CM) is present in both plastids and the cytosol, but

all arogenate dehydrogenase (ADH) and arogenate dehydratase (ADT) isoforms, which catalyze the ultimate steps of tyrosine and phenylalanine biosynthesis, respectively, are localized to plastids in Arabidopsis (Rippert *et al.,* 2009). This excludes the possibility of having tyrosine and phenylalanine synthesized in the cytosol. Given that the shikimate pathway plays divergent roles in protein production as well as the biosynthesis of plant secondary metabolites, a dual-pathway hypothesis was proposed that the plastidial shikimate pathway is responsible for the production of amino acids for protein production, while the shikimate pathway located in the cytosol is responsible for the production of precursors of secondary metabolites, which may be involved in plant defense (Herrmann, 1995; Weaver and Herrmann, 1997; Schmid and Amrhein, 1995). Consistently, two DAHP synthase isoforms have been identified in Arabidopsis with one being localized in chloroplasts and the other being cytosolic (Keith *et al*., 1991). These two isoforms are differentially expressed and regulated. The plastidial isoform is constitutively expressed, and this suggests that this enzyme may be involved in primary function (protein biosynthesis). The cytosolic isoform is highly induced by wounding and pathogen infection suggesting a role in plant secondary metabolism (Keith *et al.*, 1991; Maede and Dudareva, 2012). Despite the identification of cytosolic isoforms for some of the shikimate pathway enzymes, it still remains unclear if a whole set of the shikimate pathway enzymes are present in the cytosol.

Determination of protein subcellular localization is essential to understanding protein function. Here, a question regarding the subcellular localizations of poplar DQD/SDHs and QDHs was addressed using both online prediction tools and YFP (yellow fluorescent protein) fusion proteins. It has been shown previously that shikimate

biosynthesis is localized to plastids (Bickel *et al*., 1978; Herrmann and Weaver, 1999; Maeda and Dudareva, 2012), while the subcellular localization of quinate biosynthesis remains unclear. With its substrates (3-dehydroquinate) being produced in plastids and its derivatives (e.g. chlorogenic acids) localized to the cytosol, QDH activity is most likely to be found in either one of these two compartments. Based on subcellular localization prediction and sequence analysis, it was hypothesized that DQD/SDH1 is localized to chloroplasts while the others (DQD/SDH2, QDH1 through QDH3) are more likely to be localized to the cytosol due to the lack of clear targeting signals. Localization studies with recombinant YFP fusion proteins failed to support the plastidial localization of DQD/SDH1. All DQD/SDHs were found to be localized to the cytosol, which was unexpected since the plastidial localization of the shikimate pathway has been well established. This may be attributed to several factors including the use of heterologous transformation systems (i.e. tobacco, Arabidopsis and onion instead of poplar), erroneous gene annotation, or the presence of additional genes. Nevertheless, my data suggest that both poplar DQD/SDHs are indeed localized to the cytosol based on consistent results provided by the three transformation systems (tobacco, Arabidopsis and onion) and the two delivery methods (agroinfiltration and particle bombardment) employed. Also, all threepoplar QDHs were found to be localized to the cytosol, which is consistent with the hypothesis. Further studies including genome sequence analysis and transient expression using a different method or a different system will provide more insight into the potential localizations of poplar DQD/SDHs and QDHs in a cell.

3.2 Methods

3.2.1 N-terminal sequence analysis of DQD/SDHs (from Arabidopsis, tobacco and poplar) and QDHs (from poplar)

A multiple sequence alignment was generated using Bioedit with DQD/SDH and QDH protein sequences from Arabidopsis, tobacco and poplar. This alignment was trimmed to allow maintaining of only the N-terminal regions of all sequences. Glu124 (in Arabidopsis and the alignment) was chosen as the cut point because it had been previously identified as a key substrate binding site of the DQD domain in Arabidopsis and must therefore be maintained in the mature protein. A potential cTP cleavage site should be located within the upstream region of Glu124. The N-terminal regions of all proteins were examined manually for the presence of possible targeting signals, cTPs in particular.

3.2.2 Subcellular localization prediction of poplar QDQ/SDHs and QDHs with online tools

To infer the potential subcellular compartments where poplar DQD/SDHs and QDHs may reside, the full length amino acid sequences of these five proteins were subjected to localization prediction using the following online tools:

TargetP (http://www.cbs.dtu.dk/services/TargetP/; Emanuelsson *et al.,* 2007)

ChloroP (http://www.cbs.dtu.dk/services/ChloroP/, Emanuelsson *et al.,* 1999),

WolfPsort (http://wolfpsort.seq.cbrc.jp/; Horton *et al.,* 2007),

Plant-mPloc (http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/; Chou and shen, 2010),

BaCeILo (http://gpcr.biocomp.unibo.it/bacello/; Pierleoni *et al*., 2007),

CELLO (http://cello.life.nctu.edu.tw/; Yu *et al*., 2006)

and MultiLoc2 (http://abi.inf.uni-tuebingen.de/Services/MultiLoc2; Blum *et al*., 2009). Subcellular localizations of Arabidopsis and tobacco DQD/SDHs have been experimentally proven (Rutschow *et al*., 2007; Ding *et al*., 2007). Therefore, these sequences were included in the prediction test to assess the accuracy of these online prediction tools.

3.2.3 Molecular cloning and generation of YFP fusion constructs

To generate YFP fusion constructs (Figure 3.1), the coding sequences of poplar DQD/SDHs and QDHs were amplified from their corresponding pQE30 over-expression constructs using USER-modified gene specific primers (reverse primers were modified to allow the removal of stop codons) (Table 3.1). Proofreading Pfu Turbo Cx Hotstart DNA polymerase (Agilent) was used in PCR amplification (30 cycles, denaturation at 94 ºC for 30 sec, annealing at 56 ºC for 30 sec, extension at 72 ºC for 120 sec). PCR amplicons were then cloned into a USER modified plant over-expression binary vector with a YFP marker fused to the C terminus (pCAMBIA230035sUYFP) following the standardized USER cloning protocol (Geu-flores *et al*., 2007). Another set of constructs (Figure 3.1) containing only the N-terminal coding regions of poplar DQD/SDHs and QDHs were generated with another set of primers (Table 3.1) following the same experimental procedure. In addition, poplar polyphenol oxidase 11 (PPO11; Tran and Constabel, 2012), which served as a plastidial marker, was cloned into the same expression vector system. The plasmid pCAMBIA230035sUYFP alone was used as a cytosolic marker. All constructs were verified by DNA sequencing.

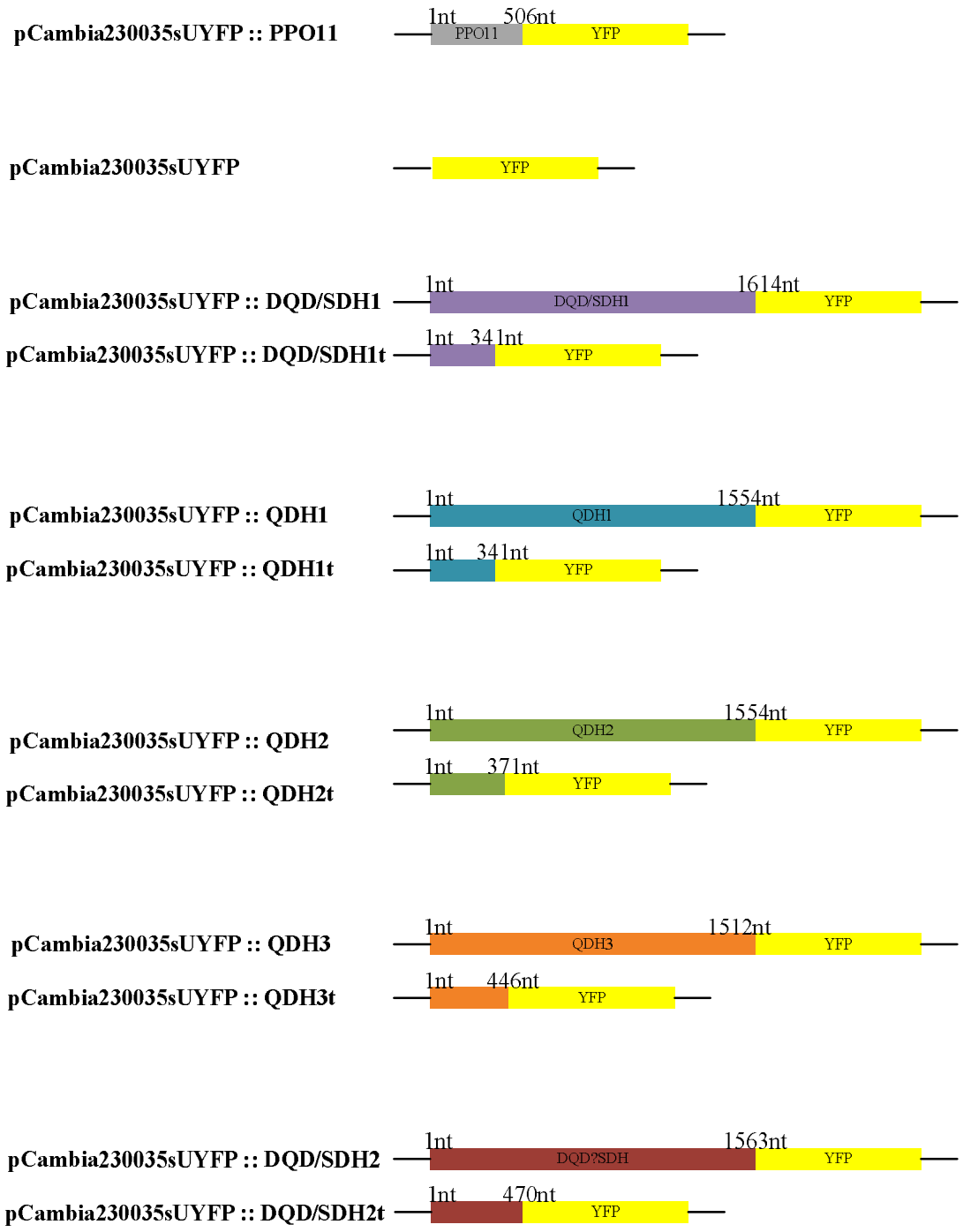


Figure 3.1: Illustration of the YFP fusion constructs (full-length and N-terminal only) used to determine the subcellular localizations of poplar DQD/SDHs and QDHs in transient transformation experiments (agroinfiltration and particle bombardment).
The cytosolic (pCambia230035sUYFP) and plastidic marker (pCambia230035sUYFP::PPO11) constructs are also included.

Table 3.1: USER-modified primers used in the localization studies of poplar DQD/SDHs and QDHs.
Two sets of YFP fusion constructs (5x2) carrying either the full-length or partial (N-terminal) coding sequences of poplar DQD/SDHs or QDHs were generated. A control construct carrying only N-terminal sequence of poplar polyphenol oxidase (PPO11) were generated and used as a plastidial marker in the localization studies. USER extensions are in bold. Sizes of expected PCR products are also provided.

Constructs	Forward Primer (5'-3')	Reverse Primer (5'-3')	Expected PCR Product Size (bp)
pCambia230035sUYFP:: **DQD/SDH1**	**GGCTTAAU**ATGGATTCTGCAAGC AACGTCC	**GGTTTAAU**CCGTACTTTGACATGATCT TCTG	1634
pCambia230035sUYFP:: **DQD/SDH1t**	**GGCTTAAU**ATGGATTCTGCAAGC AACGTCC	**GGTTTAAU**GCTAATCGAAGCGCATCC	357
pCambia230035sUYFP :: **DQD/SDH2**	**GGCTTAAU**ATGGATCTCCAAAGC GCTG	**GGTTTAAU**CCTGTATTCCTTGCTAACA CATCTC	1572
pCambia230035sUYFP :: **DQD/SDH2t**	**GGCTTAAU**ATGGATCTCCAAAGC GCTG	**GGTTTAAU**GCTCCAGTAGCTTGTATCC TG	357
pCambia230035sUYFP :: **QDH1**	**GGCTTAAU**ATGGGGCGTGCTGGG ATC	**GGTTTAAU**CCGAATTTGGCTAGAACAA TCT	1572
pCambia230035sUYFP :: **QDH1t**	**GGCTTAAU**ATGGGGCGTGCTGGG ATC	**GGTTTAAU**CGTACCAGATCAGATGCAA C	357
pCambia230035sUYFP :: **QDH2**	**GGCTTAAU**ATGGGGAGTGTTGGA GTCCTGAC	**GGTTTAAU**CCGAATTTGGCTAAAACAA TCT	1530
pCambia230035sUYFP :: **QDH2t**	**GGCTTAAU**ATGGGGAGTGTTGGA GTCCTGAC	**GGTTTAAU**CCACCATTCTGATGCTTGTT	462
pCambia230035sUYFP :: **QDH3**	**GGCTTAAU**ATGGCATTCAAGAAC AACCTCTTA	**GGTTTAAU**CCAAATTGCTCCAAGACAA G	1581
pCambia230035sUYFP :: **QDH3t**	**GGCTTAAU**ATGGCATTCAAGAAC AACCTCTTA	**GGTTTAAU**CCAGTAGACTGCATACAAG CG	486
pCambia230035sUYFP :: **PPO11**	**GGCTTAAU**ATGGCCTATAACCTT TCT	**GGTTTAAU**AGGACATTGGCTTGCTGTG	522

3.2.4 Transient expression via agroinfiltration

The ten poplar constructs carrying genes of interest, along with the two control constructs (PPO11 construct and plasmid alone), were introduced into *Agrobacterium tumefaciens* strain GV3101. Transformed bacteria were verified by colony PCR and used to inoculate 12X5mL LB broth with selective antibiotics (25 mg/L gentamycin, 25 mg/L rifanpicin and 25 mg/L kanamycin). *A. tumefaciens* cultures were allowed to grow at 28 °C overnight with vigorous shaking. Cells were then harvested by centrifugation at 5000 x G for 15 min. Cell pellets were resuspended in $MgCL_2$ solution (10 mM) containing 200 μM acetosyringone to a final OD600 = 0.4-0.8. Resuspensions were incubated at room temperature for 2 hrs before being used to infiltrate intact young *Nicotiana benthamiana* leaves (referred to as 'agroinfiltration'; Kopertekh and Schiemann, 2005). *N. benthamiana* plants had been maintained in a growth chamber for one month under short-day condition (16 hrs in dark and 8 hrs with light). Three days after infiltration, fluorescent signals due to the expressions of YFP fusion proteins were examined using spectral scanning confocal microscopy as described below. Two other plant systems including Arabidopsis (Wroblewski *et al*., 2005) and poplar were also used in the agroinfiltration experiments to test for the possibility of having proteins localized to different subcellular compartments in different organisms.

3.2.5 Transient expression via particle bombardment

Gold microcarriers with a diameter of 0.7μm were coated with 1μg of plasmid DNA containing constructs of interest (as described above) following the manufacturers' instruction (Biorad). DNA-coated gold particles were bombarded onto onion epidermal

cells using the Biorad He/PSD 1000 system at 900 psi with a shooting distance of 6 cm (Kikkert *et al*., 2005; Tran and Constabel, 2012). Tissues were maintained on MS medium at room temperature in dark after particle bombardment to allow cell recovery and the expression of YFP fusion proteins. YFP signals were examined three days after bombardment.

3.2.6 Detection of YFP signal with confocal microscopy

Transformed plant tissues were used to make wet mounts, and the subcellular localizations of poplar DQD/SDHs and QDHs were assessed by detecting YFP signals using a Nikon C1si spectral scanning confocal microscope at an excitation wavelength of 488 nm. Plant materials were viewed with a 10X objective lens (dry, plan fluor) or a 60X objective lens (oil immersion, plan fluor) depending on sizes of transformed cells that were examined (i.e. 10X lens for onion cells and 60X for tobacco and Arabidopsis cells). Multiple cells as well as different cell types (mesophyll and epidermal cells) were viewed from the same sample to ensure the accuracy of these experiments, and imaging signals were processed by eliminating signals at wavelengths other than 525nm (emission wavelength of YFP) and 630nm (chlorophyll auto-fluorescence) with EZ-C1viewer.

3.3 Results

3.3.1 N-terminal sequence analysis revealed a potential cTP for poplar DQD/SDH1

The N-terminal regions of poplar DQD/SDHs and QDHs as well as Arabidopsis and tobacco DQD/SDHs were aligned and subjected to manual inspection for possible cTPs. Arabidopsis DQD/SDH, which is suggested to have a plastidial localization based

on sequence analysis, has a long N-terminal extension (Figure 3.2). Tobacco isoform 1 (Nicta DQD/SDH1), which has been experimentally proven to be plastidial (Ding *et al*., 2007), has a much shorter cTP comparing with Arabidopsis DQD/SDH. No sequence conservation was observed between these two cTPs. Poplar DQD/SDH1 has an extension at its N terminus compared to the other four poplar isoforms, and it is about the same length as the tobacco cTP (isoform 1). However, it does not share sequence similarity with either Arabidopsis cTP or tobacco cTP. Over-representation of Alanine and Serine in this region, a distinct signature of cTPs (Zhang and Glaser, 2002), suggests that poplar DQD/SDH1 may be localized to plastids. Poplar QDH1 and QDH2 share high N-terminal sequence similarity with tobacco isoform 2 (Nicta DQD/SDH2), which had its cytosolic localization confirmed experimentally (Ding *et al*., 2007). This suggests a possible cytosolic localization for poplar QDH1 and QDH2. Also Poplar QDQ/SDH2 and QDH3 do not present any obvious sequence features indicative of any type of signaling peptide, and it is thus likely that they are localized to the cytosol.

3.3.2 Subcellular localization prediction with online tools

With rapid development of computing technologies and an increasing pool of proteins with known localizations, tools have been developed to predict the subcellular localization of a specific protein. Subcellular localizations of Arabidopsis, tobacco and poplar DQD/SDHs and QDHs were assessed with publicly available prediction tools designed for plant proteins (Table 3.2). Except for Arabidopsis DQD/SDH, no consistent results were obtained for the other proteins by different prediction tools. Notably, the experimentally proven plastidial localization of tobacco DQD/SDH1 was not supported by 'ChloroP', a tool designed for the prediction of chloroplast-targeting proteins with

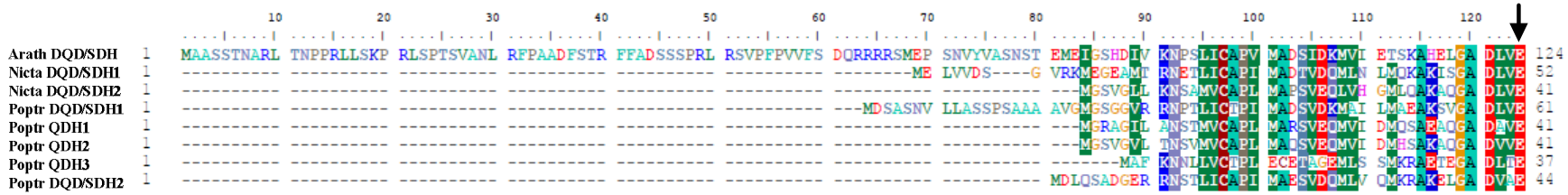


Figure 3.2: Alignment of the N-terminal regions of Arabidopsis (Arath), tobacco (Nicta) and Poplar (Poptr) DQD/SDHs and QDHs.

A multiple sequence alignment was generated using ClustalW implemented in Bioedit with DQD/SDH and QDH protein sequences from Arabidopsis, tobacco and poplar. This alignment was trimmed to allow maintaining of only the N-terminal regions of all sequences. Glu(E)124 in Arabidospsis (indicated by an arrow) has been experimentally proven to be a substrate binding site of the DQD domain, and cTP cleavage sites have to be located within the upstream region of this site (Glu124) if proteins are targeted to the chloroplast.

Table 3.2: Subcellular localization prediction of poplar DQD/SDHs and QDHs using online tools.
Arabidopsis and tobacco DQD/SDHs (gray shaded) with their localizations well established were also included in the prediction study to assess the accuracy of these online tools.

Prediction tool	Arath DQD/SDH	Nicta DQD/SDH1	Nicta DQD/SDH2	Poptr DQD/SDH1	Poptr QDH1	Poptr QDH2	Poptr QDH3	Poptr DQD/SDH2
ChloroP	chloroplast			chloroplast				
WolfPsort	chloroplast	cytoplasm	chloroplast	chloroplast	chloroplast	peroxisome	chloroplast	cytoplasm
Plant-mPloc	chloroplast	chloroplast	chloroplast	cytoplasm	mitochondria	chloroplast	chloroplast	chloroplast
BaCeILo	chloroplast	cytoplasm	cytoplasm	cytoplasm	cytoplasm	cytoplasm	cytoplasm	cytoplasm
TargetP	chloroplast		secretory pathway	chloroplast	mitochondria	secretory pathway		
CELLO	chloroplast	cytoplasm	cytoplasm	cytoplasm	mitochondria	cytoplasm	chloroplast	cytoplasm
MultiLoc2	chloroplast	secretory pathway	chloroplast	chloroplast	secretory pathway	secretory pathway	secretory pathway	cytoplasm

high accuracy (Emanuelsson *et al.,* 1999). Tobacco DQD/SDH1 was predicted to be localized to different subcellular compartments (i.e. chloroplast, cytoplasm, secretory pathway) by different tools and only 'Plant-m-Ploc' suggested a chloroplastic localization. Similarly, inconsistent results were observed for poplar DQD/SDHs and QDHs. It is worth mentioning that poplar DQD/SDH1 was predicted by 'ChloroP' to be plastidial, which is consistent with the hypothesis made based on manual N-terminal sequence analysis. The plastidial localization of poplar DQD/SDH1 was also predicted by other tools including 'WolfPsort', 'TargetP' and 'MultiLoc2'. However, this was not supported by 'Plant-mPloc' (the only tool correctly predicting the plastidial localization of tobacco DQD/SDH1), 'BaCeILo' and 'CELLO'. Due to the lack of consistency in these results, no strong hypotheses can be built in terms of the subcellular localizations of poplar DQD/SDHs and QDHs.

3.3.3 Determination of subcellular localization by transient transformation of plant materials

Sequence analysis and localization prediction by computing tools failed to provide strong evidence on the potential subcellular compartments where poplar DQD/SDHs and QDHs may be localized. Therefore, transient transformation experiments (agroinfiltration and particle bombardment) involving a series of poplar DQD/SDH- and QDH-YFP fusion constructs were set up to determine the localizations of these proteins experimentally.

3.3.3.1 Transient expression of YFP fusion constructs in tobacco, Arabidopsis and poplar intact leaves by Agroinfiltration

Two sets of YFP fusion constructs (5x2) carrying either the full-length or partial (N-terminal) coding sequences of poplar DQD/SDHs or QDHs were generated (Figure 3.1). Along with the plastidial marker (PPO11::YFP) and the cytosolic marker (YFP alone), these two sets of poplar fusion constructs were over-expressed under the control of CaMV 35S promoter in intact tobacco leaves by agroinfiltration. Three days after infiltration, YFP fluorescence was examined using confocal laser scanning microscopy. The two marker proteins (YFP only and PPO11::YFP) were sorted correctly to their designated organelles in both mesophyll and epidermal cells (Figure 3.3 and Figure 3.4). For leaves infiltrated with the ten poplar YFP fusion constructs, fluorescent signals were only detected in the cytosol regardless of cell types (mesophyll and epidermal cells) and regardless of whether the full length or the N-terminal parts of the poplar proteins were employed (Figure 3.3 shows results with the full length proteins in mesophyll cells, and Figure 3.4 shows the localization of the N-terminal fusion proteins in epidermal cells). This is contradictory to the hypothesis that at least poplar DQD/SDH1 is plastidial. To test for host effects, experiments were repeated with Arabidopsis leaves, but the same results were obtained: All poplar full length and N-terminal YFP fusion proteins were localized to the cytosol in epidermal cells, while the two marker proteins were found in the expected compartments (Figure 3.5 and Figure 3.6). Subsequently, it was attempted to introduce the twelve constructs into poplar leaves by agroinfiltration to allow the expression of YFP-fused poplar DQD/SDHs and QDHs in their native plant system. However, agroinfiltration of poplar leaves was proven to be unsuccessful, and no

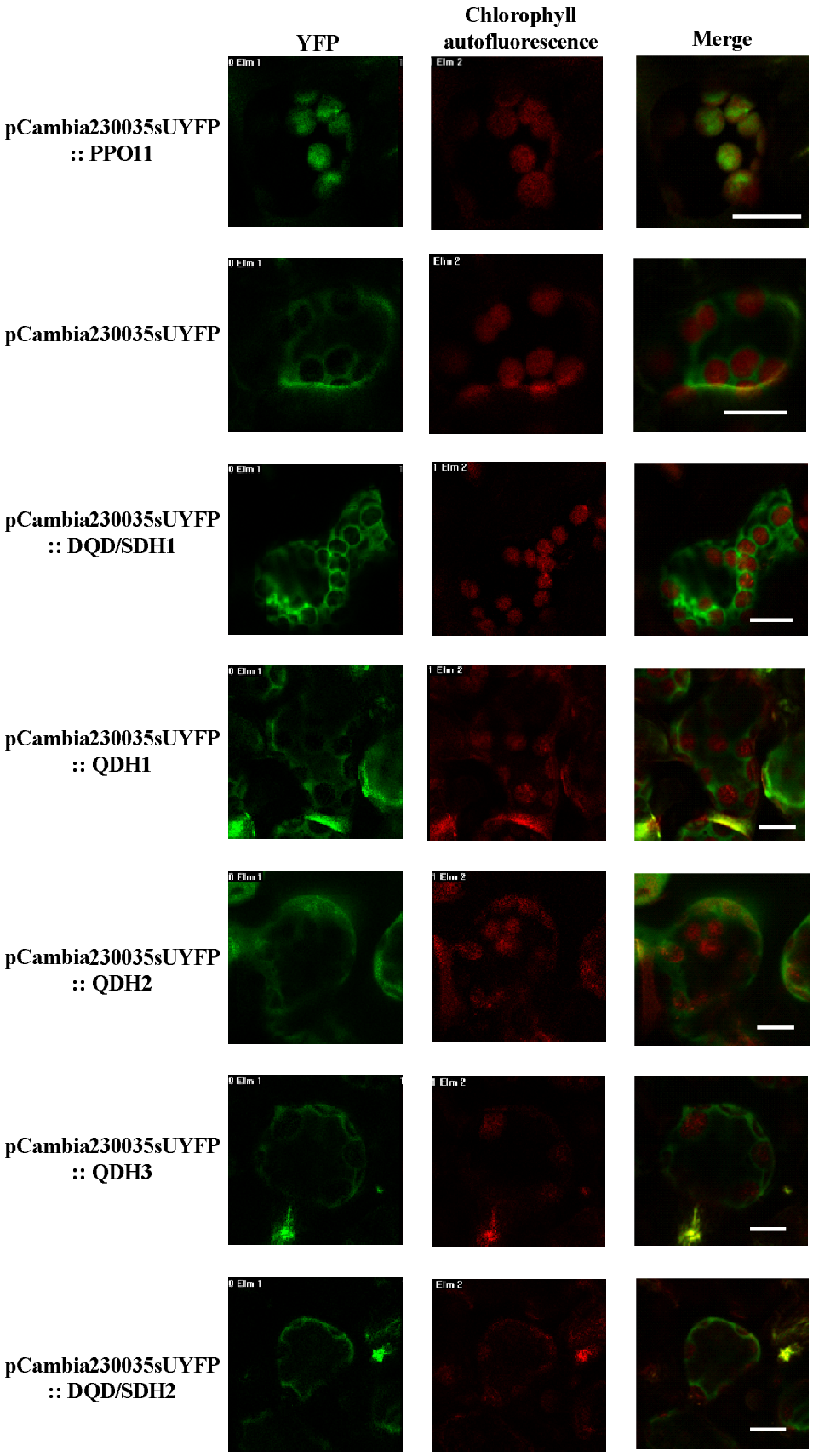

YFP
Chlorophyll autofluorescence
Merge
pCambia230035sUYFP :: PPO11
pCambia230035sUYFP
pCambia230035sUYFP :: DQD/SDH1
pCambia230035sUYFP :: QDH1
pCambia230035sUYFP :: QDH2
pCambia230035sUYFP :: QDH3
pCambia230035sUYFP :: DQD/SDH2

Figure 3.3: Targeting of Poplar DQD/SDHs and QDHs (full-length protein) to the cytosol in tobacco leaves (mesophyll cells).
Full-length YFP fusion constructs were over-expressed in tobacco leaves following agroinfiltration. Signals were assessed three days after infiltration with confocal laser microscopy. Marker proteins (YFP and PPO11+YFP) were sorted correctly to their destination organelles. *Error bar = 10μm*

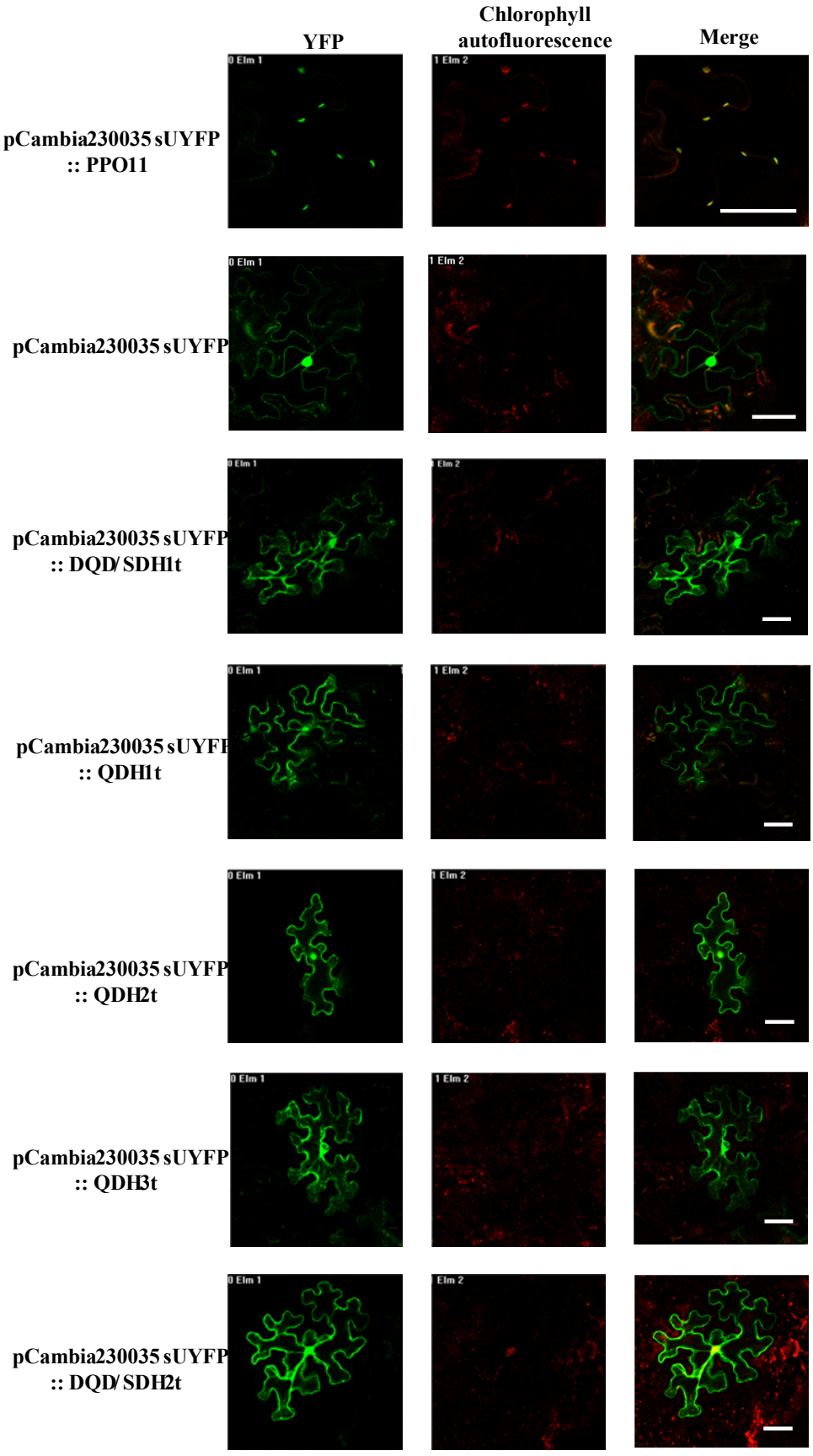

YFP
Chlorophyll autofluorescence
Merge
pCambia230035 sUYFP :: PPO11
pCambia230035 sUYFP
pCambia230035 sUYFP :: DQD/SDH1t
pCambia230035 sUYFP :: QDH1t
pCambia230035 sUYFP :: QDH2t
pCambia230035 sUYFP :: QDH3t
pCambia230035 sUYFP :: DQD/SDH2t

Figure 3.4: Targeting of Poplar DQD/SDHs and QDHs (N-terminal regions only) to the cytosol in tobacco leaves (epidermal cells).
YFP fusion constructs containing only the N-terminal regions were over-expressed in tobacco leaves following agroinfiltration. Signals were assessed three days after infiltration with confocal laser microscopy. Marker proteins (YFP and PPO11+YFP) were sorted correctly to their destination organelles. *Error bar = 50μm*

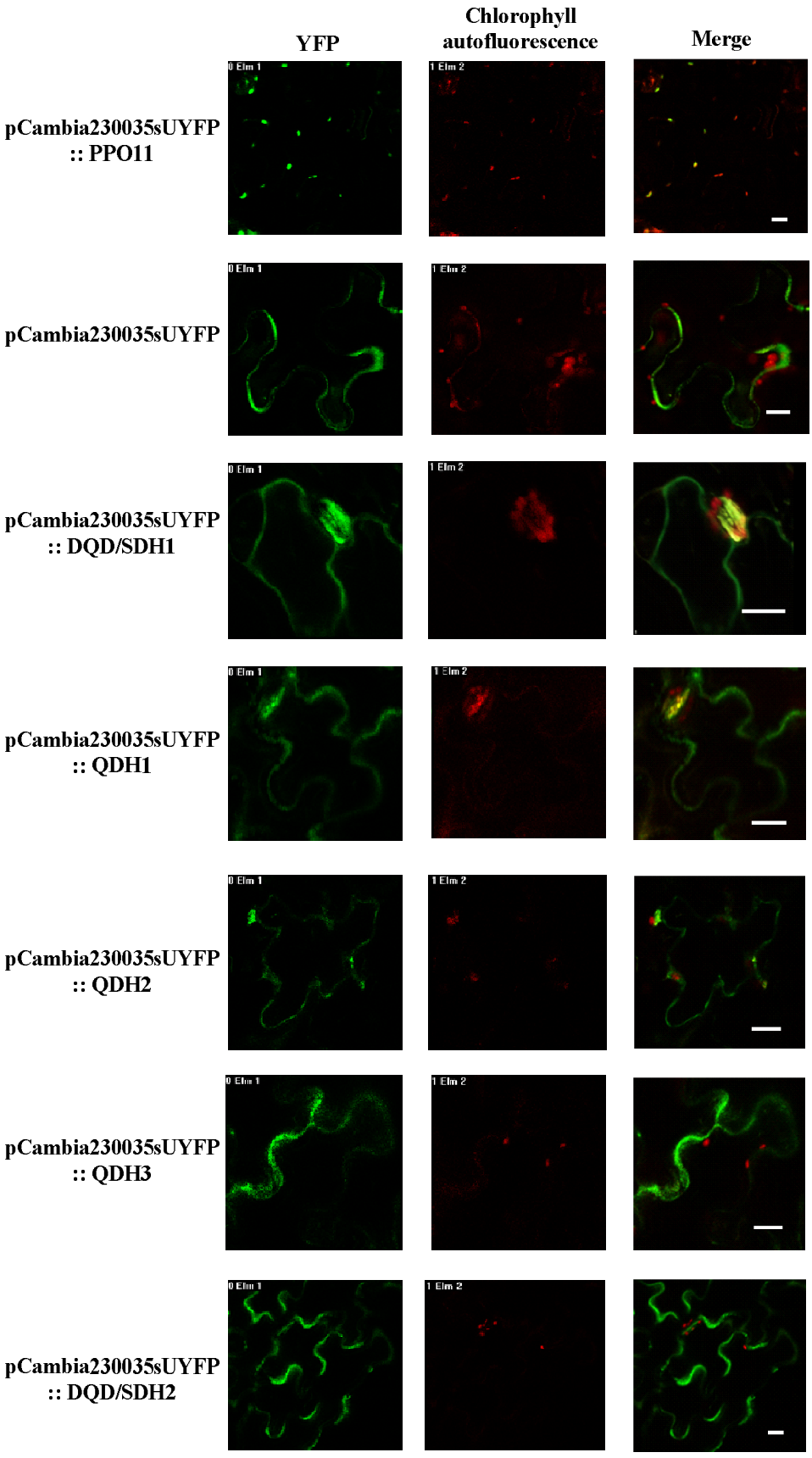

YFP
Chlorophyll autofluorescence
Merge
pCambia230035sUYFP :: PPO11
pCambia230035sUYFP
pCambia230035sUYFP :: DQD/SDH1
pCambia230035sUYFP :: QDH1
pCambia230035sUYFP :: QDH2
pCambia230035sUYFP :: QDH3
pCambia230035sUYFP :: DQD/SDH2

Figure 3.5: Targeting of Poplar DQD/SDHs and QDHs (full-length proteins) to the cytosol in Arabidopsis leaves (epidermal cells).
Full-length YFP fusion constructs were over-expressed in Arabidopsis leaves following agroinfiltration. Signals were assessed three days after infiltration with confocal laser microscopy. Marker proteins (YFP and PPO11+YFP) were sorted correctly to their destination organelles. *Error bar = 10μm*

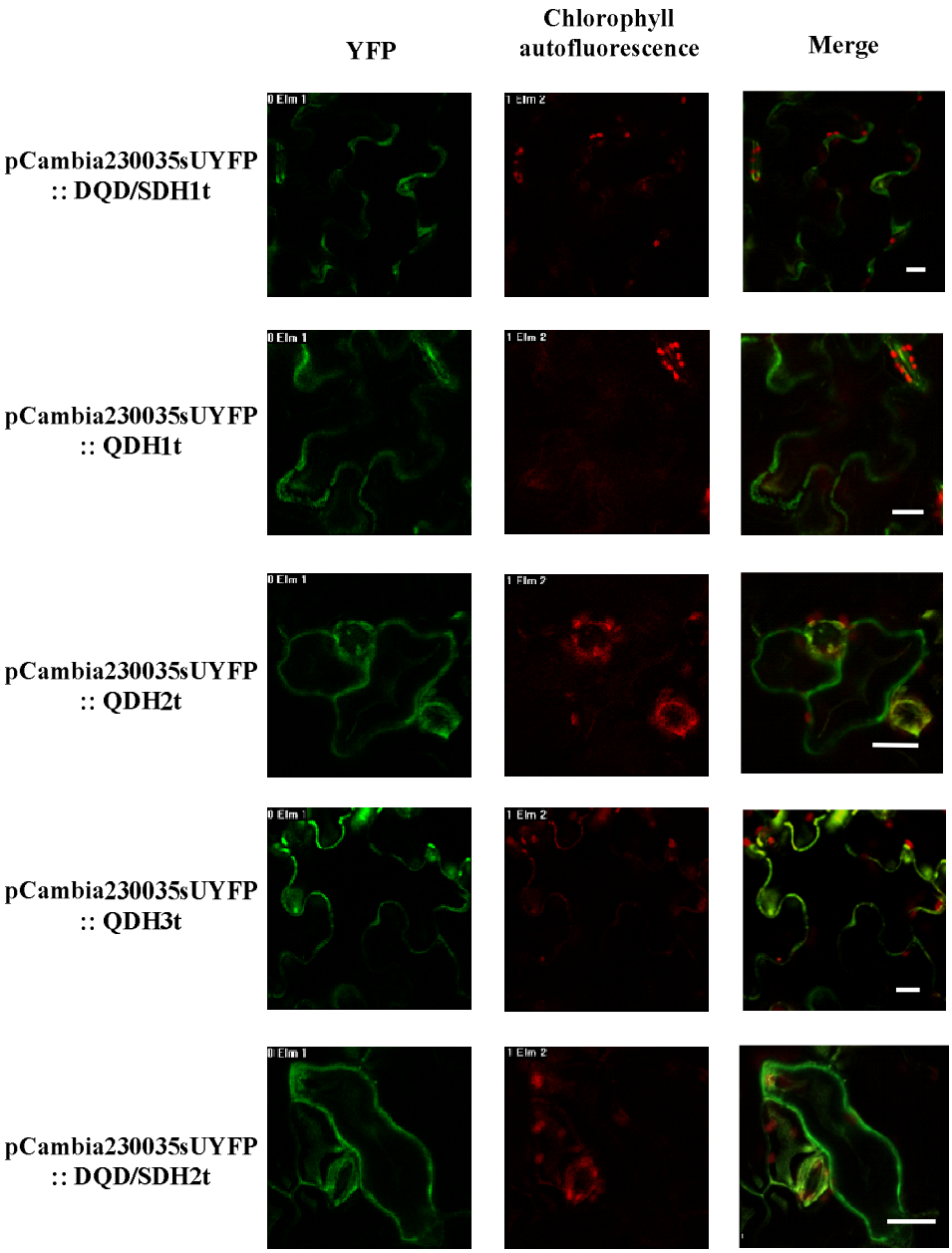


Figure 3.6: Targeting of Poplar DQD/SDHs and QDHs (N-terminal region only) to the cytosol in Arabidopsis leaves (epidermal cells).
YFP fusion constructs containing only the N-terminal regions were over-expressed in Arabidopsis leaves following agroinfiltration. Signals were assessed three days after infiltration with confocal laser microscopy. *Error bar = 10μm*

fluorescent signals were observed three days after infiltration for any construct (not shown).

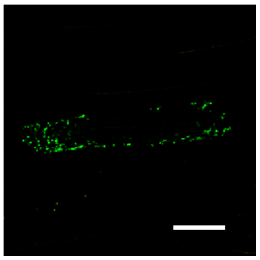
3.3.3.2 Transient expression of YFP fusion constructs in onion epidermal cells by Particle bombardment

Agroinfiltration involves bacteria in facilitating the transfer of genetic material from a binary transformation vector to the plant host nuclei. In contrast, particle bombardment allows direct delivery of plasmids into a plant host without such a carrier. In this study, the ten poplar constructs alongside with the two markers were over-expressed in onion epidermal cells following delivery by particle bombardment. Again, fluorescent signals were monitored three days after bombardment using confocal laser scanning microscopy. Expression of the marker proteins (YFP only and PPO11+YFP) led fluorescence to the expected subcellular compartments (cytosol/nuclei and plastids, respectively) (Figure 3.7). Thus, these markers are targeted to the expected compartments regardless of experimental setups. However, fluorescent signals were again only observed in the cytosol for all poplar DQD/SDHs and QDHs with some staining of nuclei in addition, and this is consistent with results obtained in the agroinfiltration experiments.

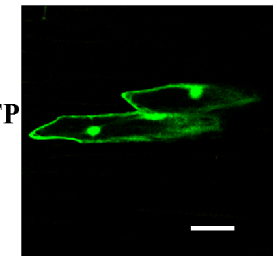
3.4 Discussion

Based on N-terminal sequence analysis, a possible cTP was revealed for poplar DQD/SDH1. For poplar DQD/SDH2 and QDH1, QDH2, and QDH3, no clear signaling peptides could be identified, which suggested a cytosolic localization for all four proteins. Subcellular localization prediction with online tools failed to provide consistent results. However, transient transformation of plants with the YFP fusion constructs did not

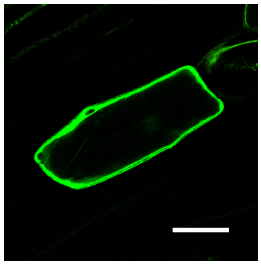
pCambia230035sUYFP :: PPO11



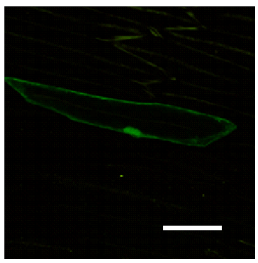
pCambia230035sUYFP



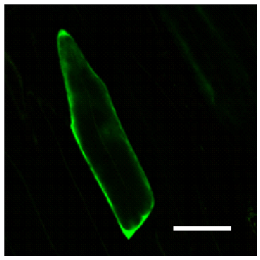
pCambia230035sUYFP :: DQD/SDH1



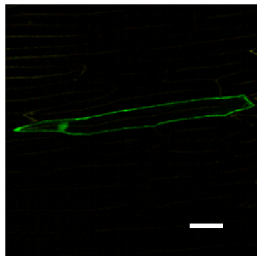
pCambia230035sUYFP :: DQD/SDH1t



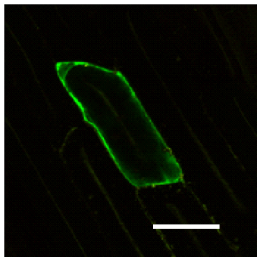
pCambia230035sUYFP :: QDH1



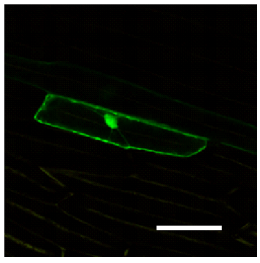
pCambia230035sUYFP :: QDH1t



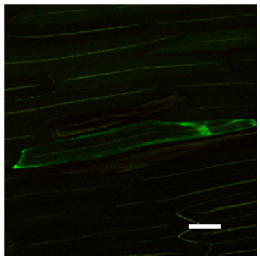
pCambia230035sUYFP :: QDH2



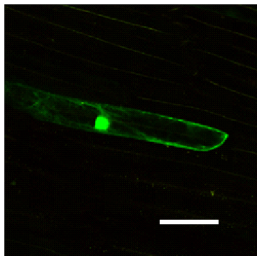
pCambia230035sUYFP :: QDH2t



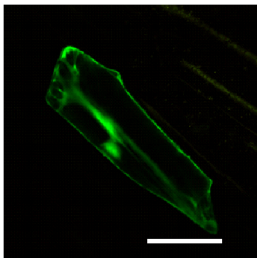
pCambia230035sUYFP :: QDH3



pCambia230035sUYFP :: QDH3t



pCambia230035sUYFP :: DQD/SDH2



pCambia230035sUYFP :: DQD/SDH2t

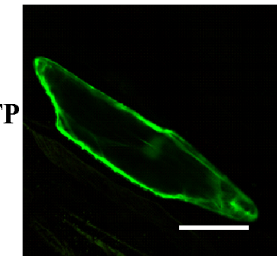


Figure 3.7: Targeting of Poplar DQD/SDHs and QDHs to the cytosol of onion epidermal cells.
YFP fusion constructs containing either complete coding regions or only the N-terminal regions of poplar DQD/SDHs and QDHs were over-expressed in onion epidermal cells following particle bombardment. Signals were assessed three days after confocal laser microscopy. Marker proteins (YFP and PPO11+YFP) were sorted correctly to their destination organelles. *Error bar = 100μm*.

support a plastidial localization of poplar DQD/SDH1; fluorescent signals were only observed in the cytosol for all poplar DQD/SDHs and QDHs regardless of experimental setups. These results are in obvious contradiction to the assumed locale of the shikimate pathway.

3.4.1 Localization of the shikimate pathway enzymes (DQD/SDHs)

Extensive experimental evidence based on both subcellular fractionation and localization experiments conclusively supports the plastidial localizations of a full set of enzymes involved in the shikimate pathway (Bickel *et al*., 1978; Bagge and Larsson, 1986; Benesova and Bode, 1992; Homeyer and Schultz, 1988; Mousdale and Coggins, 1985; Mousdale and Coggins, 1986). It is still a matter of debate if some of these biochemical steps exist in addition outside of plastids. A few localization studies demonstrated that some enzymes of the shikimate pathway (DAHP synthase, DQD/SDH and EPSP synthase) are present also in the cytosol (Schmid and Amrhein, 1995, Ding *et al*., 2007; Feierabend and Brassel, 1977; Rippert *et al.,* 2009). For example, two DAHP synthases have been identified in Arabidopsis, and these two isoforms were found to be localized in different subcellular compartments (plastids and the cytosol) and to demonstrate differential gene expression profiles and stress responses. This is strong evidence supporting the dual-pathway hypothesis, which states that the shikimate pathway present in plastids is involved in primary metabolism (protein biosynthesis) while the cytosolic shikimate pathway is responsible for the production of precursors of secondary metabolites with indispensible roles in plant defense (Herrmann, 1995; Weaver and Herrmann, 1997; Schmid and Amrhein, 1995). This dual-pathway hypothesis suggests that enzyme isoforms with different physiological roles are

potentially localized to different subcellular compartments (i.e. primary metabolism in plastids and secondary metabolism in the cytosol). If this is true, the subcellular localization of a specific shikimate pathway enzyme can be inferred if its physiological function is known. Based on expression profiling results and co-expression analysis (Chapter 2), DQD/SDH1 is most likely involved in housekeeping primary function (protein production) while DQD/SDH2 is suggested to have a role in lignin biosynthesis (secondary metabolism). Therefore DQD/SDH1 was expected to be found in plastids while DQD/SDH2 may be localized to the cytosol. A plastidial localization of DQD/SDH1 was suggested by N-terminal sequence analysis since A possible chloroplast transit peptide was predicted for poplar DQD/SDH1. However, ambiguous results were obtained by subcellular localization prediction analyses with a suite of prediction tools. Proteins were predicted to be localized to different subcellular compartments with different methods, and the plastidial localization of poplar DQD/SDH1 was rejected by some prediction methods. To test the subcellular localizations of poplar DQD/SDHs experimentally, two sets of binary constructs were created, which were transiently over-expressed in different plant systems (tobacco, Arabidopsis, Poplar and onion) using different delivery technologies (agroinfiltration or particle bombardment). Except for poplar, for which transformation did not produce YFP fusion proteins, fluorescent signals were only detected in the cytosol with the other three established plant systems for both poplar DQD/SDH1 and DQD/SDH2, which was not expected given a large amount of experimental evidence supporting the plastidial localization of the shikimate pathway.

3.4.2 Localization of quinate biosynthesis (QDHs)

Quinate biosynthetic enzymes (QDH) have been partially purified and functionally characterized in a few plant species including corn, carrot, mung bean, pea and some coniferous species (Gamborg, 1966; Boudet, 1980, Kang and Scheibe, 1993; Ossipov *et al.,* 1995; Ossipov *et al.,* 2000; Minamikawa, 1977; Refeno *et al.,* 1982; Graziana and Boudet, 1983). However, subcellular localization of QDH has not been well demonstrated. A single study suggested that QDH activity is localized in plastids in mung bean seedlings (Kang and Scheibe, 1993). The subcellular localization of QDHs from other plant species, poplar in particular, remains unknown. Based on N-terminal sequence analysis and the localization studies, poplar QDHs are more likely to be localized in the cytosol. This refines my original hypothesis of having QDHs localized to either plastids or the cytosol.

3.4.3 Difficulties in predicting plastid proteins

Subcellular localizations of Arabidopsis, tobacco and poplar DQD/SDHs and QDHs were assessed with various online prediction tools. Arabidopsis DQD/SDH has been suggested to be plastidial based on sequence analysis (Singh and Christendat, 2006) and was identified in a quantitative chloroplast proteome analysis (Rutschow *et al*., 2008). The localizations of tobacco DQD/SDH1 and isoform 2, which may encode a QDH, have been experimentally determined to be in chloroplasts and the cytosol, respectively (Ding *et al*., 2007). Localizations of the three proteins were predicted with an array of prediction tools, and results were used to assess the accuracy of these tools before using them to predict the potential subcellular localizations of Poplar DQD/SDHs and QDHs.

Arabidopsis DQD/SDH was predicted to be plastidial by all tools. However, variations did exist in the possible localizations of tobacco DQD/SDHs. Notably, tobacco DQD/SDH1 was not predicted by 'ChloroP' as a plastidial protein.

Although a few distinguishing features of chloroplast targeting peptides have been discovered (e.g. low content of acidic amino acid residues and over-representation of Serine and Alanine residues comparing with the mature protein), secondary structures of cTPs have not been well characterized, and sequence conservation is weak within the cTP region (Emanuelsson *et al.,* 2000). These features of cTPs make it very difficult to predict plastidial proteins. In addition, prediction tools are designed using different methods (i.e. 'WolfPsort' is based on amino acid composition of the first 20 residues at the N terminus whereas 'ChloroP' and 'TargetP', which integrates 'ChloroP' results in its prediction, are based on neural networks to detect transit signals and their cleavage sites) (Emanuelsson *et al.,* 2007; Emanuelsson *et al.,* 1999; Horton *et al.,* 2007). Variations in prediction methods may help explain differences observed among the predicting results. Due to the inconsistency among these programs to predict the actual localizations of known DQD/SDHs, it is difficult to judge which tool is more suitable for predicting chloroplast targeting signals in the remainder (i.e. poplar DQD/SDHs and QDHs) and no definite conclusion can be drawn from these results in terms of the subcellular compartments where poplar DQD/SDHs and QDHs are localized.

3.4.4 Protein import into chloroplast

Protein import into chloroplasts is a complex process, and this may help explain the possible miss-sorting of polar DQD/SDHs in non-native plant systems. Most plastidial

proteins are synthesized in the cytosol, and protein import starts with phosphorylation of a specific amino acid (Ser or Thr) within the cTP region of a precursor protein in the cytosol. Phosphorylation leads to the formation of a guidance complex and a distinct secondary structure of the presequence, which is important for signal recognition by the TOC complexes (Soll and Schleiff, 2004). Additionally, translocation of a plastidial protein through the TOC also involves recognitions of its cTP by a GTP-bound TOC34, a subunit of the TOC complex. Finally, translocations though the TIC is dependent on the redox potential of a specific protein, which can be sensed by the TIC complex (Hirohashi *et al*., 2007). The whole process involves multiple levels of recognition, which are essential for correct sorting of proteins into chloroplasts. Since cTPs are highly divergent in terms of their sequences, it is possible that recognition machineries are slightly different across species (e.g. a cTP from poplar may not be well recognized by the tobacco translocation complex). This may be one reason why fluorescent signals were observed in the cytosol as opposed to chloroplasts when the poplar DQD/SDH proteins were expressed in heterologous plant systems. It is also possible that the poplar DQD/SDH precursor proteins containing cTPs were modified by the non-native host post-translationally (i.e. degradation at the cTP region), which could also lead to miss-sorting of a plastidial protein in a different plant system. To test this, the YFP fusion proteins of poplar DQD/SDHs and QDHs were over-expressed in a poplar system. Unfortunately, agroinfiltration of poplar with these poplar YFP fusion constructs was not successful likely due to various reasons including toughness of poplar leaves, agrobacterial strain, infiltration condition, etc. Extensive optimization work will be

necessary to realize a successful infiltration procedure to allow proper transient expression of proteins in poplar.

The use of heterologous systems may be one possible reason for the inconsistency observed between current knowledge and the results of my localization study. However, it is worth mentioning that three different heterologous systems (tobacco, Arabidopsis and onion) were employed in the localization study. All three systems gave consistent results (i.e. fluorescent signals were only detected in the cytosol for all poplar constructs, while the plastidial marker construct yielded chloroplastic fluorescence in all three systems). Additionally, all three plant systems, tobacco in particular, have been extensively used to determine subcellullar localizations of proteins (including plastidial proteins) from a wide range of plant species including poplar (Tran and Constabel, 2012). The PPO11 construct, used as a plastidial marker in this study, is also a poplar protein and was sorted properly in the three plant systems employed. Together these may suggest that poplar DQD/SDHs, DQD/SDH1 in particular, may not be miss-sorted to the cytosol. Instead, these enzymes are indeed localized to the cytosol in poplar, which is not consistent with what is generally accepted. Notably, the plastidial localization of the shikimate pathway has only been well established in a few plant species, but no studies are available for poplar. Furthermore, most of the evidence is derived from subcellular fractionation studies. Therefore the possibility of having the shikimate pathway enzymes localized outside plastids in poplar cannot and should not be completely excluded based on our current understanding.

3.4.5 Mistakes in gene annotation?

The discrepancy between expectations and my localization experiment results could also be explained by errors in gene annotations, which may also explain the inconsistency observed in the localization prediction study. Genome annotation or identification of protein-coding genes is mainly realized by computational tools, which are susceptible to a considerable amount of errors due to the complexity of the task. In this case, the two poplar DQD/SDH gene models may not be correctly annotated, particularly DQD/SDH1. For example, it is possible that there is an upstream exon containing part of the chloroplast targeting sequence, which was not included in the current gene annotation. However, this is not likely given that the N-terminal extension of poplar DQD/SDH1 is about the same length as that of tobacco DQD/SDH1, which has a plastidial localization. In addition, there is no evidence suggesting the expression of a region upstream of Poptr1 up to the previous gene based on publicly available RNAseq experiments and searches against transcript assembly databases (Plant GDB: http://www.plantgdb.org/). Another possibility is that additional genes encoding DQD/SDH may exist in the poplar genome, which has not been identified based on the current poplar genome annotations. However, TBLASTN searches against the third version of the poplar genome scaffolds returned no additional regions (annotated or not) sharing significant similarity with the full length DQD/SDH.

3.4.6 Outlook: Alternative experimentally approaches

Traditionally, protein subcellular localization is assessed by isolating subcellular compartments through ultracentrifugation. The presence of a specific protein in a

subcellular compartment is determined by enzyme activity assays or antibody immunodetection. Localizations of the majority of the shikimate pathway enzymes are identified through this method (Bagge and Larsson, 1986; Benesova and Bode, 1992; Homeyer and Schultz, 1988; Mousdale and Coggins, 1985). It is worth noticing that constraints exist in the preparation of pure subcellular fractions, and ensuring intactness and functional integrity of organelles can be challenging (Quail, 1979). Markers are always necessary to assess the purity of a subcellular fraction. Cell fractionation methods have been widely applied in the field of protein localization studies. However, it may not be applicable in this case. Poplar contains two DQD/SDHs and three QDHs with residual SDH activities, which makes it difficult to probe the final destination of a certain form based on the presence of a certain activity in a specific subcellular fraction (i.e. presence of SDH activity in chloroplasts does not prove the plastidial localization of DQD/SDHs since QDHs still maintain some SDH activity). Nevertheless, such experiments could show if DQD/SDH activity is actually present in chloroplasts in poplar, or if it is restricted to the cytosol as suggested by the apparent localizations of all five proteins to this compartment. Separation of Poplar DQD/SDHs and QDHs using immunodetection will not be easy either due to the requirement of five highly specific antibodies targeting each of the five proteins. It will be difficult to generate antibodies that are specific enough since some of the five proteins do share high sequence similarity (e.g. QDH2 and QDH1 share 88% sequence similarity). This issue can be overcome by working with plants over-expressing His-tagged version of a protein of interest followed by subcellular fractionation and immunodetection involving antibodies targeting His-tag. However, it

has to be kept in mind that the purity of fractions is essential to protein localization determination.

In vitro chloroplast import is another alternative experimental approach to test if a protein is localized to chloroplasts. This involves incubating isolated chloroplasts with radio-labeled precursor proteins. Proteins with potential cTPs can be properly imported, which is indicated by radioactivity observed within chloroplasts (Chua and Schmidt, 1978; Highfield and Elis, 1978; Aronsson and Jarvis, 2011). Since poplar DQD/SDHs and QDHs have been hypothesized to be either plastidial or cytosolic, chloroplast import assay may be a good option in distinguishing plastidial proteins from cytosolic forms. However, pea, spinach and Arabidopsis are the three main plants commonly used in chloroplast import assay with well-established protocols. As mentioned previously, the cTP recognition complex may be species specific, and poplar protein may not be taken up properly by chloroplasts from a different plant system. As a result, optimization of the protocol is necessary to allow isolation of intact and physiologically active chloroplasts from poplar.

In conclusion, sequence analysis revealed a potential cTP for poplar DQD/SDH1. The plastidial localization of DQD/SDH1 was not supported by localization prediction with online tools, which provided ambiguous results. The YFP-based localization experiments suggested that poplar DQD/SDHs are localized to the cytosol. This contradicts our current understanding that the shikimate pathway is plastidial. However, cytosolic localizations of poplar DQD/SDHs were consistently suggested across all three heterologous systems (Arabidopsis, tobacco and onion) and two delivery methods (agroinfiltration and particle bombardment). The possibility of having poplar DQD/SDHs

localized outside of plastids should therefore not be excluded. Poplar QDHs were found in the cytosol, which is consistent with my hypothesis (poplar QDHs are localized to either plastids or the cytosol). As part of the future direction, the frequently updated poplar genomic and transcriptomic data should be continuously interrogated for the unlikely possibility of having additional genes encoding poplar DQD/SDHs or having additional upstream introns containing extended cTPs, which have not been annotated yet. In addition, transient expression using a homologous plant system (i.e. poplar) is necessary to determining the exact localizatoins of Poplar DQD/SDHs. Also, more work is required to validate the cytosolic localization of poplar QDHs.

Chapter 4 Escape from Adaptive Conflict: Evolution of quinate biosynthesis (secondary metabolism) from shikimate biosynthesis (primary metabolism) via gene duplication

*Biochemical characterization of Rhoba, Chlre, Selmo, Anc122, Pinta1 and 2 was performed by Yuriko Carrington under my immediate supervision

4.1 Introduction:

Plants are able to produce a broad array of chemicals with relatively low molecular weights, which are collectively called plant secondary metabolites, specialized compounds, or natural products (Hartmann, 2007). More than 200,000 diverse secondary metabolites are produced across the plant kingdom, but the functional aspects of plant secondary products were largely neglected because they were thought to have no direct effects on cellular survival as opposed to primary metabolites. Secondary metabolites were considered to be metabolic wastes or detoxification products for a long period of time (Hartmann, 2007). Only recently, ecological roles of secondary metabolites were recognized and appreciated as essential to maintaining plant fitness and allowing plants to respond to continuously changing environments. Plant secondary metabolites for example serve as defense compounds (against herbivores, bacteria, fungi and viruses), UV protectants and signaling molecules (e.g. in mediating plant-plant interactions as well as attracting pollinators, seed dispersers, or predators of herbivores) (Hatmenn, 2004; Wink, 1988; Wallace, 2004; Bourgaud *et al.,* 2001, Wink, 2003; Li *et al.,* 2010; Schnee; 2006, Mäntylä *et al*., 2008). They are therefore crucial to plant survival and reproductive fitness. The ability of plants to produce secondary metabolites is likely an adaptive trait, and the diversity of plant secondary metabolism is largely shaped by selection pressure

imposed by ever-changing biotic and abiotic environments during plant evolution. With the advances in plant whole genome sequencing, it became obvious that many gene families encoding metabolic enzymes are largely expanded in plants and that continuous amplifications and retractions of these families are frequent (Pichersky and Gang, 2000). It thus appears that plant chemical diversity has evolved largely through continuous duplications and functional diversifications of a core set of catalysts creating novel substrate/product specificities (Ober, 2005). Furthermore, it became clear that many of these gene families contain members encoding either primary or secondary metabolic functions (Lupien *et al*., 1999; Wang and Pichersky, 1999; Essar *et al*., 1990), which suggests that plant secondary metabolism emerged from primary metabolism. However, the molecular mechanisms, through which primary metabolic genes have evolved to become involved in secondary metabolism, have not been well analyzed.

Phenylpropanoids compose a large group of plant secondary metabolites. These compounds are derived from phenylalanine, an end product of the shikimate pathway (primary metabolism) (Vogt, 2010). The shikimate pathway connects carbohydrate metabolism with the biosynthesis of the three aromatic amino acids (Phe, Try and Tyr), which are essential protein building blocks as well as precursors of a wide array of plant secondary metabolites (Floss, 1979; Herrmann, 1995a&b; Schmid and Amrhein, 1995; Herrmann and Weaver, 1999; Weaver and Herrmann, 1997). It has been estimated that over 30% of photosynthetically fixed carbon is reallocated through this pathway (Maeda and Dudareva, 2012; Tohge *et al*., 2013). Shikimate, the central intermediate and the name giver of the shikimate pathway, is therefore an important primary metabolite. The indispensible role of shikimate in maintaining plant overall fitness is highlighted by the

embryo lethal phenotype of Arabidopsis null mutants (*emb3004*), which have their shikimate biosynthetic gene (3-dehydroquinate dehydratase/shikimate dehydrogenase: DQD/SDH) knocked out (Caruso *et al*., 2008). The lack of the ability of Arabidopsis mutant plants to survive exemplifies the strong selection pressure to maintain shikimate biosynthesis, a major component of primary metabolism. A lack of shikimate biosynthesis is also lethal in bacteria unless supplemented with aromatic amino acids (Pittard and Walace, 1966). In addition, organisms that lack the shikimate pathway (e.g. animals) must take up aromatic amino acids through diet to survive. In plants, shikimate is also required as a precursor and a cofactor in caffeoyl-CoA synthesis, which is the first committed step of G- and S-monolignol biosynthesis in the phenylpropanoid pathway (Boerjan *et al*., 2008). Not only the end products, but also almost all intermediates of the shikimate pathway are potential branch points leading to secondary metabolic pathways (Bentley and Haslam, 1990). For example, shikimate may serve as a branch point leading to the biosynthesis of cyclohexane carboxylates such as ansatrenin and asukamycin (Bentley and Haslam, 1990), which are antibiotics with anticancer properties. The CoA-ester of cyclohexane carboxylate also functions as a potential starting material for the elongation of polyketids, which can be further derivatized and modified into bioactive secondary metabolites (Bentley and Haslam, 1990). In addition, shikimate can be conjugated to various organic acids such as caffeic acid to form bioactive esters.

Quinate, a secondary metabolite sharing high structural similarity with shikimate, is also derived from the shikimate pathway intermediates. I have previously shown that quinate and shikimate biosyntheses are catalyzed by distinct members (DQD/SDHs and QDHs) of the same dehydrogenase family (Chapter 2). Quinate commonly appears as

free acid or esterified forms, which are widely distributed across the plant kingdom. The concentration of quinate can reach up to 14% of the leaf dry mass in developing tissues of some plant species (Leuschner *et al*., 1995; Ossipov *et al.,* 2000). Despite its presence in many plant species, quinate metabolism in plants remained largely enigmatic. Being a derivative of quinate, chlorogenic acid (an ester of quinate and caffeic acid) is gaining popularity as a natural plant dietary antioxidant. It is widely used as a nutraceutical with effects on preventing cancer development and lowering the risk of acquiring cardiovascular diseases (Morton *et al.,* 2000; Laranjinha *et al.,* 1994, Sawa *et al.,* 1999). As a naturally occurring phenolic compound found in many plant lineages, chlorogenic acid is also famous for its role in plant adaptation and defense. It is involved in responding to oxidative stresses (excess photoenergy) as both a direct free radical scavenger and a reducing substrate for guaiacol peroxidise (Grace and Logan, 2000). It is also clear that chlorogenoquinone, an oxidized form of chlorogenic acid, has the ability to bind free amino acids and small proteins. This reduces the bioavailability of amino acids and peptides and therefore results in starvation and fitness reduction of herbivores that ingest chlorogenic acid (Leiss *et al*., 2009). The harmful effects of chlorogenic acid have been reported in studies on different types of insect herbivores including caterpillar, leafbeetle, and leafhopper (Beninger *et al*., 2004, Jassib, 2003, Dowd and Vega, 1996). In addition to its negative effects on herbivores, chlorogenic acid is also known to have inhibitory effects on bacteria, fungi, as well as viruses (Sung and Lee, 2010; Lou *et al*., 2005; Niggeweg *et al*., 2004; Hoover *et al*., 1998).

The ability to produce chlorogenic acid and other plant secondary metabolites is considered to be a defense strategy, which has been acquired by plants during evolution.

Plants live in a complex environment shared by other organisms including herbivores, parasites, and pathogens. As a result, plants are exposed to selection pressures borne by changes in the environment and other organisms. The magnitude and direction of selection is influenced by both biotic and abiotic factors (Agrawal, 2011), and is presumably driven by a continuing process referred to as the evolutionary arms race. For example, if plants evolve a new defense compound that allows them to deter damages by herbivores, this chemical will then quickly increase in abundance. Herbivores will in response evolve a counter adaptation to this novel chemical strategy to lower its effectiveness, which leads to increased selection pressure on plants to further optimize their defense strategies (Woll *et al*., 2013; Mithofer and Boland, 2012; Leimu *et al*., 2012). According to this idea, the coevolving interactions between plants and herbivores have led to the step-by-step optimization of plant defense strategies and the formation of an increasingly complex and well-regulated chemical-based defense system (i.e. plant secondary metabolism) in plants.

Plant secondary metabolism is characterized by some distinguishing features including chemical and structural diversity and expression variations, both spatially and developmentally. Though some secondary metabolites appear wide-spread in plants (e.g. chlorogenic acid), most of these compounds are restricted to specific taxonomic groups. In addition, the production of many secondary metabolites is not constitutive but inducible by biotic and /or abiotic stresses (Hartmann, 1996; Haslam, 1994). Due to the chemical diversity of plant secondary metabolites, a large number of genes are expected to be present in plant genomes encoding enzymes responsible for the generation of those diverse structures and / or for regulating the expression patterns of these enzymes.

However, comparative genomic analyses suggest that plant gene families are highly conserved and that not many new gene families have been acquired in major clades of the plant lineages. In contrast, variation in family size does occur among different species (Rensing *et al*., 2008). It is therefore thought that the structural diversity of plant secondary metabolites results from modifications of a core set of chemicals containing key structural units, which are often intermediates of primary metabolism. All these together suggest that secondary metabolic diversity did not arise from the "creation" of a new set of genes but instead by duplication and/or specialization events of genes involved in primary metabolism (Flagel and Wendel, 2009). Gene duplication drives the recruitment of novel genes for plant secondary metabolism (Pichersky and Gang, 2000; Ober, 2005). By increasing the number of genes within a gene family, this provides plants with working materials for adaptive specializations to occur while maintaining the ancestral function. Gene duplication results in the relaxation of selection constraint on one of the duplicated copies and subsequently allows mutations to accumulate. In most cases, this will lead to a loss of function, but it also can lead to either gain of a new biochemical function or a new expression pattern (neo-functionalization), or separation of multiple existing functions (sub-functionalization) (Hahn, 2009) of a gene product. In the latter case, a "gene sharing" stage is required prior to gene duplication: A secondary function is acquired while the original function is maintained (Hughes, 1994; Schwab, 2002; Bergthorsson *et al*., 2007; Soskine and Tawfik, 2010). A pleiotropic conflict appears when the secondary function becomes beneficial, and more of the secondary function is required to better adapt to the changing environments. This functional pleiotropy can only be resolved by gene duplication event followed by further evolution

towards two genes encoding more specialized enzymes. This chain of events are known as the 'Escape from Adaptive Conflict' model (Hittinger and Carroll, 2007; Des Marais and Rausher, 2008; Conant and Wolfe, 2008; Innan and Konashov, 2010; Kroymann, 2011; Sikosek *et al*., 2012).

The role of gene duplication and divergence as a major contributing factor to evolutionary novelty and adaptation is gaining an increasing amount of attention within the biological research community. However, experimental evidence is still scarce (Des Marais and Rausher, 2008). In this dissertation, the gene family of great interest is the 3-dehydroquinate dehydratase/shikimate dehydrogenase/quinate dehydrogenase (DQD/SDH/QDH) superfamily with members encoding either shikimate or quinate biosynthetic activities. Some members of this family have exclusive specificity towards shikimate (Ding *et al*., 2007; chapter 2), while some have activity with both shikimate and quinate as substrates (gene sharing) (Ossipov *et al.,* 2000). In addition, three out of the five poplar genes, which have been previously annotated as SDHs, instead encode quinate specific QDHs (Chapter 2). All these together suggest that genes involved in quinate metabolism (secondary metabolism) have evolved from shikimate biosynthetic (primary metabolism) genes via gene duplication.

The objective of this work was to elucidate the evolutionary and functional diversification of the DQD/SDH/QDH superfamily in the plant lineages by answering the following questions: When did the gain of quinate activity occur during plant evolution (prior to or after gene duplication)? What was the fate of the daughter duplicates after gene duplication? Did functional changes evolve under positive selection?

Phylogenetic reconstruction of the DQD/SDH/QDH family with species covering the *Viridiplantae* lineage revealed two distinct clades in seed plants, which acted preferentially on either shikimate or quinate whereas lineages separated prior to the angiosperm/gymnosperm split only contain a single copy DQD/SDH. In addition, positive selection was detected on branches leading to both the SDH and the QDH clades. Ser338, Thr381 and Thr407, which may be involved in substrate binding of the SDH domain, were found to have evolved under positive selection at different time points of plant evolution. The sequence of the immediate pre-duplication ancestral DQD/SDH was estimated. Protein structures and biochemical properties of the ancestral recombinant protein (Anc122) and of enzymes from selected extant species (pre-duplication representatives: *Rhodopirellula baltica* (Rhoba), *Chlamydomonas reinhardtii* (Chlre), *Physcomitrella patens* (Phypa) and *Selaginella moellendorffii* (Selmo); after-duplication species: *Pinus taeda* (Pinta1 & Pinta2) and *Populus trichocarpa* (Poptr1 & Poptr3)) were determined. The results suggest that quinate biosynthetic activity was gained prior to gene duplication and remained low for a long period of time during plant evolution. When the need for quinate and its derivatives by plants as defense compounds increased, selection pressure acted to optimize quinate biosynthetic activity, which resulted in a pleiotropic conflict as this affected shikimate biosynthesis. This conflict was then solved by gene duplication just prior to the angiosperm/gymnosperm split followed by further optimization of the defense-related quinate biosynthetic function on one copy and "re-specialization" of the essential shikimate biosynthetic activity on the other copy. Thus, it is most likely that this gene family has diversified through the "Escape from Adaptive Conflict" (EAC) model.

4.2 Methods

4.2.1 Sequence analysis and phylogenetic reconstruction of the DQD/SDH/QDH superfamily

To assess the phylogenetic relationship among plant DQD/SDH homologues, I retrieved the amino acid sequence of functionally characterized Arabidopsis DQD/SDH from the Arabidopsis Information Resource (TAIR, http://arabidopsis.org) (Swarbreck *et al*., 2008) and used it as a query to perform BLASTp searches against various databases in order to identify DQD/SDH homologues from different plant species: NCBI non-redundant protein database, Uniprot, and Phytozome v 9.1 comprising 41 plant genome assemblies. This sequence search returned 124 protein sequences from 61 species (Appendix D) covering bacteria, red algae, green algae and the four major land plant lineages (bryophytes, lycophytes, gymnosperms and angiosperms). 36 amino acid sequences including representatives from the major groups of the green lineages as well three bacterial sequences (outgroup) were used to generate a multiple sequence alignment using Dialign (http://mobyle.pasteur.fr/cgi-bin/portal.py#welcome). This was followed by manual adjustments with BioEdit (Hall, 1999) to allow the removal of poorly conserved regions. The edited alignment was then subjected to phylogenetic reconstruction using both the maximum likelihood method (Proml, PAM model) and the parsimony method (Protpars) implemented in Phylip v3.69 (Felsenstein, 1989). Branch confidence was assessed by bootstrapping with 100 replicates and applied to all trees. An extended phylogeny (Appendix E) was generated with the complete set of 124 amino acid sequences using the same approaches.

The complete protein alignment (124 sequences) was manually trimmed. Only amino acid residues, which are potentially involved in reaction catalysis, substrate and cofactor binding of the S/QDH domains, were retained. This alignment was subdivided into seven alignments, each containing sequences from a defined sub-clade of the DQD/SDH/QDH family (bacteria, green algae, early land plants, gymnosperms DQD/SDHs & S/QDHs and angiosperms DQD/SDHs & QDHs). The seven modified alignments were then subjected to the creation of sequence logos using the 'WebLogo' tool found at http://weblogo.berkeley.edu/logo.cgi (Crooks et al, 2004).

4.2.2 Determination of signatures of positive selection

4.2.2.1 PAML (branch model)

To detect positive selection acting on a specific lineage, particularly branches leading to the DQD/SDH and the QDH clades, cDNA sequences corresponding to the 36 amino acid sequences included in the phylogeny study were aligned manually based on their protein alignment. This alignment was then manually adjusted to allow the exclusion of highly variable regions (5' end and 3'end) and gaps shared by most of the sequences. The maximum likelihood tree (generated as described above) with branches of interest (i.e. branches subtending major lineages) labeled was used as an input tree for further analysis. ω ratios (rate of non-synonymous over rate of synonymous substitutions) for all labeled branches were estimated using the branch model (seqtype = 1, CodonFreq =2, clock = 0, model = 1, NSsites = 0, icode = 0, fix_kappa = 0, fix_omega = 0, fix_alpha = 1, alpha = 0) (Yang, 2004) implemented in the Codeml program (Phylogenetic Analysis by Maximum Likelihood PAML v4.5).

4.2.2.2 Datamonkey (MEME and SLAC)

The branch model assumes that all sites in an alignment share one ω ratio. However, positive selection is frequently episodic (i.e. it only influences a few sites in an alignment or a subset of lineages). As a result, the branch model may not be the most appropriate method as it may not be able to identify episodic positive selection acting on a few sites when most of the other sites are subject to neutral or purifying selection. Positive selection was tested with a Mixed Effects Model of Evolution (MEME) method (Murrell *et al*., 2012) implemented in Datamonkey, a web-based suite of phylogenetic analysis tools (http://www.datamonkey.org/; Delport *et al*., 2010; Pond and Frost, 2005; Poon *et al*., 2009). MEME allows the ω ratio to vary across both sites and branches. The input alignment and the tree were prepared as described above, and were used to determine the most appropriate nucleotide substitution model using the automatic model selection tool (implemented in Datamonkey). With this model, sites subjected to episodic positive selection were identified based on the provided phylogeny with MEME. In addition, substitution map of each site was obtained using the Single Likelihood Ancestor Counting (SLAC) method (also implemented in Datamonkey; Pond and Frost, 2005).

4.2.3 Ancestral reconstruction

To reconstruct the ancestral DQD/SDH/QDH protein sequences, an amino acid alignment with 110 full-length DQD/SDH and QDH protein sequences was generated using Dialign (14 sequences were removed from the original collection due to incompleteness). Manual adjustments were applied to this alignment with BioEdit to allow the removal of gaps and poorly conserved regions (e.g. N-terminal and C-terminal

regions). Following this, a maximum likelihood phylogeny was performed. Subsequently, ancestral reconstructions were carried out based on this phylogeny using the Empirical Bayes (EB) method implemented in the Codeml program of the PAML package v 4.5 (Yang *et al*., 2009) (runmode = 0, seqtype = 2, CodonFreq = 2, model = 2, NSsites = 0, iCode = 0, Mgene = 0, fix_kappa = 0, fix_omega = 0, fix_alpha = 1, alpha = 0, RateAncestor = 1). The sequence data of the immediate ancestral DQD/SDH prior to gene duplication (Anc122) (Appendix F) was extracted from the Codeml output files and then reverse-translated into DNA sequence with BioEdit.

4.2.4 Protein molecular modeling

Protein structures of the ancestral DQD/SDH (Anc122) and selected extant members of the DQD/SDH family (i.e. representative pre-duplication species: *R. baltica* (Rhoba), *C. reinhardtii* (Chlre), *P. patens* (Phypa) and *S. moellendorffii* (Selmo); after-duplication species: *P. taeda* (Pinta1&2) and *P. trichocarpa* (Poptr1&3)) were determined using Phyre (http://www.sbg.bio.ic.ac.uk/~phyre/; Kelley and Sternberg, 2009). The 3D structure of Arabidopsis DQD/SDH (Protein DataBase ID: c2o7qA; http://www.rcsb.org/pdb/home/home.do) coupled with shikimate was used as a template to determine the model coordinates of proteins of interest. The generated models were further visualized and examined with Pymol (http://www.pymol.org).

4.2.5 Subcloning and recombinant protein preparation

The protein coding sequences of all genes of interest (ancestral DQD/SDHs and some extant members of the DQD/SDH superfamily) were codon-optimized for bacterial expression and chemically synthesized by Genescript (Piscataway, NJ, USA).

Recombinant plasmids (pUC57) containing the coding sequences were obtained from Genescript. Cloning by restriction enzyme digest (BamHI & HindIII) was carried out to allow the transfer of DNA fragments from their carrier vector (pUC57) to pQE30 6xHis-tagged over-expression vector. Constructs were validated by DNA sequencing (Operon) before being introduced into the *E.coli* expression system by electroporation. Transformed cells were plated onto LB agar plates with selective antibiotics (100 mg/L ampicillin and 25 mg/L kanamycin) and incubated at 37 °C overnight. Positive colonies verified by colony PCR with gene specific primers were used to inoculate 5mL LB pre-cultures with antibiotics. Pre-cultures were incubated at 37 °C overnight with vigorous shaking and used to inoculate 50 mL main cultures the next morning. Incubation was continued until OD600 = 0.6, and then IPTG was added to a final concentration of 0.06 mM. Protein induction was carried out at 19 °C with shaking for 16 hrs. Cells were harvested by centrifugation at 4000 x G for 20 min and cell pellets were frozen at -80ºC for later analysis.

6xHis-tagged recombinant proteins were extracted and purified under native condition using the Nickel-NTA agarose (Qiagen). Cell pellets were resuspended in 4 mL of the lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, NaOH, pH 8, 1mg/mL lysozyme). Cells were chemically lysed on ice for 1 hr with gentle shaking, which was followed by mechanical disruption by sonication (6X10s bursts with 59s cooling interval between each burst). Cell lysates were centrifuged at 10,000 x g for 30 min at 4°C. Soluble fractions were collected and transferred into purification columns. 1 mL of 50% Ni-NTA agarose was added to each column, and mixtures were incubated on ice for 1 hr with gentle rocking. Column flow-through was released, and nonspecific

binding was minimized with wash buffers without (50 mM NaH_2PO_4, 300 mM NaCl, NaOH, pH 8) and then with imidazole (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, NaOH, pH 8). Purified proteins were eluted with 4X 0.5mL of the elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, NaOH, pH 8) and kept on ice for immediate use.

The purity of recombinant proteins was assessed by SDS-PAGE gel electrophoresis and western blotting. Protein samples were boiled and loaded on a 10% polyacrylamide gel. 50min at 200V was allowed for proper separation of protein bands. Proteins were visualized by staining with Gel Code® Blue Stain Reagent (Thermo Scientific). The transfer of proteins onto PVDF membranes (Millipore) was carried out by electroblotting at 100V for 60min. 6xHis tagged proteins were detected by chemiluminescence using the SuperSignal® West HisProbe™ Kit (Thermo Scientific) following the manufacturer's instruction.

4.2.6 Enzymatic properties of the ancestral and selected DQD/SDHs from extant plant species

Dehydrogenase activities of recombinant enzymes with shikimate (SDH) and quinate (QDH) were determined photometrically by measuring NADPH or NADH production at 340 nm with a spectrophotometer.

4.2.6.1 Determination of enzyme substrate preference

Substrate preferences of the recombinant enzymes were determined by measuring enzyme activities with both shikimate (10 mM) and quinate (10 mM) at pH 8.5 (Trizma base-HCl buffer) and room temperature. Initial velocities (slopes) were calculated and

converted into the appropriate unit ($\mu M\ s^{-1}\ mg^{-1}$) using equation $A=c\epsilon l$, where A is absorbance, c is cofactor concentration, ϵ is NADPH extinction coefficient ($6.2\times10^3\ L\cdot mol^{-1}\cdot cm^{-1}$) and l is the length of light path (1cm). Three replicates with enzymes from three independent purifications were carried out with each substrate for all enzymes, and averages and standard deviations were calculated and plotted with Microsoft Excel 2007.

4.2.6.2 Kinetic properties

Enzymatic properties (V_{max} and K_m) of enzymes exhibiting high catalytic rates on shikimate (Rhoba, Chlre, Phypa, Selmo, Anc122, Pinta1, Pinta2, Poptr1) were determined using ten different concentrations of shikimate ranging from 0.05mM to 5mM while keeping cofactor at saturation. In addition, V_{max} and K_m values of Pinta2 and Poptr3 (quinate specific enzymes) with quinate were determined by varying the concentration of quinate (also from 0.05 mM to 5 mM). Initial velocities at each substrate concentration for each enzyme were calculated using the same method as described above. Three replicates were carried out, and means were calculated with Excel 2007.

4.2.6.3 Data analysis

Kinetic data of each enzyme were plotted and curve-fitted to the Michaelis-Menten model using plotting and curve-fitting tools implemented in MATLAB. Maximal velocity values (V_{max}) and Michaelis-Menten constants (K_m) were estimated. Confidence of curve fitting (i.e. 95% confidence bounds, R-square values and root mean square error) was also assessed with MATLAB. Enzyme substrate specificities were calculates by dividing K_m from V_{max}.

4.3 Results:

4.3.1 Sequence diversity in the DQD/SDH/QDH superfamily

Phylogenetic reconstruction of the DQD/SDH/QDH superfamily was performed with amino acid sequences from bacteria and the *Viridiplantae* lineages (Figure 4.1, Appendix E). Three planctomycete (bacteria) DQD/SDHs were included in the phylogeny study as an outgroup due to their close relatedness to the Plantae DQD/SDHs (Richards *et al*., 2006). Bacteria and all non-seed plants (e.g. *R. baltica*, *C. reinhardtii*, *P. patens* and *S. moellendorffii*) with their genomes completely elucidated only have one *DQD/SDH* gene. In contrast, multiple DQD/SDHs (i.e. Poptr1 through 5, Pinta1&2) are present in seed plants, which have diverged into two clades. Both angiosperm and gymnosperm DQD/SDH isozymes are present in both clades, so it is most likely that a gene duplication event happened prior to the angiosperm/gymnosperm split in early plant evolution (Figure. 4.1). This was then followed by gene specialization/innovation event giving rise to two phylogenetically distinct clades. One of the two clades contains all characterized DQD/SDHs while the other defines a DQD/SDH-like clade with some members being recently characterized as QDHs (Chapter 2). It is noteworthy that the DQD/SDH-like clade does not include an Arabidopsis homologue. Quinate and its derivatives are not known to accumulate in Arabidopsis while they are present in all species represented in the DQD/SDH-like clade. It is therefore plausible to assume that the DQD/SDH-like clade is indeed a QDH clade. It was also noted that some angiosperms (especially monocots) are not present in the QDH clade.

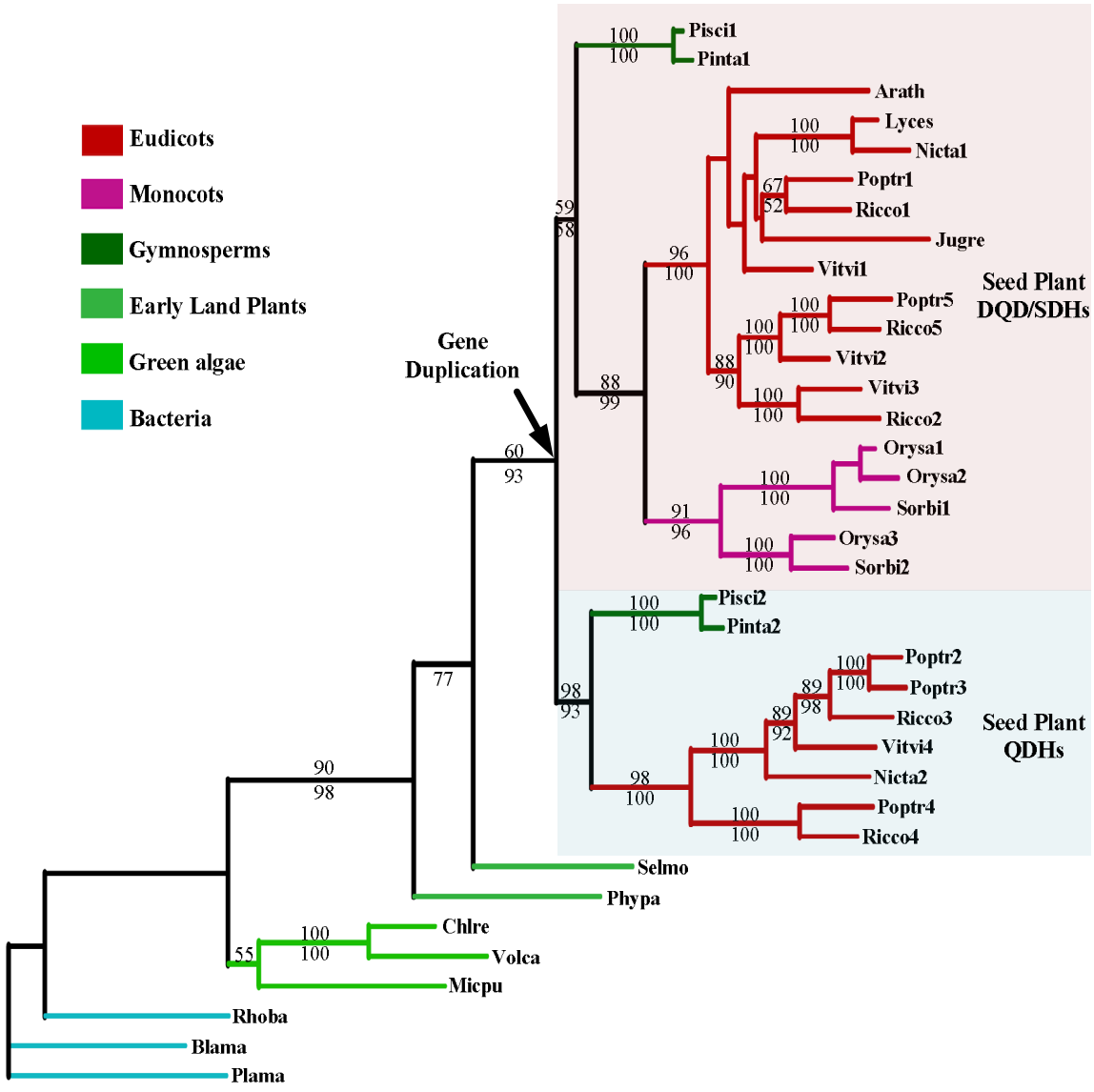


Figure 4.1: Phylogenetic reconstruction of the DQD/SDH/QDH superfamily with amino acid sequences from bacteria (outgroup) and the green lineages.
Tree topology was determined under both the maximum likelihood and the parsimony criteria with Phylip3.69. Branch confidence determined by 100 bootstrapping replicates using both methods are indicated on the tree (i.e. only values>50% are shown; values above the branch are generated with the maximum likelihood method while values below the branch are generated with the parsimony method).

Figure 4.1 abbreviations:
Arath: *Arabidopsis thaliana* DQD/SDH; **Blama**: *Blastopirellula marina* DQD/SDH; **Chlre**: *Chlamydomonas reinhardtii* DQD/SDH; **Jugre**: *Juglans regia* DQD/SDH; **Solly**/**Lyces**: *Solanum lycopersicum* DQD/SDH; **Micpu**: *Micromonas pusilla* DQD/SDH; **Nicta1**: *Nicotiana tabacum* DQD/SDH1; **Nicta2**: *Nicotiana tabacum* DQD/SDH2; **Orysa1**: *Oryza sativa* DQD/SDH1; **Orysa2**: *Oryza sativa* DQD/SDH2; **Orysa3**: *Oryza sativa* DQD/SDH3; **Phypa**: *Physcomitrella patens* DQD/SDH; **Picsi1**: *Picea sitchensis* DQD/SDH1; **Picsi2**: *Picea sitchensis* DQD/SDH2; **Pinta1**: *Pinus taeda* DQD/SDH1; **Pinta2**: *Pinus taeda* DQD/SDH2; **Plama**: *Planctomyces maris* DQD/SDH; **Poptr1**: *Populus trichocarpa* DQD/SDH1; **Poptr2**: *Populus trichocarpa* QDH1; **Poptr3**: *Populus trichocarpa* QDH2; **Poptr4**: *Populus trichocarpa* QDH3; **Poptr5**: *Populus trichocarpa* DQD/SDH2; **Rhoba**: *Rhodopirellula baltica* DQD/SDH; **Ricco1**: *Ricinus communis* DQD/SDH1; **Ricco2**: *Ricinus communis* DQD/SDH2; **Ricco3**: *Ricinus communis* DQD/SDH3; **Ricco4**: *Ricinus communis* DQD/SDH4; **Ricco5**: *Ricinus communis* DQD/SDH5; **Selmo**: *Selaginella moellendorffii* DQD/SDH; **Sorbi1**: *Sorghum bicolor* DQD/SDH1; **Sorbi2**: *Sorghum bicolor* DQD/SDH2; **Vitvi1**: *Vitis vinifera* DQD/SDH1; **Vitvi2**: *Vitis vinifera* DQD/SDH2; **Vitvi3**: *Vitis vinifera* DQD/SDH3; **Vitvi4**: *Vitis vinifera* DQD/SDH4; **Volca**: *Volvox carteri f. nagariensis* DQD/SDH

4.3.2 Sequence analysis and molecular modeling

Examination of the key amino acid residues of the S/QDH domain across different lineages revealed that the two dehydrogenase catalytic sites (Lys385 and Asp423 in Arabidopsis) are highly conserved across all SDHs and QDHs. However, variations in two main substrate binding sites (Ser338 and Thr381) were observed between different phylogenetic groups (Figure 4.2). This suggests that the ability to catalyze dehydrogenation reactions is maintained in all enzymes, while individual groups / clades may have distinct substrate preferences. All bacterial and green algal DQD/SDHs possess Ser (338) and Thr (381). However, the conversion of Ser (338) to Gly was observed in DQD/SDHs from early land plants (bryophytes and lycophytes). This leads to the appearance of an extra pocket in the active site, which potentially allows quinate to fit in (Figure 4.3, Chapter 2). This Ser (338) to Gly conversion was maintained in the gymnosperm DQD/SDH clade while Gly was converted back to Ser in the angiosperm DQD/SDH clade. Members in the QDH clade gained a second mutation (Thr (381) to Gly) while maintaining the first one (Ser to Gly). The second mutation has also led to the appearance of additional space within the reaction pocket, and this could subsequently result in further optimization of quinate activity by accommodating the less planar conformation of the hexene ring of quinate compared to the hexane ring of shikimate, where the double bond promotes a more planer conformation of the ring.

Conservation of the key amino acid residues involved in cofactor binding was also assessed through sequence analysis. The G motif (GXGGXX: Gly460, Gly462 and Gly463 in Arabidopsis), which binds to the pyrophosphate moiety of NADP(H) and NAD(H), was found to be highly conserved across all sequences analyzed (Figure 4.2).

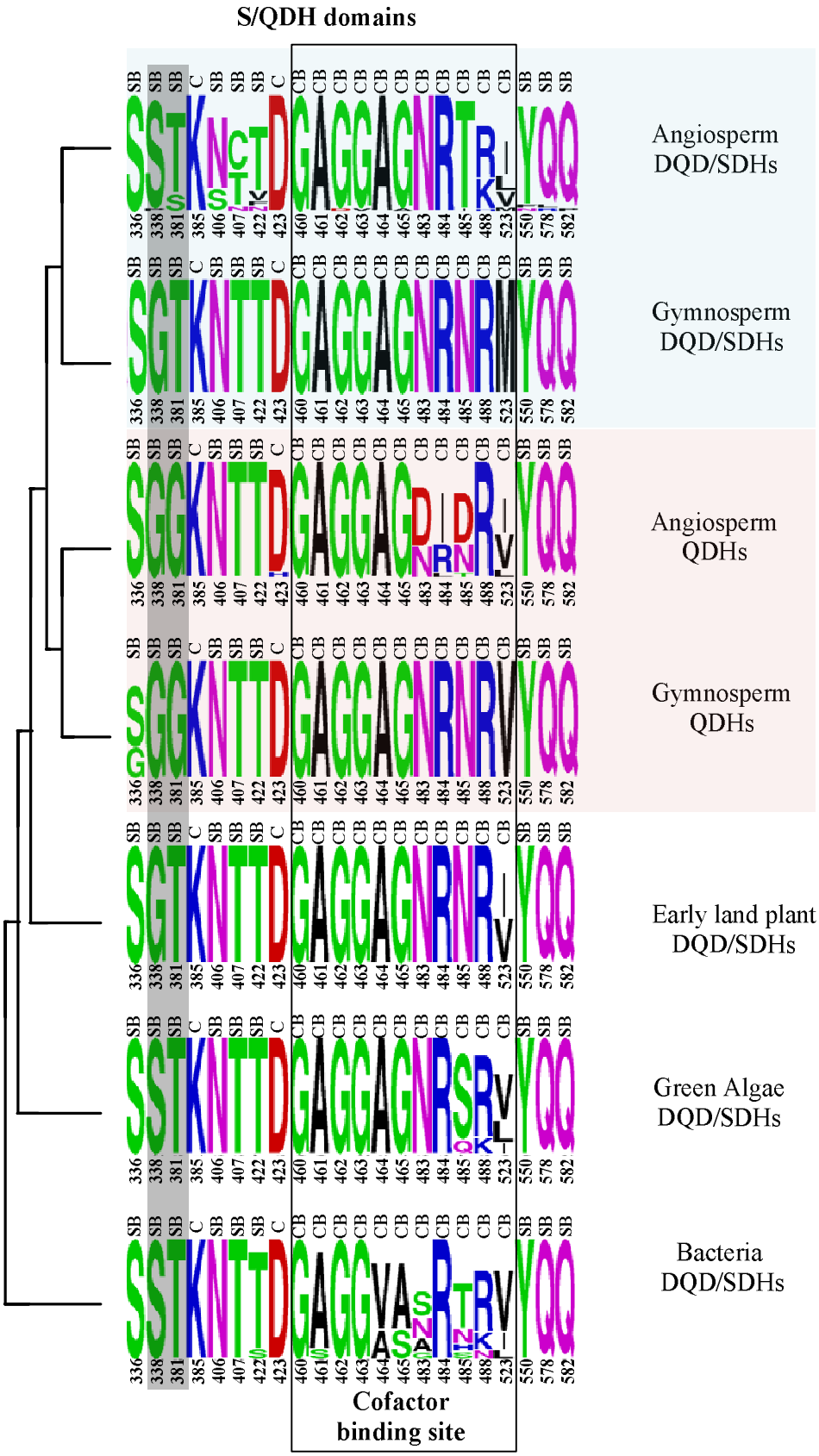

S/QDH domains
Angiosperm DQD/SDHs
Gymnosperm DQD/SDHs
Angiosperm QDHs
Gymnosperm QDHs
Early land plant DQD/SDHs
Green Algae DQD/SDHs
Bacteria DQD/SDHs
Cofactor binding site

Figure 4.2: Conservation of the key amino acids of the S/QDH domains within each major clade of the phylogeny.

The protein sequence alignment used to generate the extended phylogeny (Appendix A) was trimmed to retain the key amino acid residues, which are potentially involved in reaction catalysis (C), substrate binding (SB) or cofactor binding (CB) based on the known protein structure of Arabidopsis DQD/SDH (Singh and Christendat, 2006; Singh and Christendat, 2007). This alignment was subdivided into seven alignments, each of which only contains sequences from a sub-lineage of the DQD/SDH/QDH family (bacteria, green algae, early land plants, gymnosperms DQD/SDHs & S/QDHs and angiosperms DQD/SDHs & QDHs). These seven alignments were used to generate sequence logos with the 'WebLogo' tool. Residues were labeled based on their relative positions in Arabidopsis DQD/SDH. Key substrate binding sites of the SDH domain (Ser338 and Thr381) are highlighted in gray color. DQD/SDH clade is highlighted in blue while QDH clade is highlighted in pink

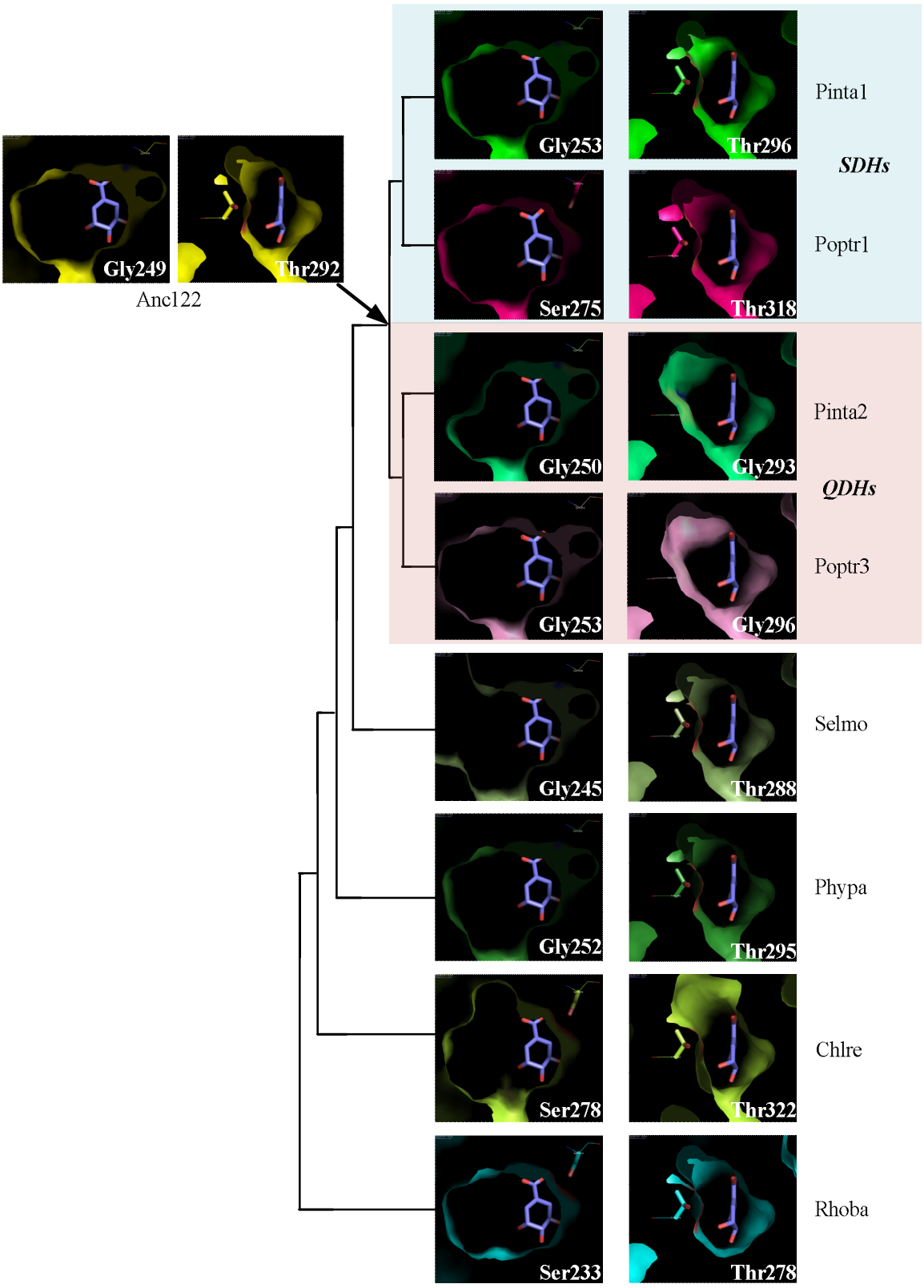

Gly249
Thr292
Anc122
Gly253
Thr296
Pinta1
SDHs
Ser275
Thr318
Poptr1
Gly250
Gly293
Pinta2
QDHs
Gly253
Gly296
Poptr3
Gly245
Thr288
Selmo
Gly252
Thr295
Phypa
Ser278
Thr322
Chlre
Ser233
Thr278
Rhoba

Figure 4.3: Ancestral reconstruction and protein structure modeling.
The amino acid sequence of the ancestral DQD/SDH (Anc122, prior to gene duplication) was reconstructed with PAML. Along with the DQD/SDHs and QDHs from selected extant plant species (*R. baltica* (Rhoba), *C. reinhardtii* (Chlre), *P. patens* (Phypa), *S. moellendorffii* (Selmo), *P. taeda* (Pinta1&2) and *P. trichocarpa* (Poptr1&3)), it was subjected to protein structure modeling using Arabidopsis DQD/SDH coupled with shikimate as a template. The impact of changes at key substrate binding sites (Ser338 and Thr 381 in Arabidopsis) is highlighted.

Conservation of the G motif suggests an critical role of cofactors (NADP(H) and NAD(H)) in reaction catalysis, and these key sites have been maintained by the action of purifying selection during plant evolution. Asn483, Arg484 and Thr485 in Arabidopsis constitute the NRT motif, which binds to the adenine phosphate and is commonly found in many NADP(H)-dependent dehydrogenases (Singh and Christendat, 2007). Sequence analysis revealed that Asn (483) and Arg (484) residues are highly conserved in major groups of the green lineages except for the angiosperms QDH clade. In some angiosperms QDHs, Asn (483, positively charged) and Arg (484, polar uncharged) were converted into negatively charged Asp and nonpolar Ile, respectively. This may explain the preference of (some) angiosperms QDHs for NAD(H), which lacks the phosphate at the adenine moiety, over NADP(H) as a cofactor. Thr (485, polar uncharged) residue was found to be specific to the angiosperm DQD/SDH clade and converted into Asn (polar uncharged) in other clades. Both Thr and Asn are poplar uncharged amino acids sharing similar structure, and the Thr-to-Asn conversion may therefore not change cofactor preference.

4.3.3 Detection of positive selection

Selection types can be inferred by the ratio of non-synonymous substitution rate (dN) over synonymous substitution rate (dS). A dN/dS ratio (ω) significantly greater than 1 is an indication of positive selection while purifying selection can be inferred when the ratio is smaller than 1. An evolution analysis of the DQD/SDH/QDH superfamily was performed using three models (the branch model, MEME and SLAC). With the branch model (implemented in PAML), ω ratios across all labeled branches were modeled. Strong signature of positive selection was identified on branches at subtending the green

lineage, the DQD/SDH clade, as well as the QDH clade (Figure 4.4). Detection of positive selection on the branch subtending the green lineages suggests that enzyme optimization leading to more DQD/SDH activities (driven by increased organismal complexity) may have occurred to the ancestor of all green organisms after the divergence of green algae and bacteria. It could also be an indication that the gain of quinate biosynthetic activity (QDH) started at this time point. It is also worth noting that bacteria and plants belong to different domains of life, and the detection of positive selection on the branch leading to the green lineages could well be due to phylogenetic divergence (i.e. bacterial sequences are quite distinct from plant sequences). Notably, the branches leading to the DQD/SDH and the QDH clades have also evolved under positive selection, which suggests that adaptive changes have occurred in both clades shortly after duplication to allow members within each clade to become specialized in one of the two original functions (either DQD/SDH or QDH).

It is worth mentioning that the branch model is limiting as it assumes that ω ratio remains constant across all sites of a protein. Therefore, a second method (MEME), which allows ω ratio to vary across both lineages and sites, was included in this study to test for positive selection (Figure 4.5, Appendix G). In total, 38 sites were found to be under positive selection at different time points of plant evolution (List in Appendix G). Focusing on active sites, three sites (Ser338, Thr381 and Thr407) were found to have evolved under positive selection on at least one branch. Ser338 and Thr381 have been previously identified as major substrate binding determinants of the SDH domain in Arabidopsis (Singh and Christendat, 2006; Singh and Christendat, 2007). In contrast, Thr407 does not interact with the substrate (shikimate) directly but is involved in

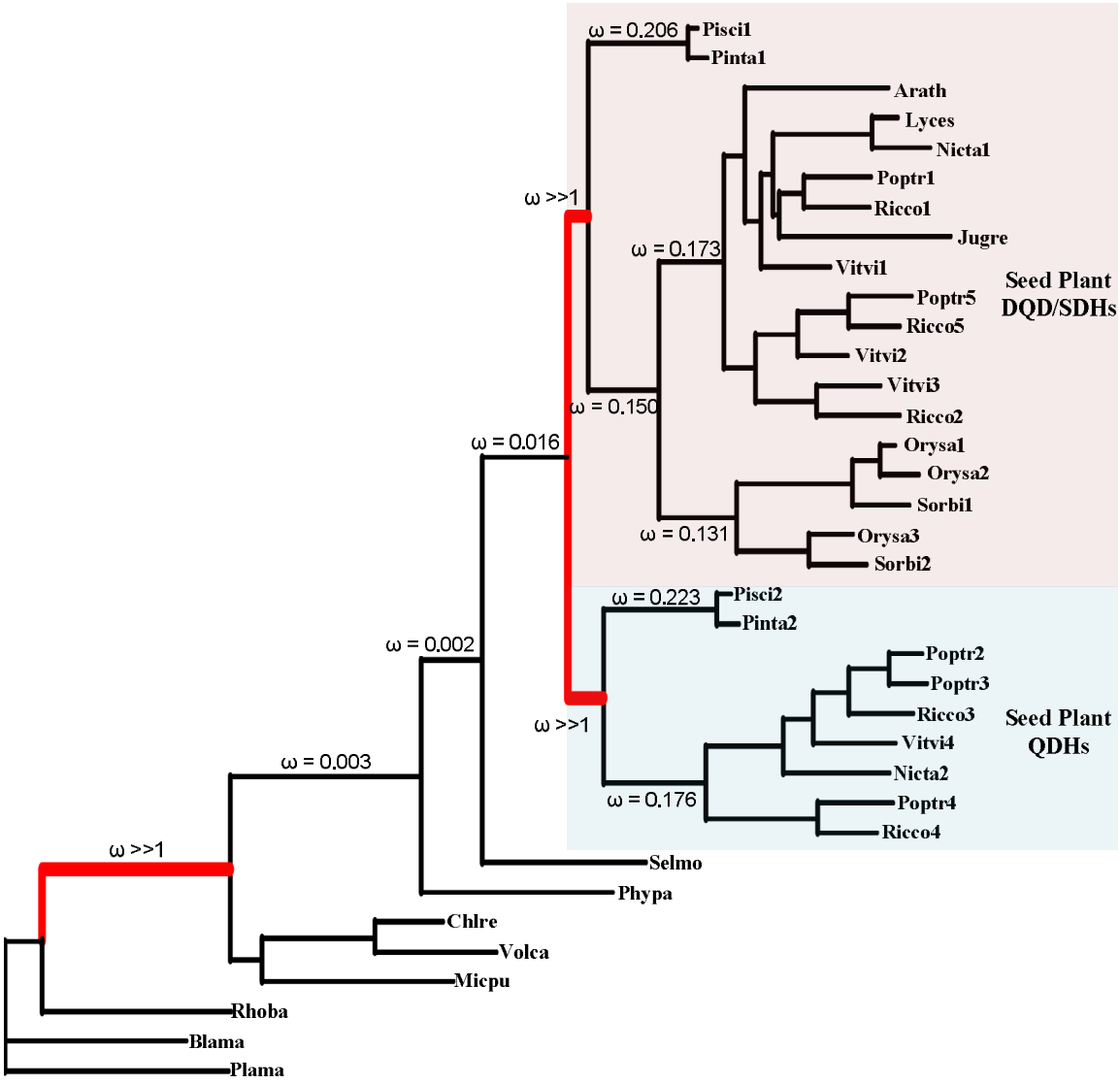


Figure 4.4: Detection of positive selection on pre-specified branches of the phylogeny.

ω ratios (non-synonymous nucleotide substitution over synonymous substitution) were determined using the branch model implemented in PAML. Strong positive selection (ω>>1) was detected on branches (thick red) subtending the green lineage, the DQD/SDH (Pink) and the QDH clades (blue).

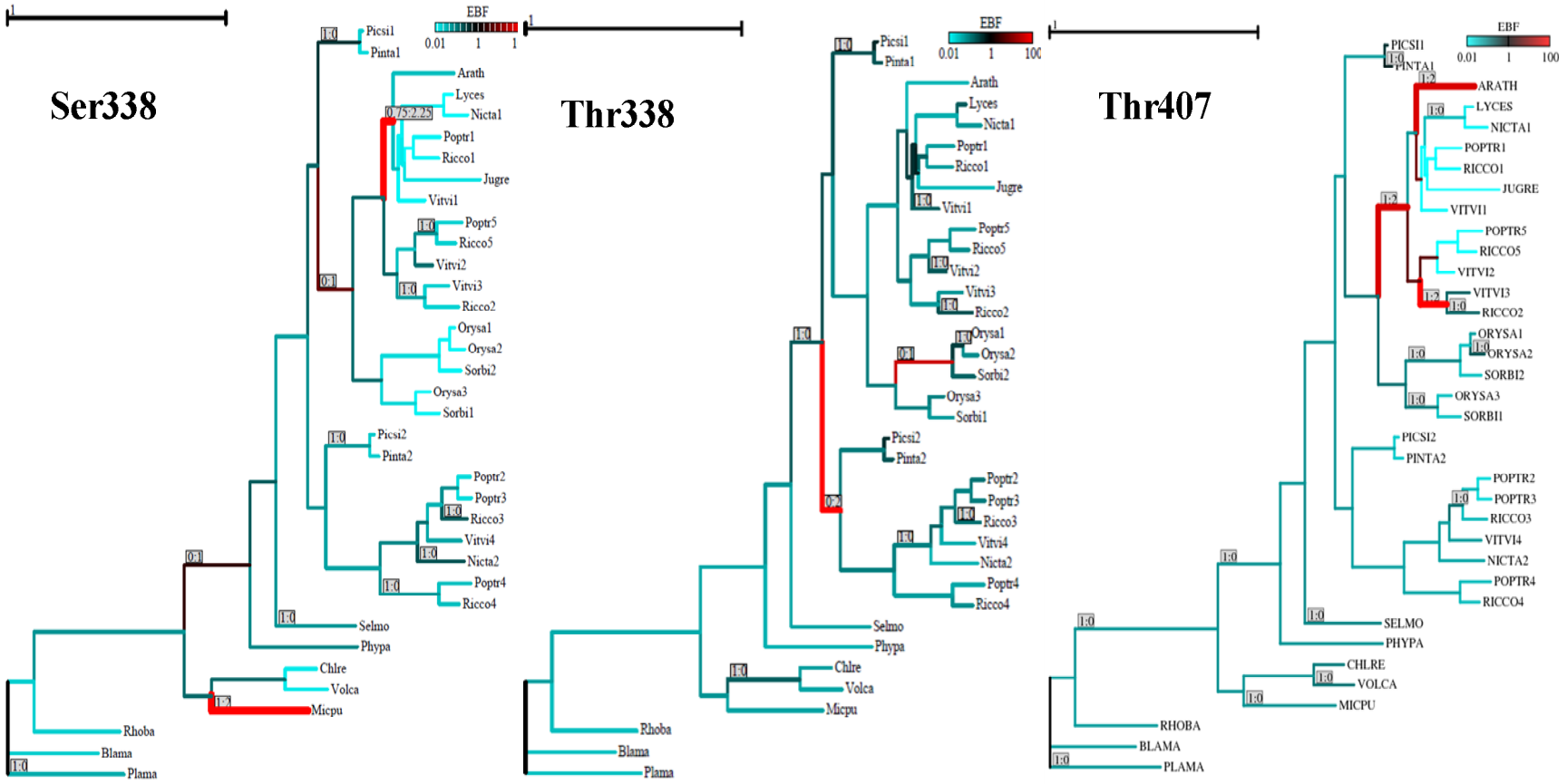


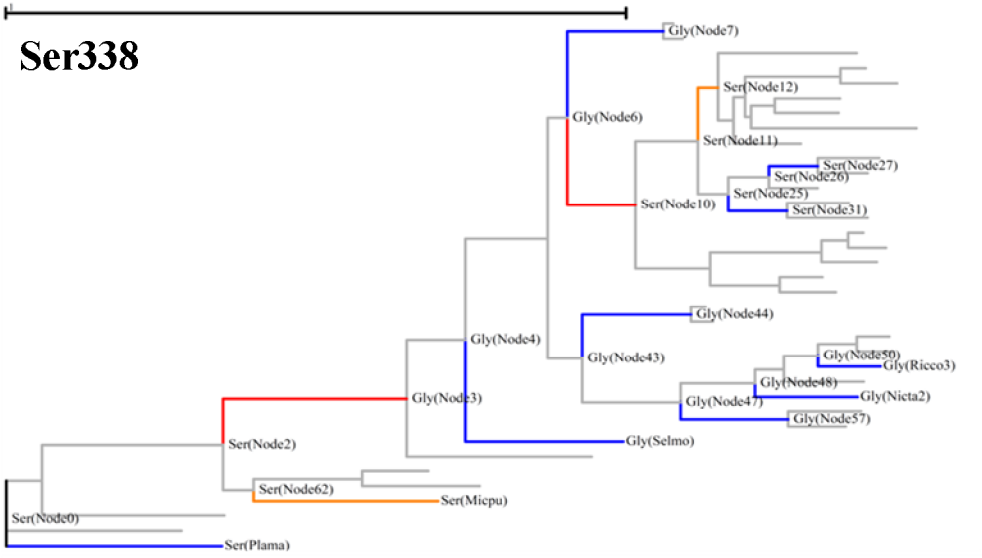
Figure 4.5: Detection of episodic diversifying selection pressure at sites Ser338 (left), Thr381 (middle) and Thr407 (right). A test for episodic positive selection was carried out with the MEME model implemented in Datamonkey. Branches that are under strong positive selection at these three sites are highlighted by red.

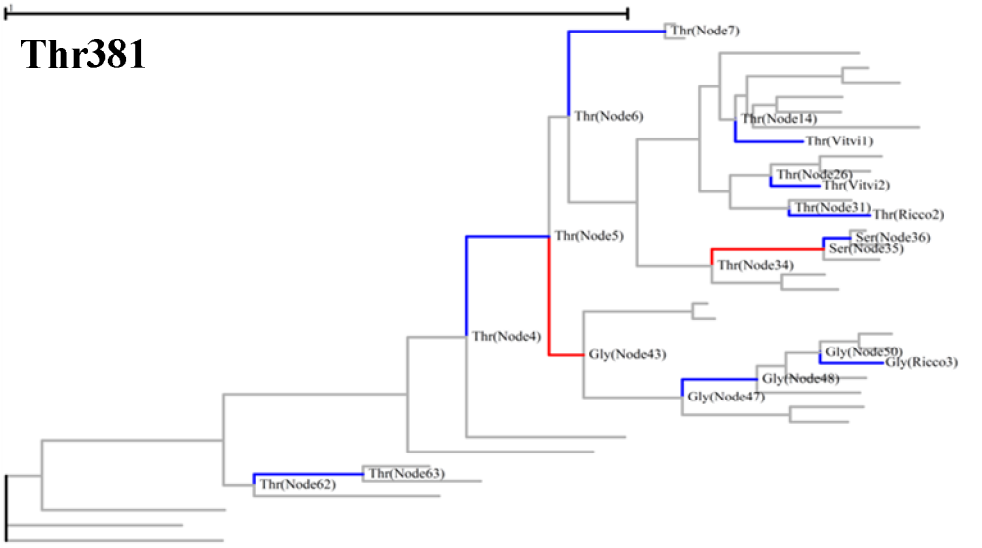
conformationally orienting Asn406, which is located proximal to the C4-hydroxyl (Singh and Christendat, 2007). For Ser338, positive selection was detected first at the branch defining all land plant sequences, where the Ser (338) is changed to Gly, and then on the branch defining the angiosperm SDH clade (where the Gly evolved back to Ser). In addition, positive selection was detected on the branch leading to the QDH clade at site Thr381, which was changed to Gly in the members of the QDH clade. For site Thr407, positive selection was detects on the branch leading to the true dicot DQD/SDH clade (Thr to Ser) as well as branches subtending some subgroups of the true dicot DQD/SDH clade.

These observations were further confirmed by substitution maps of these three sites obtained with SLAC (Figure 4.6). A non-synonymous substitution, which allowed the conversion of Ser (338) to Gly, was detected on the branch leading to the land plants. Gly was then converted back to Ser in the angiosperm DQD/SDH clade. In the QDH clade, the Ser (338) was kept, and in addition, the conversion of Thr (381) to Gly was observed on the branch leading to the QDH clade. In addition, Thr (407) to Ser conversion was observed on the branch defining the dicot DQD/SDH clade and further diversified within this clade.

4.3.4 Enzymatic properties of the DQD/SDH/QDH superfamily members across the plant kingdom

To elucidate the potential mechanism through which the DQD/SDH/QDH superfamily has evolved during plant evolution, DQD/SDHs and QDHs from bacteria, selected extant plant species and the predicted ancestor of all seed plant S/QDHs were


Ser338
Gly(Node7)
Ser(Node12)
Gly(Node6)
Ser(Node11)
Ser(Node27)
Ser(Node26)
Ser(Node25)
Ser(Node10)
Ser(Node31)
Gly(Node44)
Gly(Node4)
Gly(Node43)
Gly(Node50)
Gly(Ricco3)
Gly(Node48)
Gly(Node3)
Gly(Node47)
Gly(Nicta2)
Gly(Node57)
Gly(Selmo)
Ser(Node2)
Ser(Node62)
Ser(Micpu)
Ser(Node0)
Ser(Plama)


Thr381
Thr(Node7)
Thr(Node6)
Thr(Node14)
Thr(Vitvi1)
Thr(Node26)
Thr(Vitvi2)
Thr(Node31)
Thr(Ricco2)
Thr(Node5)
Ser(Node36)
Ser(Node35)
Thr(Node34)
Thr(Node4)
Gly(Node43)
Gly(Node50)
Gly(Ricco3)
Gly(Node48)
Gly(Node47)
Thr(Node63)
Thr(Node62)

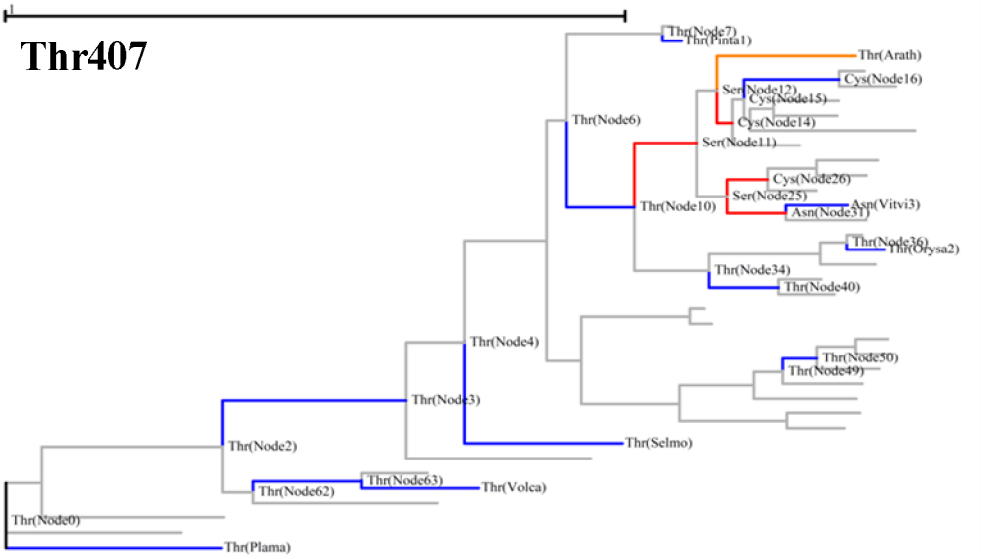

Thr407
Thr(Node7)
Thr(Pinta1)
Thr(Arath)
Cys(Node16)
Ser(Node12)
Cys(Node15)
Thr(Node6)
Cys(Node14)
Ser(Node11)
Cys(Node26)
Ser(Node25)
Asn(Vitvi3)
Thr(Node10)
Asn(Node31)
Thr(Node36)
Thr(Orysa2)
Thr(Node34)
Thr(Node40)
Thr(Node4)
Thr(Node50)
Thr(Node49)
Thr(Node3)
Thr(Selmo)
Thr(Node2)
Thr(Node63)
Thr(Node62)
Thr(Volca)
Thr(Node0)
Thr(Plama)

Figure 4.6: Codon substitution maps of sites Ser338 (top), Thr381 (middle) and Thr407 (bottom).
Substitution maps at these three sites were determined with the SLAC method implemented in Datamonkey. Tree branches are color-coded based on the type of substitutions as follows: yellow for both synonymous and non-synonymous substitutions, blue for only synonymous substitution, red for only non-synonymous substitution.

expressed as recombinant proteins in *E. coli* (Figure 4.7). Purified enzymes were assayed for enzymatic activities with both shikimate and quinate at saturating concentrations (10mM). All enzymes exhibited activities with both shikimate and quinate to some extent with apparent differences in substrate preference (Figure 4.8). It was also noted that all pre-duplication isozymes demonstrated higher affinity towards shikimate compared with quinate. Specialized QDHs only came to existence after gene duplication (the angiosperm/gymnosperm split).

4.3.4.1 QDH

QDH activity was observed in all enzymes included in this study, but to various degrees. It remained low in bacterial and green algal enzymes, and was continuously increasing during early land plant evolution (from Phypa to the pre-duplication ancestor). This suggests that minute QDH activity was present since the beginning of plant evolution. However, it might not be biological significant, and the presence of QDH activity can be explained by the fact that quinate is structurally similar to shikimate. Coinciding with the Ser (338) to Gly change, an increase in QDH activity was observed in Phypa DQD/SDH compared to the bacterial and green algal DQD/SDHs. QDH activity kept increasing from this point on and reached about 12% of the SDH activity in Anc122. After gene duplication, QDH activity decreased in the SDH clade. Pinta 1 and Poptr1 exhibited primarily shikimate specific activities (i.e. encode DHQD/SDH) with minute quinate biosynthetic activity. In contrast, QDH activity was increased in the QDH clade, which coincided with the Thr (381) to Gly conversion. Poptr3 demonstrated high specificity towards quinate while maintaining residual shikimate biosynthetic activity. Pinta2, a member of the QDH clade, was able to take both shikimate and quinate as

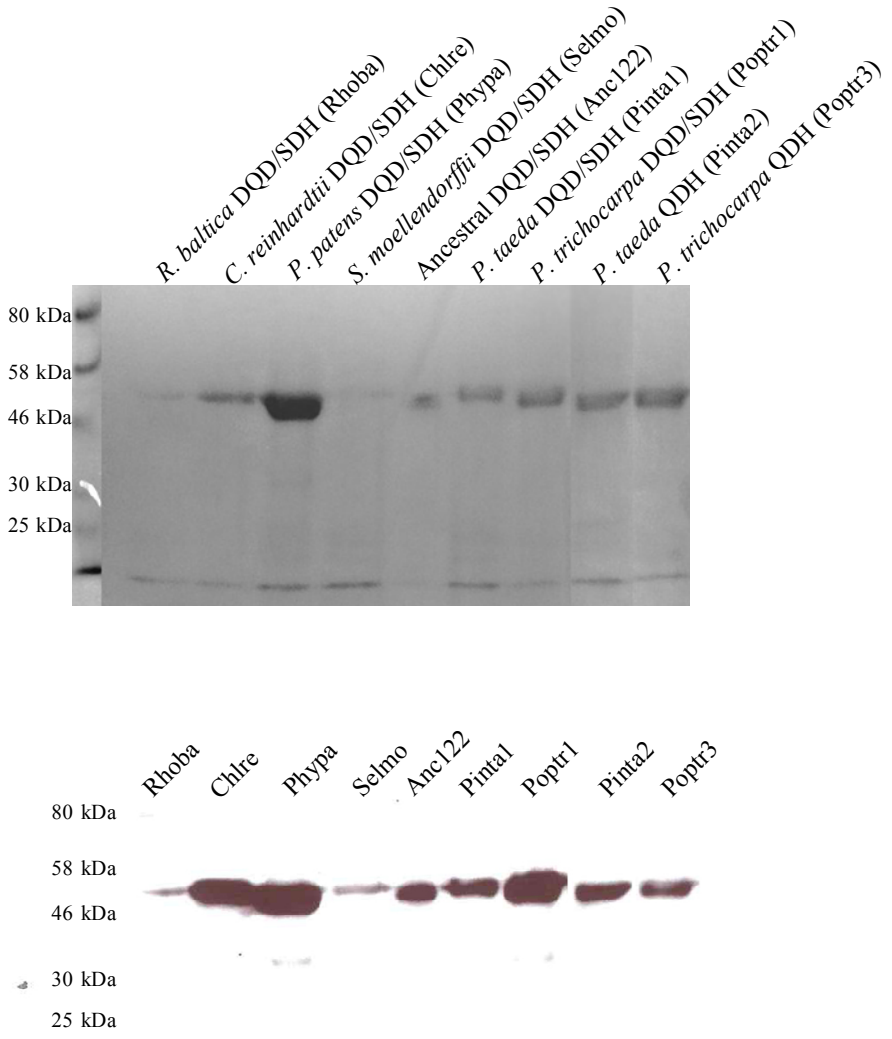
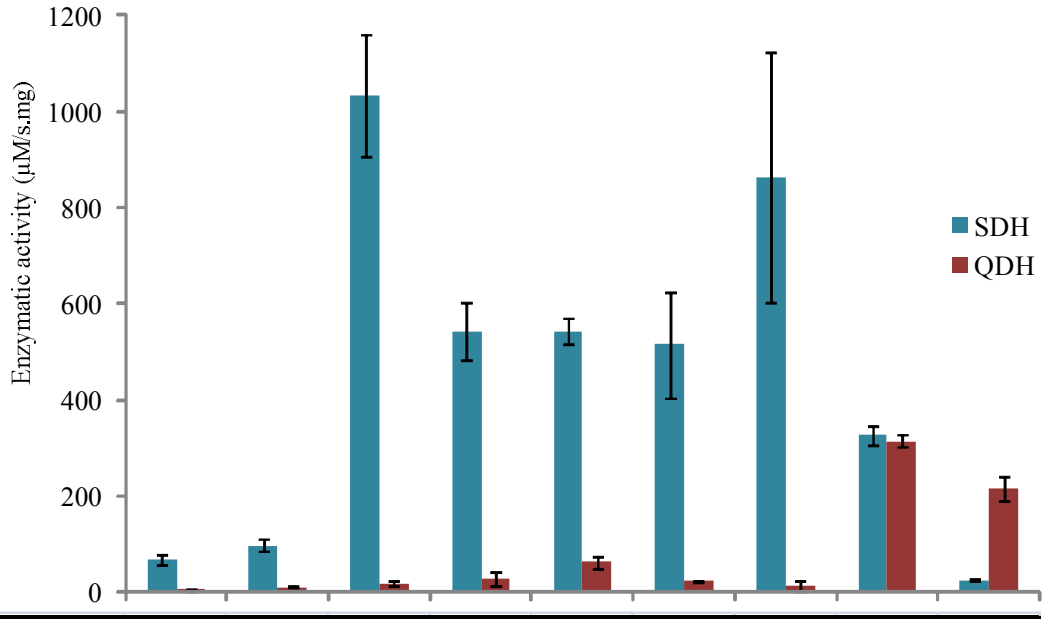


Figure 4.7: SDS-PAGE gel (Top) and Western blot (Bottom) analyses of the purified proteins of ancestral DQD/SDH (Anc122) and selected DQD/SDH/QDH superfamily members (*R. baltica* DQD/SDH (Rhoba), *C. reinhardtii* DQD/SDH (Chlre), *P. patens* DQD/SDH (Phypa), *S. moellendorffii* DQD/SDH (Selmo), *P. taeda* DQD/SDH (Pinta1), *P. taeda* QDH (Pinta2), *P. trichocarpa* DQD/SDH1 (Poptr1) *P. trichocarpa* QDH2 (Poptr3)).
Recombinant proteins were purified from *E. coli* through Ni-NTA affinity chromatography and separated on a polyacrylamide gel (top) followed by Western blotting and detection of His-tagged proteins using a HisProbe fused to horseradish peroxidase (bottom).



Enzyme specificity	Rhoba	Chlre	Phypa	Selmo	Anc122	Pinta1	Poptr1	Pinta2	Poptr3
SDH	0.70	0.56	2.65	2.11	1.63	2.52	7.98	0.24	
QDH								0.32	0.20

Figure 4.8: Enzymatic activities of the ancestral DQD/SDH (Anc122) and selected members of the DQD/SDH/QDH superfamily.

Substrate preferences of recombinant enzymes were determined with both shikimate (blue) and quinate (red) at saturating substrate concentrations (10mM). Enzyme specificities (V_{max}/K_m) were determined by varying substrate concentrations (10 concentrations ranging from 0.05 mM to 5mM) while keeping cofactor at saturation. Enzyme activities were determined by measuring NADPH or NADP formation at 340nm, and three replicates were included for each assay.

substrates equally. The presence of a less specialized enzyme in conifers after duplication is consistent with the fact that gymnosperms have evolved at a lower rate than angiosperms. Pinta2 may still be in the process of losing SDH activity and becoming a specialized QDH.

4.3.4.2 SDH

SDH activity is dominant in all pre-duplication DQD/SDH enzymes. There appears to be an initial increase in SDH activity during early land plant evolution with Phypa having much higher SDH activity and specificity (K_m/V_{max}) in comparison with bacterial and green algal DQD/SDHs. This suggests that the Ser (338)-to-Gly mutation did not interfere with the overall fitness (total dehydrogenase activity) of the enzyme (Phypa DQD/SDH). The increased level of SDH activity observed in early land plants may be driven by increased level of organismal complexity. Coinciding with the increase of QDH activity towards the pre-duplication ancestor, there appears to be a decrease in SDH specificity. After gene duplication, the SDH clade members increased their shikimate specificity again. For example, Poptr1, representing the angiosperm SDH clade with the Ser (338)-to-Gly conversion reversed and an additional Thr (407)-to-Cys mutation, has by far the highest shikimate specificity of all enzymes tested. The gymnosperm Pinta1 is characterized by increased shikimate specificity compared to the immediate pre-duplication ancestor, but not to the extent seen in the angiosperm enzyme Poptr1. This could be explained by the lower rate of evolution observed in gymnosperms compared with angiosperms.

All pre-duplication enzymes demonstrated activities with both shikimate and quinate, with shikimate biosynthetic activity being the dominant. Both SDH and QDH activities were optimized in early land plants, and this was supported by increased level of both activities in Phypa DQD/SDH in comparison to bacterial and green algal DQD/SDHs. A trend was also revealed that the increase of QDH activity was coupled with the reduction in SDH enzyme specificity (K_m/V_{max}). The ancestral protein showed a reduced level of SDH specificity but increased level of QDH activity compared to Phypa and Selmo DQD/SDHs. After gene duplication and further divergence, specialized SDHs and QDHs have evolved separately. Pinta 1 and Poptr1 exhibited primarily shikimate specific activities (i.e. encode DQD/SDH) with some quinate biosynthetic activity. In contrast, Poptr3 demonstrated specificity towards quinate while maintaining residual shikimate biosynthetic activity. Pinta2, a member of the QDH clade, was able to react with both shikimate and quinate equally suggesting that it may still be in the process of becoming specialized for QDH activity.

4.4 Discussion

Phylogenetic reconstruction revealed that a key gene duplication just prior to the angiosperm / gymnosperm split has led to the diversification of the DQD/SDH/QDH superfamily into two functionally distinct clades with one being specific to shikimate biosynthesis (primary metabolism) while the other being involved in quinate biosynthesis (secondary metabolism). Evolutionary analysis and functional characterization of the DQD/SDH and QDH enzymes from the immediate pre-duplication ancestral DQD/SDH and representative extant plant species across the green lineage suggested that functional evolution of the DQD/SDH/QDH superfamily was largely dependent on the existence of

a multi-functional ancestral DQD/SDH/QDH encoding both shikimate (SDH) and at least some quinate (QDH) biosynthetic activities. This apparently created opposing selection pressure preventing further optimization of quinate biosynthetic activity. Selection constrains were only relaxed after gene duplication to allow positive selection to act differently on the two duplicates. Ancestral functional variation was then refined to allow the appearance of genes encoding enzymes which are specialized in one of the two ancestral functions in different lineages.

4.4.1 Evolution of DQD/SDH/QDH superfamily via Escape from Adaptive Conflict (EAC)

Gene duplication has been proposed as a central mechanism for gene functional diversification and adaptation with neo-functionalization and sub-functionalization being the two most general models. Under neo-functionalization, a gene is duplicated giving rise to two functionally redundant duplicates, one of which maintains the ancestral function. The relaxed selection pressure on the other daughter copy leads to accumulation of mutations leading to a novel function (Innan and Kondrashov, 2010). In contrast, sub-functionalization states that mutations already accumulate in the original gene copy. This allows the appearance of multiple functions, which are subdivided after gene duplication (Innan and Kondrashov, 2010). Both models focus on the adaptive processes happening after gene duplication. Escape from Adaptive Conflict (EAC), a special case of sub-functionalization, takes into accounts adaptive pressures both prior to and after gene duplication (Sikosek *et al*., 2012; Des Marais and Rausher, 2008; Huang *et al*., 2012). EAC describes the situation where mutations accumulate in the ancestral gene. This leads to the arising of a second function. However, optimization of this novel function results

in reduction of the original function. This creates an adaptive conflict as optimization of one function will be at the expense of the other. Only after duplication, each copy is free to specialize and to optimize one of the two original functions. There are two defining features of EAC, which are used to distinguish EAC from other models. Under EAC, functional trade-offs (between the original trait and the novel trait) are observed in the single-copy ancestral enzyme (i.e. the evolving function cannot be optimized without compromising the existing function) (Sikosek *et al.*, 2012; Khersonsky and Tawfik, 2010). Secondly, adaptive changes happen to both duplicates after gene duplication, which allows the separation and improvements of the two original functions in different duplicates (Des Marais and Rausher, 2008).

4.4.1.1 Adaptive conflict exists prior to gene duplication

Evolutionary analysis of the DQD/SDH/QDH superfamily indicated that the single-copy ancestral DQD/SDHs have experienced relaxed purifying selection (intermediate ω ratio: 0.0157 in the branch model) before gene duplication in early land plant history. This is coincident with the initial increase of quinate biosynthetic activity after plant terrestrialization (indicated by an increased QDH activity in the moss Phypa SDH compared with the green algal and bacterial enzymes). Being involved in primary metabolism and thus necessary for cellular survival, mutations that lead to reduced shikimate biosynthetic function (original function) beyond a certain threshold should be penalized by decreased level of fitness (Sikosek *et al.*, 2012). Consistently, purifying selection acted to maintain the existing shikimate function with some degree of freedom (indicated by slightly relaxed purifying selection). Nevertheless, this allowed the accumulation of mutations leading to increased quinate biosynthetic activity to some

extent, which could be beneficial to early land plants as quinate and its derivatives are involved in plant adaptation and defense. It is important to note that the original shikimate function appears to have been compromised to some extent during this process (functional trade-offs). This was supported by functional characterization of DQD/SDHs from early land plants (Phypa and Selmo) and from the immediate pre-duplication ancestor. All three enzymes were able to take both shikimate and quinate as substrates with strong preference for shikimate, and relative activities with shikimate and quinate were negatively correlated. As quinate biosynthetic activity increased, enzyme specificity of these three enzymes towards shikimate decreased gradually. Giving the fact that the ability to produce both shikimate and quinate by plants is advantageous, an adaptive conflict appeared between specialization for shikimate or quinate biosynthetic activity (Figure 4.9). Evolutionary analysis also identified a positively selected amino acid residue (Ser338 in Arabidopsis) in pre-duplication species, which is previously identified to be involved in substrate discrimination between shikimate and quinate (Chapter 2). The conversion of Ser338 to Gly was observed in Phypa, Selmo and Ancestral DQD/SDHs. this conversion was coincident with the elevated levels of quinate biosynthetic activities in all three enzymes compared with the bacterial and green algal DQD/SDHs. With the increased levels of quinate biosynthetic activity and compromised shikimate biosynthetic activity, an adaptive conflict came to appear.

4.4.1.2 Resolution of adaptive conflict by gene duplication and gene function optimization

The adaptive conflict between shikimate and quinate biosynthetic activities was resolved by gene duplication and further specialization events. Gene duplication can

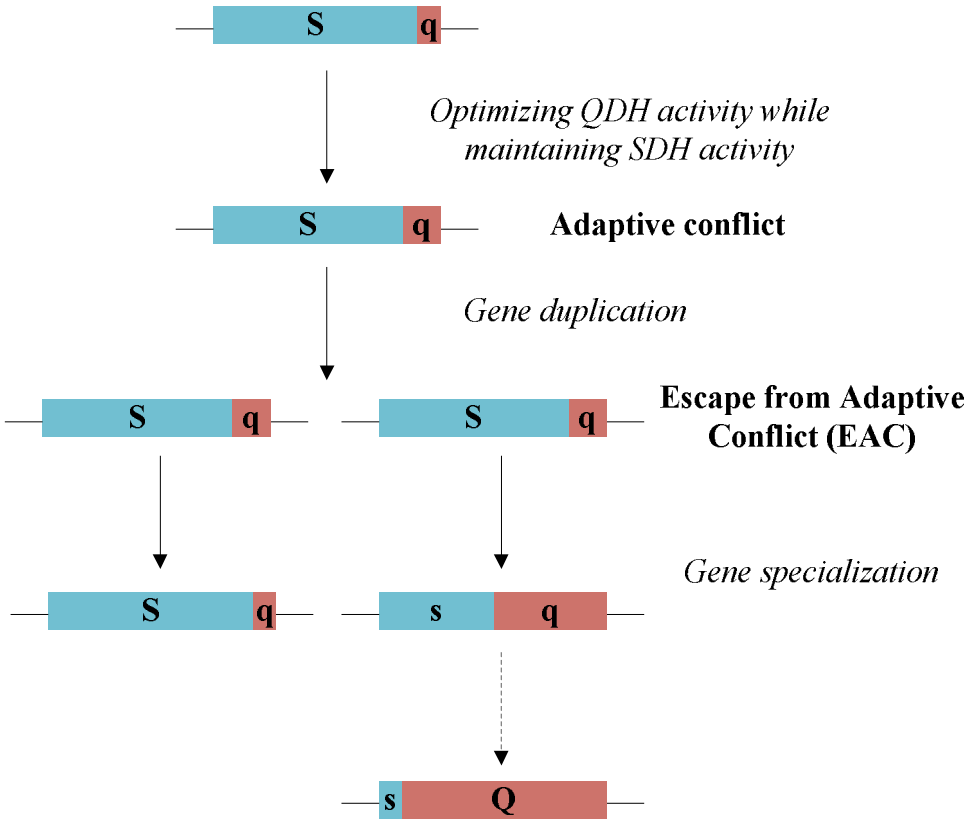


Figure 4.9: The EAC (Escape from Adaptive Conflict) model.
The ancestral gene encodes a protein with primary shikimate (S: SDH, blue) and residual quinate (q: QDH, red) biosynthetic activities. An adaptive conflict arose when quinate biosynthetic activity became beneficial, and selection favored the optimization of this activity. This conflict was resolved by gene duplication followed by gene specializations, which resulted in two genes encoding enzymes specialized in either shikimate (S) or quinate (Q) biosynthetic activities.

consequently lead to an increase in protein concentration, which may compensate for the partial loss of the original function and the low efficiency of an actively evolving function, and thereby provide immediate advantage (Soskine and Tawfik, 2010). In addition, gene duplication also provides working material for gene functional adaptation and specialization to happen. Based on phylogenetic reconstruction of the DQD/SDH/QDH superfamily, it is most likely that gene duplication occurred prior to the separation of angiosperms and gymnosperms. After gene duplication, positive selection drove each gene duplicate to specialize. This was supported by the detection of positive selection ($\omega>>1$) on the branches leading to the true DQD/SDH and the QDH clades. This result is also most compliant with the EAC model. During the EAC stage right after gene duplication, the ratios of non-synonymous substitution over synonymous substitution are expected to be greater than one due to the presence of adaptive changes (Sikosek *et al*., 2012; Innan and Kondrashov, 2010; Des Marais and Rausher, 2008). Following the separation of two functions, genes duplicates encoding specialized enzymes are constrained by weak purifying selection with ω ratios smaller than one (Innan and Kondrashov, 2010; Conant and Wolfe, 2009). Similar result was obtained in this study. Isozymes from the true DQD/SDH clade (Pinta1 and Poptr1) and the QDH clade (Pinta2 and Poptr3) were subjected to biochemical assays to determine their preferred substrates. Pinta1 and Poptr1 preferred shikimate over quinate. Both enzymes demonstrated higher substrate specificity towards shikimate than the ancestor, which indicates that shikimate biosynthetic function has been optimized in members of the true DQD/SDH clade. In contrast, Poptr3 demonstrated high activity with quinate and extremely low activity on shikimate. In addition, activity on quinate is substantially

higher for Poptr3 (post-duplication copy) than for the ancestor and pre-duplication copies. Biochemically, Pinta2 represents an intermediate between the shikimate specific enzymes and enzymes specialized for quinate biosynthesis: It exhibited equal activities on both shikimate and quinate. This may be explained by the low evolutionary rate of gymnosperms (Buschiazzo *et al.,* 2012), which would suggest that it is still being in the process of losing shikimate biosynthetic activity and becoming specialized on quinate. Together, the results suggest that optimization of gene functions occurred in both duplicates after gene duplication leading to specialized enzymes with one predominantly acting on shikimate and the other on quinate. The adaptive changes in different lineages can be best explained by EAC (Figure 4.9). Being consistent with the EAC model, a second positively selected site (Thr381in Arabidopsis) was identified, which has evolved to Gly in all QDHs. It is also notable that the Ser-to-Gly substitution occurred before gene duplication was reversed in angiosperm DQD/SDH clade leading to the highly shikimate specialized enzymes. All these together suggest an important role of both Ser338 and Thr381 residues in substrate discrimination between shikimate and quinate. The gain of quinate activity appears to have already occurred after the Ser-to Gly conversion in early land plants. However, specialized QDH enzymes only came into existence when the second mutation was gained after gene duplication. Furthermore, positive selection was also detected on the branch subtending all dicot DQD/SDHs at site Thr407 (in Arabidopsis), which may also be involved in re-specialization of DQD/SDH enzymes in shikimate biosynthetic activities.

4.4.2 Secondary metabolic genes derived from primary metabolic genes

Plants are able to produce a large diversity of chemical compounds with relatively low molecular weights, and more than 200,000 chemicals have been elucidated (Pichersky and Gang, 2000; De Luca and St Pierre, 2000). Only a small portion of these compounds are involved in primary metabolism, and they are also found to be highly conserved across different species. The remainder is classified as secondary metabolites, which demonstrate an amazing structural diversity. There has been evidence supporting that such diversity is generated by genes recruited from primary metabolic pathway via gene duplication followed by divergence (Pichersky and Gang, 2000). The present work provides an example of having quinate biosynthetic genes (QDH, secondary metabolism) derived from shikimate biosynthetic genes (SDH), which are involved in primary metabolism. The ancestral DQD/SDH enzyme before gene duplication could already recognize both shikimate and quinate as substrates with stronger catalytic rate on shikimate. Only after gene duplication, both defense-related quinate biosynthetic function and essential shikimate biosynthetic activity were optimized independently on the two duplicated copies. A similar situation, where both primary and secondary pathway genes are present in the same gene family, has been reported in a few families including cytochrome-P450-dependent monooxygenases (Lupien *et al*., 1999), o-methyltransferases (Wang and Pichersky, 1999), anthranilate synthases (Essar *et al*., 1990), diterpene synthase (Keeling *et al*., 2010). However, most of these studies are limited to describing functional diversity observed within a gene family. The molecular evolutionary perspective was not interrogated in any detail and the mode of evolution that created functional variation within a family has not been elucidated. In my study, I

demonstrated that the diversification of the DQD/SDH/QDH superfamily likely occurred through sub-functionalization via a mechanism described as "Escape from Adaptive Conflict". Due to the fact that EAC requires the presence of relaxed purifying selection prior to gene duplication to allow an adaptive conflict to appear, it is most likely to operate on gene families that are not directly involved in primary metabolism (essential to survival) (Sikosek *et al*., 2012). However, quinate shares extremely high structural similarity with shikimate. A single non-silent mutation (Ser (338) to Gly) could already result in the gain of some quinate biosynthetic activity, and shikimate function was compromised to some extent but still sufficient to maintain overall fitness. The appearance of functional trade-offs before duplication has been identified as a defining feature of EAC. This further supports that the DQD/SDH/QDH superfamily has evolved via EAC. It is also worth mentioning that gene evolution is a continuous process. Genes that have been recruited from primary metabolism for secondary metabolism may also serve as a pool for further recruitments of novel functions, which are potentially involved in another secondary metabolic pathway(Pichersky and Gang, 2000; Ober, 2005). It was noted in this study that Poptr 3 demonstrated higher catalytic rate with quinate than shikimate. However, its specificity towards quinate was lower compared to the bi-functional enzyme Pinta2, which was not expected. This may suggest that ongoing specialization towards quinate coincides with a (temporal) general loss of overall dehydrogenase activity, which can be compensated by the presence of multiple isozymes resulted from recent gene duplication events (i.e. there are three QDHs in poplar).

4.4.3 The gain of quinate biosynthetic activity as an adaptive trait

Quinate and its derivatives, such as chlorogenic acid, are widely distributed in the plant kingdom (Petersen *et al*., 2009). Chlorogenic acid accumulates potentially as a defensive compound with anti-herbivore and anti-microbial properties (Leiss *et al*., 2009). Chlorogenic acid and its isoforms may function as storage forms of the intermediates involved in lignin biosynthesis. (Boerjan *et al*., 2008, Tsai *et al*., 2006) and have a potential role in photoenergy dissipation (Grace and Logan, 2000). Chlorogenic acid is a hydroxycinnamyol ester of quinic acid. As a result, the production of quinate is essential to the biosynthesis of chlorogenic acid. Understanding the roles and distributions of chlorogenic acid and its isoforms in the plant kingdom will also provide more insight into the evolution of quinate biosynthesis by answering the following questions: when did the gain of quinate biosynthetic activity occur in nature? Why is the gain of quinate biosynthetic function beneficial?

Based on functional characterization of the DQD/SDH and QDH enzymes across the plant kingdom, it is clear that quinate biosynthetic function is present in the common ancestor of the green lineages. This is supported by the extraction of chlorogenic acid from two freshwater algae: *Anabaena doliolum* and *Spongiochloris spongiosa* (Onofrejova *et al.*, 2010). However, quinate biosynthetic activity remained low in algae due to the fact that the aquatic environment is less complex than the terrestrial environment in terms of types of stresses (i.e. less UV light, less herbivory, etc.), and the demand for the production of secondary metabolites, chlorogenic acid in particular, is comparably low. The presence of chlorogenic acid or quinate derivatives in general has been reported in some species from all major groups of early land plants including

bryophytes (Jockovic *et al.*, 2008; Erxleben *et al.*, 2012; Montenegro *et al.*, 2008), lycophytes (Lin *et al.*, 2000) and ferns (Petersen *et al*., 2009). This is also consistent with the detection of quinate biosynthetic activities in *P. patens* (Phypa) and *S. moellendorffii* (Selmo). It was also noted that quinate activities of Phypa and Selmo were higher compared to the green algal DQD/SDH (Chlre). This could be explained by the fact that land plants were facing increasing challenges (i.e. UV light, gravity, herbivores, parasites, etc) during and after terrestrialization, and being able to produce defense-related quinate derivatives was apparently advantageous in plant adaption to the terrestrial environments. It is evident that the diversity of plant secondary metabolites is largely shaped by environmental factors and interactions between plants and their enemies (Theis and Lerdau 2003). The battle between plants and their enemies is the major driving force for evolution, which explains why there was an increasing demand on the production of quinate and its derivatives already prior to gene duplication. However, the optimization of quinate biosynthetic activity was coupled with partial loss of essential shikimate activity in the single-copy DQD/SDHs. High demand for quinate derivatives and functional conflict further drove selection towards specialized enzymes after gene duplication.

4.4.4 The loss of quinate biosynthetic activity (QDH) in some plant species

Phylogenetic relationship among members of the DQD/SDH/QDH superfamily was reconstructed, and the lack of monocot and some herbaceous dicot (e.g. Arabidopsis) sequences in the QDH clade suggests that quinate biosynthetic activity may have been lost in some plant species during plant evolution. One chemical survey suggests that quinate are present in some monocot species at low levels but not in all monocots

analyzed (Yoshida *et al*., 1975). Monocots, which are able to produce quinate (derivatives), were not included in my phylogeny study due to the lack of sequence information. However, this still suggests that the loss of quinate biosynthetic activity is not at the group level (all monocots) but has only occurred multiple times at the family, genus, or species level. It was noted in this study that some of the specialized SDHs (Poptr1 and Pinta1) still maintained residual quinate biosynthetic activities. It remains possible that some of the monocot SDHs also maintain some QDH activities, which may be sufficient to be responsible for the production of quinate in these monocot species. However, a specialized QDH with residual SDH activity has been partially purified from corn seedlings (Graziana *et al*., 1980). This is not consistent with my phylogenetic reconstruction result, which placed all corn sequences in the true DQD/SDH clade. Although the corn genome is nearly complete, there remains the possibility of having genes encoding QDHs present in the un-sequenced part of the corn genome.

My work suggests that the gain of quinate biosynthetic activity was initiated by the evolutionary arm race between plants and their surrounding environments. The ability of plants to produce quinate and its derivatives was likely acquired as a defensive strategy in early plant evolution. In response to this, some plant enemies (e.g. herbivores) may have evolved a counter adaptation, which allows them to thrive even in the presence of defense-related quinate derivatives. The appearance of these specialists, which are less susceptible to quinate derivatives, may have made this quinate-based defense strategy less effective in some lineages / species and therefore less favored by selection. Consequently, this has led to the loss of quinate biosynthetic activity and potentially the gain of a new defense strategy in these species.

4.4.5 Cofactor preference variations of plant DQD/SDHs and QDHs

It is well established that all plant SDHs prefer NADPH as a cofactor (Singh and Chritendat, 2006; Bischoff *et al*., 2001; Bonner and Jenson, 1994; Ding *et al*., 2007; Singh and Christendat, 2006; Singh and Christendat, 2007, chapter 2). In contrast, variations in cofactor preference were observed in QDHs across different plant lineages. QDHs from gymnosperms (*L. sibirica*) prefer NADP(H) (Ossipov *et al.,* 1995), while QDHs from angiosperms such as mung bean, pea, carrot, corn, and poplar take NAD(H) as a cofactor (Kang and Scheibe, 1993; Minamikawa, 1977, Refeno *et al.,* 1982; Graziana and Boudet, 1983, chapter 2). This is consistent with the finding that the NRT motif, a defining feature of all NADP(H)-dependent dehydrogenases, is absent from some angiosperm QDHs (i.e. the NRT motif is replaced by DID in some QDHs from angiosperms). Variations in cofactor preference between angiosperms SDHs and QDHs may allow highly-regulated carbon flow into either shikimate or quinate biosynthesis from their common precursor 3-dehydroquinate. Shikimate plays an important role in primary metabolism (protein production) while quinate is involved in plant secondary metabolism (e.g. defense). Shikimate and quinate biosynthetic enzymes (SDHs and QDHs) are derived from one multi-functional ancestor. Enzymes with specificity towards one of the two substrates (shikimate and quinate) came to appear after gene duplication and subsequent specialization events. Differences in cofactor preference allow further separation of these two functions and therefore independent biochemical regulation of these two activities in plants.

In conclusion, the present work provides the molecular mechanism of the recruitment of a secondary metabolic gene (QDH) from a gene involved in primary

metabolism (SDH). Quinate biosynthetic activity might have become beneficial early during plant terrestrialization, and was therefore favored by natural selection. However, this promiscuous function was optimized at the expense of losing some shikimate activity, which is essential to plant survival. This created a conflict preventing both functions from being optimal. Gene duplication and further specialization events resolved this conflict, and this was supported by the presence of specialized SDHs and QDHs in seed plants. The mode of evolution underlying functional diversification and divergence in the DQD/SDH/QDH superfamily can be best described by "Escape from Adaptive Conflict". The ability to produce quinate and its derivatives (e.g. chlorogenic acid) by plants is thought to be driven by environmental changes and chemical arms race between plants and their enemies. In return, the gain of this function provides a significant selective advantage to allow them to better adapt to the changing environments and to defend themselves against biotic stresses.

Chapter 5 Conclusions and future directions

The major research work of this dissertation established the functional identity of DQD/SDHs (3-dehydroquinate dehydratse / shikimate dehydrogenase) in poplar, deciphered the genetic base of quinate biosynthesis via molecular characterization of QDHs (quinate dehydrogenase) and dissected the evolutionary mechanism that underpin the functional diversity of the DQD/SDH/QDH superfamily in the green lineages. Here, I will summarize these findings and discuss possible future research directions.

5.1 Shikimate and quinate biosynthesis in Poplar

Shikimate, a precursor of aromatic amino acids, is synthesized by a bi-functional enzyme (DQD/SDH) from 3-dehydroquinate via 3-dehydroshikimate in plants. These two reactions represent the third and fourth steps of the shikimate pathway, which is a major component of plant primary metabolism. Quinate, a defense-related plant secondary metabolite, can be formed in a single step reaction from either 3-dehydroquinate or shikimate. The two enzymes involved are QDH and QD (quinate dehydratase). The reaction mechanisms of QD and QDH resemble that of DQD and SDH, respectively. As a result, I hypothesized that quinate biosynthesis might be catalyzed by enzymes similar to DQD and/or SDH.

There are five putative DQD/SDHs (Poptr1-5) encoded in the Poplar genome. Annotations of these five genes were largely based on sequence analysis, and no experimental evidence was available in terms of their functions. The main objective of Chapter 2 was to functionally characterize the five genes and to identify isoforms that are

potentially involved in quinate biosynthesis. Sequence similarity analysis of Poptr1 to Poptr5 and functionally characterized plant DQD/SDHs revealed two groups. Poptr1 and Poptr5 were grouped with characterized plant DQD/SDHs from Arabidopsis, tomato and tobacco (NtDQD/SDH1) while Poptr2, Poptr3, and Poptr4 formed a separate group with a putative tobacco DQD/SDH isoform 2, which has reduced SDH activity. A role of Poptr2 to Poptr4 in quinate biosynthesis was suggested by protein structure modeling because they both possess an extra pocket in the active site of the SDH domain that could accommodate quinate as a substrate. This was verified by *in vitro* biochemical assays with both shikimate and quinate. Poptr1 and 5 are true DQD/SDHs while Poptr2 and 3 are indeed QDHs with only residual SDH activities. Thus, I have discovered here the molecular genetic basis of quinate biosynthesis in poplar. I showed that quinate and shikimate biosynthetic activities are encoded by distinct members of the same gene family (DQD/SDH/QDH superfamily) connecting plant primary metabolism tightly with the biosynthesis of specialized plant natural products (secondary metabolism). This permitted deciphering the molecular evolutionary history of this gene family. Identification of genes encoding QDHs allows further investigation of the roles of quinate and its derivatives in plant development and chemical ecology with reverse genetic approaches. This also enables quinate (and shikimate) biosynthesis in heterologous systems (e.g. bacteria), which potentially reduces the reliance on limited plant sources for the extraction of these valuable chiral starting molecules for (antiviral) drug production. Poptr4 was not included in the biochemical characterization due to difficulties in expressing recombinant protein in the bacterial system. Based on these enzymatic properties, Poptr1 and Poptr5 were re-named to DQD/SDH 1 and DQD/SDH2, and

Poptr2 and Poptr3 to DQH1 and QDH2, respectively. Poptr4 was assigned the name QDH3 owing to its higher similarity to QDHs than to SDHs. Biochemical characterization of QDH3 is necessary for complete characterization of the DQD/SDH/QDH family in poplar, which could be accomplished by recombinant protein expression in a different host system.

5.2 Expression variations of Poplar DQD/SDHs and QDHs and co-expression analysis

Expression patterns of poplar DQD/SDHs and QDHs across different tissue types and their induction in response to environmental cues and genetic modifications were assessed by analyzing large-scale Affymetrix microarray expression data. Expression variations were observed among DQD/SDHs and QDHs, which are correlated with their functions and potential roles in both plant development and adaptive responses to stresses. DQD/SDH1 was found to be constitutively expressed at moderate levels across a wide range of tissues (organs) and highly expressed in actively growing cells and meristematic regions. This suggests that DQD/SDH1 is most likely involved in primary function (protein production), which was further supported by co-expression analysis (e.g. co-expression of DQD/SDH1 with some shikimate pathway genes and genes involved in loading amino acids onto tRNA). DQD/SDH2 is highly expressed in vascular tissues, which may indicate that DQD/SDH2 has a role in lignin biosynthesis. QDH1 and 2 were found to be highly expressed in roots, which suggests a role of quinate and its derivatives in the rhizosphere. Though the functional characterization of QDH3 was proven unsuccessful, its predominant expression in leaves and its induction by wounding and fungal infection suggests that QDH3 may play an important role in plant defense,

which might be facilitated by the production of quinate and/or its derivatives (e.g. chlorogenic acid) in plants. QDH3's defensive role was further supported by the increased expression level of QDH3 in plants over-expressing WRKY23, a transcription factor with a regulatory role in plant defense against both abiotic and biotic stress. Co-expression analyses supported the role of DQD/SDH1 in primary metabolism, but were less efficient in pinpointing specific pathways or physiological functions for the remaining genes (DQD/SDH2, QDH1 to QDH3). The results of co-expression analysis can be further optimized by eliminating low-quality arrays from the expression data pool. Furthermore, it is noteworthy that the arrays were performed with multiple species within the genus *Populus*, and species-specific distributions of plant secondary metabolites were observed. Co-expression analyses may be interfered by distinct flows through the shikimate pathway in different species, and more careful dissection of the data set may be necessary to resolve this problem.

5.3 Subcellular localizations of Poplar DQD/SDHs and QDHs

The main objective of Chapter 3 was to determine subcellular localizations of Poplar DQD/SDHs and QDHs. It has been shown in the literature that shikimate biosynthesis is localized to plastids, while QDHs are most likely to be found in either plastids or the cytosol. Based on N-terminal sequence analysis, DQD/SDH1 was the only member with a potential chloroplast targeting peptide. Others do not show clear target signals, and are thus likely to be localized to the cytosol. Subcellular localization prediction by online tools failed to provide consistent results. Transient transformation of plants with recombinant YFP fusion proteins (i.e. two delivery system: agroinfiltration and particle bombardment; three transformation systems: tobacco, Arabidopsis and

onion) did not support a plastidial localization of DQD/SDH1. Fluorescent signals were detected in the cytosol for all Poplar DQD/SDHs and QDHs, which was not expected. This could be attributed to several factors including the use of heterologous transformation systems, incomplete gene annotation, the presence of additional genes, etc.. However, it still remains possible that Poplar DQD/SDHs and QDHs are indeed localized to the cytosol based on consistent results provided by the three transformation systems and the two delivery methods employed. Further studies including more detailed genome sequence analysis (focusing on the upstream intergenic regions) and transient expression using a homologous system (i.e. poplar) and alternative methods may be necessary to determine the localizations of Poplar DQD/SDHs and QDHs in a cell.

5.4 Evolution of the DQD/SDH/QDH superfamily

In Chapter 4, functional diversification of the DQD/SDH/QDH superfamily was studied. Biochemical characterization of DQD/SDHs and QDHs from the common ancestor of all seed plants and selected extant plant species revealed that quinate biosynthetic activity was present to minute amounts already in the common ancestor of the green lineages and remained low for a long period of time. Quinate biosynthetic function became beneficial and favored by selection during plant terrestrialization due to the role of quinate and its derivatives, such as chlorogenic acid, in plant defense and adaptation. Functional trade-offs were observed in early land plants, where a gradual increase in quinate activity coincided with a gradual decrease in shikimate specificity. This was also supported by the detection of relaxed purifying selection before gene duplication. Defense-related QDH function could not be optimized without compromising the essential shikimate biosynthetic activity. This adaptive conflict was

only resolved by gene duplication followed by specialization. Positive selection was detected on branches subtending the true DQD/SDH and QDH clades in the phylogeny. This suggested that adaptive changes occurred in both post-duplication groups, which gave rise to two groups of specialized genes encoding enzymes with one specific for shikimate biosynthesis (DQD/SDHs) and the other involved in quinate biosynthesis (QDHs). This was further supported by functional characterization of DQD/SDHs and QDHs from post-duplication species (i.e *P. trichocarpa* and *P. taeda*). Based on these results, it is most likely that the DQD/SDH/QDH superfamily has diversified via gene duplication through a mechanism described as "Escape from Adaptive Conflict" (EAC). Two key amino acid residues with roles in substrate discrimination (Ser338 and Thr381 in Arabidopsis) were identified through positive selection analysis. A first mutation (Ser to Gly) was gained by the common ancestor of all land plants, which led to the appearance of functional trade-offs observed in this study. Specialized QDH only came to exist with the gain of a second mutation (Thr to Gly) after gene duplication, while the first mutation (Ser to Gly) was reversed in true SDHs leading to highly specialized enzymes for either function. A third site Thr407 was identified and may play a role in the optimization of shikimate biosynthetic function in the dicot DQD/SDH clade.

5.5 Future directions

Overall, this dissertation demonstrated that *QDH* (a secondary metabolic gene) was recruited from *DQD/SDH*, a gene involved in primary metabolism, through EAC. This project also leaves possibilities for further investigation and extensions including:

- Arabidopsis has only one DQD/SDH, which has been functionally characterized, and homozygous null mutants in this gene (named *emb3004*) are known to be embryo lethal. The characterization of Poplar DQD/SDHs provides opportunities to test whether or not this lethal phenotype of Arabidopsis can be rescued by the introduction of Poplar DQD/SDH and if the residual DQD/SDH activity remained in QDHs is sufficient to do so. It also appears that Arabidopsis does not accumulate quinate or its derivatives such as chlorogenic acid. It will be interesting to investigate if Arabidopsis will gain the ability to produce quinate or its derivatives after being transformed with Poplar QDHs.
- Regulation of the shikimate pathway has been well studied in microbes. However, its regulation in plants has not been well established. There has been some evidence supporting that plants regulates the shikimate pathway mainly at the transcriptional level. Many transcription factors have been isolated. Identification of additional transcription factors and their interactions with previously identified regulators may facilitate better understanding of the regulation of the shikimate pathway in plants.
- Aromatic amino acids function as both protein building blocks and the precursors of plant secondary products. How carbon flow into these two sinks is regulated is still an open question. Identification of candidate genes with the same biochemical function but different physiological functions (i.e. poplar DQD/SDH1 and 2) allows deciphering the mode of action, possibly through protein-protein interactions with specific downstream proteins.

- Shikimate (and possibly quinate) plays important roles in the biosynthesis of lignin, which is a major hindrance in pulping and biofuel production. Targeting of this pathway may potentially increase the effectiveness of using poplar trees as a biofuel feedstock either by transgenic means or through marker-assisted tree breeding.

Bibliography

Agati G, Biricolti S, Guidi L, Ferrini F, Fini A, Tattini M (2011) The biosynthesis of flavonoids is enhanced similarly by UV radiation and root zone salinity in *L. vulgare* leaves. Journal of plant physiology 168: 204-212

Agati G, Stefano G, Biricolti S, Tattini M (2009) Mesophyll distribution of antioxidant flavonoids in *Ligustrum vulgare* leaves under contrasting sunlight irradiance. Annuals of Botany 104: 853-861

Agrawal AA (2000) Specificity of induced resistance in wild radish: causes and consequences for two specialist and two generalist caterpillars. Oikos 89: 493-500

Agrawal AA (2010) Current trends in the evolutionary ecology of plant defence. Functional Ecology 25 (2): 420-432

Ahuja I, Kissen R, Bones AM (2012) Phytoalexins in defense against pathogens. Trends in Plant Science 17 (2): 73-90

Aronsson H, Jarvis RP (2011) Rapid isolation of Arabidopsis chloroplasts and their use for in vitro protein import assays. Methods in Molecular Biology 774: 281-305

Astani A, Reichling J, Schnitzler P (2010) Comparative study on the antiviral activity of selected monoterpenes derived from essential oils. Phytotherapy Research 24: 673-679

Bagge P, LarssonC (1986) Biosynthesis of aromatic amino acids by highly purified spinach chloroplasts—compartmentation and regulation of the reactions. Physiologia Plantarum 68: 641–647

Bais HP, Vepachedu R, Gilroy S, Callaway RM, Vivanco JM (2003) Allelopathy and exotic plant invasion: From molecules and genes to species interactions. Science 301: 1377–1380

Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM (2006) The role of root exudates in rhizosphere interactions with plants and other organisms. Annual Reviews of Plant Biology 57: 233-266

Becker B, Marin B (2009) Streptophyte algae and the origin of embryophytes. Annals of Botany 103 (7): 999-1004

Bednarek P, Pislewska-Bednarek M, Svatos A, Schneider B, Doubsky J, Mansurova M, Humphry M, Consonni C, Panstruga R, Sanchez-Vallet A, Molina A, Schulze-Lefert P (2009) A glucosinolate metabolism pathway in living plant cells

mediates broad-spectrum antifungal defense. Science 323 (5910): 101-106

Beike AK, Rensing SA (2010) The *Physcomitrella patens* genome-a first stepping stone towards understanding bryophyte and land plant evolution. Tropical Bryology 31: 43-50

Bender SL, Mehdi S, Knowles JR (1989) Dehydroquinate synthase: the role of divalent metal cations and of nicotinamide adenine dinucleotide in catalysis. Biochemistry 28:7555–7560

Benesova M, Bode R (1992) Chorismate mutase isoforms from seeds and seedlings of *Papaver somniferum*. Phytochemistry 31: 2983–2987

Beninger CW, Abou-Zaid MM, Kistner ALE, Hallett RH, Iqbal MJ, Grodzinski B, Hall JC (2004) A flavanone and two phenolic acids from *Crysanthemum morifolium* with phytotoxic and insect growth regulating activity. Journal of Chemical Ecology 30:589–606

Bentley R, Haslam E (1990) The shikimate pathway-a metabolic tree with many branches. Critical Reviews in Biochemistry and Molecular Biology 25 (5): 307-384

Bergthorsson U, Andersson DI, Roth JR (2007) Ohno's dilemma: Evolution of new genes under continuous selection. Proceedings of the National Academy of Science of the United States of America 104 (43): 17004-17009

Bernards MA, Lewls NG (1992) Alkyl ferulates in wound healing potato tubers. Phytochemistry 31: 3409-3412

Bhushan S, Kuhn C, Berglund A, Roth C, Glaser E (2006) The role of the N-terminal domain of chloroplast targeting peptides in organellar protein import and miss-sorting. FEBS letters 580: 3966-3972

Bickel H, Palme L, Schultz G (1978) Incorporation of shikimate and other precursors into aromatic amino acids and prenylquinones of isolated spinach chloroplasts. Phytochemistry 17: 119–124

Bischoff M, Rösler J, Raesecke HR, Görlach J, Amrhein N, Schmid J (1996) Cloning of a cDNA encoding a 3-dehydroquinate synthase from a higher plant, and analysis of the organ-specific and elicitor-induced expression of the corresponding gene. Plant Molecular Biology 31:69–76

Bischoff M, Schaller A, Bieri F, Kessler F, Amrhein N, Schmid J (2001) Molecular charaterization of tomato 3-dehydroquinate dehydratase-shikimate: NADP oxidoreductase. Plant Physiology 125: 1891-1900

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inderred from age distributions of duplicated genes. The Plant Cell 16: 1667-1678

Blum T, Briesemeister S, Kohlbacher O (2009) MultiLoc2: integrating phylogeny and gene ontology terms improves subcellular protein localization prediction. BMC Bioinformatics 10: 274

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annual Review of Plant Biology 54: 519-546

Bonner CA, Jensen RA (1994) Cloning of cDNA encoding the bifunctional dehydroquinase. shikimate dehydrogenase of aromatic-amino-acid biosynthesis in *Nicotiana tabacum*. Biochemical Journal 302: 11-14

Boudet AM (1980) Studies on quinic acid biosynthesis in *Quercus pedunculata Ehrh*. seedlings. Plant & Cell Physiology 21 (5): 785-792

Boudet AM, Kajita S, Grima-Pettanati J, Goffner D (2003) Lignin and lignocellulosics: A better control of synthesis for new and improved uses. Trends in Plant Science 8:576–581

Boudet, A (1973) Les acides quinique et shikimique chez les angiospermes arborescentes. Phytochemistry 12: 363-370

Bourgaud F, Gravot A, Milesi A, Contier E (2001) Production of plant secondary metabolites: a historical perspective. Plant Science 161 (5): 839-851

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 72: 248-254

Braun M, Henstrand JM, Görlach J, Amrhein N, Schmid J (1996) Enzymatic properties of chorismate synthase isozymes of tomato (*Lycopersicon esculentum Mill.*). Planta 200 (1): 64-70

Bruce BD (2000) Chloroplast transit peptides: structure, function and evolution. Trends in Cell Biology 10: 440-447

Buschiazzo E, Ritland C, Bohlmann J, Ritland K (2012) Slow but not low: genomic comparisons reveal slower evolutionary rate and higher dN/dS in conifers compared to angiosperms. BMC Evolutionary Biology 12: 8

Byrne CE, Nagle DC (1997) Carbonization of wood for advanced materials applications. Carbon 35 (2): 259-266

Callaway RM, Thelen GC, Rodriguez A, Holben WE (2004) Soil biota and exotic plant invasion. Nature 427: 731-733

Campbell MM, Sederoff RR (1996) Variation in lignin content and composition: mechanisms of control and implications for the genetic improvement of plants. Plant Physiology 110: 3-13

Caruso M, Distefano G, Ye XR, La Malfa S, Gentile A, Tribulao E, Roose ML (2008) Generation of expressed sequence tags from carob (*Ceratonia siliqua L.*) flowers for gene identification and marker development. Tree Genetics & Genomes 4: 869-879

Chapple C, Ladisch M, Meilan R (2007) Loosening lignin's grip on biofuel production. Nature 25: 746-748

Chen F, Dixon RA (2007) Lignin modification improves fermentable sugar yields for biofuel production. Nature Biotechnology 25: 759-761

Chou KC, Shen HB (2010) Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. PLoS One 5(6): e11335

Christie PJ, Alfenito MR, Walbot V (1994). lmpact of low temperature stress on general phenylpropanoid and anthocyanin pathways: Enhancement of transcript abundance and anthocyanin pigmentation in maize seedlings. Planta 194: 541-549.

Chua N H, Schmidt GW (1978) Post- translational transport into intact chloroplasts of a precursor to the small subunit of ribulose- 1,5-bisphosphate carboxylase. Proceedings of the National Academy of Science of the United States of America 75: 6110–6114

Conant GC, Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new functions. Nature Reviews Genetics 9: 938-950

Crooks GE, Hon G, Chandonia J, Brenner SE (2004) WebLogo: A Sequence Logo Generator. Genome Research 14: 1188-1190

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Solits DE, Doyle JJ, Solits PS, Carlson JE, Arumuganathan K, Barakat A, Albert VA, Ma H, Depamphilis CW (2006) Widespread genome duplications throughout the history of flowering plants. Genome Research 15: 15

De Luca V, St Pierre B (2000) The cell and developmental biology of alkaloid biosynthesis. Trends in Plant Science 5: 168-173

De Moraes CM, Mescher MC, Tumlinson JH (2001) Caterpillar-induced nocturnal plant volatiles repel conspecific females. Nature 410: 577-580

De Sotillo DR, Hadley M, Wolf-Hall C (1998) Potato peel extract a nonmutagenic antioxidant with potential antimicrobial activity. Journal of Food Science 63: 907-910

Delaux P, Nanda AK, Mathe C, Sejalon-Delmas N, Dunand C (2011) Molecular and biochemical aspects of plant terrestrialization. Perspectives in Plant Ecology, Evolution and Systematics 14 (1): 49-59

Della-Cioppa G, Bauer SC, Klein BK, Shah DM, Fraley RT, Kishore GM (1986) Translocation of the precursor of 5-enolpyruvylshikimate-3-phosphate synthase into chloroplasts of higher plants in vitro. Proceedings of the National Academy of Science of the United States of America 83: 6873–6877

Delport W, Poon AFY, Frost SDW, Pond SLK (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26 (19): 2455-2457

Des Marais DL, Rausher M D (2008) Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature 454: 762-765

Devoto A, Ellis C, Magusin A, Chang H, Chilcott C, Zhu T, Turner JG (2005) Expression profiling reveals COI1 to be a key regulator of genes involved in wound- and methyl jasmonate-induced secondary metabolism, defense, and hormone interactions. Plant Molecular Biology 58 (4): 497-513

Dewick PM (1995) The biosynthesis of shikimate metabolites. Natural Product Reports 12: 579-607

Ding L, Hofius D, Hajirezaei MR, Fernie AR, Bornke F, Sonnewald U (2007) Functional analysis of the essential bifunctional tobacco enzyme 3-dehydroquinate dehydratase/shikimate dehydrogenase in transgenic tobacco plants. Journal of Experimental Botany 58: 2053-2067

Dixon RA, Paiva NL (1995) Stress-induced phenylpropanoid metabolism. Plant Cell 7: 1085-1097

Dowd PF, Vega FE (1996) Enzymatic oxidation products of allelochemicals as a basis for resistance against insects: effects on the corn leafhopper *Dalbulus maidis*. Natural Toxins 4: 85–91

Dudareva N, Negre F, Nagegowda DA, Orbva I (2007) Plant volatiles: Recent advances and future perspectives. Critical Reviews in Plant Sciences 25 (5): 417-440

Durrant WE, Dong X (2004) Systemic acquired resistance. Annual Review of Phytopathology 42: 185-209

Dyer WE, Henstrand JM, Handa AK, Herrmann KM (1989) Wounding induces the first enzyme of the shikimate pathway in *Solanaceae*. Proceedings of the National Academy of Science of the United States of America 86 (19): 7370-7373

Edwards D, KerpH, Hass H (1998) stomata in early land plants: an anatomical and ecophysiological approach. Journal of Experimental Botany 49: 255-278

Emanuelsson O, Nielsen H, Brunak S, Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. Journal of Molecular Biology 300: 1005-1016

Emanuelsson O, Nielsen H, von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science 8 (5): 978-984

Emanuelsson O, von Heijne G (2001) Prediction of organellar targeting signals. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research 1541: 114-119

Emiliani G, Fondi M, Fani R, Gribaldo (2009) A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. Biology Direct 4

Emiliani J, Grotewold E, Ferreyra F, Lorena M, Paul C (2013) Flavonols protect arabidopsis plants against UV-B deleterious effects. Molecular Plant 6 (4): 1376-1379

Entus R, Poling M, Herrmann KM (2002) Redox regulation of Arabidopsis 3-deoxy-d-*arabino*-heptulosonate 7-phosphate synthase. Plant Physiology 129: 1866–1871

Erxleben A, Gessler A, Verviliet-Scheebaum M, Reski R (2012) Metabolite profiling of the moss *Physcomitrella patens* reveals evolutionary conservation of osmoprotective substances. Plant Cell Reports 31 (2): 427-436

Essar QW, Eberly L, Hadero A, Crawford IP (1990) Identification and characterization of genes for a second anthranilate synthase in *Pseudomonas aeruginosa*: interchangeability of the two anthranilate synthases and evolutionary implications. Journal of Bacteriology 172 (2): 884-900

Farah A, Monteiro M, Donangelo CM, Lafay S (2008) Chlorogenic acids from green coffee extracts are highly bioavailable in humans. Journal of Nutritions 138 (12): 2309-2315

Fawcett JA, Maere S, Van de Peer Y (2009) Plants with double genomes might have had a better chance to survive the Cretacceous –Tertiary Extinction event. Proceedings of the National Academy of Science of the United States of America 106: 5737-5742

Feierabend J, Brassel D (1977) Subcellular localization of shikimate dehydrogenase in higher plants. Zeitschrift für Pflanzenphysiologie 82 (4): 334-346

Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5: 164-166

Felton GW, Korth KL, Bi JL, Wesley SV, Huhman DV, Mathews MC, Murphy JB, Lamb C, Dixon RA (1999) Inverse relationship between systemic resistance of plants to microorganisms and to insect herbivory. Current Biology 9: 317-320

Fisher RF, Long SR (1992) Rhizobium – plant signal exchange. Nature 357: 655–660

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytologist 183: 557-564

Floss HG (1979) The shikimate pathway. Recent Advances in Phytochemistry 12: 59-89

Floyd SK, Bowman JL (2007) The ancestral developmental tool kit of land plants. International Journal of Plant Science 168: 1-35

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151: 1531-1545

Forlani G, Parisi B, Nielsen E (1994) 5-enol-pyruvyl-shikimate 3-phosphate synthase from Zea mays cultured cells. Plant Physiology 105: 1107–1114

Fucile G, Falconer S, Christendat D (2008) Evolutionary diversification of plant shikimate kinase gene duplicates. PLoS Genetics 4: e1000292

Fucile G, Garcia C, Carlsson J, Sunnerhagen M, Christendat D (2011) Structural and biochemical investigation of two Arabidopsis shikimate kinases: The heat-inducible isoform is thermostable. Protein Science 20: 1125–1136

Gamborg OL (1966) Aromatic metabolism in plants III. Quinate dehydrogenase from mung bean cell suspension culture. Biochimica et Biophysica Acta (BBA) - Enzymology and Biological Oxidation 128 (3): 483-491

Gamborg OL (1967) Aromatic metabolism in plants—IV: The interconversion of shikimic acid and quinic acid by enzymes from plant cell cultures. Phytochemistry 6 (8): 1067-1073

Ganson RJ, D'Amato TA, Jensen RA (1986) The two-isozyme system of 3-deoxy-d-*arabino*-heptulosonate 7-phosphate synthase in *Nicotiana silvestris* and other higher plants. Plant Physiology 82: 203–210

Gasser CS, Klee HJ (1990) A *Brassica napus* gene encoding 5-enolpyruvylshikimate 3-phosphate synthase. Nucleic Acids Research 18:2821

Gasser CS, Winter JA, Hironaka CM, Shah DM (1988) Structure, expression, and evolution of the 5-enolpyruvylshikimate 3-phosphate synthase genes of petunia and tomato. The Journal of Biological Chemistry 263: 4280–4289

Gautier L, Cope l, Bolstad BM, Irizarry RA (2003) affy—analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20 (3): 307-315

Gavel Y, von Heijne G (1990) A conserved cleavage-site motif in chloroplast transit peptides. FEBS Letters 261 (2): 455-458

Gensel PG (2008) The earliest land plants. Annual Review of Ecology, Evolution, and

Systematics 39(1): 459-477

Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA (2007) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucleic Acids Research 35: e55

Gibson RA, Schneider EA, Wightman F (1972) Biosynthesis and metabolism of indol-3yl-acetic acid. Journal of Experimental Botany 23(1): 152–170

Givry S, Bliard C, Duchiron F (2007) Selective ketopentose analysis in concentrate carbohydrate syrups by HPLC. Elsevier 342: 859-864

Glaser E, Soll J (2004) Targeting signals and import machinery of plastids and plant mitochondria. In Molecular Biology and Biotechnology of Plant Organelles: Chloroplasts and mitochondria, Daniell H and Chase C (eds.), Kluwer Academic Publishers, pp 385-418

Goldschmidt O, Quimby GR (1964) The role of quinic acid. Tappi 47: 528–533

Gora V, Konig J, Lunderstadt J (1994) Physiological defence reactions of young beech trees (*Fagus sylvatica*) to attack by *Phyllaphis fagi*. Forest Ecol. Manag. 70(1-3): 245-254

Görlach J, Raesecke H, Rentsch D, Regenass M, Roy P, Zala M, Keel C, Boller T, Amrhein N, Schmid (1995) Temporally distinct accumulation of transcripts encoding enzymes of the prechorismate pathway in elicitor-treated, cultured tomato cells. Proceedings of the National Academy of Science of the United States of America 92: 3166–3170

Görlach J, Schmid J, Amrhein N (1993) Differential expression of tomato (*Lycopersicon esculentum L.*) genes encoding shikimate pathway isoenzymes. II. Chorismate synthase. Plant Molecular Biology 23: 707–716

Görlach J, Schmid J, Amrhein N (1994) Abundance of transcripts specific for genes encoding enzymes of the prechorismate pathway in different organs of tomato (*Lycopersicon esculentum L.*) plants. Planta 193: 216–223

Grace SC, Logan BA (2000) Energy dissipation and radical scavenging by the plant phenylpropanoid pathway. Philosophical Transactions of the Royal Society of London Series B355: 1499–1510

Graziana A, Boudet A, Boudet AM (1980) Association of the quinate: NAD+ oxidoreductase with one dehydroquinate hydro-lyase isoenzyme in corn seedlings. Plant Cell Physiology 21 (8): 1163-1174

Graziana A, Boudet AM (1983) La quinate hydro-lyase, nouvelle enzyme du métabolisme des acides alicycliques. In: Boudet AM, Ranjeva R (eds) Groupe Polyphénols Journées internationales d'études et assemblées générales 11: 120–

125

Grotewold E (2006) The genetics and biochemistry of floral pigments. Annual Review of Plant Biology 57: 761–780

Grotkopp E, Rehmanek M, Rost TL (2002) Toward a causal explanation of plant invasiveness: Seedling growth and life-history strategies of 29 pine (Pinus) species. The American Naturalist 159: 396-419

Gunatilaka AL (2008) Natural Products in Plants : Chemical Diversity. Wiley Encyclopedia of Chemical Biology, 1–17

Hahn MW (2009) Distinguishing among evolutionary models for the maintenance of gene duplicates. Journal of Heredity 00 (5): 605-617

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95-98

Hamberger B, Ehlting J, Barbazuk B, Douglas CJ (2006) Comparative genomics of the shikimate pathway in Arabidopsis, *Populus trichocarpa* and *Oryza sativa*: shikimate pathway gene family structure and identification of candidates for missing links in phenylalanine biosynthesis. Recent Advances in Phytochemistry 40: 85-113

Hanson JR (2003) Natural products: the secondary metabolites. Royal Society of Chemistry,UK

Hartmann T (1996) Diversity and variability of plant secondary metabolism: a mechanistic view. Entomologia Experimentalis et Applicata 80 (1): 177-188

Hartmann T (2004) Plant-derived secondary metabolites as defensive chemicals in herbivorous insects: a case study in chemical ecology. Planta 219 (1): 1-4

Hartmann T (2007) From waste products to ecochemicals: Fifty years research of plant secondary metabolism. Phytochemistry 68 (22-24): 2831-2846

Haruta M, Major IT, Christopher ME, Patton JJ, Constabel CP (2001) A Kunitz trypsin inhibitor gene family from trembling aspen (*Populus tremuloides Michx.*): cloning, functional expression, and induction by wounding and herbivory. Plant Mol Biol 46: 347–359

Haslam E (1994) Secondary metabolism—evolution and function: Products or processes? Chemoecology 5-6 (2): 89-95

Haworth m, Elliott-Kingston C, McElwain JC (20110) Stomatal control as a driver of plant evolution. Journal of Experimental Botany 62 (8): 2419-2423

Heineke D, Riens B, Grosse H, Hoferichter P, Peter U, Flugge U, Heldt HW (1991) Redox transfer across the inner chloroplast envelope membrane. Plant Physiology 95 (4): 1131-1137

Herrmann KM (1995a) The shikimate pathway: early steps in the biosynthesis of aromatic compounds. The Plant Cell 7: 907-919

Herrmann KM (1995b) The shikimate pathway as an entry to aromatic secondary metabolism. Plant Physiology 107 (1): 7-12

Herrmann KM, Weaver LM (1999) The shikimate pathway. Annual Reviews of Plant Physiology and Plant Molecular Biology 50: 473-503

Highfield PE, Ellis RJ (1978) Synthesis and transport of the small subunit of chloro-plast ribulose bisphosphate carboxylase. Nature 271: 420–424

Hirohashi T, Hase T, Nakai M (2001) Maize non-photosynthetic ferredoxin precursor is missorted to the intermembrane space of chloroplasts in the presence of light. Plant Physiology, 125: 2154–2163

Hittinger CT, Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. Nature 449: 677-681

Hoffmann L, Besseau S, Geoffroy P, Rotzenthaler C, Meyer D, Lapierre C, Pollet B, Legrand M (2004) Silencing of Hydroxycinnamoyl-Coenzyme A Shikimate/Quinate Hydroxycinnamoyltransferase Affects Phenylpropanoid Biosynthesis. The Plant Cell 16 (6): 1446-1465

Homeyer U, Schultz G (1988) Activation by light of plastidic shikimate pathway in spinach. Plant Physiology and Biochemistry 26: 365–370

Hoover K, Alaniz SA, Yee JL, Rocke DM, Hammock BD, Duffey SS (1998) Dietary protein and chlorogenic acid effect on baculovirus disease of noctuid (*Lepidoptera: Noctuidae*) larvae. Environmental Entomology 27: 1264–1272

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) Wolf PSORT: protein localization predictor. Nucleic Acid Research 35 (2): 585-587

Hossaert-McKey M, Soler C, Schatz B, Proffit M (2010) Floral scents: their roles in nursery pollination mutualisms. Chemoecology 20: 75–88

Huang R, Hippauf F, Rohrbeck D, Haustein M, Wenke K, Feike J, Sorrelle n, Piechulla B, Barkman TJ (2012) Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. Proceedings of the National Academy of Science of the United States of America 109 (8): 2967-2971

Hughes AL (1994) The Evolution of Functionally Novel Proteins after Gene Duplication.

Proceedings of the Royal Society B: 256 (1346): 119-124

Huminieki I, Wolfe KH (2004) Divergence of spatial gene expression profiles following species-specific gene duplication in human and mouse. Reviews of Microbiology 30: 409-425

Iason G (2005) The role of plant secondary metabolites in mammalian herbivory: ecological perspectives. Proceedings of the Nutrition Society 64: 123-131

Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nature Reviews Genetics 11: 97-108

Jarvis P, Robinson C (2004) Mechanisms of protein import and routing in chloroplasts. Current Biology 14 (24): 1064-1077

Jassbi AR (2003) Secondary metabolites as stimulants and antifeedants of *Salix integra* for the leaf beetle *Plagiodera versicolora*. Z Naturforsch C 58: 573–579

Jasson S, Douglas CJ (2007) *Populus*: A Model System for Plant Biology. Annual Review of Plant Biology 58: 435-458

Javis P, Soll J (2001) Toc, Tic, and chloroplast protein import. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1541 (1-2): 64-79

Jockovic N, Andrade PB, Valentao P, Sabovljevic M (2008) HPLC-DAD of phenolics in bryophytes *Lunularia cruciata*, *Brachytheciastrum velutinum* and *Kindbergia praelonga*. Journal of the Serbian Chemical Society 73 (12): 1161-1167

Kang X, Scheibe R (1993) Purification and characterization of the quinate: Oxidoreductase from *Phaseolus mungo* sprouts. Phytochemistry 33 (4) 769-773

Kasai K, Kanno T, Akita M, Ikejiri-Kanno Y, Wakasa K, Tozawa Y (2005) Identification of three shikimate kinase genes in rice: characterization of their differential expression during panicle development and of the enzymatic activities of the encoded proteins. Planta 222: 438–447

Keegstra K, Froehlich JE (1999) protein imports into chloroplasts. Current Opinions on Plant Biology 2: 471-476

Keeling CI, Dullat HK, Yuen M, Ralph SG, Jancsik S, Bohlmann J (2010) Identification and Functional Characterization of Monofunctional *ent*-Copalyl Diphosphate and *ent*-Kaurene Synthases in White Spruce Reveal Different Patterns for Diterpene Synthase Evolution for Primary and Secondary Metabolism in Gymnosperms. Plant Physiology 152 (3): 1197-1208

Keith B, Dong XN, Ausubel FM, Fink GR (1991) Differential induction of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase genes in *Arabidopsis thaliana* by wounding and pathogenic attack. Proceedings of the National Academy of

Science of the United States of America 88 (19): 8821-8825

Kelly LA, Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nature Protocols 4: 363-371

Kenrick P, Crane PR (1997) The origin and early evolution of plants on land. Nature 389: 33–39

Khersonsky O, Tawfik DS (2010) Enzyme promiscuity: a mechanistic and evolutionary perspectives. Annual review of Biochemistry 79: 471-505

Kikkert JR, Vidal JR, Reisch BI (2005) Stable transformation of plant cells by particle bombardment/biolistics. Methods in Molecular Biology 286: 61-78

Klee HJ, Muskopf YM, Gasser CS (1987) Cloning of an *Arabidopsis thaliana* gene encoding 5-enolpyruvylshikimate-3-phosphate synthase: sequence analysis and manipulation to obtain glyphosate-tolerant plants. Molecular & General Genetics 210: 437–442

Koch BM, Sibbesen O, Halkier BA, Svendsen I, Moller BL (1995) The primary sequence of cytochrome P450tyr, the multifunctional N-hydroxylase catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. Archives of Biochemistry and Biophysics 323(1): 177-186

Kopertakh L, Schiemann J (2005) Agroinfiltration as a tool for transient expression of cre recombinase in vivo. Transgenic Research 14: 793-798

Kroymann J (2011) Natural diversity and adaptation in plant secondary metabolism. Current Opinion in Plant Biology 14 (3): 246-251

Künzler M, Paravicini G, Egli CM, Irniger S, Braus GH (1992) Cloning, primary structure and regulation of the ARO4 gene, encoding the tyrosine-inhibited 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from *Saccharomyces cerevisiae*. Gene 113: 67–74

Kutchan TM (1995) Alkaloid Biosynthesis: The Basis for Metabolic Engineering of Medicinal Plants. The Plant Cell 7 (7): 1059-1070

Laranjinha JA, Almeida LM, Madeira VM (1994) Reactivity of dietary phenolic acids with peroxyl radicals: antioxidant activity upon low density lipoprotein peroxidation. Biochemical Pharmacology 48: 487–494

Lee DW, Kim JK, Lee S, Choi S, Kim S, Hwang I (2008) Arabidopsis Nuclear-Encoded Plastid Transit Peptides Contain Multiple Sequence Subgroups with Distinctive Chloroplast-Targeting Sequence Motifs. The Plant Cell 20 (6): 1603-1622

Leimu R, Muola A, Laukkanen L, Kalske A, Prill N, Mutikainen P (2012) Plant-

herbivore coevolution in a changing world. Entomologia Experimentalis et Applicata 144: 3-13

Leiss KA, Maltese F, Choi YH, Verpoorte R, Klinkhamer PGL (2009) Identification of chlorogenic acid as a resistance factor for thrips in *Chrysanthemum*. Plant Physiology 150 (3): 1567-1575

Lemieux C, Otis C, Turmel M (2007) A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. BMC Biology 5: 2

Leuschner C, Schultz G (1991) Uptake of shikimate-pathway intermediates in chloroplasts. Phytochemistry 30: 2203-2207

Leuschner K, Herrmann KM, Schultz G (1995) The metabolism of quinate in pea roots (purification and partial characterization of a quinate hydrolyase). Plant Physiology 108 (1): 319-325

Lewis L, McCourt RM (2004) Green algae and the origin of land plants. American Journal of Botany 91: 1535-1556

Li HM, Chiu CC (2010) Protein transport into chloroplast. Annual Review of Plant Biology 61: 157-180

Li LG, Zhou YH, Cheng XF, Sun JY, Marita JM, Ralph J, Chiang VL (2003) Combinatorial modification of multiple lignin traits in trees through multigene cotransformation. Proceedings of the National Academy of Science of the United States of America 100: 4939-4944

Li Z, Wang Q, Ruan X, Pan C, Jiang D (2010) Phenolics and plant allelopathy. Molecules 15 (2): 8933-8952

Lin L, Kuo Y, Chou C (2000) Cytotoxic bioflavonoids from *Selaginalle delicatula*. Journal of Natural Products 63 (5): 627-630

Lindner HA, Guy N, Allan M, Gurvan M, Robert m, Miroslaw C (2005) Site-directed mutagenesis of the active site region in the quinate/shikimate 5-dehydrogenase YdiB of *Escherichia coli*. The Journal of Biological Chemistry 280 (8): 7162-7169

Lois R (1994) Accumulation of UV-absorbing flavonoids induced by UV-B radiation in *Arabidopsis thaliana L*. I. Mechanisms of UVresistance in Arabidopsis. Planta 194: 498-503

Lou Z, Wang H, Zhu S, Ma C, Wang Z (2011) Antibacterial activity and mechanism of action of chlorogenic acid. Journal of Food Science 76 (6): M398-M403

Lowry B, Lee D, Hebant C (1980) The origin of land plants: a new look at an old

problem. Taxon 29 (2/3): 183-197

Lu J, Bao J, Wu G, Xu, W, Huang M, Chen X, Wang Y (2013) Quinones derived from plant secondary metabolites as anti-cancer agents. Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry - Anti-Cancer Agents) 13 (3): 456-463

Lupien S, Karp F, Wildung M, Croteau R (1999) Regiospecific cytochrome p450 limonene hydroxylases from mint (Mentha) species: cDNA isolation, characterization, and functional expression of (2)-4S-limonene-3-hydroxylase and (2)-S-limonene-6-hydroxylase. Archives of Biochemistry and Biophysics 368 (1): 181-192

Lytovchenko O, Melin J, Schulz C, Kilisch M, Hutu DP, Rehling P (2013) Signal recognition initiates reorganization of the presequence translocase during protein import. The EMBO Journal 32: 886-898

Ma CM, Kully M, Khan JK, Hattori M, Daneshtalab M (2007) Synthesis of chlorogenic acid derivatives with promising antifungal activity. Bioorganic & Medicinal Chemistry 15: 6830-6833

Macel M (2011) Attract and deter: a dual role for pyrrolizidine alkaloids in plant–insect interactions. Phytochemistry Reviews 10 (1): 75-82

Maclean J, Ali S (2003) The structure of chorismate synthase reveals a novel flavin binding site fundamental to a unique chemical reaction. Structure 11:1499–1511

Maeda H, Dudareva N (2012) The shikimate pathway and aromatic amino acid biosynthesis in Plants. Annual Review of Plant Biology 63: 73-105

Maere S, De Bodt S, Raes J, Casneuf T, Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proceedings of the National Academy of Science of the United States of America 102: 5454-5459

Maher EA, Bate NJ, Ni W, Elkind Y, Dixon RA, Lamb CJ (1994) Increased disease susceptibility of transgenic tobacco plants with suppressed levels of preformed phenylpropanoid products. Proceedings of the National Academy of Science of the United States of America 91: 7802-7806

Malik S, Cusido RM, Mirjalili MH, Moyano E, Palazon J, Bonfill M (2011) Production of the anticancer drug taxol in *Taxus baccata* suspension cultures: A review. Processes Biochemistry 46 (1): 23-34

Mäntylä E, Alessio GA, Blande JD, Heijari J, Holopainen JK, Laaksonen T, Piirtola P, Klemola T (2008) From plants to birds: higher avian predation rates in trees responding to insect herbivory. PLoS ONE 3: p. e28322810.1371/Journal.pone.0002832

Marsh KB, Boldingh HL, Shilton RS, Laing WA (2008) Changes in quinic acid metabolism during fruit development in three kiwifruit species. Functional Plant Biology 36 (5): 463-470

Mattocks AR (1986) Toxicity of pyrrolizidine alkaloids. Nature 217:723–728

Matsuno M, Compagnon V, Schoch GA, Schmitt M, Debayle D, Bassard JE, Pollet B, Hehn A, Heintz D, Ullmann P, Lapierre C, Bernier F, Ehlting J, Werck-Reichhart D (2009) Evolution of a novel phenolic pathway for pollen development. Science 325: 1688–1692

May T, Soll J (2000) 14–3–3 proteins form a guidance complex with chloroplast precursor proteins in plants. Plant Cell 12: 53-63

McCandliss RJ, Herrmann KM (1978) Iron, an essential element for biosynthesis of aromatic compounds. Proceedings of the National Academy of Science of the United States of America 75:4810–4813

Meinke D, Muralla R, Sweeney C, Dickerman A (2008) Identifying essential genes in Arabidopsis thaliana. Trends in Plant Science 13: 483-491

Michel G, Roszak AW, Sauve V, Maclean J, Matte A, Coggins JR, Cygler M, Lapthorn AJ (2003) Structures of shikimate dehydrogenase AroE and its paralog YdiB. A common structural framework for different activities. The Journal of Biological Chemistry 278: 19463-19472

Miller R, Owens SJ, Rorslett B (2011) Plants and color: flowers and pollination. Optics and Laser Technology 43: 282-294

Minamikawa T (1977) Quinate: NAD oxidoreductase of germinating *Phaseolus mungo* seeds: Partial purification and some properties. Plant and Cell Physiology 18 (4): 743-752

Minamikawa T, Yoshiba S, Hasegawa M (1969) Alicyclic acid metabolism in plants 3. Fate of 14C-shikimate and 14C-quinate in mung bean plants. Plant Cell Physiology 10 (2): 283-289

Mithofer A, Boland W (2012) Plant defense against herbivores: chemical aspects. Annual Review of Plant Biology 63: 431-450

Montenegro G, Portaluppi MC, Salas FA, Diaz MF (2008) Biological properties of the Chilean native moss *Sphagnum magellanicum*. Biological Research 42 (2): 233-237

Morton LW, Caccetta RA, Puddey IB, Croft KD (2000) The chemistry and biological effects of dietary phenolic compounds: Relevance to cardiovascular disease. Clinical and Experimental Pharmacology and Physiology 27 (3): 152–159

Mousdale DM, Coggins JR (1985) High-performance liquid chromatography of shikimate pathway intermediates. Journal of Chromatography A 329: 268-272

Mousdale DM, Coggins JR (1985) Subcellular localization of the common shikimate-pathway enzymes in *Pisum sativum L*. Planta 163: 241–249

Mousdale DM, Coggins JR (1986) Detection and subcellular localization of a higher plant chorismate synthase. FEBS Letters 205: 328–332

Muir RM, Ibanez AM, Uratsu SL, Ingham ES, Leslie CA, McGranahan GH, Batra N, Goyal S, Joseph J, Jemmis ED, Dandekar AM (2011) Mechanism of gallic acid biosynthesis in bacteria (*Escherichia coli*) and walnut (*Juglans regia*). Plant Molecular Biology 75: 555-565

Muller KO, Borger H (1940) Experimentelle Untersuchungen über die Phythophthora-Resistenz der Kartoffel. Zugleich ein Beitrag zum Problem der 'erworbenen Resistenz' im Pflanzenreich. Arbeiten der Biologischen Reichsanstalt für Land- und Forstwirtschaft 23: 189–231

Murrell B, Werthrim JO, Moola S, Weighill T, Scheffler K, Pond SLK (2012) Detecting Individual Sites Subject to Episodic Diversifying Selection. PLoS Genetics 8 (7): e1002764

Nasvall J, Sun L, Roth JR, Andersson DI (2012) Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence. Science 338: 384-387

Newman DJ, Cragg GM, Snader KM (2003) Natural products as sources of new drugs over the period 1981-2002. Journal of Natural Products 66: 1022-1037

Newton EL, Bullock JM, Hodgson DG (2009) Glucosinolate polymorphism in wild cabbage (*Brassica oleracea*) influences the structure of herbivore communities. Oecologia 160: 63-76

Nie L, Shi X (2008) A novel asymmetric synthesis of oseltamivir phosphate (Tamiflu) from (−)-shikimic acid. Tetrahedron: Asymmetry 20 (1): 124-129

Niggeweg R, Michael AJ, Martin C (2004) Engineering plants with increased levels of the antioxidant chlorogenic acid. Nat Biotechnol 22: 746–754

Nilsson I, Whitley P, von Heijne G (1994) The COOH-terminal ends of internal signal and signal-anchor sequences are positioned differently in the ER translocase. The Journal of Cell Biology 126 (5): 1127-1132

Nishida R, Tan KH, Wee SL, Hee AKW, Toong YC (2004) Phenylpropanoids in the fragrance of the fruit fly orchard, *Bulbophyllum cheiri*, and their relationship to the pollinator, *Bactrocera papayae*. Biochemical Systematics and Ecology 32: 245-252

O'Callaghan D, Maskell D, Liew FY, Easmon CS, Dougan G (1988) Characterization of aromatic- and purine-dependent *Salmonella typhimurium*: attention, persistence, and ability to induce protective immunity in BALB/c mice. Infection and Immunity 56 (2): 419-423

Ober D (2005) Seeing double: gene duplication and diversification in plant secondary metabolism. Trends in Plant Science 10 (9): 444-449

Onofrejova L, Vasickova J, Klejdus B, Stratil P, Misurcova L, Kracmar S, Kopecky J, Vacek J (2010) Bioactive phenols in algae: The application of pressurized-liquid and solid-phase extraction techniques. Journal of Pharmaceutical and Biomedical Analysis 51 (2): 464-470

Ossipov V, Bonner C, Ossipova S, Jensen R (2000) Broad-specificity quinate (shikimate) dehydrogenasefrom *Pinus taeda* needles. Plant Physiology and Biochemistry 38 (12): 923-928

Ossipov V, Chernov A, Zrazhevskaya G, Shein I (1995) Quinate: NAP(P)+-oxidoreductase from *Larix sibirica*: purification, characterization and function. Trees 10 (1): 46-51

Ossipov VI, Shein IV (1986) Role of quinate dehydrogenase in quinic acid metabolism in conifers. Biochemistry 51(2): 184-190

Paravicini G, Braus G, Hütter R (1988) Structure of the ARO3 gene of *Saccharomyces cerevisiae*. Molecular & General Genetics 214: 165–169

Paravicini G, Schmidheini T, Braus G (1989) Purification and properties of the 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (phenylalanine-inhibitable) of *Saccharomyces cerevisiae*. European journal of biochemistry 186: 361–366

Paynter NP, Yeh HC, Voutilainen S, Schmidt MI, Heiss G, Folsom AR, Brancati FL, Kao WHL (2006) Coffee and sweetened beverage consumption and the risk of type 2 diabetes mellitus: the atherosclerosis risk in communities study. American Journal of Epidemiology 164: 1075–1084

Petersen M, Abdullah Y, Benner J, Eberle D, Gehlen K, Hucherig S, Janiak V, Kim KH, Sander M, Weitzel C, Wolfers S (2009) Evolution of rosmarinic acid biosynthesis. Evolution of Metabolic Diversity 70: 1663-1679

Phillips DA, Dakora FD, Sands E, Joseph CM, Zon J (1994) Synthesis, release, and transmission of alfalfa signals to rhizobial symbionts. Plant and Soil 161 (1): 69-80

Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends in Plant Science 5 (10): 439-445

Pierleoni A, Marteli PL, Fariselli P, Casadio R (2006) BaCelLo: a balanced subcellular

localization predictor. Bioinformatics 22 (14): e408-e416

Pinto JEBP, Suzich JA, Herrmann KM (1986) 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase from potato tuber (*Solanum tuberosum L.*). Plant Physiol. 82: 1040–1044

Pires ND, Dolan L (2012) Morphological evolution in land plants: new design with old genes. Philosophical Transactions of the Royal Society B 367: 508-518

Pittard J, Wallace BJ (1966) Distribution and function of genes concerned with aromatic biosynthesis in Escherichia coli. Journal of Bacteriology 91: 1494-1508

Poelman EH, Van Loon JA, Van Dam NM, Vet LM, Dicke M (2010) Herbivore-induced plant responses in *Brassica oleracea* prevail over effects of constitutive resistance and result in enhanced herbivore attack. Ecological Entomology 35: 240-247

Pond SLK, Frost SDW (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21 (10): 2531-2533

Poon AFT, Frost SDW, Pond K (2009) Detecting signatures of selection from DNA sequences using datamonkey. Methods in Molecular Biology 537: 163-183

Popper ZA, Michel G, Herve C, Domozych DS, Willats WGT, Tuohy MG, Kloareg B, Stengel DB (2011) Evolution and diversity of plant cell walls: from algae to flowering plants. Annual Review of Plant Biology 62: 567-590

Qiu YL (2008) Phylogeny and evolution of charophytic algae and land plants. Journal of Systematics and Evolution 46: 287–306

Quail PH (1979) Plant cell fractionation. Annual Review of Plant Physiology 30: 425-484

Quevillon-Cheruel S, Leulliot N, Meyer P, Graille M, Bremang M, Blondeau K, Sorel I, Poupon A, Janin J, van Tilbeurgh H (2004) Crystal structure of the bifunctional chorismate synthase from *Saccharomyces cerevisiae*. The Journal of Biological Chemistry 279: 619–625

Radwanski ER, Last RL (1995) Tryptophan biosynthesis and metabolism: biochemical and molecular genetics. The Plant Cell 7: 921-934

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairnney J, Eckert CA, Frederick Jr. WJ, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T (2006) The pathway forward for biofuels and biomaterials. Science 311: 484-489

Ranjeva R, Refeno G, Boudet AM, Marme D (1983) Activation of plant quinate:NAD+ 3-oxidoreductase by Ca2+ and calmodulin. Proceedings of the National

Academy of Science of the United States of America 80 (17): 5222-5224

Refeno G, RanJeva R, Boudet AM (1982) Modulation of quinate: NAD+ oxidoreductase activity through reversible phosphorylation in carrot cell suspensions. Planta 154 (3): 193-198

Rensing SA, Lang D, Zimmer A, Terry A, Salamov A, Shapiro H, Mishiyama T, Perroud P, Lindquist EA, Kamisugi Y, Tanahas T, Sakakibara K, Fujita T, Oishi K, Shin T, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Suhano S, Kohara Y, Fuijiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle m, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hick KA, Hughes J, Lohr M, Mayer K, MelKozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu S, Stueber K, Theodoulou FL, Tu H, de Peer YV, Verrier PJ, Waters E, Wood A, Yang l, Cove D, Cuming AC, Hasebe m, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL (2008) The Physcomitrella Genome Reveals Evolutionary Insights into the Conquest of Land by Plants. Science 319 (5859): 64-69

Riano-Pachon DM, Correa LGG, Trejos-Espinosa R, Mueller-Roeber B (2008) Green transcription factors: A Chlamydomonas overview. Genetics 179: 31-39

Richardt S, Lang D, Reski R, Frank W, Rensing SA (2007) PlanTAPDB, a phylogeny-based resource of plant transcription-associated proteins. Plant Physiology 143: 1452-1466

Rippert P, Puyaubert J, Grisollet D, Derrier L, Matringe M (2009) Tyrosine and phenylalanine are synthesized within the plastids in arabidopsis. Plant Physiology 149 (3): 1251-1260

Roberts F, Roberts CW, Johnson JJ, Kyle DE, Krell T, Coggins JR, Coombs GH, Milhous WK, Tzipori S, Ferguson DJP, Chakrabarti D, McLeod R (1998) Evidence for the shikimate pathway in apicomplexan parasites. Nature 393: 801-805

Rogiers SY, Knowles NR (1997) Physical and chemical changes during growth, maturation, and ripening of saskatoon (*Amelanchier alnifolia*) fruit. Canadian Journal of Botany 75 (8): 1215-1225

Rushton PJ, Somssich IE, Ringler P, Shen QJ (2010) WRKY transcription factors. Trends in Plant Science 15 (5): 247-258

Rutschow H, Ytterberg AJ, Friso G, Nilsson R, van Wijk KJ (2008) Quantitative proteomics of a chloroplast SRP54 sorting mutant and its genetic interactions with CLPC1 in Arabidopsis. Plant Physiology 148 (1): 156-175

Sanchez-Vallet A, Ramos B, Bednarek P, López G, Piślewska-Bednarek M, Schulze-

Lefert P, Molina A (2010) Tryptophan-derived secondary metabolites in *Arabidopsis thaliana* confer non-host resistance to necrotrophic *Plectosphaerella cucumerina* fungi. The Plant journal : for cell and molecular biology 63(1): 115-127

Sannigrahi P, Ragauskas AJ (2010) Poplar as a feedstock for biofuels: A review of compositional characteristics. Biofuels, Bioproducts, Biorefining 4: 209-226

Sawa T, Nakao M, Akaike T, Ono K, Maeda H (1999) Alkylperoxyl radical- scavenging activity of various flavonoids and other phenolic com- pounds: implications for the anti-tumor-promoter effect of vegetables. Journal of Agricultural Food Chemistry 47: 397–402

Schaller A, van Afferden M, Windhofer V, Bülow S, Abel G, Schmid J, Amrhein N(1991) Purification and characterization of chorismate synthase from *Euglena gracilis*: comparison with chorismate synthases of plant and microbial origin. Plant Physiology 97: 1271–1279

Schmid J, Amrhein N (1995) Molecular organization of the shikimate pathway in higher plants. Phytochemistry 39 (4): 737-749

Schmid J, Schaller A, Leibinger U, Boll W, Amrhein N (1992) The *in-vitro* synthesized tomato shikimate kinase precursor is enzymatically active and is imported and processed to the mature enzyme by chloroplasts. The Plant Journal 2: 375–383

Schmidt CL, Danneel HJ, Schultz G, Buchanan BB (1990) Shikimate kinase from spinach chloroplasts: purification, characterization, and regulatory function in aromatic amino acid biosynthesis. Plant Physiology 93: 758–766

Schmidt CL, Griindemann D, Groth G, Miiller B, Hennig H, Schultz G (1991) Shikimate pathway in non-photosynthetic tissues. Identification of common enzymes and partial purification of dehydroquinate hydrolyase, shikimate oxidoreductase and chorismate mutase from roots. Journal of Plant Physiology 138: 51-56

Schnee C, Kollner TG, Held M, Turlings TCJ, Gershenzon J, Degenhardt J (2006) A maize terpene synthase contributes to a volatile defense signal that attracts natural enemies of maize herbivores. Proceedings of the National Academy of Sciences of the United States of America 103: 1129-1134

Schwab W (2003) Metabolome diversity: too few genes, too many metabolites? Phytochemistry 62 (6): 837-849

Semon M, Wolfe KH (2008) preferential Subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. Proceedings of the National Academy of Science of the United States of America 105: 8333-8338

Shah DM, Horsch RB, Klee HJ, Kishore GM, Winter JA, Tumer NE, Hironaka CM, Sanders PR, Gasser CS, Aykent S, Siegel NR, Rogers SG, Fraley RT (1986)

Engineering herbicide tolerance in transgenetic plants. Science 233:478–481

Shi LX, Theg SM (2013) The chloroplast protein import system: from algae to trees. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research 1833 (2): 314-331

Shneier A, Harris J, Kleanthous C, Coggins JR, Hawkins AR, Abell C (1993) Evidence for opposite stereochemical courses for the reactions catalyzed by type I and type II dehydroquinases. Bioorganic & Medicinal Chemistry Letters 3: 1399–1402

Sikosek T, Chan HS, Bornberg-Bauer E (2012) Escape from Adaptive Conflict follows from weak functional trade-offs and mutational robustness. Proceedings of the National Academy of Science of the United States of America 109 (37): 14888-14893

Singh SA, Christendat D (2006) Structure of Arabidopsis dehydroquinate dehydrataseshikimate dehydrogenase and implications for metabolic channeling in the shikimate pathway. Biochemistry 45: 7787-7796

Singh SA, Christendat D (2007) The DHQ-dehydroshikimate-SDH-shikimate-NADP(H) Complex: Insights into Metabolite Transfer in the Shikimate Pathway. Crystal Growth Design 7 (11): 2153-2160

Smeekens S, Weisbeek P, Robinson C (1990) Protein transport into and within chloroplasts. Trends in Bichemical Science 15 (2): 73-76

Soll J (2004) Protein import into chloroplasts. Nature Reviews Molecular Cell Biology 5: 198-208

Soskine M, Tawfik DS (2010) Mutational effects and the evolution of new protein function. Nature Reviews Genetics 11: 572-582

Steinrucken HC, Amrhein N (1980) The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimic acid-3-phosphate synthase. Biochemical and Biophysical Research Communications 94 (4): 1207-1212

Sterck L, Rombauts S, Vandepoele K, Rouze P, Van de Peer Y (2007) How many genes are there in plants, and why are they there? Current Opinion in Plant Biology 10: 199–203

Stettler RF, Zsuffa L, Wu R (1996) The role of hybridization in the genetic manipulation of Populus. In RF Stettler, HD Bradshaw Jr, PE Heilman, TM Hinckley, eds, Biology of Populus and its Implications for Management and Conservation. NRC Research Press, National Research Council of Canada, Ottawa, ON, pp 87–112

Strother PK, AlHajri S, Traverse A (1996) Clarification of the genus nematothallus-lang.

Journal of Paleontology 67: 1090–1094

Stuart F, Hunter IS (1993) Purification and characterization of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase from *Streptomyces rimosus*. Biochimica et biophysica acta, Protein structure and molecular enzymology 1161: 209–115

Sung WS, Lee DG (2010) Antifungal action of chlorogenic acid against pathogenic fungi, mediated by membrane disruption. Pure and Applied Chemistry 82 (1): 219-226

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hermandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L, Radenbaugh A, Singh S, Swing V, Tissier C, Zhang P, Huala E (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. Nucleic Acids Research 36: D1009-D1014

Taylor TN, Taylor El, Krings M (2009) Paleobotany: the biology and evolution of fossil plants, 2nd ed. San Diego, CA: American Press

Theis N, Lerdau M (2003) The evolution of function in plant secondary metabolites. International Journal of Plant Science 164: S93-S102

Timperio Am, D'Alessandro A, Fagioni M, Magro P, Zolla L (2012) Production of the phytoalexins trans- resveratrol and delta-viniferin in two economy-relevant grape cultivars upon infection with *Botrytis cinerea* in field conditions. Plant Physiology and Biochemistry 50: 65-71

Tohge T, Watanbe M, Hoefgen R, Fernie A (2013) Shikimate and phenylalanine biosynthesis in the green lineage. Frontiers in Plant Science 4: 62

Tomescu AMF, Pratt LM, Rothwell GW, Strother PK, Nadon GC (2009) Carbon isotopes support the presence of extensive land floras pre-dating the origin of vascular plants. Palaeogeography Palaeoclimatology Palaeoecology 283: 46–59

Toufighi K, Brady SM, Austin RLE, Provart NJ (2005). The Botany Array Resource: e-Northerns, expression angling, and promoter analyses. The Plant Journal 43: 153-163

Tran LT, Constabel CP (2012) The polyphenol oxidase gene family in poplar: phylogeny, differential expression and identification of a novel, vacuolar isoform. Planta 234(4):799-813

Tsai C, Harding SA, Tschaplinski TJ, Lindroth RL, Yuan Y (2006) Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. New Phytologist 172 (1): 47-62

Tsuji J, Jackson EP, Gage DA, Hammerschmidt R, Somerville SC (1992) Phytoalexin Accumulation in *Arabidopsis thaliana* during the Hypersensitive Reaction to *Pseudomonas syringae pv syringae*. Plant Physiology 98: 1304-1309

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

Unsicker SB, Kunert G, Gershenzon J (2009) Protective perfumes: the role of vegetative volatiles in plant defense against herbivores. Current Opinion in Plant Biology 12 (4): 479-485

Vaast P, Bertrand B, Perriot J, Guyot V, Genard M (2006) Fruit thinning and shade improve bean characteristics and beverage quality of coffee (*Coffea arabica L.*) under optimal conditions. Journal of the Science of Food and Agriculture 86 (2): 197-204

Vandepoele K, Simillion C, Van de Peer Y (2003) Evidence that rice and other cereals are ancient aneuploids. The Plant Cell 15: 2192-2202

Vanholme R, Morreel K, Ralph J, Boerjan W (2008) Lignin engineering. Current Opinion in Plant Biology 11 (3): 278-285

Veljanoski V, Constabel CP (2013) Molecular cloning and biochemical characterization of two UDP-glycosyltransferases from poplar. Phytochemistry 91: 148-157

Vogt T (2010) Phenylpropanoid biosynthesis. Molecular Plant 3 (1): 2-20

Von Heijne G, Steppuhn J, Herrmann RG (1989) Domain structure of mitochondrial and chloroplast targeting peptides. European Journal of Biochemistry 180 (3): 535-545

Waegemann K, Soll J (1996) Phosphorylation of the transit sequence of chloroplast precursor proteins. Journal of Biological Chemistry 271: 6545-6554

Wallace RJ (2004) Antibacterial properties of plant secondary metabolites. Proceedings of the Nutrition Society 63: 621-629

Wang J, Pichersky E (1999) Identification of specific residues involved in substrate discrimination in two plant O-methyltransferases. Archives of Biochemistry and Biophysics 368: 172-180

Weaver LM, Herrmann KM (1997) Dynamics of the shikimate pathway in plants. Trends in Plant Science 2 (9): 346-351

Webby CJ, Jiao W, Hutton RD, Blackmore NJ, Baker HM, Baker RN, Jameson GB, Parker EJ (2010) Synergistic allostery, a sophisticated regulatory network for the control of aromatic amino acid biosynthesis in *Mycobacterium tuberculosis*. The Journal of biological chemistry 285: 30567–30576

Weinstein LH, Porter CA, Laurencot HJ (1961) Role of quinic acid in aromatic biosynthesis in higher plants. Contributions from Boyce Thompson Institute 21: 201–214;

Weisshaar B, Jenkins GI (1998) Phenylpropanoid biosynthesis and its regulation. Current opinion in Plant Biology 1: 251-257

Wellman Ch, Osterloff PL, Mohiuddin Y (2003) Fragments of the earliest land plants. Nature 425: 282–285

Weng JK, Chapple C (2010) The origin and evolution of lignin biosynthesis. New Phytologist 187(2): 273-285

Wimp GM, Martinsen GD, Floate KD, Bangert RK, Whitham TG (2005) Plant genetic determinants of arthropod community structure and diversity. Evolution 59: 61-69

Wink M (1988) Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores. Theoretical and Applied Genetics 75 (2): 225-233

Wink M (2003) Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry 64 (1): 3-19

Wink M (2009) Functions and biotechnology of plant secondary metabolites. Annual Plant Reviews 39: 1-20

Woll S, Kim SH, Creten HJ, Efferth T (2013) Animal plant warfare and secondary metabolite evolution. Natural Products and Bioprospecting 3 (1): 1-7

Wroblewski T, Tomczak A, Michelmore R (2005) Optimization of Agrobacterium - mediated transient assays of gene expression in lettuce, tomato and Arabidopsis. Plant Biotechnology Journal 3: 259-273

Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. The Society for Molecular Biology and Evolution 24 (8): 1586-1591

Yang Z, Kumar S, Nei M (1995) A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141: 1641-1650

Yoshida S (1969) Biosynthesis and conversion of aromatic amino acids in plants. Annual Review of Plant Physiology 20: 41-62

Yoshida S, Tazaki K, Minamikawa T (1975) Occurrence of shikimic and quinic acids in angiosperms. Phytochemistry 14: 195-197

Yu CS, Chen YC, Lu CH, Hwang JK (2006) Prediction of protein subcellular localization. Proteins: Structure. Function and Bioinformatics 64: 643-651

Zhang X, Glaser E (2002) Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone. Trends in Plant Science 7 (1): 14-21

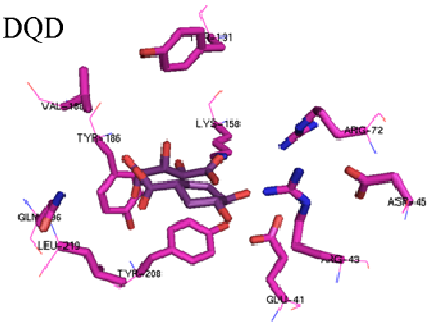
Zhang XL, Zhang SB, Hao F, Lai XH, Yu HD, Huang YS, Wang HH (2005) Expression, purification and properties of shikimate dehydrogenase from *Mycobacterium tuberculosis*. Journal of Biochemistry and Molecular Biology 38: 624-631

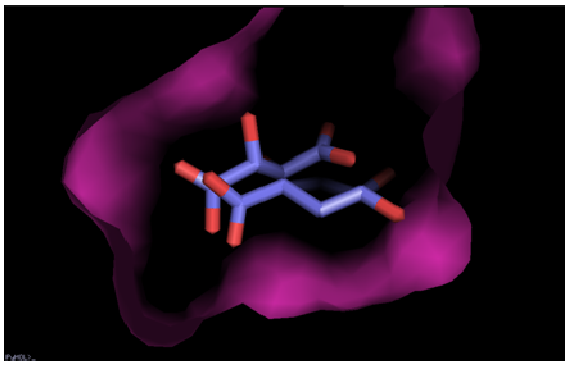
Zhao J, Weaver LM, Herrmann KM (2002) Translocation of 3-deoxy-D-*arabino*-heptulosonate 7- phosphate synthase precursor into isolated chloroplasts. Planta 216: 180–186

Appendices

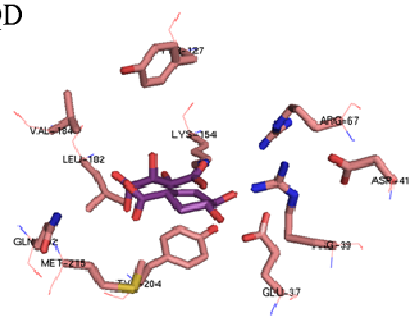
Appendix A: Predicted structure models of Poptr3, Poptr4 and Poptr5

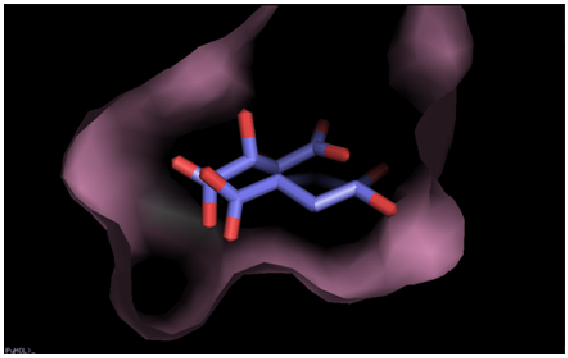
Poptr3 DQD



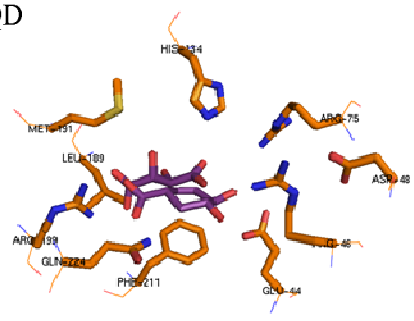


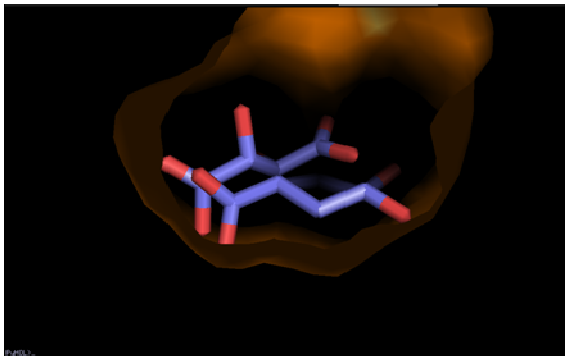
Poptr4 DQD



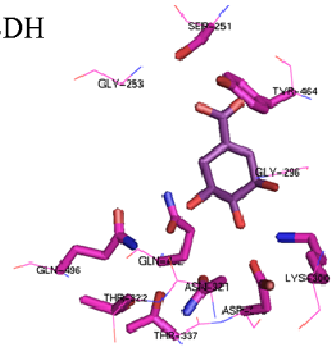


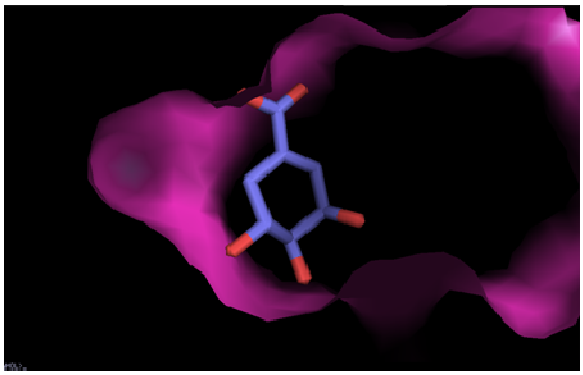
Poptr5 DQD



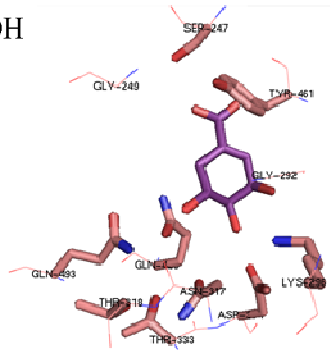


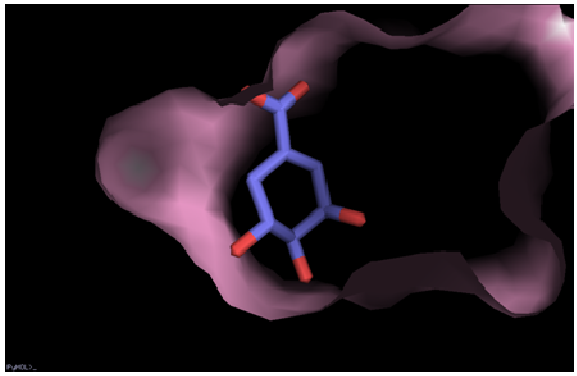
Poptr3 SDH



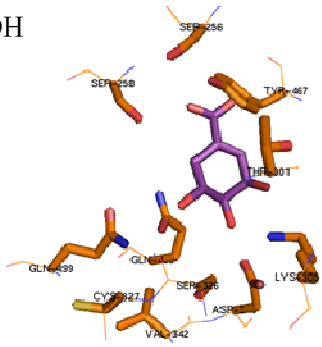


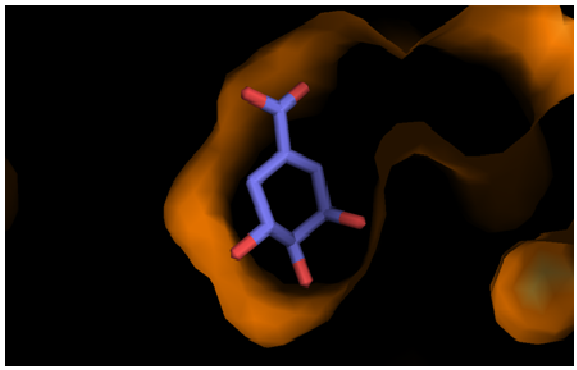
Poptr4 SDH



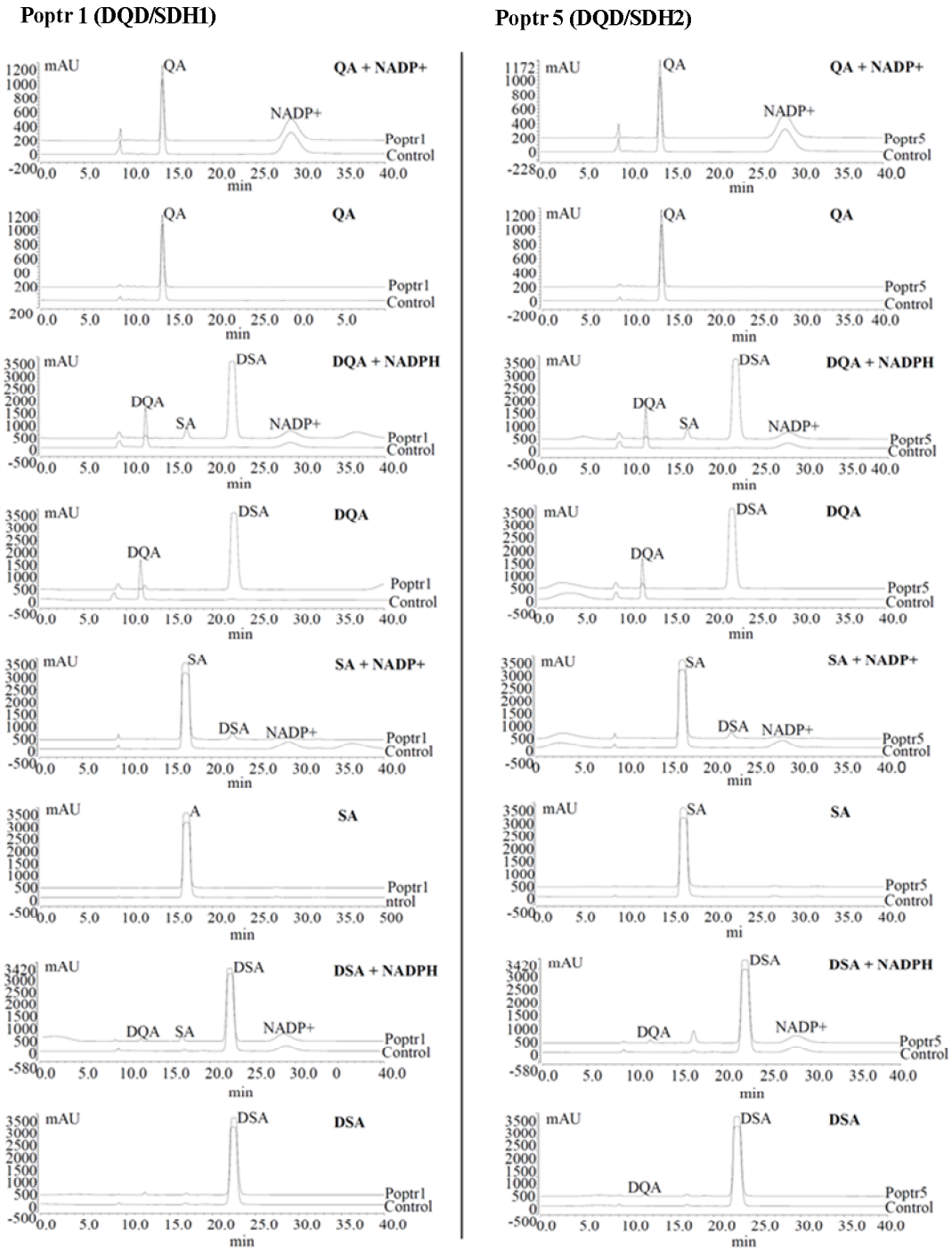


Poptr5 SDH

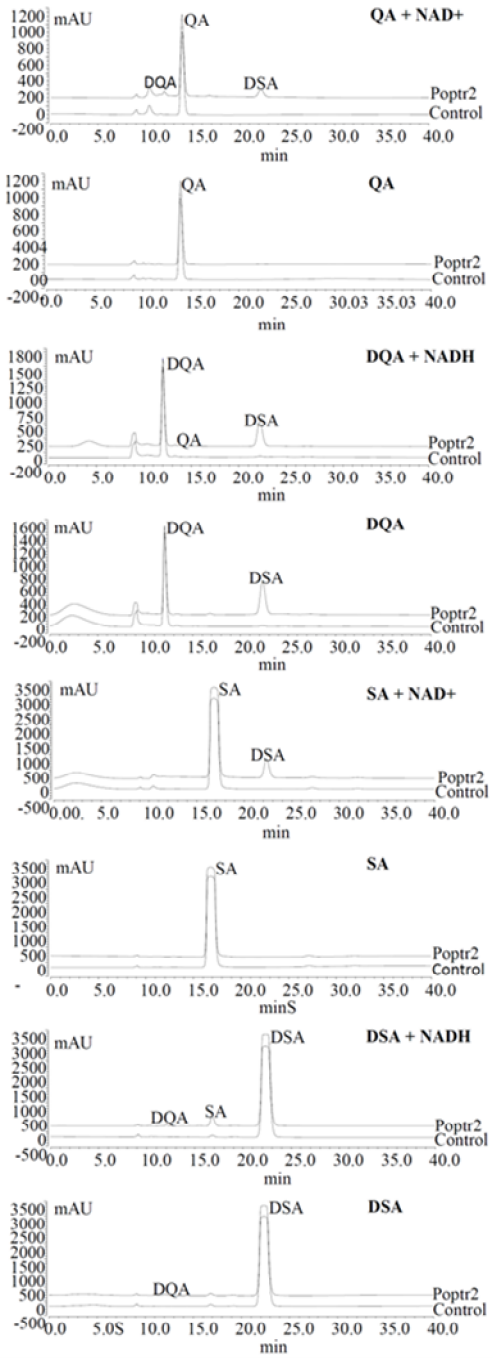




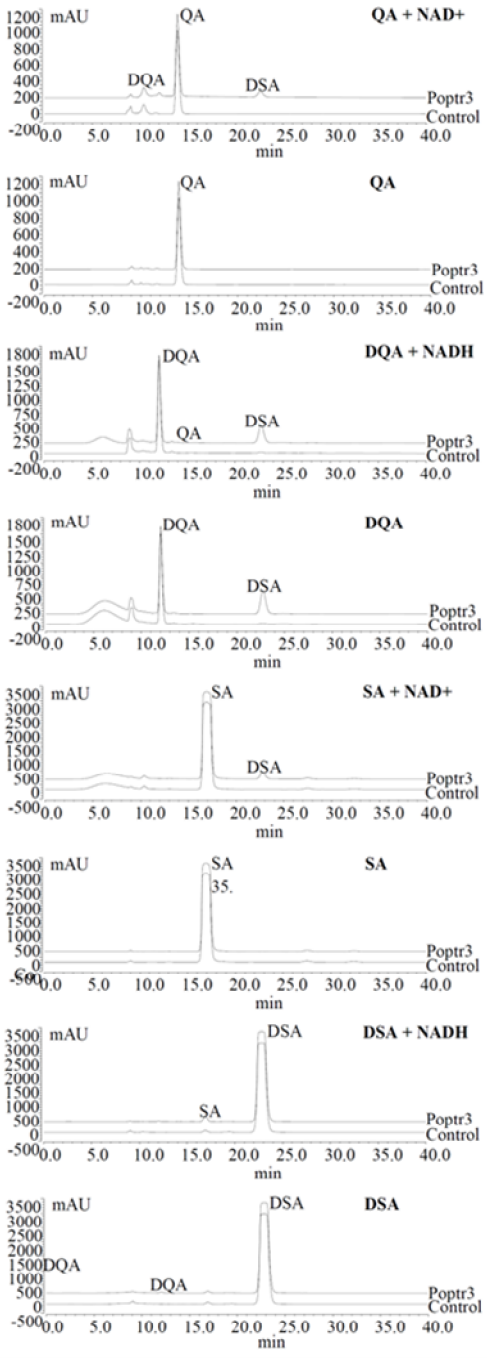
Appendix B: HPLC elution profiles of enzyme assays catalyzed by recombinant DQD/SDHs (Poptr1 and Poptr5) and QDHs (Poptr2 and Poptr3)



Poptr2 (QDH1)

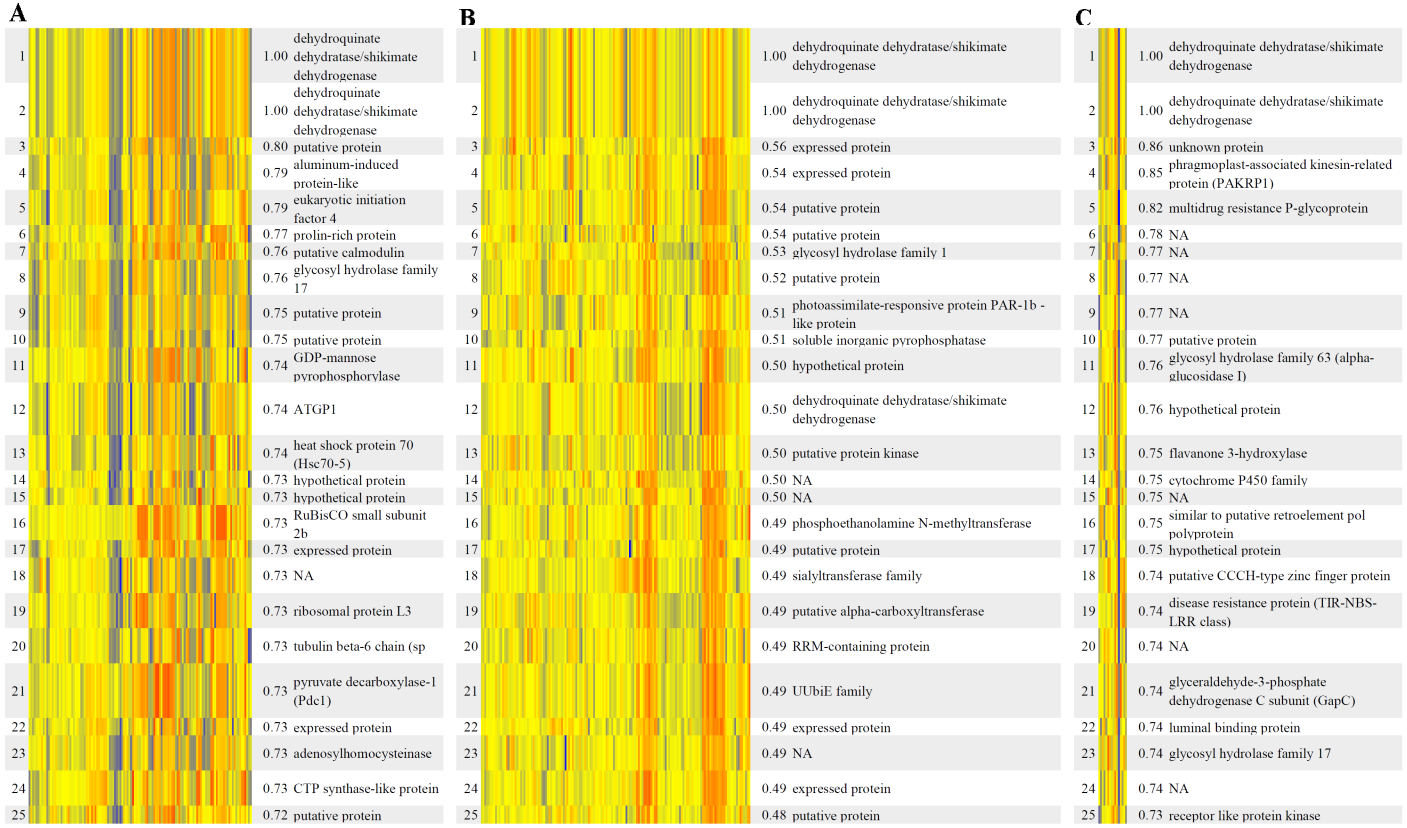


Poptr3 (QDH2)



Appendix C: Lists of genes that are coexpressed with DQD/SDH2 (Poptr5) and QDHs (Poptr2, Poptr3 and Poptr4) based on co-expression analysis with large-scale microarray expression data (A: organ-specific; B: treatment; C: transgenic)

QDH1 (Poptr2)



A

1	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
2	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
3	0.80	putative protein
4	0.79	aluminum-induced protein-like
5	0.79	eukaryotic initiation factor 4
6	0.77	prolin-rich protein
7	0.76	putative calmodulin
8	0.76	glycosyl hydrolase family 17
9	0.75	putative protein
10	0.75	putative protein
11	0.74	GDP-mannose pyrophosphorylase
12	0.74	ATGP1
13	0.74	heat shock protein 70 (Hsc70-5)
14	0.73	hypothetical protein
15	0.73	hypothetical protein
16	0.73	RuBisCO small subunit 2b
17	0.73	expressed protein
18	0.73	NA
19	0.73	ribosomal protein L3
20	0.73	tubulin beta-6 chain (sp
21	0.73	pyruvate decarboxylase-1 (Pdc1)
22	0.73	expressed protein
23	0.73	adenosylhomocysteinase
24	0.73	CTP synthase-like protein
25	0.72	putative protein

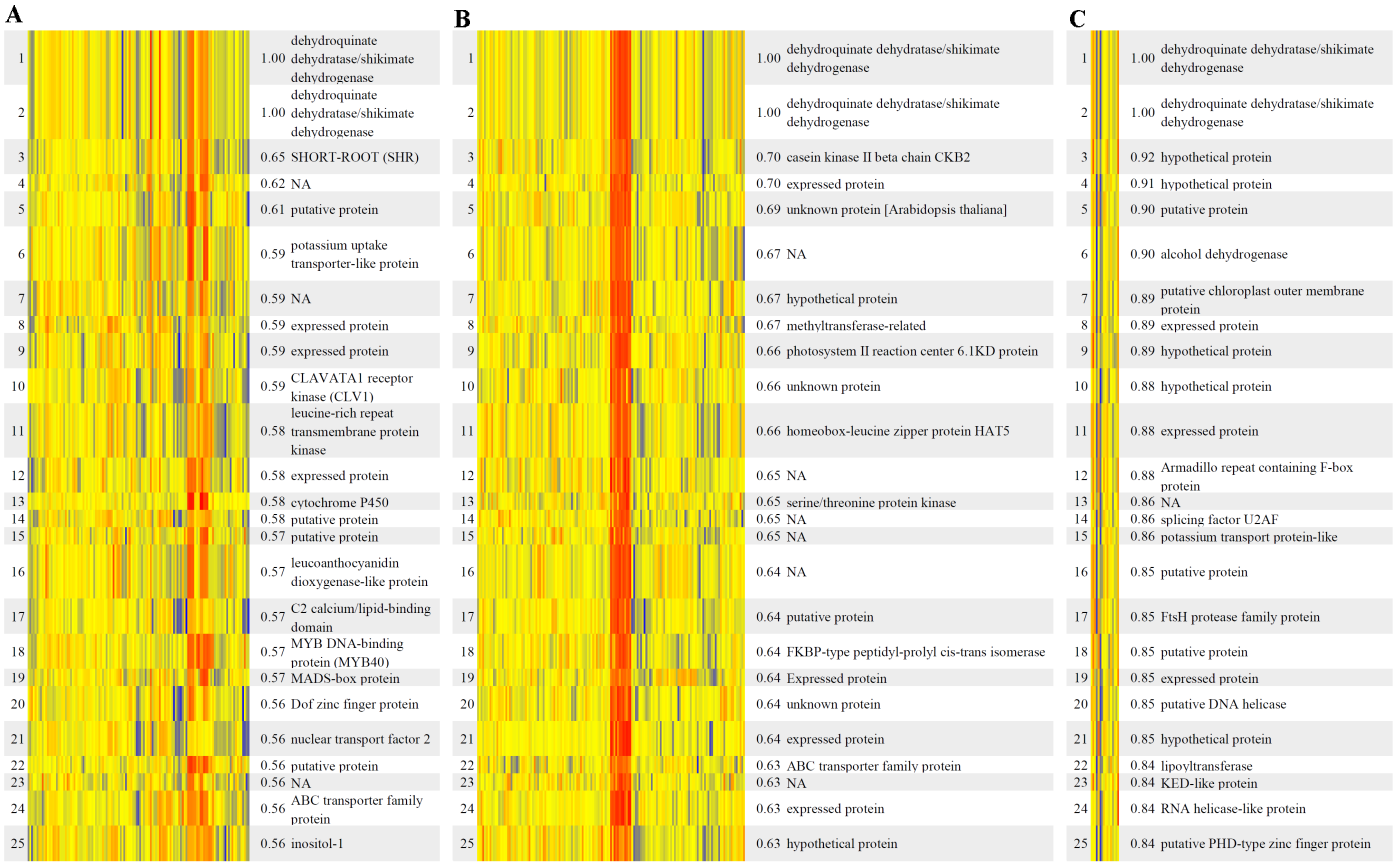
B

1	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
2	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
3	0.56	expressed protein
4	0.54	expressed protein
5	0.54	putative protein
6	0.54	putative protein
7	0.53	glycosyl hydrolase family 1
8	0.52	putative protein
9	0.51	photoassimilate-responsive protein PAR-1b - like protein
10	0.51	soluble inorganic pyrophosphatase
11	0.50	hypothetical protein
12	0.50	dehydroquinate dehydratase/shikimate dehydrogenase
13	0.50	putative protein kinase
14	0.50	NA
15	0.50	NA
16	0.49	phosphoethanolamine N-methyltransferase
17	0.49	putative protein
18	0.49	sialyltransferase family
19	0.49	putative alpha-carboxyltransferase
20	0.49	RRM-containing protein
21	0.49	UUbiE family
22	0.49	expressed protein
23	0.49	NA
24	0.49	expressed protein
25	0.48	putative protein

C

1	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
2	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
3	0.86	unknown protein
4	0.85	phragmoplast-associated kinesin-related protein (PAKRP1)
5	0.82	multidrug resistance P-glycoprotein
6	0.78	NA
7	0.77	NA
8	0.77	NA
9	0.77	NA
10	0.77	putative protein
11	0.76	glycosyl hydrolase family 63 (alpha-glucosidase I)
12	0.76	hypothetical protein
13	0.75	flavanone 3-hydroxylase
14	0.75	cytochrome P450 family
15	0.75	NA
16	0.75	similar to putative retroelement pol polyprotein
17	0.75	hypothetical protein
18	0.74	putative CCCH-type zinc finger protein
19	0.74	disease resistance protein (TIR-NBS-LRR class)
20	0.74	NA
21	0.74	glyceraldehyde-3-phosphate dehydrogenase C subunit (GapC)
22	0.74	luminal binding protein
23	0.74	glycosyl hydrolase family 17
24	0.74	NA
25	0.73	receptor like protein kinase

QDH2 (Poptr3)



A

1	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
2	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
3	0.65	SHORT-ROOT (SHR)
4	0.62	NA
5	0.61	putative protein
6	0.59	potassium uptake transporter-like protein
7	0.59	NA
8	0.59	expressed protein
9	0.59	expressed protein
10	0.59	CLAVATA1 receptor kinase (CLV1)
11	0.58	leucine-rich repeat transmembrane protein kinase
12	0.58	expressed protein
13	0.58	cytochrome P450
14	0.58	putative protein
15	0.57	putative protein
16	0.57	leucoanthocyanidin dioxygenase-like protein
17	0.57	C2 calcium/lipid-binding domain
18	0.57	MYB DNA-binding protein (MYB40)
19	0.57	MADS-box protein
20	0.56	Dof zinc finger protein
21	0.56	nuclear transport factor 2
22	0.56	putative protein
23	0.56	NA
24	0.56	ABC transporter family protein
25	0.56	inositol-1

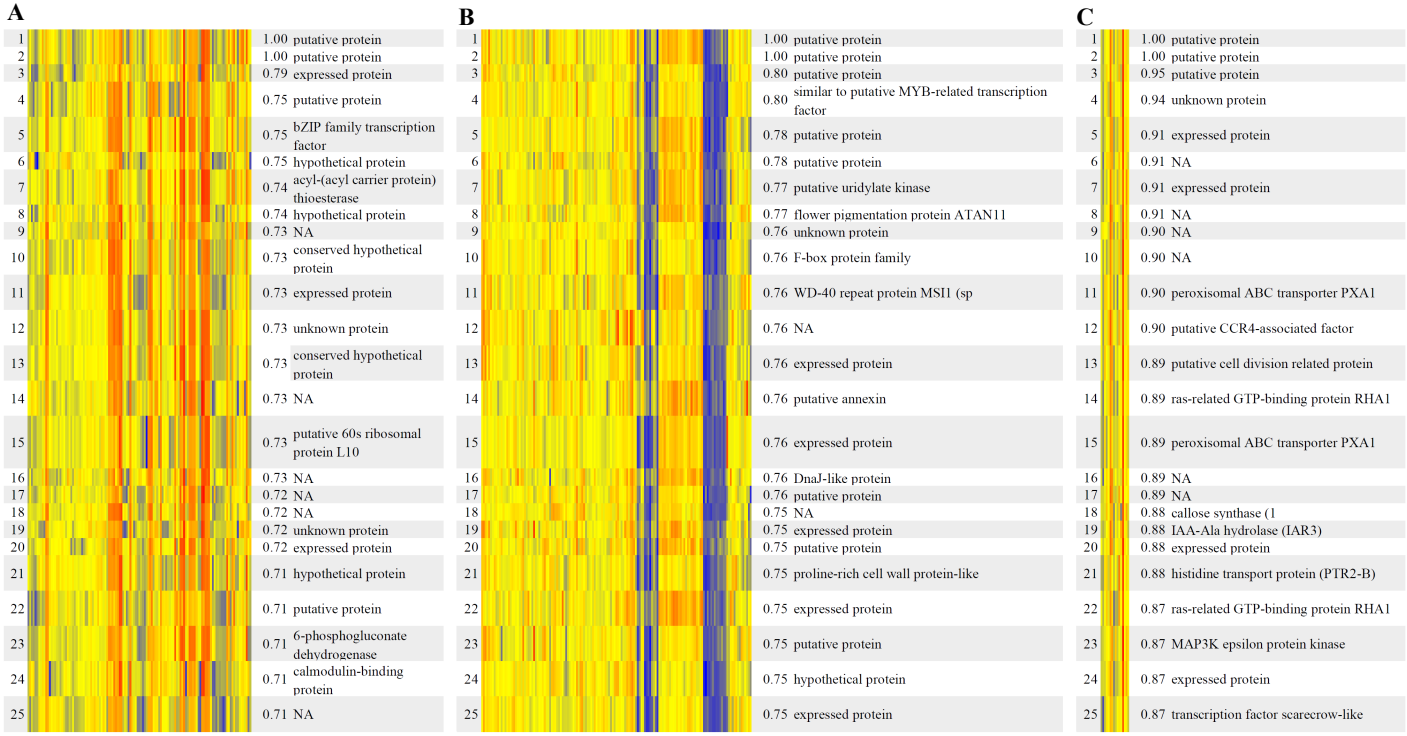
B

1	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
2	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
3	0.70	casein kinase II beta chain CKB2
4	0.70	expressed protein
5	0.69	unknown protein [Arabidopsis thaliana]
6	0.67	NA
7	0.67	hypothetical protein
8	0.67	methyltransferase-related
9	0.66	photosystem II reaction center 6.1KD protein
10	0.66	unknown protein
11	0.66	homeobox-leucine zipper protein HAT5
12	0.65	NA
13	0.65	serine/threonine protein kinase
14	0.65	NA
15	0.65	NA
16	0.64	NA
17	0.64	putative protein
18	0.64	FKBP-type peptidyl-prolyl cis-trans isomerase
19	0.64	Expressed protein
20	0.64	unknown protein
21	0.64	expressed protein
22	0.63	ABC transporter family protein
23	0.63	NA
24	0.63	expressed protein
25	0.63	hypothetical protein

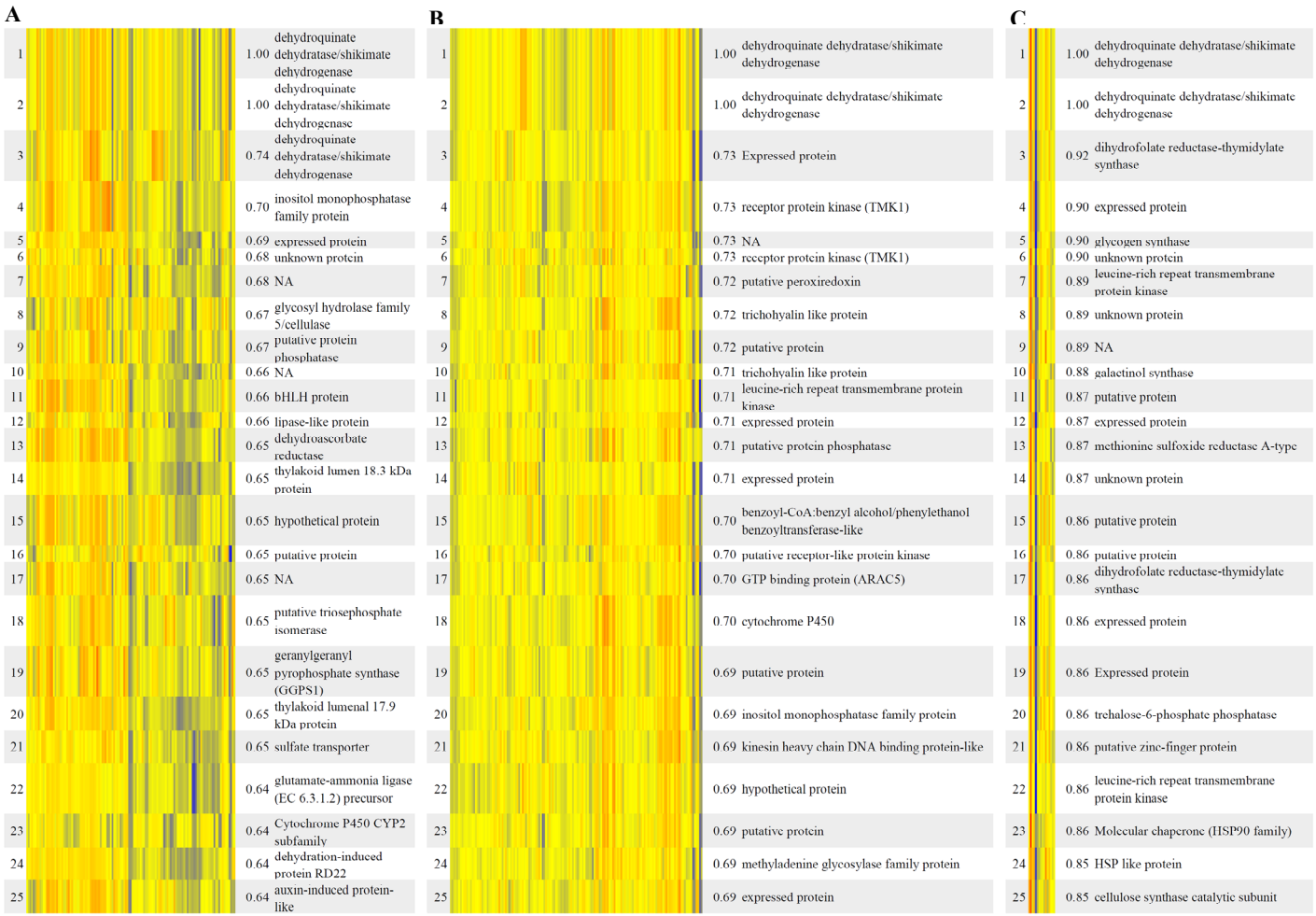
C

1	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
2	1.00	dehydroquinate dehydratase/shikimate dehydrogenase
3	0.92	hypothetical protein
4	0.91	hypothetical protein
5	0.90	putative protein
6	0.90	alcohol dehydrogenase
7	0.89	putative chloroplast outer membrane protein
8	0.89	expressed protein
9	0.89	hypothetical protein
10	0.88	hypothetical protein
11	0.88	expressed protein
12	0.88	Armadillo repeat containing F-box protein
13	0.86	NA
14	0.86	splicing factor U2AF
15	0.86	potassium transport protein-like
16	0.85	putative protein
17	0.85	FtsH protease family protein
18	0.85	putative protein
19	0.85	expressed protein
20	0.85	putative DNA helicase
21	0.85	hypothetical protein
22	0.84	lipoyltransferase
23	0.84	KED-like protein
24	0.84	RNA helicase-like protein
25	0.84	putative PHD-type zinc finger protein

QDH3 (Poptr4)


A
1 1.00 putative protein
2 1.00 putative protein
3 0.79 expressed protein
4 0.75 putative protein
5 0.75 bZIP family transcription factor
6 0.75 hypothetical protein
7 0.74 acyl-(acyl carrier protein) thioesterase
8 0.74 hypothetical protein
9 0.73 NA
10 0.73 conserved hypothetical protein
11 0.73 expressed protein
12 0.73 unknown protein
13 0.73 conserved hypothetical protein
14 0.73 NA
15 0.73 putative 60s ribosomal protein L10
16 0.73 NA
17 0.72 NA
18 0.72 NA
19 0.72 unknown protein
20 0.72 expressed protein
21 0.71 hypothetical protein
22 0.71 putative protein
23 0.71 6-phosphogluconate dehydrogenase
24 0.71 calmodulin-binding protein
25 0.71 NA
B
1 1.00 putative protein
2 1.00 putative protein
3 0.80 putative protein
4 0.80 similar to putative MYB-related transcription factor
5 0.78 putative protein
6 0.78 putative protein
7 0.77 putative uridylate kinase
8 0.77 flower pigmentation protein ATAN11
9 0.76 unknown protein
10 0.76 F-box protein family
11 0.76 WD-40 repeat protein MSI1 (sp
12 0.76 NA
13 0.76 expressed protein
14 0.76 putative annexin
15 0.76 expressed protein
16 0.76 DnaJ-like protein
17 0.76 putative protein
18 0.75 NA
19 0.75 expressed protein
20 0.75 putative protein
21 0.75 proline-rich cell wall protein-like
22 0.75 expressed protein
23 0.75 putative protein
24 0.75 hypothetical protein
25 0.75 expressed protein
C
1 1.00 putative protein
2 1.00 putative protein
3 0.95 putative protein
4 0.94 unknown protein
5 0.91 expressed protein
6 0.91 NA
7 0.91 expressed protein
8 0.91 NA
9 0.90 NA
10 0.90 NA
11 0.90 peroxisomal ABC transporter PXA1
12 0.90 putative CCR4-associated factor
13 0.89 putative cell division related protein
14 0.89 ras-related GTP-binding protein RHA1
15 0.89 peroxisomal ABC transporter PXA1
16 0.89 NA
17 0.89 NA
18 0.88 callose synthase (1
19 0.88 IAA-Ala hydrolase (IAR3)
20 0.88 expressed protein
21 0.88 histidine transport protein (PTR2-B)
22 0.87 ras-related GTP-binding protein RHA1
23 0.87 MAP3K epsilon protein kinase
24 0.87 expressed protein
25 0.87 transcription factor scarecrow-like

DQD/SDH2 (Poptr5)



Appendix D: List of DQD/SDH and QDH protein sequences (from bacteria, red algae and major groups of the green lineages) that are included in phylogenetic analysis

Names	Database	Accession	Species
Bacteria			
Blama	NCBI	ZP_01093851.1	*Blastopirellula marina DSM 3645*
Gemob	NCBI	ZP_02732082.1	*Gemmata obscuriglobus UQM 2246*
Isopa	NCBI	YP_004177782.1	*Isosphaera pallida ATCC 43644*
Pirst	NCBI	YP_003372532.1	*Pirellula staleyi DSM 6068*
Plabr	NCBI	YP_004271124.1	*Planctomyces brasiliensis DSM 5305*
PlaKSU	NCBI	ZP_10101094.1	*planctomycete KSU-1*
Plali	NCBI	YP_003629868.1	*Planctomyces limnophilus DSM 3776*
Plama	NCBI	ZP_01854559.1	*Planctomyces maris DSM 8797*
Rhoba	NCBI	gb|EGF26198.1	*Rhodopirellula baltica WH47*
Sinac	NCBI	ZP_09572955.1	*Singulisphaera acidiphila DSM 18658*
Red Algae			
Cyame	*Cyanidioschyzon merolae* Genome Project	CMK269C	*Cyanidioschyzon merolae*
Green Algae			
Chlre	NCBI	XP_001694346.1	*Chlamydomonas reinhardtii*
Chlva	NCBI	gb|EFN57759.1	*Chlorella variabilis*
Cocsu	NCBI	gb|EIE23903.1	*Coccomyxa subellipsoidea C-169*
Micpu	NCBI	XP_003055405.1	*Micromonas pusilla CCMP1545*
Micsp	NCBI	XP_002507566.1	*Micromonas sp. RCC299*
Ostlu	NCBI	XP_001420686.1	*Ostreococcus lucimarinus CCE9901*
Ostta	NCBI	XP_003082412.1	*Ostreococcus tauri*

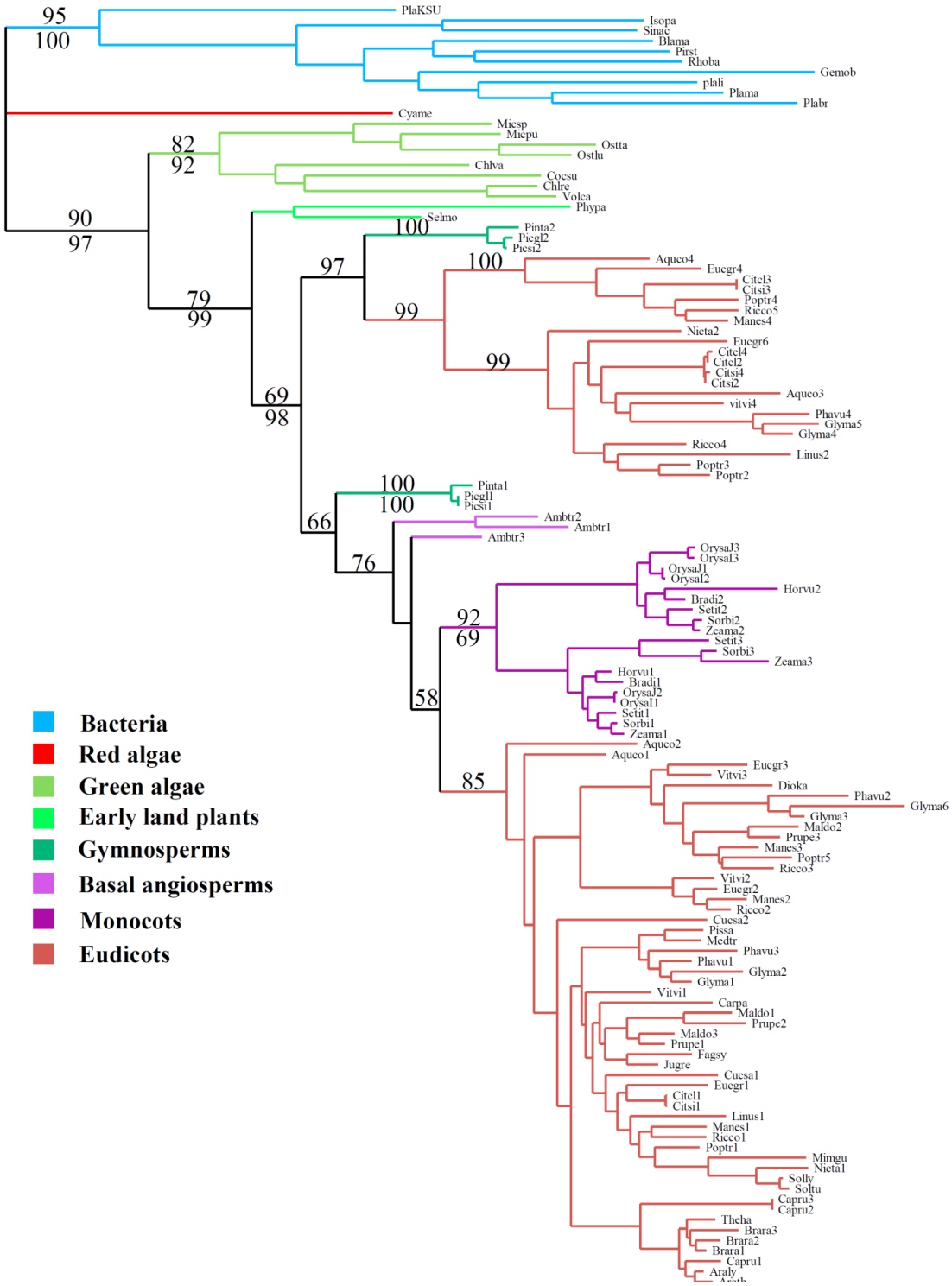
Volca	NCBI	XP_002952960.1	*Volvox carteri f. nagariensis*
Bryophyte			
Phypa	NCBI	XP_001761240.1	*Physcomitrella patens subsp. patens*
Lycophyte			
Selmo	NCBI	XP_002990348.1	*Selaginella moellendorffii*
Gymnosperms			
Picgl1	Plant GDB	PUT-163a-Picea_glauca-21137	*Picea glauca*
Picgl2	NCBI	ABK24375.1	*Picea glauca*
Picsi1	NCBI	gb|ABR16592.1	*Picea sitchensis*
Picsi2	NCBI	gb|ABK24375.1	*Picea sitchensis*
Pinta1	TIGR Plant Transcript Assembly	TA6261_3352	*Pinus taeda*
Pinta2	TIGR Plant Transcript Assembly	TA15690_3352	*Pinus taeda*
Basal Angiosperms			
Ambtr1	Amborella Genome Database	lcl|evm_27.model.AmTr_v1.0_scaffold00057.123	*Amborella trichopoda*
Ambtr2	Amborella Genome Database	lcl|evm_27.model.AmTr_v1.0_scaffold00057.124	*Amborella trichopoda*
Ambtr3	Amborella Genome Database	lcl|evm_27.model.AmTr_v1.0_scaffold00057.122	*Amborella trichopoda*
Monocots			
Bradi1	NCBI	XP_003575813.1	*Brachypodium distachyon*
Bradi2	NCBI	XP_003567707.1	*Brachypodium distachyon*
Horvu1	NCBI	dbj|BAJ99066.1	*Hordeum vulgare subsp. vulgare*
Horvu2	NCBI	dbj|BAJ97275.1	*Hordeum vulgare subsp. vulgare*
OrysaI1	NCBI	gb|EEC69419.1	*Oryza sativa Indica Group*
OrysaI2	NCBI	gb|EEC70675.1	*Oryza sativa Indica Group*
OrysaI3	NCBI	gb|EAY74111.1	*Oryza sativa Indica Group*
OrysaJ1	NCBI	dbj|BAD61388.1	*Oryza sativa Japonica Group*
OrysaJ2	NCBI	gb|EEE53338.1	*Oryza sativa Japonica Group*

OrysaJ3	NCBI	dbj|BAD61389.1	*Oryza sativa Japonica Group*
Setit1	Phytozome	Si021701m	*Setaria italica*
Setit2	Phytozome	Si000995m	*Setaria italica*
Setit3	Phytozome	Si021763m	*Setaria italica*
Sorbi1	NCBI	XP_002443289.1	*Sorghum bicolor*
Sorbi2	NCBI	XP_002455659.1	*Sorghum bicolor*
Sorbi3	NCBI	XP_002443290.1	*Sorghum bicolor*
Zeama1	NCBI	gb|ACG40194.1	*Zea mays*
Zeama2	NCBI	NP_001152357.1	*Zea mays*
Zeama3	NCBI	NP_001140289.1	*Zea mays*
Diocots			
Aquco1	Phytozome	Aquca_017_00293.1	*Aquilegia coerulea*
Aquco2	Phytozome	Aquca_003_00904.1	*Aquilegia coerulea*
Aquco3	Phytozome	Aquca_017_00294.1	*Aquilegia coerulea*
Aquco4	Phytozome	Aquca_003_00901.1	*Aquilegia coerulea*
Araly	NCBI	XP_002884573.1	*Arabidopsis lyrata subsp. lyrata*
Arath	NCBI	NP_187286.1	*Arabidopsis thaliana*
Brara1	Phytozome	Bra040225	*Brassica rapa*
Brara2	Phytozome	Bra001195	*Brassica rapa*
Brara3	Phytozome	Bra020756	*Brassica rapa*
Capru1	Phytozome	Carubv10013287m	*Capsella rubella*
Capru2	Phytozome	Carubv10013176m	*Capsella rubella*
Capru3	Phytozome	Carubv10013429m	*Capsella rubella*
Carpa	Phytozome	evm.model.supercontig_16.98	*Carica papaya*
Citcl1	Phytozome	clementine0.9_007196m	*Citrus clementina*
Citcl2	Phytozome	clementine0.9_007541m	*Citrus clementina*
Citcl3	Phytozome	clementine0.9_007574m	*Citrus clementina*
Citcl4	Phytozome	clementine0.9_009894m	*Citrus clementina*

Citsi1	Phytozome	orange1.1g009615m	*Citrus sinensis*
Citsi2	Phytozome	orange1.1g010050m	*Citrus sinensis*
Citsi3	Phytozome	orange1.1g010101m	*Citrus sinensis*
Citsi4	Phytozome	orange1.1g012866m	*Citrus sinensis*
Cucsa1	Phytozome	Cucsa.250300.3	*Cucumis sativus*
Cucsa2	Phytozome	Cucsa.101380.1	*Cucumis sativus*
Dioka	NCBI	dbj|BAI40147.1	*Diospyros kaki*
Eucgr1	Phytozome	Eucgr.J00263.1	*Eucalyptus grandis*
Eucgr2	Phytozome	Eucgr.H04427.1	*Eucalyptus grandis*
Eucgr3	Phytozome	Eucgr.H04428.1	*Eucalyptus grandis*
Eucgr4	Phytozome	Eucgr.H01214.1	*Eucalyptus grandis*
Eucgr5	Phytozome	Eucgr.B01770.2	*Eucalyptus grandis*
Fagsy	NCBI	gb|ABA54867.1	*Fagus sylvatica*
Glyma1	NCBI	XP_003516831.1	*Glycine max*
Glyma2	NCBI	XP_003556937.1	*Glycine max*
Glyma3	NCBI	XP_003555589.1	*Glycine max*
Glyma4	NCBI	XP_003520825.1	*Glycine max*
Glyma5	NCBI	XP_003554674.1	*Glycine max*
Glyma6	NCBI	XP_003535355.1	*Glycine max*
Jugre	NCBI	gb|AAW65140.1	*Juglans regia*
Linus1	Phytozome	Lus10033981	*Linum usitatissimum*
Linus2	Phytozome	Lus10001725	*Linum usitatissimum*
Maldo1	Phytozome	MDP0000170005	*Malus domestica*
Maldo2	Phytozome	MDP0000176374	*Malus domestica*
Maldo3	Phytozome	MDP0000306786	*Malus domestica*
Manes1	Phytozome	cassava4.1_030256m	*Manihot esculenta*
Manes2	Phytozome	cassava4.1_004545m	*Manihot esculenta*
Manes3	Phytozome	cassava4.1_005332m	*Manihot esculenta*
Manes4	Phytozome	cassava4.1_008122m	*Manihot esculenta*

Medtr	NCBI	XP_003608198.1	*Medicago truncatula*
Mimgu	Phytozome	mgv1a004563m	*Mimulus guttatus*
Nicta1	NCBI	gb|AAS90325.1	*Nicotiana tabacum*
Nicta2	NCBI	gb|AAS90324.2	*Nicotiana tabacum*
Phavu1	Phytozome	Phvulv091013901m	*Phaseolus vulgaris*
Phavu2	Phytozome	Phvulv091006763m	*Phaseolus vulgaris*
Phavu3	Phytozome	Phvulv091001374m	*Phaseolus vulgaris*
Phavu4	Phytozome	Phvulv091013701m	*Phaseolus vulgaris*
Pissa	NCBI	gi|7488788	*Pisum sativum*
Poptr1	Phytozome	Potri.010G019000.2	*Populus trichocarpa*
Poptr2	Phytozome	Potri.005G043400.1	*Populus trichocarpa*
Poptr3	Phytozome	Potri.014G135500.3	*Populus trichocarpa*
Poptr4	Phytozome	Potri.013G029800.1	*Populus trichocarpa*
Poptr5	Phytozome	Potri.013G029900.2	*Populus trichocarpa*
Prupe1	Phytozome	ppa004026m	*Prunus persica*
Prupe2	Phytozome	ppa004515m	*Prunus persica*
Prupe3	Phytozome	ppa004574m	*Prunus persica*
Ricco1	Ricinus communis Blast	29726.m003976	*Ricinus communis*
Ricco2	Ricinus communis Blast	30128.m008684	*Ricinus communis*
Ricco3	Ricinus communis Blast	30128.m008683	*Ricinus communis*
Ricco4	Ricinus communis Blast	29647.m002035	*Ricinus communis*
Ricco5	Ricinus communis Blast	32108.m008685 + 32108.m008686	*Ricinus communis*
Solly/Lyces	NCBI	NP_001234051.1	*Solanum lycopersicum*
Soltu	NCBI	gb|ADA57640.1	*Solanum tuberosum*
Theha	Phytozome	Thhalv10020322m	*Thellungiella halophila*
Vitvi1	NCBI	XP_002277395.2	*Vitis vinifera*
Vitvi2	NCBI	XP_002270188.1	*Vitis vinifera*
Vitvi3	NCBI	XP_002270232.1	*Vitis vinifera*
Vitvi4	NCBI	XP_002270055.1	*Vitis vinifera*

Appendix E: Extended phylogeny of DQD/SDHs and QDHs from bacteria, red algae and the plant kingdom



Appendix F: Protein sequence of the immediate pre-duplication ancestral DQD/SDH (Anc122)

```
                  10          20          30          40          50
              ....|....| ....|....| ....|....| ....|....| ....|....|
Anc122  1   MQNGTLICTP LVAETVEEML SQMQKAKASG ADCVELRLDY LSNFQPRVDL  50

                  60          70          80          90         100
              ....|....| ....|....| ....|....| ....|....| ....|....|
Anc122  51  ERLLKARPLP AIVTYRPKWE GGQYEGDEET RLDALRLAME LGADYIDVEL  100

                 110         120         130         140         150
              ....|....| ....|....| ....|....| ....|....| ....|....|
Anc122  101 QVASDFIASQ SNKKSSNSKI IVSNHNYQNT PSTEELGHLV ARMQATGADI  150

                 160         170         180         190         200
              ....|....| ....|....| ....|....| ....|....| ....|....|
Anc122  151 VKIVTTATDI TDVARIFHLL SHCQVPIIAL VMGERGLISR LLCPKFGGYL  200

                 210         220         230         240         250
              ....|....| ....|....| ....|....| ....|....| ....|....|
Anc122  201 TFGTLESGKE SAPGQPTLTD LRNVYKIRQI NRDTKVFGII GNPVGHSKGP  250

                 260         270         280         290         300
              ....|....| ....|....| ....|....| ....|....| ....|....|
Anc122  251 ILHNPAFREV GFNAVYVPLL VDDLKEFLNV YSSPDFAGFS VTIPHKEAAL  300

                 310         320         330         340         350
              ....|....| ....|....| ....|....| ....|....| ....|....|
Anc122  301 RCCDEVDPVA KSIGAVNTIV RRPSDGKLVG YNTDCEGAIS AIEDGLRGSH  350

                 360         370         380         390         400
              ....|....| ....|....| ....|....| ....|....| ....|....|
Anc122  351 SSGSSSSSPL AGKLFVVIGA GGAGKALAFG AKQRGARVVI ANRNYERAKA  400

                 410         420         430         440         450
              ....|....| ....|....| ....|....| ....|....| ....|....|
Anc122  401 LANSVGGEAI SLEELDSFRP ETGMILANTT SVGMHPNVDE TPISKEALKS  450

                 460         470         480         490         500
              ....|....| ....|....| ....|....| ....|....| ....|....|
Anc122  451 YAVVFDAVYT PKETRLLREA KEAGAVVVSG LEMFIRQAIG QFELFTGLPA  500

                 510
              ....|....| ...
Anc122  501 PEELMREIVL KKT 513
```

Appendix G: Sites that are subject to episodic diversifying selection

Codon Position**	α	β^-	Pr[β*=β^-]	β^+	Pr[β=β^+]	p-value	q-value
99	0.633	0.000	0.969	37.774	0.031	0.007	0.422
109	0.564	0.150	0.815	4.763	0.185	0.050	1.000
117	0.536	0.146	0.848	606.179	0.152	0.006	0.439
118	0.361	0.361	0.792	4.832	0.208	0.064	1.000
131	0.596	0.228	0.902	10.949	0.098	0.085	1.000
136	0.693	0.227	0.787	62.987	0.213	0.017	0.790
141	0.634	0.000	0.596	3.982	0.404	0.001	0.197
150	0.313	0.095	0.953	144.705	0.047	0.067	1.000
161	0.644	0.000	0.961	3.967	0.039	0.096	1.000
190	0.000	0.000	0.142	0.285	0.858	0.048	1.000
193	0.471	0.049	0.618	2.037	0.382	0.097	1.000
205	0.272	0.225	0.915	63.299	0.085	0.051	1.000
231	0.570	0.000	0.700	3.866	0.300	0.021	0.879
234	0.563	0.000	0.876	7.194	0.124	0.009	0.498
262	0.806	0.160	0.913	17.805	0.087	0.063	1.000
278	0.854	0.027	0.943	54.112	0.057	0.030	1.000
338	0.632	0.032	0.966	348.310	0.034	0.002	0.261
340	0.403	0.013	0.650	2.709	0.350	0.081	1.000
380	0.312	0.031	0.962	78.805	0.038	0.002	0.224
381	0.473	0.009	0.975	10.452	0.025	0.085	1.000
400	0.313	0.024	0.770	4.081	0.230	0.034	1.000
401	0.349	0.194	0.869	6.367	0.131	0.026	1.000
407	1.094	0.000	0.953	1394.590	0.047	0.039	1.000
425	0.258	0.196	0.930	231.662	0.070	0.043	1.000
429	0.552	0.000	0.740	3.743	0.260	0.045	1.000
443	0.326	0.000	0.739	4.260	0.261	0.001	0.343
447	0.208	0.000	0.226	2.277	0.774	0.005	0.434
497	0.752	0.000	0.925	4.911	0.075	0.080	1.000
513	0.229	0.085	0.774	4.601	0.226	0.015	0.755
515	0.470	0.000	0.923	1359.110	0.077	0.071	1.000
517	0.494	0.000	0.983	20.340	0.017	0.087	1.000
553	0.620	0.066	0.958	13.625	0.042	0.084	1.000
564	0.524	0.000	0.618	2.400	0.382	0.029	1.000
566	0.991	0.011	0.804	4.148	0.196	0.085	1.000
594	0.324	0.081	0.736	1.921	0.264	0.087	1.000
600	1.288	0.182	0.930	685.200	0.070	0.036	1.000

* Two categories of β: $\beta^- < \alpha$ and β^+ (unrestricted)

** Relative positions in Arabidopsis DQD/SDH