Biochemistry and evolution of the shikimate dehydrogenase/quinate dehydrogenase gene family in plants

by

Yuriko Carrington
B.Sc. (Honours), University of Victoria, 2012

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

DOCTOR OF PHILOSOPHY

in the Department of Biology

Supervisory Committee

Biochemistry and evolution of the shikimate dehydrogenase/quinate dehydrogenase gene family in plants

by

Yuriko Carrington
B.Sc. (Honours), University of Victoria, 2012

Supervisory Committee

Dr. Jürgen Ehlting, (Department of Biology)
Supervisor

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

Dr. Louise Page, (Department of Biology)
Departmental Member

Dr. Alisdair Boraston, (Department of Biochemistry and Microbiology)
Outside Member

Abstract

Gene duplication and functional diversification is a central driving force in the evolution of plant biochemical diversity. However, the latter process is not well understood. Here the diversification of the plant shikimate/quinate dehydrogenase (S/QDH) gene family was investigated in order to shed light on how duplicate genes functionally diversify. The shikimate pathway is the major biosynthetic route towards the aromatic amino acids, linking vital protein biosynthesis with the production of aromatic secondary metabolites. Dehydroquinate dehydratase/shikimate dehydrogenase (SDH) encodes the central enzyme of this pathway, catalyzing the production of shikimate. Quinate is a secondary metabolite synthesized using the same precursors as shikimate by quinate dehydrogenase (QDH). Gene duplication prior to the gymnosperm / angiosperm split generated two distinct clades in seed plants separating SDH and QDH functions whereas non-seed plants have a single copy SDH. *In vitro* biochemical characterization of a reconstructed ancestral enzyme was performed alongside extant members separated prior to duplication (from a lycopod, a bryophyte, and a chlorophyte) and afterwards (from a gymnosperm and an angiosperm). This revealed that novel quinate biosynthetic activity was gained in seed plants, providing evidence for the diversification of gene function via neofunctionalization. However, the ability to use both NAD(H) and NADP(H) seems to have developed in both SDH and QDH clade members of angiosperms. Finally, a method is described for analysing quinate and its derivative, chlorogenic acid in transgenic *Arabidopsis*.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I would like to extend my utmost gratitude to my primary supervisor, Dr. Jürgen Ehlting, for granting me the opportunity to pursue my scientific curiosity and work in his laboratory. Through countless trial and error, I feel that I have grown stronger as a person. Along with my supervisor, I would like to extend my deepest appreciation to the members of my committee, Dr. C. Peter Constabel, Dr. Louise Page and Dr. Alisdair Boraston for always supporting me and helping me to reach my research goals. Outside of research, I am particularly grateful for the fact they pushed me to step outside my comfort zone and join a speech club, Toastmasters International, where I learned invaluable leadership and communication skills, and made many friends. As an added bonus, I have gained a few trophies from competing in speech competitions on the island. Prior to joining Toastmasters, I never would have imagined even competing!

I would like to deeply thank Dr. Jia Guo, for sharing her extraordinary project with me; Drs. Lan Tran and Cuong Hieu Le for their mentorship in the laboratory; and Dr. Ori Granot for his support and advice on LC-MS analyses. I extend my gratitude to all past and present members of the Ehlting and Constabel labs as well as to Forest Biology as a whole: your advice and friendship will always be appreciated.

Finally, from the bottom of my heart, I thank my dear family, Richard, Tomoko, Ayumi, Mina, Haku and Nana Carrington for making sure I stayed healthy (and alive) as I have a tendency to get lost in my work. I could not have taken on this challenge without your love and support for which I am always grateful for.

Chapter 1

1.1 Plant adaptations on a grand scale

Diversity is the spice of life in the plant kingdom whose estimated 300,000 species members including floating bladderworts, low-lying moss, giant sequoias, and snap-trap flowers differ uniquely in their morphological and physiological abilities. Plants have successfully invaded all corners of the earth, from the Mediterranean Seabed, concrete suburbs and north to the arctic tundra. Their rich biodiversity attests to a competition between plants and their natural enemies. For plants, life presents challenges including damaging UV-B radiation, pests, pathogens, and herbivorous animals. Unlike humans and other animals, plants cannot simply uproot themselves and move to more ideal locations when their lives are endangered. To survive, plants have instead evolved a complex armoury of defense strategies. These may be shared or taxon-specific; physical or chemical; indirect or direct and above all else, they have helped plants stay successful (Kroymann, 2011; Weng, 2014; Wink, 2003).

Physical barriers of defence provide protection to plants against herbivorous insects and vertebrates. Prominent examples of structural adaptations are the thorns, prickles and spines of cacti (*Cactacea*e) and other prickly plants (e.g. thistles; e.g. of *Asteraceae*) (Charles-Dominique *et al*., 2017; Ronel and Lev-Yadun, 2012; War *et al*., 2012). These protuberances discourage feeding by inflicting wounds or restricting the bite size and rate of herbivorous grazers. Even in the absence of such appendages, the physical architecture of plants is an effective means of defense against predators (Charles-Dominique *et al*., 2017; Ronel and Lev-Yadun, 2012). In New Zealand, for example, the youths of some tree species have interlaced cage-like stems and scanty leaves. These are thought to have been adapted to discourage feeding by a now-extinct species of giant flightless birds called moas (Bond *et al*., 2004). Plants may also fortify their defenses microscopically. Through the deposition of lignin, a complex and durable polymer found prominently in wood, they can thicken their secondary cell walls, forming an-almost impregnable barrier (Hanley *et al*., 2007; War *et al*., 2012). Thus, an effective strategy of plants to stay alive is to reduce their appeal and digestibility to hungry foragers (Bond *et al*., 2004; Hanley *et al*., 2007;

War *et al*., 2012). Not all enemies are predators, however. Plants, like humans, are susceptible to diseases. As one pre-emptive measure, lignin can be synthesized post-wounding to seal off potential sites of infection by fungi or bacteria (Dixon and Paiva, 1995; Liu *et al.*, 2018).

Returning to the example of New Zealand's moa-resistant trees, the adults have broad leaves and simpler branching patterns compared to the juveniles. They rely on another means of counterattack to defend against enemies. The mature trees store a greater concentration of defense-related phenylpropanoids (e.g. condensed tannins and other phenolics) in their leaves and stems (Bond *et al*., 2004). Studies on domestic herbivores suggest tannins bind to proteins and digestive enzymes in the animal's gut to restrict digestion of plant tissues (Barbehenn and Constabel, 2011). Such a mode of action may have worked against the extinct moas although this theory is difficult to prove without the birds in question. In North America, willow species *Salix eriocephala* and *S. sericea* are believed to deploy condensed tannins and phenolic glycosides as antiherbivore defense akin to New Zealand's trees. However, seedlings less than six weeks old lack such defense compounds and are commonly eaten by the slug *Arion subfuscus*. Unlike New Zealand's flightless birds, *A. subfuscus* is an exotic herbivore that has been introduced from Europe. It is possible this relatively recent interaction has caught the willow seedlings off-guard in a "surprise slug attack" such that they have not had enough time to evolve a response (Fritz *et al*., 2001).

Although a single plant, such as the common garden flower, may be short-lived, collectively they stand the test of time. Angiosperms alone are a lineage that is over a hundred million years old (Amborella Genome Project, 2013). Historically, each individual has endured and continues to endure a life-time of struggle, much like the scenarios depicted above using a combination of morphological, physiological and biochemical strategies (Kroymann, 2011; Wink, 1999, 2003). Resultantly, some like the flowering trees of New Zealand, have managed to out-compete their competitors (Bond *et al*., 2004). Overall, plants dominate almost all corners on land, from equatorial rainforests to the arctic tundra. How can plants be so successful at life? Insights into this puzzle can be gained by studying the evolution of plant secondary metabolites, because these small chemicals play pivotal roles in many local adaptations. This provides an advantage over studying other adaptive traits as the underlying genetics of plant secondary metabolites are reasonably candid; often a single gene creates a single phenotype. In contrast, other (physiological

and morphological) characteristics frequently result from multiple genes interacting together such that their singular functions become masked (Weng, 2014).

1.2 Plant chemical adaptations—an overview

As plant secondary metabolites are the heart and bones of this work, it is useful to start the next section with a briefing of their functions. Conventionally a line has been drawn between secondary metabolites and their primary metabolic counterparts. The long-held view is that primary metabolites (e.g. sugars, amino acids, nucleic acids and lipids) carry out functions necessary for growth and development. In contrast, secondary metabolites are described as facilitating ecological interactions, having been shaped during evolution by competing plants, animals, and pathogens (Bourgaud *et al*., 2001; Wink, 2003). A major example is the class of nitrogen-containing alkaloids that serve as feeding deterrents; many of which are toxic to invertebrates and vertebrates (Hartmann, 1996; Matsuura and Fett-Neto, 2015; Wink, 2003). Traditionally, primary metabolites are believed to carry out essential functions whereas secondary metabolites involved in signaling and defense are dispensable (Hartmann, 1996; Wink, 2003). This distinction holds true for the most part but is not without loopholes. For example, lignin functions in defense, but also provides rigidity and mediates water transport to promote growth of lofty trees (Croteau *et al*., 2000) and may thus be considered to be either a primary or secondary metabolite. Both normal and extreme environmental conditions (e.g. UV-B radiation, climate, salinity, poor soil nutrient content, etc.) can also pose as challenges to plants; and as such, many secondary metabolites represent adaptations towards abiotic stresses (Croteau *et al*., 2000; Hartmann, 2007; Wink, 2003). Most phenylpropanoids (e.g. lignin, tannins and chlorogenic acid), for example, maximally absorb light in the UV-B range, enabling photosynthesis to occur without damaging leaf mesophyll. Remarkably, plants harness the sun's energy to produce their own food: it is almost no surprise that they have also developed their own sun-blocks (Clé *et al*., 2008; Solecka, 1997; Weng and Chapple, 2010).

With some exceptions, phenylpropanoids such as lignin and its hydroxycinnamic acid precursors are common among plants: however, others are rarer. Because secondary metabolites have been largely moulded by external forces, their distribution can vary uniquely among species (Bourgaud

et al., 2001; Croteau *et al*., 2000; Wink, 2003). It is hardly surprising that nitrogen containing alkaloids are characteristic of nitrogen-fixing legumes (Fabaceae): however, within the Fabaceae, the tribe Genistae of subfamily Papilionoideae is particularly abundant in antiherbivorous (Frick *et al*., 2017), quinolizidine alkaloids (Wink, 2003). Another example and one that is more pertinent to this work, is the distribution of the phenylpropanoid chlorogenic acid (see also chapters three and five). This hydroxycinnamic ester is widespread among many land plant lineages, accumulating to high levels in solanaceous species (e.g. tomato, tobacco, eggplant) as well as in coffee, pears, plums, and apples. However, its distribution is not universal as it is absent in some lineages such as the thale-crest, *Arabidopsis thaliana* (Niggeweg *et al*., 2004). The metabolic "fingerprints" of plants provide exciting prospects for elucidating the historical trajectories of secondary biosynthetic pathways (Theis and Lerdau, 2003). Of course, care must be taken when evaluating possible reasons for their existence, since their original and contemporary functions may differ. In some cases, the selective forces acting on secondary metabolites have dissolved into history like the moas, making them even more difficult to discern (Bond *et al*., 2004).

1.2.1 Briefing—a short ode to the chemical defense strategies of non-plants

While many new and exciting examples of plant interactions with their environments have been made over the last decades (Bourgaud *et al*., 2001; War *et al*., 2012), before continuing on this topic, it is worth mentioning that plants are not alone in their reliance on chemical defense responses. In temperate and tropical coral reef communities, astonishing arrays of bioactive compounds are produced by a far-distant relative of plants (Hay and Fenical, 1988). Red, brown and green macroalgae are loosely lumped together as "seaweeds" but differ from "true plants" (Embryophyta) as they lack several of their defining characteristics including double membrane chloroplasts (Solymosi, 2013) and the presence of xyloglucan in the cell wall (Popper and Tuohy, 2010). Brown algae are the furthest removed, sharing more biochemical and molecular similarities to diatoms (Michel *et al*., 2010; Raven and Giordano, 2014): however, they produce high concentrations of a special class of tannins, called phlorotannins, that function similarly to tannins of terrestrial plants. Together with bromophenols, these have been shown to deter grazing by the sea snails *Turbo cornutus* (Shibata *et al*., 2014) and *Littorina sitkana* (Geiselman and McConnell, 1981) in laboratory assays.

In a surprising finding, Australasian kelp has a greater concentration of phlorotannins than those living in the North Pacific (Steinberg *et al*., 1995). Trends in the biogeographical distribution of algal defense compounds match those of terrestrial plants. Polyphenolics as mentioned above, are widely distributed whereas alkaloids are scarcer (Bourgaud *et al*., 2001). These findings provide information about the extraordinary evolutionary histories of different organisms. A difference in phlorotannin concentrations between Australasian and North Pacific kelp systems is explained by corresponding differences in their trophic systems. New Zealand's kelp forests have traditionally consisted of a two-tier trophic system with autotrophs (e.g. kelp) as producers and marine herbivores (e.g. sea urchins) as consumers. In the North Pacific, a third member enters this food chain: sea otters. Due to the predatory effects of sea otters on bottom grazers, the selection pressures imposed by the latter on macroalgae are lessened (Steinberg *et al*., 1995).

The use of bioactive compounds to deter feeding herbivores also occurs in the animal kingdom. In laboratory feeding experiments, the organic extracts of marine sponges display antifeedant effects against wrasse fish and the sea urchin *Diadema setosum* (Burns *et al*., 2003). Thus secondary metabolites not only co-occur across phylogenetically distinct groups, but apparently function analogously under the sea and above it. Nonetheless, there is a greater body of understanding about chemical interactions of terrestrial as opposed to marine communities. The obvious reason for this could be because the latter are obscured underwater. But also, secondary metabolites of marine systems tend to be unstable once isolated (Hay and Fenical, 1988).

1.3 Structural links between primary and secondary metabolism

Overlaps in the structures, chemistry, and biosynthetic pathways of secondary metabolites enable them to be categorized in a variety of ways, such as according to their biosynthetic pathways (Verpoorte *et al*., 2000). Their biosynthetic routes generate what are essentially the 'bare bones' of the complex medley of compounds we observe. Specific compounds are assembled by biosynthetic enzymes modifying these backbones in a Lego®-like manner via acylation, prenylation, hydroxylation and isomerization reactions to name a few (Verpoorte *et al*., 2000). The phenylpropanoids (e.g. lignin, tannins, and chlorogenic acid) have in common an aromatic

ring structure with an attached three carbon chain (cinnamic acid) derived from phenylalanine. The deamination of phenylalanine to cinnamate is catalyzed by phenylalanine ammonia-lyase (PAL). Simple modifications to the C6-C3 skeleton including the addition of hydroxyl and methyl groups give rise to the hydroxycinnamic acids such as caffeic acid (a precursor of chlorogenic acid), ferulic acid, sinapic acid and p-coumaric acid. In turn, these can serve as building blocks to make more complex phenylpropanoid polymers such as lignin, or condensed tannins and their monomers, which are derived by modifying p-coumaric acid (Dixon and Paiva, 1995; Vogt, 2009).

Plant metabolism is labyrinthine in nature, consisting of numerous biosynthetic pathways converging and diverging like a road map. Upon reading this map, clues are revealed about the underlying origins of secondary metabolism: namely, that secondary metabolic pathways tend to branch from primary metabolism like the side roads off a highway (Weng, 2014) (Figure 1.1). Phenylpropanoids, as mentioned previously, are all derived from phenylalanine (Dixon and Paiva, 1995; Vogt, 2009). Many alkaloids are also of amino acid origin (tryptophan, lysine and ornithine) except for those using nucleosides as precursors (e.g. caffeine and other purine alkaloids) (Verpoorte *et al*., 2000; Weng, 2014). Phenylalanine, tyrosine and tryptophan are products of the shikimate pathway, a primary metabolic pathway that has been thus nick-named the bridge connecting primary and secondary metabolism (Herrmann and Weaver, 1999).

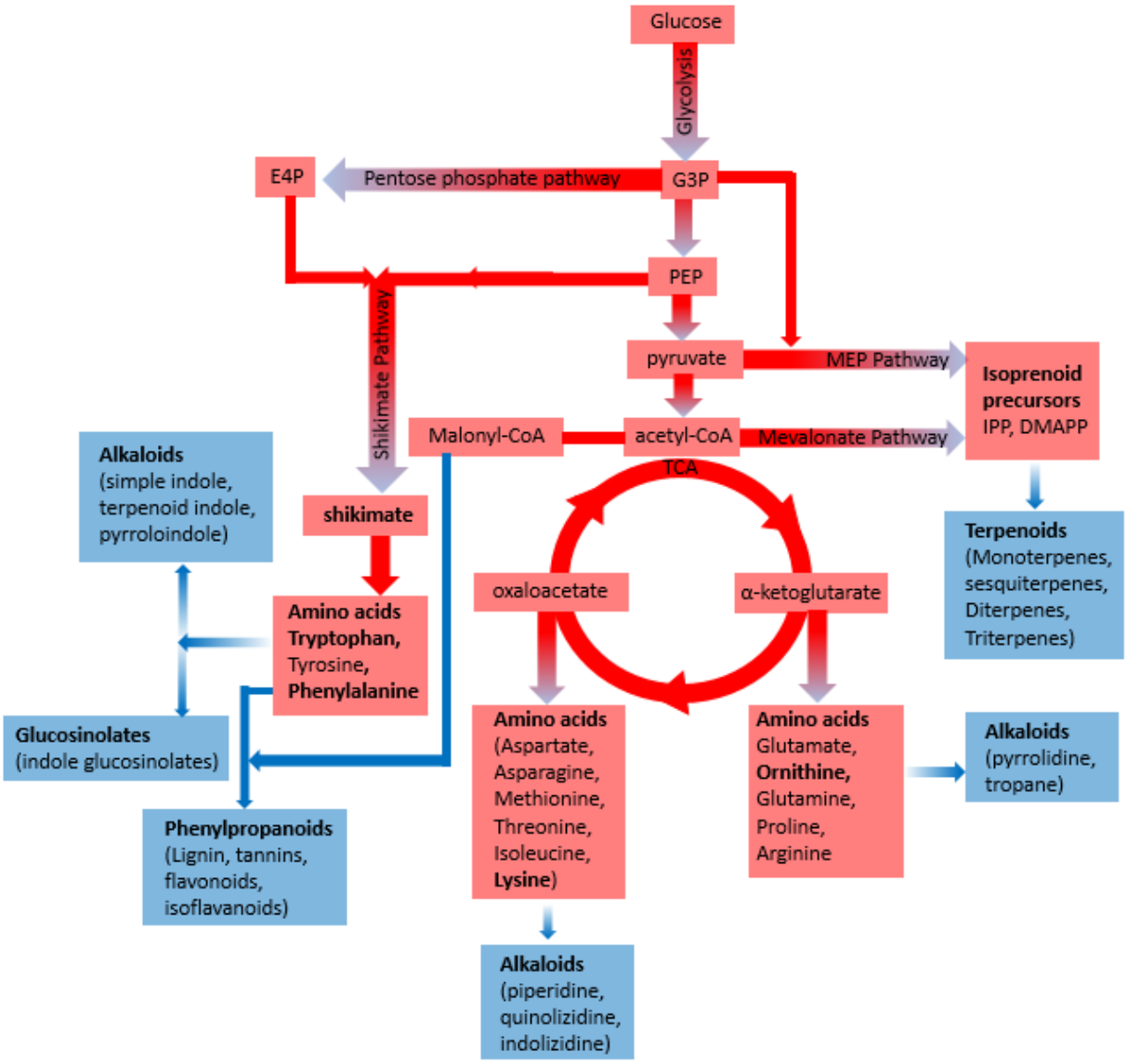


Figure 1.1: Interconnectedness of plant primary and secondary metabolic pathways. Many plant secondary metabolic pathways (blue) branch from and derive their carbon skeletons from core primary metabolism (red). Note not all known pathways are shown.

1.3.1 The shikimate pathway

In a series of seven enzymatic steps, the shikimate pathway converts phosphoenolpyruvate and erythrose 4-phosphate to chorismate, the common precursor of the aromatic amino acids: phenylalanine, tyrosine, and tryptophan. The shikimate pathway is present in plants, fungi, and bacteria albeit with characteristic differences among the participating enzymes: however, it is absent in animals. Lack of this pathway in animals makes it ideal for drug and pesticide development, and in effect it has become popularized in research and business sectors as the target of Monsanto's hotly debated herbicide, glyphosate. The shikimate pathway initially proceeds with the conversion of phosphoenolpyruvate and erythrose 4-phosphate from glycolysis and the pentose phosphate pathway respectively, to 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP). This condensation reaction is catalyzed by DAHP synthase. Next, 3-dehydroquinate synthase eliminates a phosphate group, converting DAHP to the cyclic intermediate 3-dehydroquinate. Note that at this point, a structure similar to the ultimate end products of shikimate and phenylpropanoid metabolism, namely a six-carbon ring, is casted. The third and fourth steps of the shikimate pathway involve the dehydration of dehydroquinate to dehydroshikimate and the reduction of dehydroshikimate to shikimate (Figure 1.2 and 1.3). In plants, these are catalyzed by a single bifunctional enzyme, dehydroquinate dehydratase/shikimate dehydrogenase (DQD/SDH). This enzyme adopts a central position in the shikimate pathway as the generator of its core intermediate and the source of its name. Not only is it important for the shikimate pathway, but DQD/SDH is also the crux of this thesis. However, as focus is placed on the enzyme's oxidoreductase activity present in its carboxy terminal domain, it will be henceforth referred to as simply shikimate dehydrogenase (SDH). In the fifth step of the shikimate pathway, shikimate is phosphorylated by shikimate kinase to produce shikimate-3-phosphate. Subsequently, condensation of shikimate-3-phosphate with a second molecule of phosphoenolpyruvate generates 5-enolpyruvylshikimate-3-phosphate (EPSP) (Herrmann, 1995; Herrmann and Weaver, 1999). This penultimate reaction is catalyzed by EPSP synthase (EPSPS), which doubly serves as the target of glyphosate (Funke *et al*., 2006). Finally, EPSP acts as the substrate for the seventh and last enzyme of this pathway, chorismate synthase, which eliminates the phosphate group yielding chorismate. Altogether the last five steps introduce two C=C bonds and a side chain to the heterocyclic 3-dehydroquinate. The final product, chorismate serves as the common precursor of the aromatic amino acids (Herrmann, 1995; Herrmann and Weaver, 1999).

Depending on the pathway, chorismate may be converted to prephenate or anthranilate en route to phenylalanine and tyrosine, or tryptophan biosynthesis respectively via steps that will not be discussed here for brevity reasons (Tzin and Galili, 2010). All of the studied enzymes in plants have been observed to possess a plastid transit peptide, pointing to the localization of this pathway in chloroplasts (Herrmann, 1995; Herrmann and Weaver, 1999). About 20% of photosynthetically fixed carbon in a plant flows through the shikimate pathway before being converted into a rich diversity of compounds. Both the main trunk and side-branches of the pathway culminates in the production of numerous diverse aromatic secondary metabolites alongside primary metabolic proteins. It therefore adopts a fundamental role in the growth, development, health, and resilience of plants against biotic and abiotic stresses. The significance of this pathway is clearly demonstrated by the weed-killer glyphosate, which blocks EPSPS (Herrmann, 1995; Herrmann and Weaver, 1999). In addition, a homozygous loss of function mutation of the *SDH* gene, as demonstrated by a line of publicly available T-DNA *Arabidopsis* insertion mutants result in an embryo lethal phenotype (TAIR; http://arabidopsis.org).

As earlier discussed, *trans*-cinnamic acid derived from phenylalanine is the precursor of diverse phenylpropanoids. The aromatic plant hormone, salicylic acid which mediates defense responses towards pathogens and abiotic stresses is one of its derivatives (Dempsey *et al*., 2011; Vogt, 2009). Hydrolysable tannins are complex phenylpropanoid defense compounds typically consisting of glucose esterified to multiple gallic acid groups (Barbehenn and Constabel, 2011).

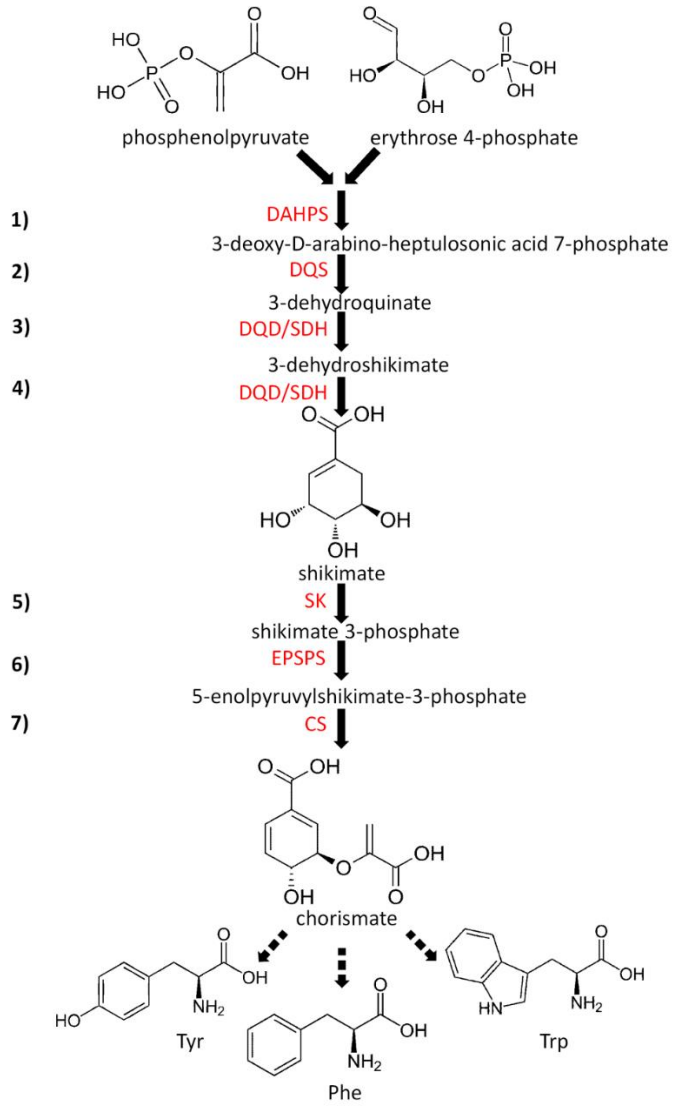


Figure 2.2: Schematic representation of the plant shikimate pathway. In a series of seven enzyme-catalyzed reaction and starting with the condensation of phosphoenolpyruvate and erythrose 4-phosphate, this pathway is used to generate chorismite, which is converted to the aforementioned amino acids in downstream reactions denoted by broken arrows. DAHPS, 3-deoxy-d-arabino-heptulosonate 7-phosphate synthase; DQS, 3-dehydroquinate synthase; DQD/SDH, dehydroquinate dehydratase/shikimate dehydrogenase; SK, shikimate kinase; EPSPS, 5-enolpyruvylshikimate-3-phosphate; CS, chorismate mutase

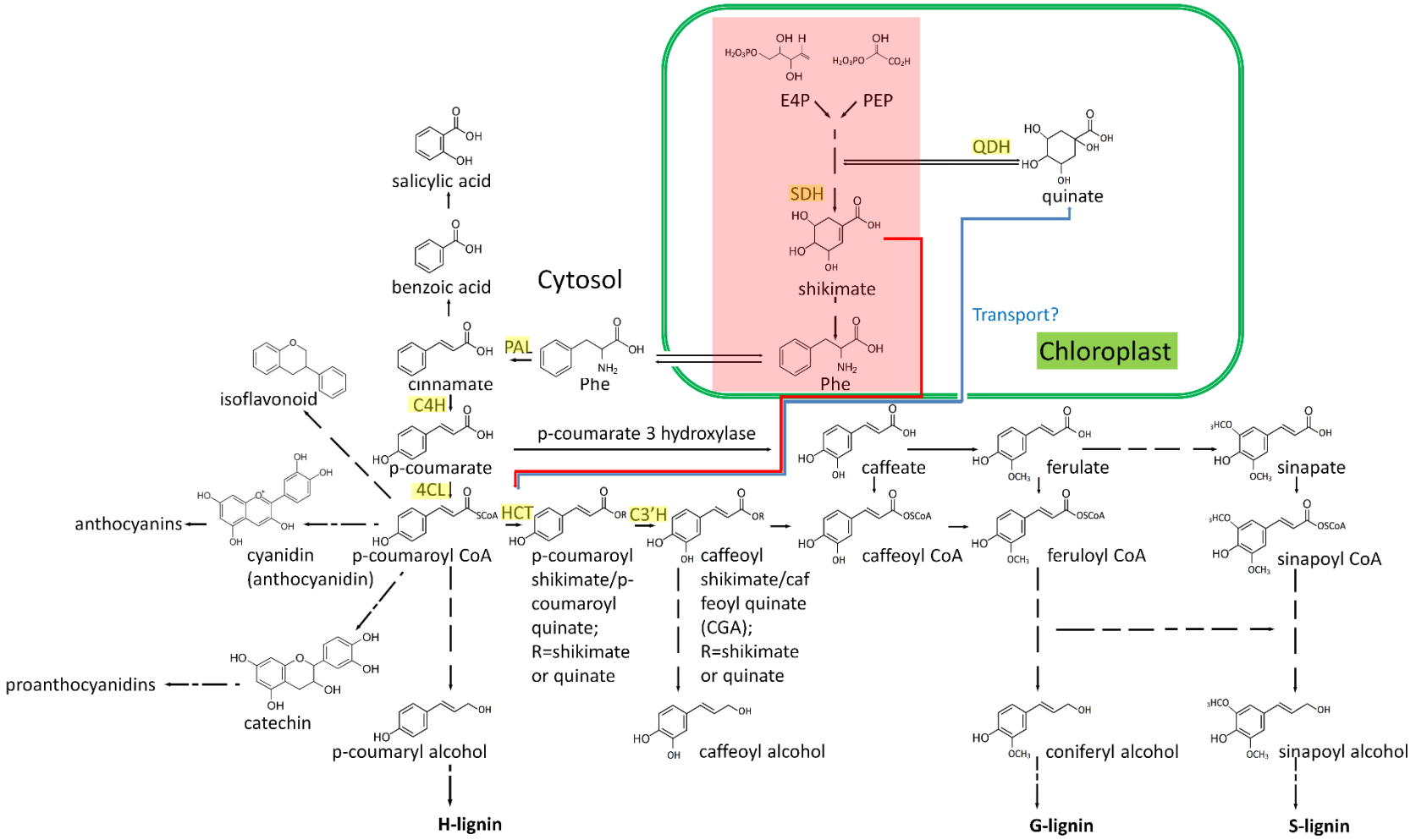

E4P
PEP
QDH
quinate
SDH
shikimate
Cytosol
Transport?
Chloroplast
salicylic acid
benzoic acid
PAL
Phe
Phe
cinnamate
isoflavonoid
C4H
p-coumarate
p-coumarate 3 hydroxylase
caffeate
ferulate
sinapate
anthocyanins
4CL
HCT
C3'H
cyanidin
(anthocyanidin)
p-coumaroyl CoA
p-coumaroyl shikimate/p-coumaroyl quinate; R=shikimate or quinate
caffeoyl shikimate/caffeoyl quinate (CGA); R=shikimate or quinate
caffeoyl CoA
feruloyl CoA
sinapoyl CoA
proanthocyanidins
catechin
p-coumaryl alcohol
caffeoyl alcohol
coniferyl alcohol
sinapoyl alcohol
H-lignin
G-lignin
S-lignin

Figure 1.3: Interconnectivity of phenylpropanoid metabolism and the shikimate pathway. The primary metabolic shikimate pathway, highlighted in pink, is localized in the chloroplast where it fuels the production of aromatic compounds derived from phenylalanine such as proanthocyanidins (condensed tannins), isoflavonoids and lignin. The common carbon backbone of flavonoids and isoflavonoids requires the combination of p-coumaroyl CoA with three molecules of malonyl CoA (not shown). Intermediates of the shikimate pathway can also serve as precursors of secondary metabolites, including quinate, derived from 3-dehydroquinate. QDH, quinate dehydrogenase; SDH, shikimate dehydrogenase (synonymous with dehydroquinate dehydratase/shikimate dehydrogenase); PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase (CYP73); 4CL, p-coumarate CoA-ligase; HCT, hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase; CGA; chlorogenic acid; C3'H, p-coumaroyl quinate/shikimate 3'-hydroxylase (CYP98A3)

In particular, the latter are derived from the shikimate pathway intermediate 3-dehydroshikimate (Ossipov *et al*., 2003). In some plants, [e.g. *Quercus mongolica* and *Q.* myrsinifolia *(*Ishimaru *et al*., 1987) and *Pistacia lentiscus* (Romani *et al*., 2002)] glucose may be replaced as the central moiety in gallotannins by other polyols including shikimate and quinate (Hagerman, 2002). The latter is also a secondary metabolite synthesized in a side branch of the shikimate pathway. As it is important to this work, it is worth noting here that the reversible reduction of 3-dehydroquinate to quinate in seed plants is catalyzed by quinate dehydrogenase (QDH) (Guo *et al*., 2014) (Figure 1.3). True to the propulsive nature of plant secondary metabolism, the generation of quinate leads to still more diverse compounds including chlorogenic acid (see also chapters three and five) (Guo *et al*., 2014; Niggeweg *et al*., 2004). Yet the mechanisms underlying the production of the majority of plant secondary metabolites remain elusive (Pichersky and Gang, 2000).

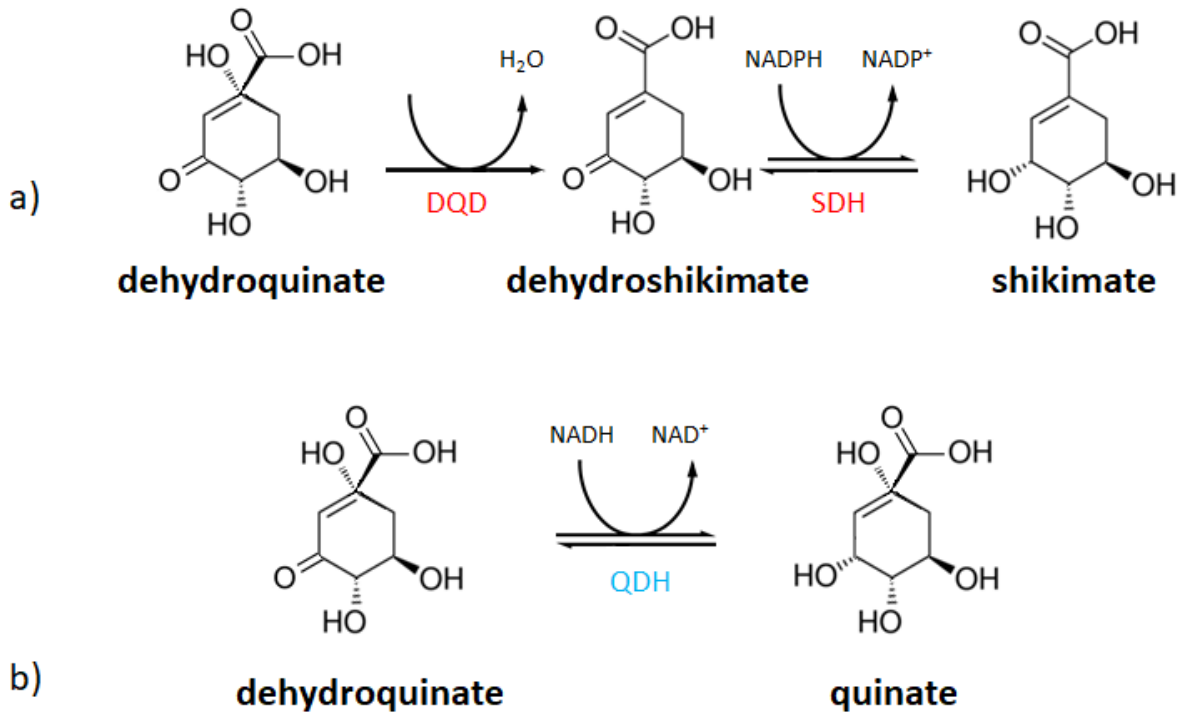


Figure 1.4: Reactions catalyzed by shikimate dehydrogenase (SDH) and quinate dehydrogenase (QDH). The dehydratase domain of bifunctional plant DQD/SDH is found at the amino terminal of the protein and catalyzes the dehydration of dehydroquinate to dehydroshikimate while the subsequent reduction of dehydroshikimate to shikimate is catalyzed by the SDH domain at the carboxy terminal of the protein and involves the transfer of a hydride ion to the cofactor NADPH. Quinate dehydrogenase (QDH) is proposed to use a similar reaction mechanism as SDH to catalyze the reversible reduction of dehydroquinate to quinate.

1.4 Duplicated genes in plants—an overview

Genes encoding SDH and QDH proteins are a striking example of paralogous sister genes arising from gene duplication. They are characterized by similar gene architectures and often functions of the encoded enzymes. Plants are master gene hoarders. Much of their genome consists of superfluous genetic elements and about 65% of their genomes are duplicated (Panchy *et al*., 2016) whereas only 38% and 30% of genes are duplicated in humans and yeast respectively (Zhang, 2003). The presence of gene copies provides relief to geneticists and bioinformatics researchers because it means a plant's genetic closet is not a complete tumultuous jumble of genes. Instead,

the latter can be organized into families based on common descent (Kroymann, 2011; Panchy *et al*., 2016; Zhang, 2003). Understanding the evolution of whole biosynthetic pathways is challenging due to the fact they typically consist of multiple, interweaving enzyme catalyzed steps that are difficult to tease apart. However, comparisons between gene family pairs placed in a phylogenetic framework can provide important hints (Boudet, 2007; Kroymann, 2011; Ober, 2005). A textbook example of a gene family is the cytochrome P450 monooxygenase (CYP) superfamily with members carrying out hydroxylation reactions necessary for phenylpropanoid biosynthesis. These include cinnamate-4-hydroxylase (C4H or CYP73) and p-coumaroyl quinate/shikimate 3'-hydroxylase (C3'H or CYP98A3) which catalyze hydroxylation of *trans*-cinnamic acid and p-coumaric esters of shikimate and quinate respectively (Boudet, 2007; Mahesh *et al*., 2007). Albeit their shared sequence similarity, the enzymes prefer distinct substrates indicating they have sequentially (and therefore functionally) diverged since their time of birth, *i.e.* were duplicated. Such observations shed fascinating light on the molecular switches in the form of amino acid changes needed to alter gene and enzyme functions. The expansion of gene families via duplication and divergence is undoubtedly a driver of metabolic plasticity (Ober, 2005; Pichersky and Gang, 2000; Weng, 2014). In coffee, for example, a CYP98A ancestor was duplicated creating CYP98A35 and CYP98A36. Among them, CYP98A35 showed higher affinity for p-coumaroyl quinic acid than CYP98A36 (Boudet, 2007; Mahesh *et al*., 2007). The ability to synthesize both caffeoylquinic acid (chlorogenic acid) and caffeoylshikimic acid may have optimized lignin biosynthesis or defense responses involving chlorogenic acid (see also chapter three and five) in *Coffea* plants (Mahesh *et al.*, 2007). Above all else, it likely increased their biosynthetic capacity by promoting new paths using caffeoylquinic acid as a substrate or intermediate. The presence of a p-coumaroyl quinate specific isoform in coffee, and not for example in *Arabidopsis* (Mahesh *et al*., 2007) provides timely clues as to when during land plant evolution caffeoylquinic acid biosynthesis developed as long as taxonomic relationships between the plants are established (Pichersky and Gang, 2000; Weng, 2014).

1.4.1 Unearthing the roots of plant secondary metabolism

As illustrated by the case of CYPs, plants rarely create something from scratch. Secondary metabolic genes undoubtedly serve as a pool from which still more genes of secondary metabolism can evolve (Ober, 2005; Pichersky and Gang, 2000). Yet given that primary metabolism is shared

by all plants, and that specialized metabolism branches from it in a species-specific manner, it is very likely that the root of all secondary metabolic genes can be traced back to primary metabolism (Ober, 2005; Pichersky and Gang, 2000; Weng, 2014). This idea is supported by shared homology between primary and secondary metabolic genes. As earlier mentioned, comparisons between members of a gene family provide important clues about the establishment of novel biosynthetic pathways. Since many secondary metabolic genes have limited taxonomic distribution (Bourgaud *et al*., 2001; Croteau *et al*., 2000; Wink, 2003), comparing them provide insights about lineage-specific events. In contrast, when gene families also include members of primary metabolism which is shared across all plant taxa (Hartmann, 2007; Wink, 2003), broader comparisons can be made (Pichersky and Gang, 2000; Weng, 2014). Notably, this paints a possible picture of the stepwise events leading to the establishment of certain metabolic pathways in the context of land plant evolution. Few examples in the literature describe such expansive gene family trees. The aforementioned CYP superfamily producing phenylpropanoids in an assembly line manner is one example since it also includes members involved in the biosynthesis of phytohormones. The latter include abscisic acid and gibberellins that promote growth and development and are therefore classified as primary metabolites (Croteau *et al*., 2000; Mizutani, 2012; Theis and Lerdau, 2003). Gibberellin biosynthesis also involves terpene synthases, which like CYPs, constitute a large superfamily; some of whose members biosynthesize defense-related compounds (e.g. abietic acid found in conifer resin) (Chen *et al*., 2011; Croteau *et al*., 2000; Theis and Lerdau, 2003). Core to this work, is a third example of a gene family involved in both primary and secondary metabolism: the S/QDH family (Guo *et al*., 2014).

1.5 The shikimate/quinate dehydrogenase (*S/QDH*) gene family

The S/QDH family which provides the bedrock of this thesis work, spans major taxonomic lineages of green plants including green algae (Chlorophyta), mosses (Bryophyta), lycopods (Lycophyta), gymnosperms, and angiosperms. Note that this phylogenetic tree (Figure 1.3) was built using amino acid sequences, the advantage of which being they are more conserved than DNA sequences. Genes encoding S/QDH exists as a single copy in the aquatic bacterial phylum, Planctomycetes which serves as an outgroup (Richards *et al*., 2006). Single copy *S/QDH* genes are also found in non-seed plants used in the phylogeny (chlorophytes, bryophytes, and lycopods). This observation points to the maintenance of single-copy *S/QDH* genes throughout early land

plant evolution. In contrast, multiple *S/QDH* gene copies are found in most (but not all) seed plants due to a gene duplication event in the common ancestor of seed plants (>300 MYA). These form two major clades within angiosperms and gymnosperms. All biochemically characterized SDH enzymes [*i.e.* from *A. thaliana* (Singh and Christendat, 2006), *Juglans regia* (Muir *et al*., 2011), *Nicotiana tabacum* (Bonner and Jensen, 1994; Ding *et al*., 2007), *Solanum lycopersicum* (Bischoff *et al*., 2001), *Vitis vinifera* (Bontpart *et al*., 2016), *and Populus trichocarpa* (Guo *et al*., 2014)] cluster closely into one of the angiosperm clades, earning it the title of angiosperm "SDH" clade. Less is known about the second angiosperm clade, which includes only two previously biochemically characterized sequences. This clade was denoted the angiosperm "QDH" clade because its two characterized sequences [from *P. trichocarpa* (Poptr*QDH and PoptrQDH2*) (Guo *et al*., 2014)] exhibited mostly QDH activity *in vitro*. Nothing is known about the biochemical functions encoded by S/QDH sequences of gymnosperms that are sisters to the angiosperm SDH's and QDH's. Although bifunctional S/QDH enzymes from loblolly pine (*Pinus taeda*) have been reported (Ossipov *et al*., 2000), it is unknown if their sequences correspond to those used in the S/QDH family. Like a real living tree, the branches of this phylogeny continue to slough off and bifurcate. Members of the seed plant SDH and QDH clades have undergone additional lineage-specific deletions and duplications; the latter giving rise to clearly separated subclades in each group (Gritsunov *et al*., 2018). Credit for the hard work of constructing this phylogeny goes to Drs. Jia Guo, Jürgen Ehlting and Cuong Hieu Le (Carrington *et al*., 2018).

1.6 Research objectives

Phylogenetic sequence alignments provide a useful snapshot of evolution, but they alone are insufficient to describe changes in the functions of genes (and therefore proteins). As mentioned earlier in the case of CYPs, closely related enzymes often show distinct substrate profiles (Boudet, 2007). Terpene synthases showing greater than 70% sequence identity are known to catalyze distinct reactions (Theis and Lerdau, 2003). A more detailed understanding of evolution is gained by combining sequence analyses with functional characterization of the encoded enzymes (Pichersky and Gang, 2000). Their roles must then be evaluated at the cellular and organismal levels to appreciate the significances of genetic/enzymatic changes on a larger scale. Analysis of the *in vivo* functions of secondary metabolic genes in plants will in turn help direct future studies

delving into more challenging issues such as their ecological roles, and the reason(s) behind their existence and maintenance (Theis and Lerdau, 2003).

The core objective of this work is to characterize the evolutionary diversification of the plant *S/QDH* gene family. This will be done by biochemically characterizing the activities of pre- and post duplication *S/QDH* family members in green plants. The characterization of gene activities among early and late derived plants is expected to provide insights into changes that have occurred to S/QDH protein function in a stepwise context. Both changes to substrate specificity and cofactor specificity will be examined. In addition to determining the functions of *S/QDH* genes in extant plant taxa that diverged prior to the duplication event and which are assumed to represent the ancestral state, ancestral reconstruction will also be used to analyze the functions of an *S/QDH* gene that belonged to a hypothetical ancestor of seed plants that existed before the duplication event (>300 MYA). Understanding the pre- and post duplication activities of *S/QDH* genes will be used to determine the model of evolution that best represents gene duplication and functional diversification of this gene family (see Chapter 2). Specifically, these experiments will involve heterologous expression of recombinant SDH and QDH proteins in *E. coli* followed by affinity purification and enzyme assays. *In vitro* characterization will be combined with mutagenesis work to pinpoint the specific role(s) of amino acid substitutions that are predicted to have played a part in facilitating alterations to substrate profiles based on positive selection tests. Lastly, this thesis describes attempts to elucidate the *in vivo* functions of angiosperm *QDH* genes order to validate *in vitro* studies and to shed light on their biological importance. This will be done using *Arabidopsis* plants overexpressing *QDH* from *P. trichocarpa* and by searching for novel production of quinate and quinate-derived compounds in mutant plants compared to wildtype. While it could not be completed here, future testing of transgenic, *QDH* overexpressing plants will provide an invaluable opportunity to elucidate the physiological relevance of QDH proteins, which to this day remains shrouded in mystery and hence poses exciting new research opportunities.

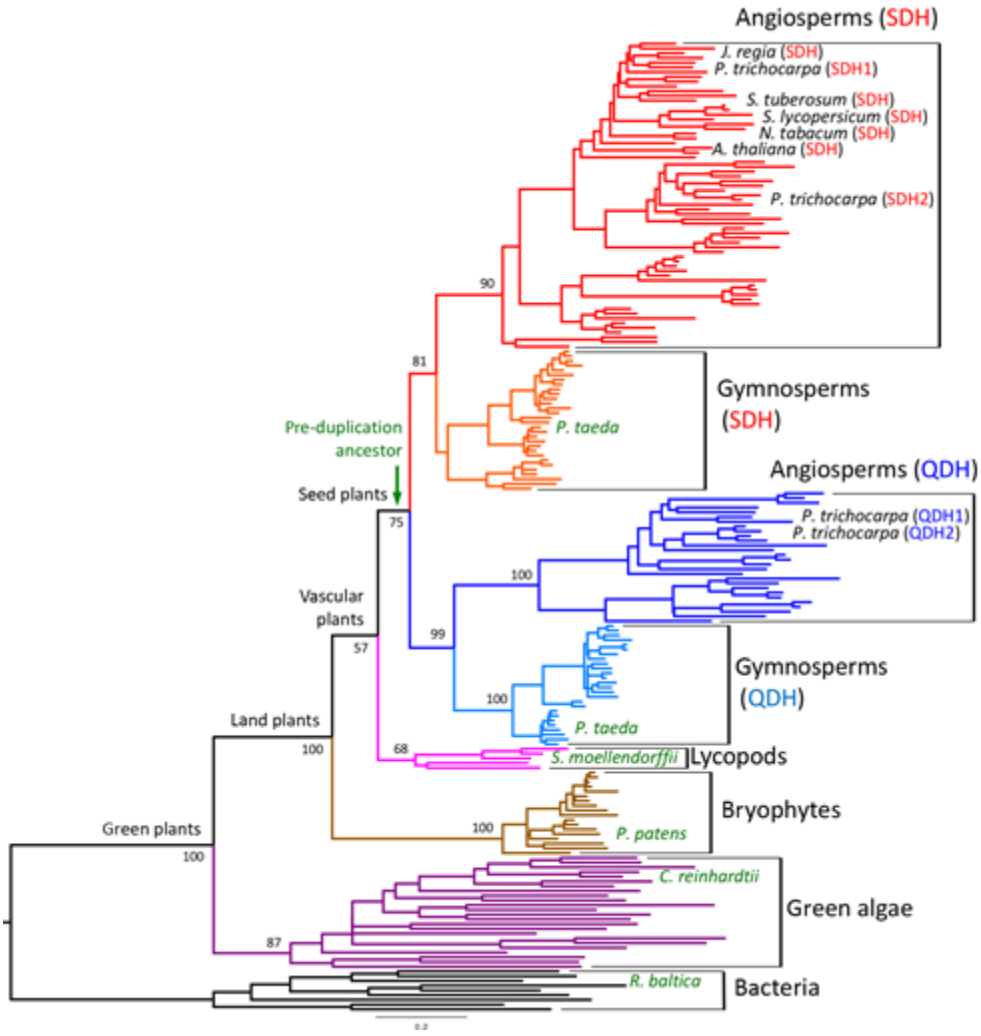


Figure 1.5: Maximum-likelihood phylogeny of plant S/QDH protein sequences. Bootstrap values (from 1,024 replicates) are given in percent for branches leading into the major clades only. Clades depicting taxonomic groups are indicated and color-coded. Proteins previously biochemically characterized are shown by species name and biochemical function (in brackets); proteins characterized here are shown by species name in green (Carrington *et al*., 2018).

Chapter 2

2.1 Gene duplication in plants

As mentioned in the previous section, an overarching theme of this work is to identify the model of evolution that best describes diversification of the *S/QDH* gene family. Since there are a number of different evolutionary models that have been described in the literature, the purpose of this chapter therefore is to briefly describe the most popular ones. However, first it will be useful to start this section with a brief overview of gene duplication prior to diversification. The idea that more genes enable more mutational opportunities is seemingly straightforward, yet the actual phenomena of gene duplication and evolution are riddled with mysteries. There are several explanations as to *how* gene duplications occur. Duplications of singleton genes for instance, may occur inadvertently if they reside close to autonomous retrotransposons. The latter come fully equipped with the transcriptional machinery needed to copy themselves in the genome. However, this process is somewhat lax such that RNA polymerase occasionally reads past their polyadenylation signal and includes downstream genes, copying them along in the process (Lynch, 2007). Segmental duplications might occur by unequal crossover and whole genome duplications (WGDs) can result from meiotic nondisjunction and the fusion of unreduced gametes (Otto, 2007; Sankoff and Zheng, 2018). Both autopolyploidy (containing chromosomes from a single species) and allopolyploidy (containing chromosomes from different species as a result of hybridization) are common in wild and bred plant populations (Hegarty *et al*., 2013; Sattler *et al*., 2016). Well-known polyploids include crop species such as triploid seedless watermelon (*Citrullus vulgaris*), tetraploid cotton (*Gossypium*) and hexaploid bread wheat (*Triticum aestivum*) (Sattler *et al*., 2016). Polyploidy can be induced artificially using colchicine, the gem of many plant breeders, which arrests cell division without halting DNA replication (Hegarty *et al*., 2013; Sattler *et al*., 2016). On the other hand, fewer cell cycle check-points in plants compared to animals likely contribute to the high frequency of observed polyploid plants in the wild (Wijnker and Schnittger, 2013). In flowering plants, where a doubling, tripling and quadrupling of the genome is unexceptional, the potential for evolving novel traits is high (Hegarty *et al*., 2013; Panchy *et al*., 2016; Sattler *et al*., 2016).

2.1.1 Models of gene duplication

About 65% of plant genes are duplicated and it is unknown why plants show a tendency towards keeping extra gene copies. The presence of extra genetic materials is thought to place an energetic burden on cells as ATP is used to replicate, transcribe and translate them (Lynch and Marinov, 2015). Not only are they energetically costly to keep, but given their relaxed selectional constraints, the probability that they will accumulate deleterious mutations and be rendered non-functional (a process called pseudogenization) is high. Indeed, flowering plants including *Arabidopsis thaliana* have experienced genomic downsizing following WGD. Still the high prevalence of duplicated genes in *A. thaliana* and many other (especially angiosperm) species largely point to an alternative fate (Panchy *et al*., 2016). Gene copies may be maintained in the genome if they provide a selective advantage to the host that outweighs their costs. There are two general ways in which a gene can benefit a host: 1) by enhancing pre-existing activities or 2) by bestowing novel adaptations (Chen *et al*., 2013; Pichersky and Gang, 2000; Weng, 2014). Several explanations exist under the first category including the "gene dosage," "duplication degeneration and complementation" and "gene balance" models, which are roughly summarized below:

The **gene dosage model** points to the benefits of a quantitative increase in useful gene products following gene duplication. For example, amplification of glycolytic pathway genes in yeast increases the efficiency of anaerobic energy production when glucose availability in the environment is high (Panchy *et al*., 2016). Gene dosage effects are also an evolutionary advantage of drug resistant pathogens like *Plasmodium falciparum*, the causative agent of malaria. *P. falciparum*'s resistance to melfoquine is enhanced by increased copy number of a multidrug resistance gene (Conant and Wolfe, 2008).

The **duplication degeneration and complementation (DDC) model** refers to a phenomenon under the umbrella term **subfunctionalization** in which a subset of the functions carried out by an ancestral gene is (un- or) equivocally lost ("degenerated") amongst its gene copies due to random mutations. This can occur, for example, if the parent gene encoded a multifunctional protein and the daughter genes each adopt only a subset of these functions. Alternatively, changes in gene regulatory regions may cause a partitioning of parental gene expression. In either case, both copies are required to "complement" each other and fulfill their parent's job(s) (Force *et al*., 1999). A

distinguishing factor of this model above others is that fixation of duplicated genes is a stochastic process so that there is no net gain of new functions.

The **gene balance model** states that "dosage-sensitive genes," or those involved in molecular interactions (e.g. in the form of structural complexes or signalling cascades) are preferentially retained after gene duplication. This is because the loss of one or more participating members of such networks could compromise their functionality (Flagel and Wendel, 2009; Panchy *et al*., 2016; Thomas *et al*., 2006). In support of this theory, in *Arabidopsis*, genes whose products are involved in ribosomal, transcriptional or signalling complexes have been selectively retained as duplicates since WGD whereas their non-interacting homoeologues were lost (Thomas *et al*., 2006). [Note that the term "homoeologues" is distinct from "homologues" in that it refers to pseudo-pairs of chromosomes derived from different species as a result of past hybridization. Such chromosomes may or may not pair during meiosis (Gaeta and Pires, 2010; Glover *et al*., 2016)].

Despite initial hurdles faced by young polyploids (e.g. reproductive isolation), WGD has been associated with the colonization of newly opened habitats, and to a much greater effect, speciation (Brochmann *et al*., 2004; Otto, 2007; Ramsey, 2011; Wertheim *et al*., 2013). This is because many polyploid plants (e.g. the arctic grass *Dupontia*) are pioneer species of extreme latitudes or recently unglaciated regions that are unoccupied by their diploid counterparts (Brochmann *et al*., 2004; Otto, 2007; Ramsey, 2011; Wertheim *et al*., 2013). The adaptive prowess of polyploids has been attributed to novel, useful traits. The latter may arise when extra gene copies diversify and obtain new functions benefitting the host and, in turn, allowing the duplicated genes to be maintained in the genome and flee pseudogenization (Hegarty *et al*., 2013; Panchy *et al*., 2016; Sattler *et al*., 2016). Two well-known models that attempt to explain the retention of gene copies via adaptive specialization are **neofunctionalization** and **subfunctionalization** summarized below:

The **neofunctionalization model** describes the gain of new gene/protein functions in one gene copy as a result of rare mutation events. Mutations can occur at protein-coding or regulatory regions as long as vital ancestral activities are retained by one gene copy (Matsuno, *et al*., 2009; Moore and Purugganan, 2005; Zhang, 2003). A prominent example of neofunctionalization is the evolution of two cytochrome P450 (CYP) genes, CYP98A8 and CYP98A9 leading to the

development of a novel N^1,N^5-di(hydroxyferuloyl)-N^{10} biosynthesis pathway in *Arabidopsis*. The parent gene, CYP98A3 catalyzes the formation of lignin precursors via meta-hydroxylation of p-coumaroyl shikimate in vascularized tissues (flowers, stems, and roots). Following duplication and mutation, its daughter genes gained novel expression patterns in reproductive organs and meta-hydroxylase activity with tricoumaroylspermidine, a precursor of pollen (Matsuno, *et al*., 2009). Although the current study focuses on plants, it is worthwhile to note that cases of neofunctionalization are also found in the animal kingdom. For example, resistance of Asian brown planthoppers (*Nilaparvata lumens*) to the insecticide imidacloprid evolved via duplication and divergence of yet another CYP gene, CYP6ER1, shared by non-resistant strains. In particular, two amino acid substitutions in the substrate recognition site of the encoded protein conferred the ability to bind to and metabolize imidacloprid whereas ancestral CYP6ER1 cannot (Zimmer *et al*., 2018). Note that in both cases, the ancestral functions (of CYP98A3 and CYP6ER1) are retained by one gene copy, opening a window of mutational opportunities for another.

Subfunctionalization models of adaptive evolution include **Escape from Adaptive Conflict (EAC)** and **Innovation Amplification** (described subsequently) **and Diversification** in which, like DDC, the responsibilities of a multifunctional parent gene are divided amongst its daughter genes. Despite heavy overlaps across these processes, they are nevertheless given unique names and if they have unique names then they must be different processes. One distinction between the DDC and EAC model is that the latter describes optimizing mutations in the daughter genes, whereas DDC is assumed to be an evolutionary neutral process (Des Marais and Rausher, 2008). EAC assumes that a "conflict" is created when opposing selection pressures prevent simultaneous optimization of dual functions encoded by a single gene. By duplicating and partitioning its functions among its daughter genes, the parent gene "escapes" this conflict and its sub-functions are free to evolve independently (Deng *et al*., 2010; Des Marais and Rausher, 2008; Sikosek *et al*., 2012). This situation has been called the "Babe Ruth effect," referring to the years in which the athlete was a phenomenal pitcher and later hitter and fielder: however, his skills dropped to subpar (in the eyes of fans) when he served as both pitcher and position player for the Boston Red Sox (Hughes, 2005)—of course Xeroxing Ruth to solve the issue was out of question.

So far, compelling evidence for EAC is lacking due to its multiple stringent requirements. According to Des Marais and Rausher (2008) who are credited for the concept, EAC occurs if 1) a conflicted multifunctional parent gene undergoes gene duplication and 2) positive selection acts on *both* daughter genes to optimize both (or all of) the parent's (sub)functions (Barkman and Zhang, 2009). Unfortunately, the case in point presented by Des Marais and Rausher (2008) was criticized on a number of grounds. While they claimed that dihydroflavonol-4-reductases (DFR) in the common morning glory (*Ipomoea purpurea*) evolved via EAC, they could not detect positive selection acting on one of the daughter genes nor determine their functions making it difficult to determine whether gene duplication helped solve an adaptive conflict (Barkman and Zhang, 2009). It is worthwhile now to take another look at other eukaryotes: although they do not provide foolproof support of EAC, they do provide similar scenarios under which it may occur.

The evolution of the animal eye lens provides a remarkable tale of Swiss-army knife proteins with highly disparate functions. For example, duck δ-crystallin and arginosuccinate lyase are encoded by two dual-functioning genes. The encoded protein acts either as a structural component of the eye or as a metabolic enzyme based on where it is expressed. Chickens also have two δ-crystallin paralogs arising from gene duplication. Unlike in ducks, one of them has become specialized such that it is predominantly expressed in the eyes with negligible arginosuccinate lyase activity. Such specialization may have helped resolve conflicting selection pressures for the structural and metabolic activities of the parent gene (represented by duck δ-crystallin/arginosuccinate lyase) (Piatigorsky, 1991; Piatigorsky *et al*., 1988; Wistow, 1993). However, since selection tests were not performed it is unclear if evolution of chicken δ-crystallin genes occurred adaptively. It is also a mystery whether the ancestor's lyase activity was optimized.

Innovation, amplification, and divergence (IAD) is a more recent model of evolution that falls into the grey zone between neofunctionalization and subfunctionalization. Confusingly, some publications use the term synonymously with adaptive radiation. Based on enzyme promiscuity, IAD highlights the tendency of enzymes to carry out minor activities alongside their evolved roles. Following gene duplication, these latent skills become amplified through gene dosage effects, reaching a level where this activity becomes physiologically relevant and can be the target of selection. An increase in gene population size resultantly increases the probability that at least one

gene copy will mutate, becoming optimized for a minor activity. One of its sister genes will carry out the parent gene's fulltime job while all others can be shed (Conant and Wolfe, 2008; Khersonsky *et al*., 2006; Näsvall *et al*., 2012). This model was proposed by Näsvall *et al*. (2012) who witnessed evolution in action in stressed bacteria. The authors grew *Salmonella* mutants defective in the tryptophan biosynthetic enzyme TrpF on minimal media lacking both tryptophan and histidine. However, they did possess a functioning copy of the histidine biosynthetic enzyme, HisA with shared ancestry with TrpF. Notably the two enzymes are catalytically similar such that HisA could biosynthesize tryptophan under the said conditions albeit at low levels. Over many generations HisA proliferated and mutated until bona fide TrpF alongside HisA genes evolved. Due to a certain degree of substrate permissiveness, some metabolites may be formed 'randomly' as a result of enzyme activities towards non-native substrates. In the case of secondary metabolism, these products, termed "metabolic noise" presumably serve no initial purpose in an organism. However, they may confer fitness advantages when environmental conditions change such that they become favoured by natural selection. In this way, novel functions are thought to arise, not only *de novo* or from pre-existing activities, but also from broad catalytic activities (either through using multiple substrates or catalytic mechanisms (Peisajovich and Tawfik, 2007; Weng and Noel, 2012; Weng, 2014). A summary of these models is provided in Table 1.1.

Table 1.1: A summary of evolutionary models. Classical models of gene duplication and evolution may overlap or differ in the following: the 1) presence or absence of a promiscuous ancestor gene; 2) loss or preservation of duplicated genes in the genome; 3) functions carried out by duplicated genes; and 4) selection type(s) acting on duplicated genes. Note "Na" is used when unspecified. Different publications show a conflict regarding the selection pressures partaking in EAC. Some describe it as "relaxed purifying" selection which enables daughter genes to optimize their ancestral functions without losing them whereas others simply call it "positive" selection. Note the terms "a" and "b" arbitrarily refers to the two gene copies after duplication.

Model	Moon-lighting ancestor?	Surviving duplicates?	Fate(s) of duplicated genes	Selection on each gene duplicate
Pseudogenization	NA	No	a) Loss of function	None/neutral
			b) Retains ancestral functions	Negative/purifying
Neofunctionalization	No	Yes	a) Gains novel functions	Positive
			b) Retains ancestral functions	Negative/purifying
Gene dosage	NA	Yes	a) Retains ancestral functions	NA
			b) Retains ancestral functions	NA
DDC	Yes (dual functional)	Yes	a) Retains partial ancestral functions	None/neutral
			b) Retains partial ancestral functions	None/neutral
EAC	Yes (dual functional)	Yes	a) Improves partial ancestral functions	Relaxed purifying and positive
			b) Improves partial ancestral functions	Relaxed purifying and positive

Chapter 3

Data from this chapter as well as written information in the methods, results and (partly) from the discussion has been published in Carrington *et al*. (2018). Construction of the phylogenetic tree was done by Drs. Jürgen Ehlting, Jia Guo and Cuong Hieu Le. Positive selection tests [the methods of which are not described here but are published in Carrington *et al*. (2018)] were performed by Drs. Jürgen Ehlting and Jia Guo. Ancestral reconstruction and cloning of *Poptr*SDH and QDH was performed by Dr. Jia Guo.

3.1 Introduction

The contributions of each model towards describing evolution in real life are unknown, and it is highly possible they act in concert to functionally diversify duplicated genes. To validate these models, it is necessary to apply them to actual scenarios but so far, unambiguous molecular evidence for them is largely lacking. Analyses of gene families such as the CYPs and terpene synthases carrying out multifarious reactions in both primary and secondary metabolism provided unprecedented insights into the mechanisms underlying their expansion and functional diversification (Chen *et al*., 2013; Pichersky and Gang, 2000; Weng, 2014). In spite of the large amount of useful data applicable for bioengineering and evolutionary biology that can be extracted from studying such large families, their sizes and complexities can make a complete profiling of their members seem like a daunting prospect. For example, the CYP superfamily includes 245 members in *Arabidopsis* alone (Weng, 2014) even though its genome is relatively small for a flowering plant (Sena *et al.*, 2014). Compared to CYPs and terpene synthases, the S/QDH superfamily includes fewer members but has received little attention. For example, only a single copy *S/QDH* gene is found in *Arabidopsis* and no more than five copies are found in *Populus trichocarpa* (Guo *et al*., 2014), whose genome is more than double the size of *Arabidopsis* (Stival Sena *et al.*, 2014). Its relative modest size, combined with well-established protocols for measuring dehydrogenase activities *in vitro* and the available crystal structure of *Arabidopsis* SDH (Singh and Christendat, 2006) makes the S/QDH family an ideal platform to study the evolutionary fates of duplicated genes.

3.1.2. Experimental objectives: characterization of *S/QDH* across taxonomic representatives of green plants representing pre- and post-duplication enzyme activities

A problem that arises during phylogenetic analyses is the absence of sub-optimal intermediate forms that have been lost over evolution (Darwin, 1859). They create "gaps" in an evolutionary timeline, making it difficult to accurately track the complete progression of events from an ancestral gene to its extant descendants. Ancestral activities can be comparatively inferred from sister clades (Kroymann, 2011; Pichersky and Gang, 2000; Weng, 2014); in this case, from the activities encoded by unduplicated *S/QDH* genes of non-seed plants. An alternative method of studying how genes have changed over time is ancestral reconstruction (Huang *et al*., 2012; Voordeckers *et al*., 2012). This method relies on phylogenetic relationships to predict the most likely nucleotide (or amino acid) sequence at a nodal position of a tree (Cai *et al*., 2004). For this study, the S/QDH sequence of the immediate pre-duplication ancestor of seed plants (from more than 300 MYA), dubbed "Anc122SDH," was reconstructed by Dr. Jia Guo, a retired protein necromancer (Carrington *et al*., 2018). The complete list of species used in this work includes (also depicted in Figure 3.1):

a) the green algae *Chlamydomonas reinhardtii* (*ChlreSDH*)
b) the bryophyte *Physcomitrella patens* (*PhypaSDH*)
c) the lycopod *Selaginella moellendorffii* (*SelmoSDH*)
d) the reconstructed ancestor of seed plants (Anc122*SDH*)
e) the angiosperm *P. trichocarpa* (*Poptr*SDH, and *Poptr*QDH)
f) the gymnosperm *P. taeda* (*PintaSDH* and *PintaS/QDH*)

The first goal of this work is to biochemically characterize S/QDH sequences of non-seed plants (a-c), the resurrected ancestor (d) and seed plants (e-f) *in vitro* representing the pre-duplicated (a-d) and post-duplicated (e-f) states of *S/QDH* genes respectively. The obtained data will be used to determine the evolutionary model(s) that best describes functional diversification of the *S/QDH* gene family. Given that shikimate is needed for protein biosynthesis (Herrmann and Weaver, 1999), it is expected that ancient proteins acted on shikimate *i.e.* had SDH activity. However, it is unknown whether they also encoded at least some QDH activity that was later augmented in one of its gene copies *or* if QDH activity was gained exclusively in some gene copies of seed plants

after gene duplication. Support for these alternative hypotheses would confirm either evolution by subfunctionalization (EAD and/or IAD) or neofunctionalization respectively.

3.1.3 Experimental objectives: mutagenesis of S338G and T381G in wildtype SDH from *P. trichocarpa*

Analysis of the crystal structure of *Arabidopsis* SDH has helped identify key active site residues. Notably, Ser338 and Thr381 are required for substrate orientation; the former binds to the C1 carboxylate of shikimate (Singh and Christendat, 2006). These residues are highly conserved at the homologous positions of all other angiosperm SDH's analyzed, confirming their importance for catalysis (Carrington *et al*., 2018; Gritsunov *et al*., 2018; Guo *et al*., 2014). In contrast, both Ser and Thr are replaced by Gly at the corresponding positions of angiosperm QDH's. Modeling of poplar SDH and QDH proteins revealed that substitution with Gly deepens the QDH binding pocket relative to SDH, and enables grasping to a new substrate that is structurally alike but bulkier than shikimate, *i.e.* quinate (Guo *et al*., 2014). Together with the different activities displayed by poplar SDH and QDH enzymes *in vitro* (Guo *et al*., 2014), these observations have prompted mutagenesis to test the effects of the aforementioned amino acid substitutions on substrate binding affinities. Statistical tests previously performed by Drs. Jia Guo and Jürgen Ehlting identified signatures of positive selection acting on these residues (Carrington *et al*., 2018). However, the power of selection tests to find adaptive sites is increased when evidence is provided for the effects of their substitution on protein function(s). The second objective of this experiment is therefore to determine if the replacement of Ser275 and Thr318 (corresponding to positions 338 and 381 in *Arabidopsis* SDH respectively) with Gly leads to a shift in substrate specificity from shikimate to quinate in a shikimate-specific, poplar SDH isoform using site directed mutagenesis. In order to investigate the individual and combined effects of each mutation, three mutant poplar SDH constructs were made and examined:

a) single mutant, Ser275Gly
b) single mutant, Thr318Gly
c) double mutant, Ser275Gly/Thr318Gly

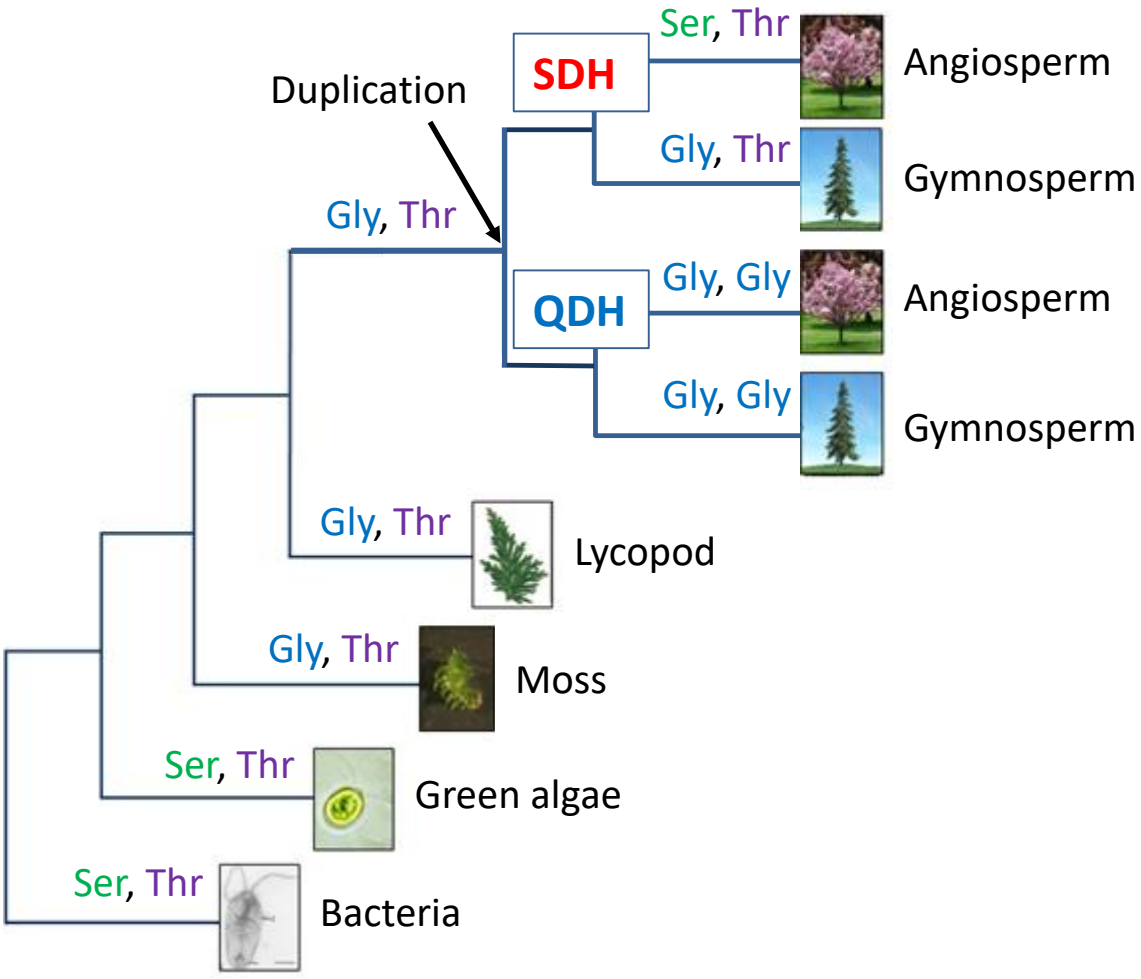


Figure 3.1: Simplified representation of the plant S/QDH superfamily. Taxonomic representatives of the plant S/QDH family characterized in this study are shown. Among them include single-copy sequences from *C. reinhardtii* (green algae), *P. patens* (moss), *S. moellendorffii* (lycopod) and separate SDH and QDH's from *P. taeda* (gymnosperm) and *P. trichocarpa* (angiosperm). Angiosperms have multiples *SDH* gene copies that cluster into either the "SDH" or "QDH" clade based on whether they show activity towards shikimate or quinate respectively *in vitro*. Each clade has a gymnosperm sister groups although the biochemical activities of the latter have not been characterized prior to this work. The hypothetical S/QDH sequence of the common ancestor of seed plants (Anc122) which diverged prior to gene

duplication ~300 MYA (indicated by the arrow) is also characterized in this work as is the planctomycete bacteria, *R. baltica* which serves as an outgroup. The amino acids corresponding to positions 338 (left) and 381 (right) of *Arabidopsis* SDH, belonging to the angiosperm SDH clade, are shown.

3.2 Methods

3.2.1 Homolog fishing

SDH homolog sequences were obtained by performing a BLASTP search against Phytozome v11, the 1KP transcriptome assembly and NCBI's non-redundant protein databases (Goodstein *et al*., 2012; Matasci *et al*., 2014; O'Leary *et al*., 2016) using the amino acid sequence of the characterized DQD/SDH from *A. thaliana* (AT3G06350) or the DQD/SDH and QDH sequences from *P. trichocarpa* [Potri.010G019000 (SDH1), Potri.013G029900 (SDH2), Potri.005G043400 (QDH1), Potri.014G135500 (QDH2)] as bait.

3.2.2 Ancestral reconstruction (performed by Dr. Jia Guo)

Ancestral reconstruction Ancestral S/QDH protein sequences were reconstructed based on a maximum likelihood phylogeny of 110 amino acid sequences. Ancestral character states were reconstructed based on the resulting phylogeny using the empirical Bayes (EB) method implemented in Codeml (PAML v4.5; Yang, 2007) (runmode = 0, seqtype = 2, CodonFreq = 2, model = 2, NSites = 0, iCode = 0, Mgene = 0, fix_kappa = 0, fix_omega = 0, fix_alpha = 1, alpha = 0, Rate Ancestor = 1). The ancestral sequence at the node just prior to gene duplication, Anc122, was extracted from the Codeml output files and reverse translated with BioEdit to obtain a DNA sequence for gene synthesis (see below).

3.2.3 Gene cloning and recombinant protein purification

Plasmid constructs with either *PoptrSDH* or *PoptrQDH* from *P. trichocarpa* in the pQE30 expression vector (Qiagen, http:// www.qiagen.com/) were used from our previous work (Guo *et al*., 2014). The open reading frames of Anc122 (reconstructed pre-duplication ancestor), *Rhoba* (from *R. baltica*), *Chlre* (from *C. reinhardtii*), *Phypa* (from *P. patens*), *Selmo* (from *S. moellendorfii*), *PintaSDH* and *PintaQDH* (both from *P. taeda*) were optimized for *E. coli* expression and chemically synthesized by Genescript and obtained in the pUC57 vector. These

sequences were subcloned into the pQE30 overexpression vector (Qiagen) using BamHI and HindIII sites added to the open reading frames and present in the pQE30 vector prior to transformation into *E. coli* DH5a. Sanger sequencing was performed by Sequetech (http://sequetech.c om/) to validate sequence integrity. Recombinant protein expression was performed as described previously (Guo *et al*., 2014). In brief, recombinant pQE30 constructs were transformed into M15 *E. coli* and positive colonies were grown in liquid culture to an optical density (OD600) of 0.4–0.6. Protein expression was induced with 0.06 mM isopropyl-1-1-thio-β-D-galactoside (IPTG) for 24 h at 19°C. Cells were harvested by centrifuging at 4000 X g and stored at -80°C for at least 1 hr. Recombinant proteins were purified by Ni-NTA affinity chromatography. Frozen cell pellets were resuspended in 4 ml of lysis buffer (50 mM NaH_2PO_3, 300 mM NaCl, 10 mM imidazole and 1 mg/mL lysozyme) and incubated on ice for 1 h with gentle rocking prior to sonication (5 X 10 sec on ice). The lysate was centrifuged at 10 000 X g at 4°C for 30 min to collect the supernatant fraction. Soluble lysate was incubated with 50% Ni-NTA agarose beads (Qiagen) for 1 h on ice with gentle rocking. The Ni-NTA lysate was washed twice with 10 ml of wash buffer 1 (50 mM NaH_2PO_3 and 300 mM NaCl) followed by three washes with 4 ml of wash buffer 2 (50 mM NaH_2PO_3, 300 mM NaCl and 20 mM imidazole) and eluted four times with 0.5 ml of elution buffer (50 mM NaH_2PO_3, 300 mM NaCl and 250 mM imidazole). The first elution was discarded while the remaining three were combined for subsequent experiments. Protein concentrations were determined using the Bradford assay (Bradford, 1976).

3.2.4 SDS PAGE and Western blotting

A fraction of each protein elution was analysed either by SDS PAGE or Western blotting to assess their purities. For SDS-PAGE, elutions were boiled for 20 min in 2X crack buffer prior to gel electrophoresis. Protein samples were separated on a 10% polyacrylamide gel and visualized by staining with Gel Code Blue Stain Reagent (Thermo Fisher Scientific, https://www.thermofisher.com/) according to the manufacturer's protocol except for an extended 15-hr incubation period with the dye. Western blotting was performed by electroblotting proteins onto PVDF membranes (60 min at 100 V). Bands were detected using the SuperSignal West His Probe^{TM} Kit (Thermo Fisher Scientific) following the manufacturer's recommended methods.

3.2.5 Spectrophotometric measurement of SDH and QDH activities

Dehydrogenase activities were measured by monitoring the reduction of $NADP^+$ (or NAD^+) spectrophotometrically at 340 nm using a UV/VIS spectrophotometer (Shimadzu, https://www.shimadzu.com/) under computerized control of the Shimadzu UV probe personal software. Reaction mixtures consisting of 100 mM Trizma base–HCl pH 9, 0.2 mM $NADP^+$ or 0.5 mM NAD^+ and substrate (see below) were carefully mixed in 1-cm path length quartz cuvettes before adding enzyme to start the reaction. A catalytic amount of enzyme (6–20 µg) was used per reaction depending on the protein sample and velocity of the observed reaction. Reactions were carried out for 90 sec at room temperature (20ºC). Enzyme activities initially obtained in units of Abs sec^{-1} , were converted to concentration units (µmol NADPH min^{-1} mg^{-1}) using the extinction coefficient of NADPH at 340 nm (6.22 X 10^{-3} l mol^{-1} cm^{-1}) and normalized for the amount of enzyme used. Kinetic properties were determined by testing multiple (typically 10) shikimate or quinate concentrations with the appropriate cofactor. The apparent K_M value and maximum velocities (V_{max}) of three replicates (independent protein purifications) were modelled to the Michaelis–Menten equation using the 'drm' package implemented in R.

3.2.6 Site-directed mutagenesis

Two codon sites found to be under positive selection in QDH proteins were introduced into the shikimate-specific SDH of *P. trichocarpa*. Two single-mutant constructs harbouring both the Ser338Gly or Thr381Gly substitutions and a double mutant construct containing both were generated using the protocol adapted from the QuikChange® Site-directed Mutagenesis Kit (Agilent Technologies, https://www.agilent.com/) with wild-type PoptrSDH cloned into pQE30 (Guo *et al*., 2014) as a template. Following the validation of desired mutations by Sanger sequencing (Sequetech), the mutant PoptrSDH plasmids were electro-transformed into *E. coli* M15 cells. These were cultured for expression of recombinant His6-tagged wild-type and mutant PoptrSDH. Induction, purification as well as both protein characterizations (i.e. SDS PAGE and Western Blotting) and kinetic analyses of recombinant proteins were carried out as described above.

3.3 Results

3.3.1 SDH and QDH activity across the green plant lineage

All biochemically characterized SDH or QDH enzymes with sequence information available are from angiosperms. To follow enzymatic specificity throughout the plant lineage, members representing each major clade were biochemical characterized. Species were selected based on available sequence information with preference to species with completely elucidated genomes. The two proteins from *Pinus taeda* were chosen to represent the post-duplication SDH and QDH clades from gymnosperms. The single-copy genes from *Selaginella moellendorffii, Physcomitrella patens* and *Chlamydomonas reinhardtii* were selected to represent the pre-duplication lycopod, bryophyte and green algal clades, respectively. In addition to these extant species, we reconstructed the sequence of the immediate pre-duplication ancestor forming the node into the seed plant clade. We included only the most likely reconstructed ancestral sequence for gene synthesis and biochemical characterization. Recombinant His6- tagged proteins were heterologously expressed in *E. coli* and purified by affinity chromatography. Purified SDH or QDH enzymes had sizes consistent with expectations based on the DNA constructs employed (Figure 3.2b,c). Enzymatic activities and cofactor preferences were first determined by incubating enzymes with presumably saturating concentrations (10 mM) of either shikimate or quinate using both $NADP^+$ or NAD^+ as a cofactor. The reconstructed pre-duplication ancestor and all enzymes from extant species that diverged prior to the duplication exhibited high activities with shikimate and $NADP^+$ as expected based on their presumed involvement in the shikimate pathway, but no appreciable activity with quinate (Figure 3.2a). The only exceptions may have been the enzyme from *S. moellendorffii* and the reconstructed seed plant ancestor, both of which showed very minute QDH activities at very high substrate concentrations and when using large amounts of protein. These activities were too low to determine kinetic properties and were very close to or within the limit of detection (mean activity of boiled enzyme plus three standard deviations of the mean). This indicates that SDH is the primary activity of enzymes prior to the duplication and that QDH activity is not present at levels suggestive of a physiological function in SDH enzymes from non-seed plants. Purified SDHs and QDHs displayed typical Michaelis– Menten kinetics (Appendix A). Shikimate dehydrogenase from the green alga *C. reinhardtii* and from the planctomycete *Rhodopirellula baltica* displayed lower maximal velocities towards shikimate than

land plant SDHs (Figure 3.2, Table 3.1), but apparent affinities were similar across all SDHs with K_M values ranging from 100 to 280 µM. Within seed plants, the pine protein representing the SDH clade (PintaSDH) displayed high activity and specificity for shikimate but no detectable activity with quinate, comparable to angiosperm SDHs previously described. Representing the QDH clade from gymnosperms, PintaQDH reacted equally well with both shikimate and quinate (Figure 3a, Table 3.1). PintaQDH has similar apparent affinities and maximal velocities for both quinate and shikimate with a slightly (1.7-fold) higher specificity for quinate than shikimate (based on V_{MAX}/K_M; Table 1). Like all SDHs tested here, PintaQDH is dependent on $NADP^+$ as a cofactor and showed negligible activities close to the detection limit (not exceeding 0.4 µmol mg^{-1} min^{-1}) when NAD^+ was used as a cofactor instead for both SDH and QDH activities. Showing the opposite trend, poplar QDHs preferred NAD^+ to $NADP^+$ as a cofactor with either shikimate or quinate as a substrate, as previously described (Guo *et al*., 2014).

3.3.2 Repeating evolutionary history: site-directed mutagenesis

As a complement to the positive selection tests, the Ser338 to Gly mutation, the Thr381 to Gly mutation or both were introduced into the highly shikimate-specific PoptrSDH background using site-directed mutagenesis. Ser338 and Thr381 in *A. thaliana* correspond to positions 275 and 318 in *Poptr*SDH, respectively. The recombinant mutant enzymes were purified and analysed using SDS PAGE and Western blotting to determine successful purification (Figure 3.3a). Based on Michaelis–Menten kinetic analyses (Figure 3.3b,c), the Ser275Gly mutant showed only slightly reduced maximum activity with shikimate (91 $\pm$ 5 µmol NADPH mg^{-1} min^{-1}) compared with wild-type PoptrSDH (103 $\pm$13 µmol mg^{-1} min^{-1}). The K_M also appeared to be relatively unaffected. Notably, this mutant had no detectable activity with quinate. The Ser275Gly change is thus not sufficient to enable gain of quinate activity, consistent with the enzymatic properties of non-seed plant SDHs that also contain a Gly at this position but lack detectable activity with quinate. The Thr318Gly mutant yielded only a very small amount of enzyme, and in consequence the relative amounts of co-purified proteins from *E. coli* are high (Figure 3.3). This could reflect a destabilizing effect of the mutation on the active site of the protein or its overall 3D structure. In support of this conjecture, none of the sequences analysed here have only the Thr to Gly without the Ser to Gly substitution. Despite low yields, we were able to measure the activities of the Thr318Gly mutant at comparably high substrate concentrations (0.6– 5 mM) of shikimate and quinate. At these

concentrations, Thr318Gly displayed low but clearly detectable activities with both shikimate and quinate (Figure 3.3b,c). In contrast, the Ser275Gly/Thr318Gly double mutant was well expressed in *E. coli* and showed bona fide QDH activity besides its original SDH activity, which was severely reduced. Although the Ser275Gly/Thr318Gly double mutant is clearly sufficient to confer gain of activity with quinate, its activity was lower than the QDH activities of *Pinta*QDH and *Poptr*QDH activity. This, and the very high K_M value for quinate (2351 $\pm$ 1468 µM), suggests that other mutations were probably required to refine conversion of SDH, highly optimized for shikimate biosynthesis, to QDH.

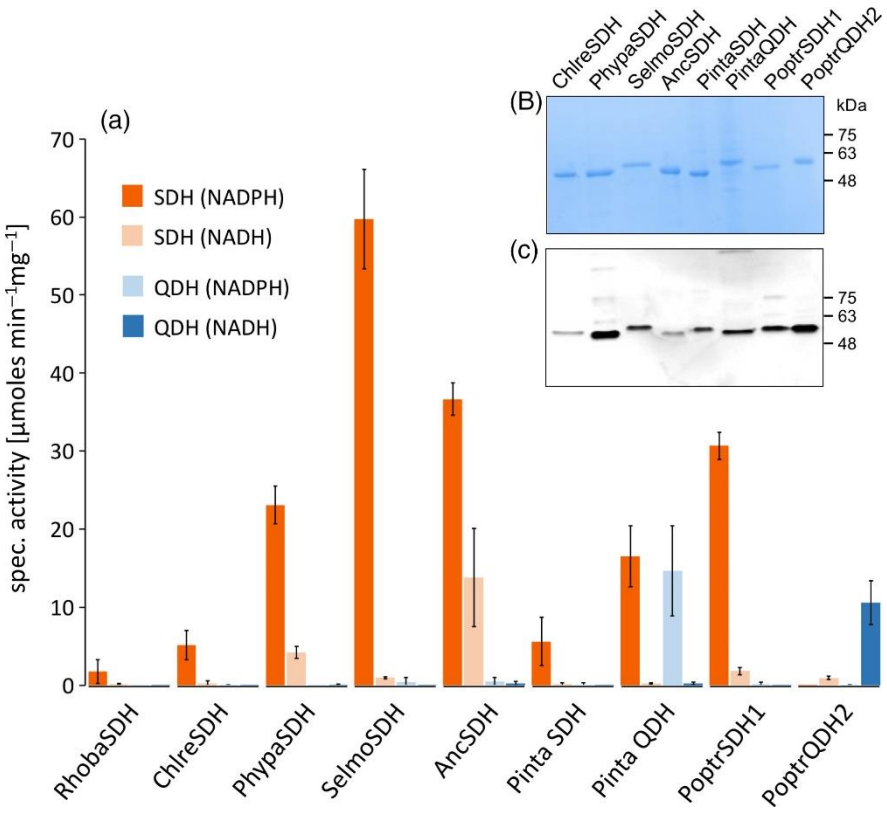


Figure 3.2: Enzyme activities with shikimate and quinate Histidine-tag purified proteins were separated by SDS PAGE and stained with GelCode Blue (b). A separate gel was blotted and probed with a West His Probe™ Kit to detect His-tagged proteins (c). Enzyme activities were monitored spectrophotometrically measuring NADPH or NADH production in the presence of either shikimate or quinate as substrate to determine shikimate dehydrogenase (SDH) or quinate dehydrogenase (QDH) activity, respectively (a). Enzyme activity was normalized to the protein amount used. Shown is the mean of three replicates (independent protein purifications); error bars denote standard deviation. Species abbreviations: *Pinus taeda* (Pinta), *Selaginella moellendorfii* (Selmo), *Physcomitrella patens* (Phypa), *Chlamydomonas reinhardtii* (Chlre), *Rhodopirellula baltica* (Rhoba). Anc122 represents the pre-duplication ancestor sequence reconstructed from the phylogeny. SDH and QDH from *Populus trichocarpa* (Poptr) (Guo *et al.*, 2014) were included as controls for comparison.

Table 3.1: Enzymatic properties based on Michaelis–Menten kinetic analysis

Enzyme	Substrate	Cofactor	V_{MAX}[a] (µmol min^{-1} mg^{-1})	K_M[a](µM)	V_{MAX}/K_M
RhobaSDH	Shikimate	NADP+	5.1 ± 0.3	101 ± 24	n.a
ChlreSDH	Shikimate	NADP+	4.8 ± 0.2	120 ± 24	n.a
PhypaSDH	Shikimate	NADP+	51.8 ± 2.9	239 ± 47	n.a
SelmoSDH	Shikimate	NADP+	36.2 ± 2.0	79 ± 51	n.a
Anc122SDH	Shikimate	NADP+	90.5 ± 4.2	43 ± 40	n.a
PintaSDH	Shikimate	NADP+	32.4 ± 2.6	218 ± 63	n.a
PintaQDH	Shikimate	NADP+	12.8 ± 1.0	820 ± 162	15.6
PintaQDH	Quinate	NADP+	17.7 ± 2.3	677 ± 229	25.1
PoptrSDH	Shikimate	NADP+	101.9 ± 3.8	168.2 ± 24.8	n.a
PoptrQDH	Shikimate	NAD+	1.1 ± 0.0	404.0 ± 32.8	2.7
PoptrQDH	Quinate	NAD+	27.7 ± 2.0	545.8 ± 108.0	50.8

[a] Based on three replicate experiments from independent protein purifications. For each replicate 10 substrate concentrations ranging from 0.05 to 5 mM were used. Kinetic constants were modelled using non-linear regression to the Michaelis–Menten equation; standard errors are provided.

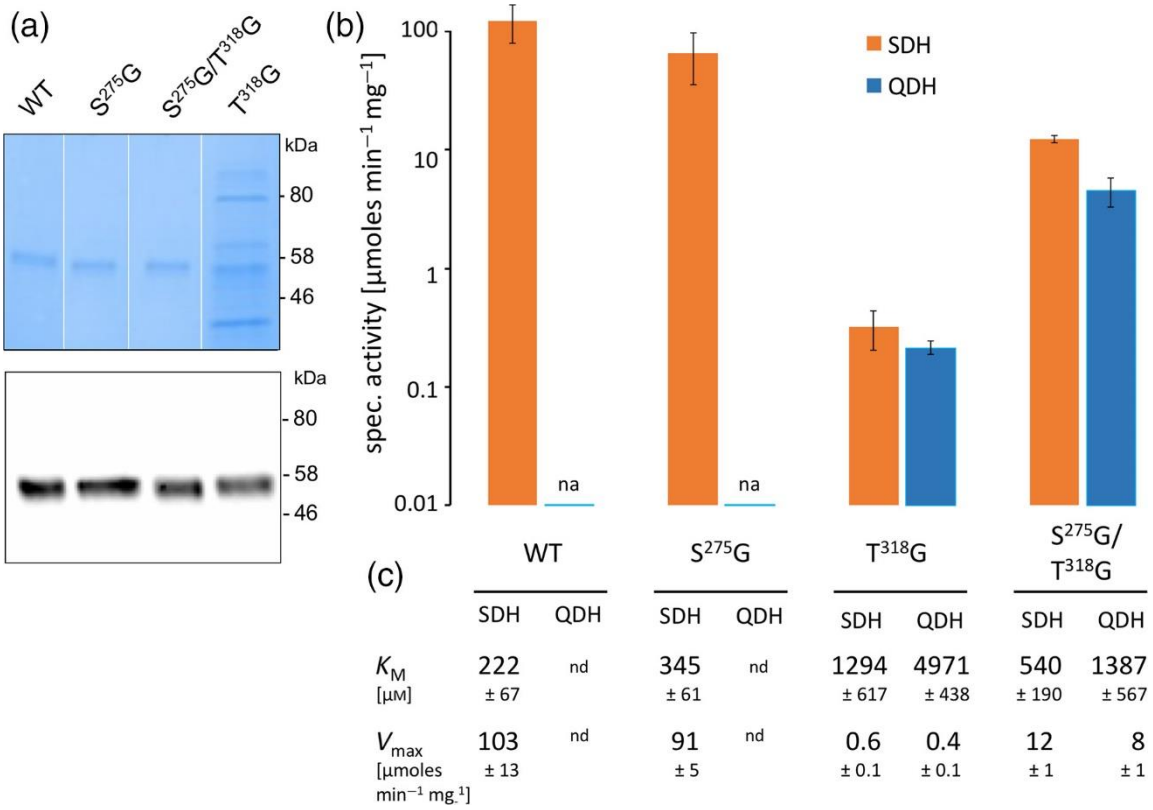


Figure 3.3: Activities of mutant Populus trichocarpa shikimate dehydrogenase (*Poptr*SDH) with shikimate and quinate. Amino acid changes at position 275 (Ser to Gly) or 318 (Thr to Gly) or both were introduced through site-directed mutagenesis into the PoptrSDH protein. Affinity-purified recombinant His_6-tagged enzymes were separated by polyacrylamide gel electrophoresis and stained with GelCode Blue (a, top; empty lanes removed from image) or blotted and probed with a West His Probe™ Kit (a, bottom). Each enzyme was incubated with shikimate or quinate in the presence of $NADP^+$ to determine SDH or quinate dehydrogenase (QDH) activity, respectively. Specific activities at a substrate concentration of 5 mM are shown as bar graphs (b). Kinetic constants were determined from three replicate purifications using at least nine substrate concentrations ranging from 0.05 to 5 mM (except for T318G, where substrate

concentrations ranged from 0.6 to 5 mM) (c). Kinetic properties of mutant enzymes in comparison to wild type (WT) *Poptr*SDH enzyme; na, no activity detectable; nd, activity too low to be determined

3.4 Discussion

3.4.1 Summary

Gene duplication has been well-accepted as a driver of functional novelty; by contributing raw tinkering materials for natural selection it has contributed to a dizzying array of specialized metabolic pathways in plants. Selectional constraints on one copy of a gene are removed, opening a window of mutational opportunities (Kroymann, 2011; Weng, 2014; Wink, 2003). However, the genetic switches leading to the functional diversification of duplicated genes is relatively unknown. Several hypotheses have been developed to explain such mechanisms although they lack empirical evidence. Here comparative analysis of the enzyme activities encoded by *S/QDH* members from major plant taxa and a reconstructed ancestor was combined with mutagenic data to shed light on this topic. From these results it appears diversification of S/QDH proteins is best described by a combination of the EAC, IAD and neofunctionalization models.

3.4.2 S/QDH evolution—an overview

One way of distinguishing between evolutionary models is based on the type(s) of selection pressures they involve. In neofunctionalization, for instance, one gene copy maintains ancestral functions under purifying (negative) selection while the other copy, liberated from fitness constraints, may functionally diversify under positive selection (Matsuno, *et al*., 2009; Moore and Purugganan, 2005; Zhang, 2003). A tenet of the EAC model on the other hand, is that both gene copies evolve under positive selection to optimize activities encoded by a multifunctional parent gene (Deng *et al*., 2010; Des Marais and Rausher, 2008; Sikosek *et al*., 2012). Previously it was found that the substitution of two active site residues, namely a Ser and Thr required for binding shikimate (Singh and Christendat, 2007), were positively selected for in members of the QDH clade of seed plants (Carrington *et al*., 2018). The replacement of Ser to Gly appears to have been selected for early during the evolution of land plants based on the presence of this substitution in the S/QDH sequences of bryophytes and lycopods. However, the Thr to Gly substitution is

exclusive to the QDH clade members of seed plants (which also kept the Ser to Gly substitution). Positive selection was also detected in the branch subtending the green lineage and (unexpectedly) the SDH clade of seed plants. The former is suggestive of the optimization of SDH function in plants, which is surprising since, as an essential primary metabolic enzyme, its functions were presumed to have been maintained by negative selection. Steps 3 and 4 of the bacterial shikimate pathway, namely the dehydration of 3-dehydroquinate to 3-dehydroshikimate and reduction of 3-dehydroshikimate to shikimate, are performed by separate enzymes, *AroD* and *Aro*E respectively. These were fused in an ancestral prokaryotic genome, presumably related to Planctomycetes, represented here by *R. baltica*, and the resulting *Aro*DE fusion gene was likely passed to plants by horizontal gene transfer (HGT) (Richards *et al*., 2006). The *R. baltica* enzyme clearly confers SDH activity and is devoid of detectable QDH activity. Positive selection may have driven the incorporation of *Aro*DE into the early streptophyte genome, where it played an essential role in amino acid biosynthesis (Degnan, 2014) and replaced the original enzyme, which should have come vertically from its ancestor. Most branches across the phylogeny showed signatures of strong purifying selection, indicating the need to maintain the shikimate pathway in plants. The previous detection of positive selection in the branch subtending the SDH clade was puzzling to see since SDH represents the ancestral function in this model. In one of these clades, namely the angiosperm SDH clade, the Ser to Gly substitution was reversed, possibly because shikimate binding is more optimal when the ancestral Ser residue is present. This is evident by the fact angiosperm SDH (represented here by *Poptr*SDH) displayed higher SDH activity (indicated by a ~6 fold greater V_{max}) than gymnosperm SDH (*Pinta*SDH), which maintained the Ser to Gly substitution. Nevertheless, based on a checklist of (selectional) requirements, the observation that positive selection acted on both gene duplicates of seed plants is congruent with the EAC model.

3.4.3. Genetic and biochemical changes from SDH to QDH

Mutagenesis of Ser275 (homologous to *Arabidopsis* Ser338) to Gly in a PoptrSDH protein did not lead to a detectable increase in affinity for quinate but instead caused a (slight) reduction in V_{MAX} (and raised K_M) for shikimate. It is therefore surprising to see that this change occurred under positive selection (as suggested by the branch site model) before reverting back to Ser in the angiosperm SDH clade. Unfortunately, the T381G single mutant yielded poor protein yields in the recombinant *E. coli* system employed. Nevertheless, the datum suggests that this mutation

alone may have sown the seeds for QDH activity. Supporting this idea, in a similar study published in parallel, the introduction of the Thr381Gly mutation into wildtype *Arabidiopsis* SDH increased quinate biosynthetic activity *in vitro* from 0 to 8.8±0.7 s^{-1} (Gritsunov *et al*., 2018). It is interesting to note that while the Ser to Gly substitution is present in extant S/QDH sequences (of bryophytes and lycopods for example) none of the analyzed sequences contained the Thr to Gly substitution alone without the Ser to Gly substitution. Given its early occurrence during the evolution of land plants preceding the Thr to Gly substitution, the latter may have helped stabilize the active site, enabling further changes that would optimize binding to quinate. Based on the mutagenesis results, it appears that both mutations are necessary for clear-cut activity with quinate, which was observed for the *Poptr*SDH double mutant. However, this gain in QDH activity came at the expense of notably reduced SDH activity (as indicated by a ~10 fold reduction in V_{MAX} for shikimate). It is worth mentioning that although the mutant displayed QDH activity, its high K_M for quinate likely means further cycles of mutation and natural selection were required for optimizing this activity as observed in extant QDH proteins. Still it is extraordinary to think that merely flipping a couple of amino acid switches in a superefficient molecular catalyst that has been evolving for millennia were sufficient to rewire its activities. Similar conclusions were obtained in a study on the diversification of plant SABATH (salicylic acid/benzoic acid/theobromine) gene family members. Following initial duplication of an ancestral protein presumed to have high preference for benzoic acid, the substitution of a single active site residue (His201Met) caused a functional shift from benzoic acid to salicylic acid methyltransferase activity. In some of its descending lineages, this substitution was reversed such that ancestral activities with benzoic acid were regained and even increased, in a case resembling the evolution of angiosperm SDH enzymes (Huang *et al*., 2012). Proteins therefore demonstrate remarkable pliable capacity reminiscent of Play-Doh® rather than concrete structures, amenable to mutations (Khersonsky *et al*., 2006; Weng and Noel, 2012). Of course, mutations must fall within the fitness landscapes of a protein so that its solubility and activity are not jeopardized (Trudeau and Tawfik, 2019). In *Arabidopsis*, Lys385 and Asp423 have been identified as the major catalytic residues of the SDH active site (Singh and Christendat, 2006). These are highly conserved across all members of the S/QDH protein family, attesting to their functional relevance in mediating dehydrogenase catalytic activity.

3.4.4 The perplexity of poplar and pine QDH proteins

The gain in QDH activity and concomitant loss of SDH activity observed in the *Poptr*SDH double mutant were also reflected by the gymnosperm QDH protein, *Pinta*QDH. *Poptr*QDH showed higher preference for quinate than shikimate (Guo *et al*., 2014) whereas *Pinta*QDH showed roughly equal specificity towards both compounds. This is in consistency with work by Ossipov *et al*. (2000) who characterized a broad-specificity SDH from *P. taeda* utilizing both shikimate and quinate as substrates. The dual functioning *Pinta*QDH analyzed here likely corresponds to this isozyme purified from pine needle tissues. Notably, the observed trade-offs between SDH and QDH activity in the double mutant and *Pinta*QDH conforms to predictions made by the EAC model of evolution (Deng *et al*., 2010; Des Marais and Rausher, 2008; Sikosek *et al*., 2012). Active site modeling of poplar SDH and QDH enzymes suggests that the Ser to Gly and Thr to Gly substitutions helped widen the substrate binding pocket, accommodating binding to quinate (Guo *et al*., 2014). On the flip side, a widening of the active site would concomitantly decrease activity with shikimate since a widening of the substrate binding pocket induced by the T381G and S338G mutations is expected to decrease structural complementarity of the enzyme relative to its substrate or transition state (Robinson, 2015; Petsko and Ringe, 2004). Both *Pinta*SDH and *Pinta*QDH preferred $NADP^+$ over NAD^+ as a coenzyme. Similar trends have been obtained for well-characterized angiosperm SDH's preferring $NADP^+$ over NAD^+, fitting their description as NADPH-dependent dehydrogenases (Ding *et al*., 2007; Muir *et al*., 2011; Singh and Christendat, 2006). Like other studied angiosperm QDH's (Kang and Scheibe, 1993; Minamikawa, 1977; Refeno *et al*., 1982), *Poptr*QDH preferred NAD^+ over $NADP^+$. As discussed in the following chapter, coenzyme specificity may help separate regulation of metabolic pathways (Cahn *et al*., 2017)—in particular, shikimate and quinate metabolism in flowering plants (Gritsunov *et al*., 2018). However, while $NADP^+$ and NAD^+ have been associated with anabolic and catabolic reactions respectively, this rule is not absolute and plant enzymes exhibiting equal preferences for both pyridine nucleotides (e.g. aldehyde dehydrogenases) have been reported (Perozich *et al*., 2000). Nevertheless, *Pinta*QDH appears to represent an intermediate that has the ability to act both on shikimate and quinate but still requires $NADP^+$ as a coenzyme.

3.4.5 Evolutionary differences between flowering plants and gymnosperms

Analyses of orthologous genes between angiosperms and gymnosperms have shed light on the molecular forces carving their evolutionary paths which may provide possible reasons for characteristic differences between *Poptr*QDH and *Pinta*QDH enzymes. These studies point to slower evolutionary rates in gymnosperms compared to angiosperms based on their lower rates of synonymous (dS) and nonsynonymous (dN) substitutions. It is therefore possible that within a given amount of time, poplar QDH enzymes experienced a greater number of mutations that altered their functionality relative to pine QDH enzymes. However, the idea that gymnosperm evolution lags behind their sister clade is only speculation which is complicated by the surprising revelation that they have experienced greater rounds of positive selection. Thus while gymnosperms evolve slower *i.e.* have lower substitution rates, a greater proportion of their substitutions are adaptive compared to angiosperms as indicated by higher dN/dS ratios (Buschiazzo *et al*., 2012; De la Torre *et al*., 2017). Why these sister lineages evolve at different rates is a knotty puzzle to solve as the latter is influenced by a string of intermingling genetic and ecological factors (Augusto *et al*., 2014; Buschiazzo *et al*., 2012; De La Torre *et al*., 2017). Substitution rates are dependent on genes and genomes (nuclear, mitochondrial or chloroplastic). Typically genes whose products are involved in mediating responses towards abiotic and biotic stresses (e.g. kinases, phosphatases and transcription factors) and that exist in multiple copies are likely to become fixed by positive selection (have high dN/dS ratios) (Buschiazzo *et al*., 2012; De La Torre *et al*., 2017). Another general trend for seed plants is that substitution rates are lower in mitochondrial than in chloroplastic and nuclear genomes, which is the opposite of what is observed in mammals (Buschiazzo *et al*., 2012; Guo *et al*., 2016; Palmer *et al*., 2000). External selection pressures include biotic and abiotic stresses that drive adaptive changes giving plants an edge for survival. Given that forces operating across genes are population-specific and vary over time, the differences between those acting on angiosperms and gymnosperms whose lineages diverged ~300 MYA are likely grand and consequently hard to boil down *(*Augusto *et al*., 2014).

Exactly why *Poptr*QDH and *Pinta*QDH enzymes differ in their activities is unknown but the aforementioned points nevertheless make for an interesting discussion. As mentioned above, differences between *Poptr*QDH and *Pinta*QDH enzymes may reflect differences in their

substitution rates or divergence patterns or both. In general, substitution rates are higher in fast-growing than slow-growing species. As outlined in the "generation time hypothesis" this is because the probability of accumulating mutations increases with increasing frequency of DNA replication and cell division. Thus, substitution rates are inversely proportional to generation time. In support of this theory, substitution rates are observably higher in the small herbaceous annual *Arabidopsis* than in woody perennial poplar trees (Buschiazzo *et al*., 2012). Extant angiosperms on average grow faster than long-lived gymnosperms under optimal conditions (Augusto *et al*., 2014; De La Torre *et al*., 2017). Faster growth rates have been a driving theme during the evolution of angiosperms, whose fossil record documents changes in water conducting (vascular) tissues and leaf surface area to name a few that have translated to increased photosynthetic capacity and faster growth. In contrast, the photosynthetic capacities of most gymnosperms have remained relatively unchanged over time (Augusto *et al*., 2014), and today they are lower than those reported for angiosperm trees (Becker, 2000). Nutrient loading and unloading is also slower in gymnosperms, whose narrow sieve tubes (relative to angiosperms) generate greater hydraulic resistance and lower phloem translocation (Liesche *et al*., 2015). As mentioned above, it is possible that *Pinta*QDH is simply evolving slower than *Poptr*QDH. If this is true, then it is expected that over time it will develop greater specificity towards quinate and even NAD^+. Most interestingly and in support of this idea, a QDH from *Pinus sylvestris* has been shown to use both $NADP^+$ and NAD^+ albeit having greater activity with the latter (Osipov and Shein, 1987).

It should be pointed out that seed plants differ on a number of aspects apart from their growth rates including carbon allocation strategies, mating systems and genome structures (Augusto *et al*., 2014; De la Torre *et al*., 2017) and it is erroneous to conclude that angiosperms are simply evolving faster than gymnosperms. Divergent evolutionary patterns of angiosperms and gymnosperms may reflect differences in their environmental selection pressures in combination with variations between their molecular evolutionary rates as described above. Gymnosperms are generally observed to dwell in low-nutrient and low-temperature habitats (Becker, 2000). A common example that comes to mind are cold-adapted boreal forests withstanding biting temperatures as low as -70˚C (Roden *et al*., 2009). This trend may have been set during the late Cretaceous Period when gymnosperms were driven to sub-optimal locations (e.g. high altitudes and low temperatures) by fast-growing and outcompeting angiosperms (Becker, 2000; Crisp and Cook,

2011). If angiosperms and gymnosperms have evolved under different sets of selection pressures as a result of living in different environments, then it is safe to expect that they have developed their own independent chemical and physiological adaptations towards them. It is well known that chlorogenic acid, a major derivative of quinic acid mitigates biotic and abiotic stresses in angiosperms. For example, increased levels of chlorogenic acid have been observed in *Mahonia repens* (Grace *et al*., 1998) and *Solanum lycopersicum* (Clé *et al*., 2008) in response to sunlight suggesting its roles as a UV protectant. Higher levels of chlorogenic acid in thrip-resistant, as opposed to susceptible chrysanthemums, point to its functions as an insect deterrent in these plants *(*Leiss *et al*., 2009): and chlorogenic acid has also been shown to act as a free radical scavenger (Niggeweg *et al*., 2004). Less is known about chlorogenic acid in gymnosperms than in angiosperms. Chlorogenic acid has been observed in *Podocarpus*, *Gnetum* and Douglas-fir although its levels are reportedly low in mature (10 year old) Douglas-fir trees (Radwan, 1972, 1975). A surprising finding that is in stark contrast to its suggested protective functions in angiosperms is that there is a positive correlation between chlorogenic acid and susceptibility of Douglas fir to deer browsing (Radwan, 1975). As for its precursor, radiolabeling studies have revealed quinic acid accumulates in young conifer needles before being converted to shikimic acid. These observations point to a role of quinate not in defense but rather as a carbon reserve for synthesizing shikimate and related aromatic compounds (Boudet, 2012). If excess shikimate and its storage form are required by gymnosperms then this may possibly explain why *Pinta*QDH dually acts on shikimate and quinate. Of course, why gymnosperms would require a surplus of shikimate and what purposes chlorogenic acid serves in these plants introduces yet another can of worms that will not be opened here.

3.4.6 Conclusions

The absence of QDH activity in SDH enzymes of extant non-seed plants and the hypothetical seed plant ancestor (Anc122) conforms to the neofunctionalization model of gene evolution, where one gene copy stays optimized for ancestral function(s) and the other functionally diversifies under positive selection (Zhang, 2003; Moore and Purugganan, 2005; Matsuno, *et al*., 2009). This observation fits well with the notion that metabolic pathways became progressively expanded during the evolution of land plants (Ober, 2005; Pichersky and Gang, 2000; Jing-Ke Weng, 2014).

However, the data does not perfectly conform to its criteria since ancestral activities were found to be under positive, rather than purifying selection, in the seed plant SDH clade. Interestingly, and unlike poplar enzymes which show a clear split in preferences for either shikimate or quinate, *Pinta*QDH accepts both as substrates. Its lower activities with respect to the poplar enzymes point to functional trade-offs between the two sub-functions, which may have also been experienced by young, emerging QDH genes of angiosperms on their evolutionary path provided that gymnosperms are slower evolving than angiosperms (Buschiazzo *et al*., 2012; De la Torre *et al*., 2017). Finally, but unlikely to be of any physiological relevance, the single-copy SDH from *S. moellendorffii* and the resurrected ancestor showed minute QDH activities very close to the detection limit. Promiscuous activities are common to all enzymes, albeit to varying extents (Copley, 2015; Tawfik, 2010; Weng and Noel, 2012). These may occur, for example, if the native and non-native substrates are structurally similar, as in the case of shikimate and quinate (Guo *et al*., 2014) and benzoic acid and salicylic acid (Huang *et al*., 2012) which differ by a single hydroxyl group. Larger, non-native molecules may be permitted to loosely fit in the active site by partially jutting out—at least until mutations make the substrate-binding pocket more accommodating (Copley, 2015). Non-specific activities towards shikimate-like molecules may have provided a take-off point for the evolution of QDH genes via neofunctionalization. To sum up, diversification of the *S/QDH* gene family is at best described by a combination of predictions made by neofunctionalization and co-option type models. Such mixed-model hypotheses of evolution may extend to describing functional diversification that also has shaped other gene families (Deng *et al*., 2010).

Chapter 4

4.1 Introduction

Proteins execute diverse biological processes including gene expression, transport, and metabolism as well as provide structural support to cells. Indeed, their functional diversity and abundance has earned them a reputation as cellular "work horses" (Bloom and Arnold, 2009). Most cell proteins are enzymes catalyzing diverse biochemical reactions (Lodish *et al*., 2000) on a wide variety of substrates [from glucose to plastics (Palm *et al*., 2019)]. Despite the diversity of reactions they participate in, enzymes show recurrent structures and functions as shown here, in the story of plant S/QDH enzymes. Phylogenetically related enzymes catalyze similar types of reactions and homology between them is inferred by similarities in active site architecture and residues (Galperin and Koonin, 2012; Martínez Cuesta *et al*., 2015; Ribeiro *et al*., 2020). These point to a common theme about enzyme evolvability: de novo catalytic functions are rarely observed in nature. This is evident through the sparse number of functional categories used to describe enzymes: transferases, hydrolases, lyases, isomerases, ligases and oxidoreductases (Lesk, 2010; Verpoorte *et al*., 2000). The relatively few types of reactions catalyzed by enzymes can be explained at least in part by the fact that protein architecture is restricted to 20 amino acids each of which show differing levels of chemical reactivity. This seemingly creates a barrier towards the development of new metabolic pathways. However, one solution to this problem is the use of organic and inorganic cofactors to aid in enzyme catalysis (Begley, 2006; Broderick, 2001; Ribeiro *et al*., 2020). Returning to the specific example of S/QDH proteins, these and many other oxidoreductases catalyze cellular oxidations and reductions using the nicotinamide coenzymes NAD(H) and NADP(H) (Figure 4.1) to transfer electrons between species (Begley, 2006; Broderick, 2001; Chánique and Parra, 2018). Similar to the catalytic sites of homologous enzymes, the dinucleotide-binding domains of nicotinamide-dependent dehydrogenases and related enzymes, are highly conserved (Laurino *et al*., 2016).

4.1.2 NAD(P)-binding domains

The core structure of this domain (Rossmann fold) consists of repeated βαβ folds; the first of which is involved in interacting with adenosine monophosphate (AMN) of nucleotide-derived coenzymes (Hanukoglu, 2015). Such interaction is mediated by a glycerine rich loop contained within this

initial βαβ turn that has the consensus sequence GX(X)GXXG(or A) (where X is any amino acid) (Chánique and Parra, 2018; Kallberg and Persson, 2006; Lesk, 1995). The actual amino acid sequence hints at the coenzyme preference [either NAD(H) or NADP(H)] of nicotinamide-dependent dehydrogenases: however, this is subject to exceptions (Baker *et al*., 1992; Lauvergeat *et al*., 1995; Lunzer *et al*., 2005). Coenzyme specificity is often defined by the nature of the amino acid residue towards the C-terminal region of this motif (Chánique and Parra, 2018; Kallberg and Persson, 2006; Lesk, 1995). An Asp residue [Asp223 in horse liver alcohol dehydrogenase (ADH), which was one of the first of this family to have its structure elucidated (Chen *et al*., 1991; Lesk, 1995; Rosell *et al*., 2003)] forms hydrogen bonds with the ribosyl moiety of NAD(H). In contrast, NADP(H)-binding enzymes have a characteristic basic residue, often Arg (Chánique and Parra, 2018; Chen *et al*., 1991; Kallberg and Persson, 2006) or ones with small, uncharged side chains (e.g. Ser, Ala or Gly) (Lauvergeat *et al*., 1995; Rosell *et al*., 2003). The latter are likely more favourable for fitting the 2'-phosphate group of NADP(H) (Chánique and Parra, 2018; Chen *et al*., 1991; Rosell *et al*. 2003) (Figure 4.1).

The elucidation of the crystal structure of *Arabidopsis* SDH (Singh and Christendat, 2006; Singh and Christendat, 2007) was the gateway towards understanding its catalytic and cofactor-binding mechanisms. The NADP(H)-binding domain of this enzyme includes residues 484 to 486, (Asn-Arg-Thr or NRT) residing near the C-terminal end of the canonical GX(X)GXXG sequence (residues 460 to 465) (Gritsunov *et al*., 2018; Singh and Christendat, 2007). As mentioned above, the presence of either an Arg or Asp commonly dictates specificity for either NADP(H) or NAD(H) respectively (Chánique and Parra, 2018; Chen *et al*., 1991; Kallberg and Persson, 2006). In *Arabidopsis*, the NRT trio interacts with the 2' phosphate group of the AMP moiety of NADP(H). Coenzyme binding also requires help from neighbouring residues, namely Met523 and Gly462 and 463 that interact with the purine and pyrophosphate groups of NADP(H) respectively (Singh and Christendat, 2007). Like *Arabidopsis* SDH (Singh and Christendat, 2006; Singh and Christendat, 2007), *P.trichocarpa* SDH (*Poptr*SDH) (Carrington *et al*., 2018; Guo *et al*., 2014) and those described elsewhere are active with NADP(H) (e.g. Bischoff *et al*., 2001; Ding *et al*., 2007; Ossipov *et al*., 2000), adhering to their role as NADP(H)-dependent dehydrogenases and this also holds true for non-seed plant orthologues (Carrington *et al*., 2018).

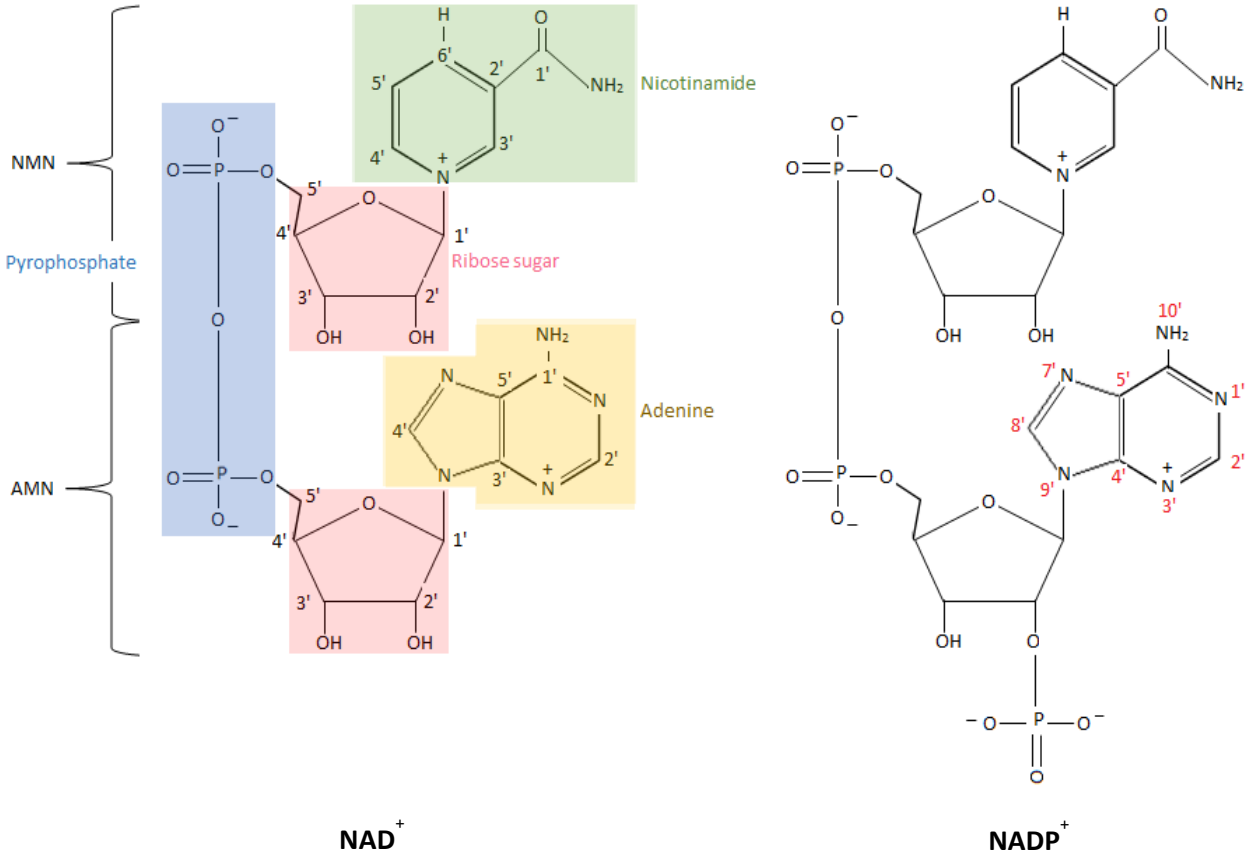


Figure 4.1: Simplified structures of nicotinamide adenine dinucleotide (NAD+) and nicotinamide adenine dinucleotide phosphate (NADP+). Each cofactor consists of two nucleotides, an adenosine monophosphate (AMN) and nicotinamide mononucleotide (NMN) joined by a phosphate bridge. The only difference between them is the presence of an extra phosphate group at the 2' position of the ribosyl moiety of the adenosine mononucleotide of $NADP^+$. For ease of visualization, the different chemical groups are only shown on NAD^+ although they also apply to $NADP^+$. For the same reason, carbon atoms are labeled on NAD^+ while $NADP^+$ is used to show labeling of the adenine ring. In both NAD^+ and $NADP^+$, the nicotinamide ring participates in electron transfer. Acceptance of a hydrogen ion (H^+) leads to the production of NADH and NADPH respectively while loss of H^+ from NADH and NADPH regenerates their oxidized forms.

In contrast, the NRT motif is replaced by Asp-Ile-Asp (DID) at its equivalent position in angiosperm QDHs (Figure 4.2) (Gritsunov *et al*., 2018). As shown in the previous results, QDH from *P.trichocarpa* (*Poptr*QDH) primarily accepts QDH and NAD^+ as substrate (also described in Carrington *et al*., 2018; Guo *et al*., 2014) which is in consistency with other reports on plant QDHs (e.g. Kang and Scheibe, 1993; Refeno *et al*., 1982; Gritsunov *et al*., 2018). The replacement of Arg by Ile, and the gain of two negatively charged residues within the nucleotide-binding site may have favoured interactions with NAD(H) while simultaneously increasing electrostatic repulsion against the 2'-phosphate group of NADP(H) (Chánique and Parra, 2018; Kallberg and Persson, 2006; Lesk, 1995). Surprisingly, angiosperm sequences that are active with quinate and NADP(H) have been found even though they phylogenetically cluster with shikimate and NADP(H)-specific SDH enzymes. These include SDH isoforms in *Brassica napus* and *B. rapa* that, unlike other characterized flowering plant NAD(H)-binding QDH proteins, have maintained the NRT motif and NADP(H) specificity (Gritsunov *et al*., 2018). This observation led Gritsunov *et al*. (2018) to conclude that the NRT versus DID pattern is predictive of NADP(H) and NAD(H) specificity.

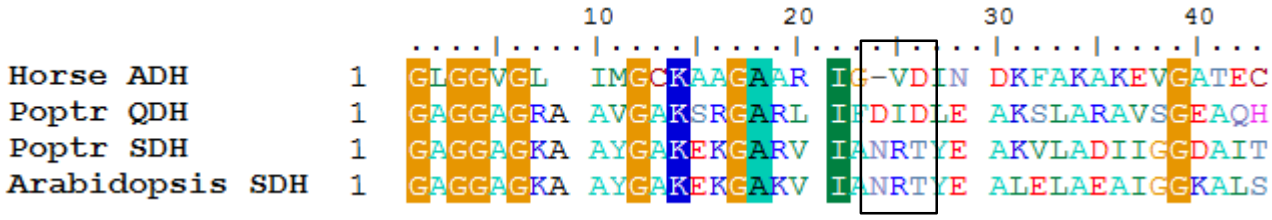


Figure 4.2: Multiple sequence alignment of conserved residues in the Rossmann folds. This alignment uses sequences from biochemically characterized dehydrogenases: horse liver alcohol dehydrogenases (ADH), *P*. trichocarpa quinate dehydrogenase (QDH), *P. trichocarpa* shikimate dehydrogenase (SDH) and *Arabidopsis thaliana* SDH. The DID/NRT motif discerning NAD(H) or NADP(H) specificity respectively is boxed.

4.2.2 Experimental objectives: characterization of cofactor preferences among S/QDH enzymes across taxonomic representatives of green plants

Extending on this idea and the previous hypothesis of chapter one, the purpose of the current experiment is to determine how cofactor specificity evolved among plant S/QDH proteins. Specifically, it was asked if NAD(H)-preference is a novel characteristic of angiosperm QDH proteins or if ancestral SDH proteins encoded minute NAD(H) binding activity that was later augmented in some of its descending lineages. Based on their description as NADP(H)- dependent SDH proteins, it is predicted that the single-copy *SDH* genes of from *C. reinhardtii* (green algae), *P. patens* (moss), *S. moellendorffii* (lycopod) encode an Arg or small-side chain amino acid downstream of the GXGXXG motif that discriminates against NAD(H). However, while the primary sequence of a protein encodes important information about its function, it is not an absolute predictor. As an example, NAD^+-dependent isocitrate dehydrogenase from *E.coli* has Arg near the GXGXXG fingerprint sequence even though Arg is a hallmark of most NADP(H)-dependent enzymes (Lunzer *et al*., 2005). In contrast, glyceraldehyde-3-phosphate dehydrogenase from maize chloroplasts has the C-terminal Asp but is dually specific for NAD^+ and $NADP^+$ (Baker *et al*., 1992). The power of Asp (or Arg) to define cofactor specificity is not definitive. It should be pointed out, Gritsunov *et al* (2018)'s conclusions above are based on a rigorous combination of sequence and kinetic data. The latter is important since most enzymes have promiscuous side activities that are nearly impossible to predict using sequence information alone (Colin *et al*., 2015; Tawfik, 2010). The objectives of this experiment were to confirm NADP(H)-specificity of SDH and QDH proteins from seed plants and earlier derived lineages *in vitro* as predicted by their cofactor binding fingerprints. If NAD(H)-binding is observed or not observed among the single copy SDH proteins, then the evolution of cofactor preference can be said to have undergone subfunctionalization and neofunctionalization respectively.

4.2 Methods

See chapter two for a detailed description of the cloning, protein purification and spectrophotometric assay procedures. To determine kinetic properties with NAD^+ and $NADP^+$, either shikimate or quinate was used as substrate at saturating concentrations while variable cofactor concentrations (at least six) were tested (from 0.002 to 2 mM).

4.2.1 Gene mining

The NAD(P)(H) binding regions of SDH and QDH proteins from *R. baltica*, *C. reinhardtii*, *P. patens*, *S. moellenodorffii*, Anc122, *P. taeda* and *P. tricocharpa* were obtained as described as before and aligned using Clustal W in Bioedit v.7.0 (www.mbio.ncsu.edu/BioEdit/bioedit.html)

4.2.2 Protein modelling and in silico mutagenesis

Structural and binding differences between angiosperm and non-flowering plant SDH proteins were analyzed using PyMOL (PyMOL Molecular Graphics System, Version 2.3. Schrödinger, LLC.). The structure of *Arabidopsis* SDH (Protein Data Bank ID: 2O7S) published by Singh and Christendat (2007) was downloaded and used as a template for mutagenesis of Thr485 to Asn485.

Results

4.3.1 Cofactor binding motifs of non-seed plants

The single copy SDH proteins from *R. baltica*, *C. reinhardtii*, *P. patens*, and *S. moellendorffii* encode SRT, NRS, NRN and NRN motifs respectively (Figure 4.3). The latter is also shared by SDH and QDH enzymes from *P. taeda*. The absence of a DID motif point, and the presence of positively charged Arg points to NADP(H)-specificity of these enzymes.

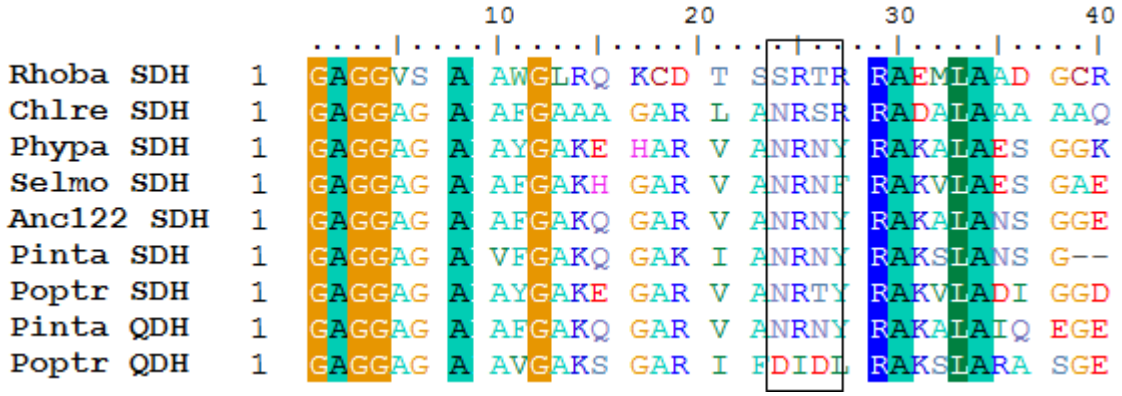


Figure 4.3: NAD(P)(H)-binding domains of proteins used in this study.

4.3.2 Cofactor affinities across the S/QDH family tree

All single copy SDH proteins as well as those from *P. trichocarpa* and *P. taeda* were active with SA and $NADP^+$. Although activities were measurable with NAD^+ and shikimate (except

*Pinta*SDH) and the observed V_{MAX} values were higher than (three standard deviations above the mean of) boiled controls, these proteins showed very low affinity for NAD^+ ($K_M > 1000\mu M$). Only one trial could be performed for *Pinta*SDH and NAD^+. Since this enzyme shows activity with shikimate and $NADP^+$, a lack of activity was not associated with inherent problems with the enzyme. Instead *Pinta*SDH likely has very low specificity towards NAD^+. Based on apparent V_{MAX}/K_M ratios, these enzymes prefer $NADP^+$ over NAD^+ by a factor of 74 to 1780 fold. The gymnosperm QDH, *Pinta*QDH was similarly more efficient using $NADP^+$ than NAD^+ regardless of the substrate used. An exception to this trend was *Poptr*SDH, which could perform equally well with $NADP^+$ and NAD^+ using shikimate as substrate. At the opposite end of the spectrum, *Poptr*QDH was the only tested protein with higher affinity for NAD^+ than $NADP^+$. Unlike *Pinta*QDH, *Poptr*QDH displayed a 100-fold preference for NAD^+ using quinate as a substrate. These results are summarized in Table 4.1.

4.3.3 Prediction of NADP(H)-binding in non-flowering plant SDH proteins

SDH sequences of *Selmo*, *Phypa* and *Pinta*SDH have an NRN motif downstream of the canonical glycine-rich loop instead of NRT that is common to all angiosperm SDH proteins. The side chain of Arg484 interacts with the N7 nitrogen of adenine while the amide group of Asn483 and hydroxyl group of Thr485 interact with the electronegative oxygen atoms of the terminal phosphate group as previously described (Singh and Christendat, 2007). H-bonding also appears to occur between the said oxygen atom and the amine group of Thr485. Interactions between NADP and the third asparagine of the non-flowering plant NRN motif was predicted by simulating mutagenesis of Thr485 to Asn in the *Arabidopsis* SDH. Replacement with Asn485 does not appear to impede interactions with NADP sterically or electrically. Thr485's hydroxyl- and Asn485's amide group are about equal distances from the aforementioned oxygen (2.6 Å and 2.8 Å respectively) (Figure 4.4**)**.

Table 4.1: Kinetic properties of SDH and QDH proteins from taxonomic representatives of green plants and a bacterial outgroup. Activities of recombinant proteins were measured spectrophotometrically for 90 seconds, following the production of NADH or NADPH at 340nm. Substrates (either shikimate or quinate) were held constant at saturating concentrations while different cofactor (either $NADP^+$ or NAD^+) concentrations were tested. From 0.001 to 0.008mg of protein was used for each reaction which was conducted at room temperature and at pH9. For enzymes with very low activities with NAD^+ kinetic parameters had to be determined by extrapolation.

Shikimate as substrate

	NAD^+			$NADP^+$			$NADP^+/NAD^+$
Protein	**V_{MAX} (µM/mg*min)**	**K_M (µM)**	**$V_{MAX/KM}$ (mg^{-1} min^{-1})**	**V_{MAX}**	**K_M**	**$V_{MAX/KM}$**	**Eff.* Δfold**
Rhoba	0.8	1253.0	0.0	0.9	20.9	0.0	74
Chlre	4.0	5435.7	0.0	9.8	24.2	0.4	552
Phypa	108.8	5265.4	0.0	555.1	157.5	3.5	171
Selmo	56.8	39470.7	0.0	61.5	19.7	3.1	2169
Anc122	3349.1	23239.7	0.1	332.7	25.7	13.0	90
Pinta(SDH)	9.4	2348.2	0.0	122.1	36.7	3.3	833
Pinta(QDH)	7.6	3620.9	0.0	116.6	31.3	3.7	1779
Poptr(SDH)	64.7	25.1	2.6	33.3	12.9	2.6	1
Poptr(QDH)	1.0	266.9	0.0	n.d.	n.d.	n.d.	n.d.

Quinate as substrate

	NAD^+			$NADP^+$			$NADP^+/NAD^+$
Protein	**V_{MAX} (µM/mg*min)**	**K_M (µM)**	**$V_{MAX/KM}$ (mg^{-1} min^{-1})**	**V_{MAX}**	**K_M**	**$V_{MAX/KM}$**	**Eff.* Δfold**
Rhoba	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Chlre	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Phypa	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Selmo	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Anc122	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Pinta(SDH)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Pinta(QDH)	62.9	11186.7	0.0	157.8	54.7	2.9	513
Poptr(SDH)	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Poptr(QDH)	7.0	30.4	0.2	2.8	1300.3	0.0	0

Catalytic efficiencies of *Poptr* and *Pinta*QDH

	Efficiency (NAD^+ SA/NAD^+ QA)	**Efficiency ($NADP^+$ SA/$NADP^+$ QA)**
Pinta(QDH)	n.d.	1.3
Poptr(QDH)	0.2	n.d.

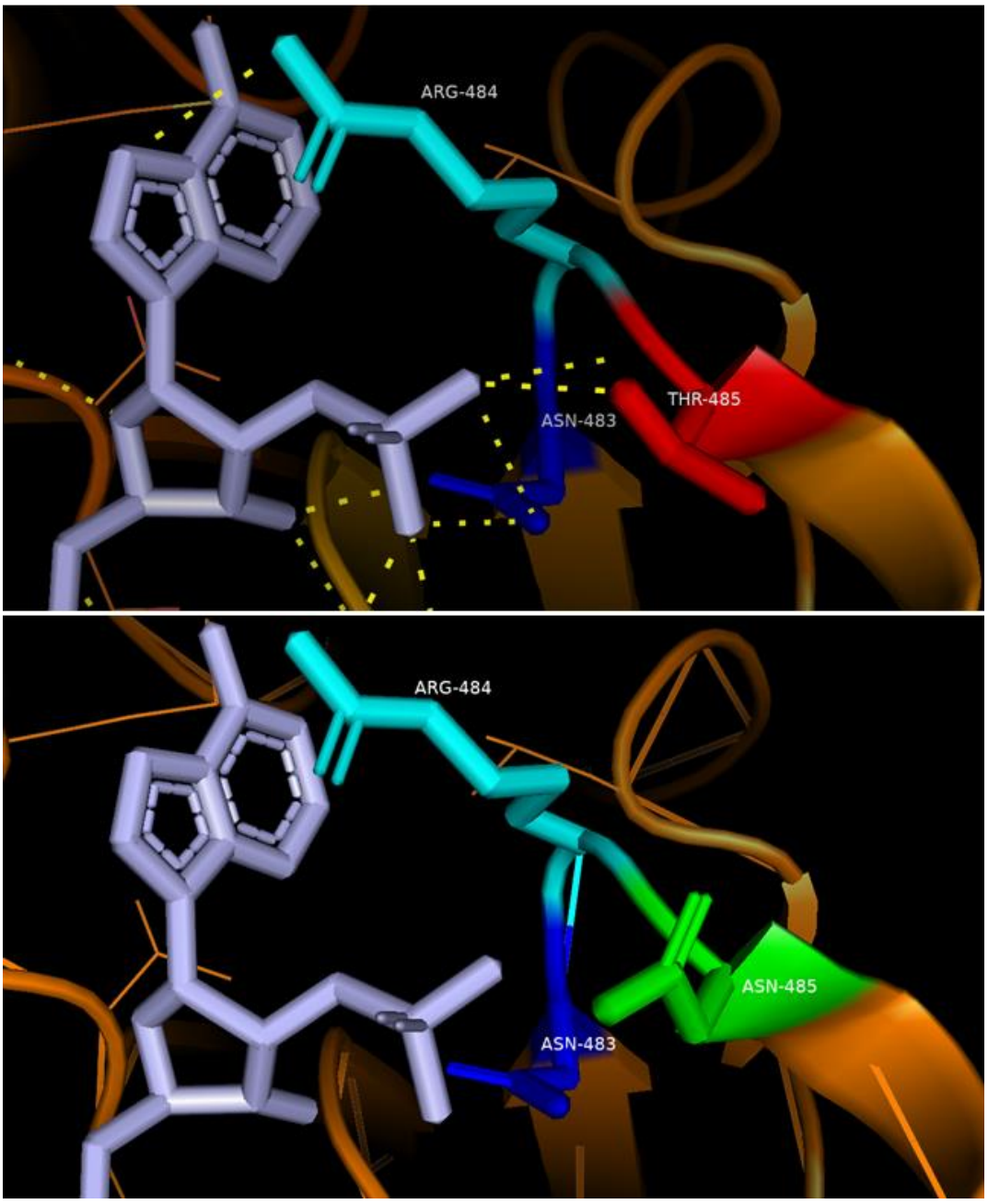


Figure 4.4: Simulated mutagenesis of Arabidopsis. SDH (Top) Hydrogen bonding between $NADP^+$ (white) and the NRT motif in *Arabidopsis* SDH. Only the side chains of Asn483, Arg484 and Thr485 are shown for ease of visualization. Hydrogen bonding between the groups is indicated by dashed yellow lines. The rest of the protein backbone (shown in yellow-orange) was partially removed for clarity. The wild-type structure was compared to the simulated Thr485Asn mutant (bottom).

4.4 Discussion

4.4.1 Summary

Nicotinamide-nucleotide dependent dehydrogenases have been extensively examined. The cofactor binding domains of such enzymes have a common Rossmann fold structure albeit with variations in the surrounding structures and residues. This fold features a conserved glycine-rich loop containing the canonical GXGXXG fingerprint sequence that interacts with the pyrophosphate group of NAD(P)(H) (Chánique and Parra, 2018; Kallberg and Persson, 2006; Lesk, 1995). The presence of an Asp or Arg residue downstream of this motif typically determines specificity towards NAD(H) or NADP(H) respectively (Chánique and Parra, 2018; Kallberg and Persson, 2006; Lesk, 1995). Plant SDH clade members have been described as NADP(H)-dependent dehydrogenases (Herrmann, 1995; Herrmann and Weaver, 1999). Among other features, they share a common NRT motif that predictably interacts with adenine phosphate (Gritsunov *et al*., 2018; Singh and Christendat, 2007). In contrast, in the QDH enzyme from *P. trichocarpa* (*Poptr*QDH) and other angiosperms NAD(H)-dependent QDH enzymes (Carrington *et al*., 2018; Gritsunov *et al*., 2018; Guo *et al*., 2014) the conserved NRN (or NRT in the case of angiosperm SDH's) motif is exchanged for DID, which may play a role in discriminating against NADP(H) (Gritsunov *et al*., 2018). Such differences prompted testing of cofactor specificities among early and late derived SDH lineages. Specifically, it was asked whether they could promiscuously act on both NADP(H) and NAD(H) or whether NAD(H)-specificity is a novel trait of flowering plant QDHs.

4.4.2 Differential use of NAD(H) and NADP(H) by SDH and QDH's

SDH proteins from early derived lineages including *C. reinhardtii*, *P. patens*, and *S. moellendorffii* showed specificity for shikimate and $NADP^+$. Interestingly, these showed measurable activity with NAD^+ but also had low affinity for it suggesting it is not a physiological substrate. However, it should be pointed out that substrate concentrations did not plateau for these enzymes and kinetic parameters had to be extrapolated beyond the collected data, which introduces errors. Yet given their description as chloroplastic, NADP(H)-dependent dehydrogenases (Herrmann, 1995; Herrmann and Weaver, 1999), this result was intuitively expected despite variation of the NRT

motif to NRN. Levels of NAD(H) are almost twice greater than NADP(H) in the cytosol while the opposite holds true in chloroplasts (Gakière *et al*., 2018). Moreover, cytosolic NAD(H) exists mainly in the oxidized form (low NADH/NAD ratio) whereas in chloroplasts NADP(H) is highly reduced (high NADPH/NADP ratio) (Gakière *et al*., 2018; Lunzer *et al*., 2005; Maruta *et al*., 2016). Generally, NAD^+ is an oxidant for catabolic, ATP-generating pathways while NADPH provides reducing power driving biosynthetic (anabolic) reactions (Carugo and Argos, 1997; Rosell *et al*., 2003). The excess of reducing equivalents in the chloroplast provides a suitable stage for biosynthesizing primary metabolites (Hashida *et al*., 2009; Maruta *et al*., 2016) including shikimate (Herrmann, 1995; Herrmann and Weaver, 1999).

In contrast to plastid-bound SDH enzymes following a relatively straightforward distribution, quinate metabolizing enzymes localize to a grey zone. Plant QDH's have been isolated from the cytosol *(*Ding *et al*., 2007; Leuschner *et al*., 1995) as well as from chloroplasts (Gritsunov *et al*., 2018). The cytosolic and chloroplastic QDH's are active with NAD^+ and $NADP^+$ *in vitro* respectively (Gritsunov *et al*., 2018). Based on the presumption that most cytosolic NAD(H) exists in the oxidized state and participates in degradatory pathways, it is believed that NAD(H)-dependent QDH's in the cytosol mostly contributes towards CGA catabolism (Gritsunov *et al*., 2018). However, the *in vivo* functions of QDH proteins are yet to be discovered.

4.4.3 Dual NAD^+ and $NADP^+$ specificities of angiosperm SDH's

A surprising finding of this study was the high affinity of *Poptr*SDH for both $NADP^+$ (K_M 12.9 $\pm$ 2.2 µM) and NAD^+ (25.1 $\pm$ 2.3 µM). The apparent K_M and V_{MAX} values are comparable to each other suggesting equal efficiency towards either $NADP^+$ or NAD^+. In contrast, *Pinta*SDH, the other seed plant SDH tested here, showed barely measurable activity with NAD^+. Despite previous reports on NAD(H)-dependent seed plant QDHs (e.g. Gritsunov *et al*., 2018; Guo *et al* 2014; Refeno *et al*., 1982), *Pinta*QDH behaved similarly to SDH proteins. Based on catalytic efficiency ratios (V_{MAX}/K_M) it prefers $NADP^+$ over NAD^+ by ~1700 and ~440 fold using shikimate and quinate as substrates respectively, thereby indicating strong preference for $NADP^+$ regardless of the substrate used. *Poptr*QDH was unique among others as it was primarily active with NAD^+. The specificity ratio of *Poptr*QDH was 100-fold in favour of NAD^+ over $NADP^+$ and notably, this

enzyme showed no activity with $NADP^+$ when shikimate was used as substrate. This finding is particularly fascinating as it demonstrates a loss of ancestral functions and a concomitant gain of new functions following gene duplication. Differential cofactor uses by *Poptr*QDH and *Pinta*QDH may point to differences in their physiological roles. In conifer needles for example, QDH enzymes are believed to contribute towards the biosynthesis of secondary metabolites such as protocatetchuic and gallic acid while quinate biosynthesized by QDH in the xylem is thought to serve as a carbon source for lignin biosynthesis (Ossipov *et al*., 1995). Unlike pine QDH's playing roles in anabolic reactions, NAD^+-dependent angiosperm QDH's are believed to participate in the catabolism of chlorogenic acid in the cytosol (Gritsunov *et al*., 2018). On a side note, the observation that both *Poptr*SDH and *Poptr*QDH were active with NAD^+ indicates that NAD(H)-binding affinity developed in both SDH and QDH lineages of angiosperms.

4.4.4 NRN versus NRT

Although *Poptr*SDH had greater affinity for both $NADP^+$ and NAD^+ than SDH enzymes from *C. reinhardtii, S. moellendorffii, P. patens* and *P. taeda*, it is unlikely this enzyme uses NAD(H) *in vivo* due to low concentrations of NADH in the chloroplast (Gakière *et al*., 2018). At the structural level, similarities between NAD(H) and NADP(H) may permit at least, loose fitting of NAD(H) into the NADP(H)-binding pocket of *Poptr*SDH (Chánique and Parra, 2018; Rosell *et al*., 2003). However, then the inviting question becomes this: what enables *Poptr*SDH to accept both $NADP^+$ and NAD^+ as cofactors, at least *in vitro*, but not the other SDH proteins? Notably their cofactor-binding domains share high homology and, as mentioned above, the only notable difference is a threonine (NR<u>T</u>) in angiosperm sequences as opposed to asparagine (NR<u>N</u>) in non-angiosperms. Given the similar properties and orientations of these amino acids, it is unclear what accounts for the lower $K_{M(NAD)}$ of *Poptr*SDH than *Selmo, Phypa* and *Pinta*SDH. In particular, the hydroxyl group of *A. thaliana* Thr485 and the amide side chain of Asn485 of the simulated mutant are approximately at equal distances from the terminal phosphate group, suggesting that Asn485 also participates in hydrogen-bonding to 2' phosphate. It is possible that H-bonding between the 2' OH group of NAD(H) and the hydroxyl group of Thr485 is more favourable than between 2' OH and the amide group of Asn485. Or perhaps the branched side chain of threonine provides stabilizing effects within the β-sheet (Petsko and Ringe, 2004). However, in the absence of a

crystallographic model, these reasons are speculatory at best. On the other hand, the fact that *Poptr*SDH can use both $NADP^+$ and NAD^+ whereas *Poptr*QDH is specific for NAD^+ suggests that while the DID motif discriminates against $NADP^+$-binding, NRT allows binding to both. However, binding to NAD^+ has yet to be tested with other NADPH-using angiosperm SDH enzymes to confirm this prediction; it is possible poplar SDH is an exception to the norm.

4.4.5 Multiple amino acids working in concert define catalytic activities

As a third possibility, the above-mentioned amino acids may exert their effects outside of the cofactor-binding domain. After all, tweaking of the Rossmann fold has been shown to alter catalytic functions and substrate preferences of enzymes (Cahn *et al*., 2017), as documented by numerous protein engineering studies (Brinkmann-Chen *et al*., 2013; Lauvergeat *et al*., 1995; Rosell *et al*., 2003). In an attempt to switch coenzyme preference of a NADP(H)-dependent Eucalyptus cinnamoyl alcohol dehydrogenase, Lauvergeat *et al*. (1995) mutated a conserved Ser to Asp mimicking the Rossmann loop of NAD(H)-binding dehydrogenases. While the single mutant could react with NAD(H), it displayed only $1/25^{th}$ the activity of the wild-type protein with NADP(H). A similar experiment was performed by Rosell *et al*. (2003) who substituted Gly223 (note the position denotes that of horse ADH) for Asp, which is found at the homologous position in an NADP(H)-dependent amphibian alcohol dehydrogenase (ADH8). The mutation only partially lowered activity with NADP(H) but did not affect activity with NAD(H). On the other hand, cofactor specificity was successfully rewired in an ADH8 triple mutant in which two nearby residues, namely Thr224 and His225 were mutated alongside Gly223. In native ADH8 these three residues help form a cavity fitting the terminal phosphate group of NADP(H) (Rosell *et al*., 2003). Although the interactions between NAD(H) or NADP(H) and the variable residues around the Gly-rich loop of SDH proteins are unknown, it is expected that they have an additive effect on cofactor-specificity, and may also account for the higher affinity of *Poptr*SDH for both NAD(H) and NADP(H).

Recently, four SDH sequences from *Camellia sinensis* were characterized by Huang *et al*. (2019). One of them, *Cs*DQD/SDHb has characteristic angiosperm QDH "fingerprints" including the S338G and T381G substitutions compared to *A. thaliana* SDH as well as the DID motif. It also showed lower activity with shikimate compared to the other isoforms (although the authors did

not test for QDH activity). A Gly338Ser/Gly381Thr/Asp483Asn/Leu484Arg/Asp485Thr quintuplet mutant of *Cs*DQD/SDHb was generated that showed 6-fold higher SDH activity than the wild-type. While the effects of the single mutations, and changes to cofactor specificity were not analyzed, it is assumed that all of them contributed to SDH activity.

4.4.6 Conclusions

SDH from non-seed plants and seed plants possess a defining Arg in their cofactor binding domains that has been characterized in NADP(H)-dependent angiosperm SDH's *(*Gritsunov *et al*., 2018; Singh and Christendat, 2007) such that $NADP^+$ specificity among these enzymes was expected. Thus, like changes in substrate preferences, NAD^+-binding likely represents a novel trait of seed plants. Unlike *Poptr* QDH, which prefers NAD^+ over $NADP^+$, *Pinta* QDH was specific for $NADP^+$. As mentioned earlier in chapter three, such differences may be explained by differential evolutionary rates between angiosperms and gymnosperms (Augusto *et al*., 2014; De La Torre *et al*., 2017) and if this were the case, then over time *Pinta* QDH may be expected to evolve preference for NAD^+. The differential use of $NADP^+$ and NAD^+ by angiosperm SDH and QDH enzymes may permit separate regulation of anabolic and catabolic pathways respectively (Gritsunov *et al*., 2018). On the other hand, why seed plant SDH (*Poptr* SDH) proteins can dually act on both cofactors is a question that deserves further investigation. Constructions of a Thr485Asn and Asn483Asp/Arg484Ile/Thr485Asp (again using the same nomenclature as the *Arabidopsis* protein) *Pinta* QDH single or triple mutant respectively, may help pinpoint the exact contributions of the NRT and DID motif in defining cofactor specificity. On a similar note, the crystal structure of NAD(H)-dependent QDH proteins with the DID motif have not been examined. Its elucidation should provide further insights into the structure-ligand binding interactions of these amino acids and, the evolutionary path walked by NAD^+-dependent QDH enzymes.

Chapter 5

5.1 Introduction

Following a rise in available genome-based resources, recent years has seen an increase in sequence information for a wide collection of plant species. Sequence similarities and enzyme homologies provide a first glimpse into the activities of newly discovered genes and of the metabolic pathways they participate in. However, sequence similarities alone do not guarantee the encoded functions of a protein (Croteau *et al*., 2000), and *in silico* data must be paired with protein characterization *in vitro* followed by a confirmation of their roles *in vivo*. The latter is necessary since it is difficult to accurately mimic complex cellular environments in a test tube (Baruch *et al*., 2004). In the case of plant *QDH* genes, recombinant DNA technology has so far been used to study the *in vitro* activities of QDH enzymes cloned from a few angiosperm species including *P. trichocarpa* (Guo *et al*., 2014; Carrington *et al*., 2018); *N. tabacum*, and *S. lycopersicon* (Gritsunov *et al*., 2018). *In vivo* experiments using these sequences is a logical next step for the understanding of their metabolic roles in plants. Transgenic approaches interfering with QDH activity appear straight forward for *in vivo* analyses. Negatively interfering with QDH activity through RNAi or CRISPR mediated approaches in poplar have been attempted as part of this thesis but were plagued with experimental challenges that prevented the completion of these experiments. An alternative gain-of function approach was instead taken that over-expressed poplar QDH into a plant system that lacks endogenous QDH and is unable to produce quinate or its derivatives. Before detailing this experimental approach, it is important to introduce previous *in planta* biochemical work and discuss the subcellular localization of quinate biosynthesis and downstream metabolic pathways.

5.1.1 Studies on the activities of QDH isolated from plants

While so far, *in vivo* investigations into *QDH* gene functions have not been undertaken, other groups have described the activities of QDH enzymes purified from crude plant extracts. Although the sequences of these enzymes are unknown, their activities coincide nicely with those of the genetically and biochemically characterized *QDH* sequences mentioned above. For example, QDH isolated from mung beans have been observed to reduce NAD^+ in the presence of quinate (Gamborg, 1966). There are also several reports of QDH's from gymnosperms. A dual

functioning S/QDH capable of catalyzing the oxidation of quinate and shikimate using $NADP^+$ was isolated from *P. taeda* needles (Ossipov *et al*. 2000) and a QDH capable of reducing $NADP^+$ in the presence of quinate has also been reported in the needles and xylem extracts of *Larix siberica* (Ossipov *et al*., 1995).

In stopped assays using HPLC-CAD, Guo *et al*. (2014) identified dehydroquinate as an end-product of reaction mixtures containing two QDH isoforms from *P. trichocarpa* (*Poptr*QDH1 and *Poptr*QDH2). Gamborg *et al*. (1966) identified dehydroshikimic acid produced from reaction mixtures containing QDH and SDH from mung beans using radio-TLC. While dehydroquinate could not be observed because the separation method could not distinguish between dehydroquinate and quinate, the authors predicted it was readily converted to dehydroshikimate (Gamborg *et al*., 1966). In addition to dehydroquinate and dehydroshikimate which are both intermediates of the shikimate pathway, hydroxybenzoic acids (gallic acid and protocatechuic acid) have also been identified as end-products of alkaline reaction mixtures containing QDH from *L. siberica* (Ossipov *et al*., 1995). Thus, in plants expressing endogenous *QDH* genes it is expected that a combination of quinate, dehydroquinate and hydroxybenzoic acids will be present. Given that quinate derived-chlorogenic acid also accumulates to high levels in *P. trichocarpa* and solanaceous plants (Niggeweg *et al*., 2004), which possess both *QDH* and hydroxycinnamoyl-CoA quinate hydroxycinnamoyl transferase (*HQT*), the latter catalyzing the reversible acylation of quinate using cafeoyl CoA as a donor group thereby leading to the production of chlorogenic acid (Niggeweg *et al*., 2004). It is therefore also expected that *QDH* genes also confers chlorogenic acid biosynthetic activities to these plants.

5.1.2 Possible roles of QDH in lignin and chlorogenic acid biosynthesis

Radiolabeling studies have shown the conversion of ^{14}C-labelled quinate to hydroxybenzoic acids and lignin in developing needles and xylem respectively of *P. sylvestris* (Ossipov *et al*., 1995). Caffeoylquinic esters (including chlorogenic acid) follows a similar trend, being observed to accumulate in the developing vasculature of young *Coffea canephora*. Chlorogenic acid itself may serve as an intermediate in lignin biosynthesis by providing monolignol precursors (Mondolot *et al*., 2006). This idea is based on the negative correlation between chlorogenic acid and lignin levels, for examples in developing hypocotyls and cotyledons of *Capsicum annuum* and *Coffea arabica* seedlings, respectively (Silva *et al*., 2019). However, the idea that chlorogenic acid is an

intermediate of lignin biosynthesis is somewhat at odds with other findings since silencing of HQT in tomatoes causes a reduction in chlorogenic acid content without affecting lignin content (Niggweg *et al*., 2004). Moreover, while the reversibility of HQT-catalyzed reactions is well demonstrated *in vitro* (Niggweg *et al*., 2004), the enzyme appears to be more efficient at catalyzing the production of caffeoyl-quinate rather than its breakdown (Silva *et al*., 2019). A homolog of HQT, hydroxycinnamoyl CoA shikimate hydroxycinnamoyl transferase (HCT) may play a more prominent role in lignin biosynthesis (Li *et al*., 2010; Sonnante *et al*., 2010; Hoffmann *et al*., 2004), and silencing of HCT leads to decreased lignin content in *Arabidopsis* (Hoffmann *et al*., 2004; Li *et al*., 2010) and *N. benthamiana* (Hoffmann *et al*., 2004). Interestingly, *Panicum virgatum* accumulates chlorogenic acid and possess three copies of HCT (PvHCT1a, PvHCT2a and PvHCT-Like1) but lacks genes encoding HQT. PvHCT-Like1 could catalyze activation of quinate with either *p*-coumaroyl CoA or caffeoyl-CoA, the latter leading to the production of chlorogenic acid. Alternatively, *p*-coumaroyl CoA could be hydroxylated by two shikimate 3' hydroxylase enzymes (PvC3'H1 and PvC3'H2) to generate chlorogenic acid via a different route (Escamilla-Treviño *et al*., 2014). However, unlike tobacco HQT, PvHCT-Like1 is unable to breakdown chlorogenic acid *in vitro*. Thus, another enzyme may be responsible for the hydrolysis of chlorogenic acid to provide caffeoyl CoA for lignin biosynthesis. Gritsunov *et al*., (2018) suggests a chlorogenic acid esterase might serve this role. While such an enzyme has not yet been discovered, the presence of a caffeoyl shikimate esterase (CSE) in *Arabidopsis* capable of hydrolyzing caffeoyl-shikimate esters (Escamilla-Treviño *et al*., 2014) points to a promising search. On the other hand, the ability of *Arabidopsis* to accept both caffeoyl-shikimate and *p*-coumaroyl shikimate as substrates suggests it has broad specificity and may also act on caffeoyl-quinate (Silva *et al*., 2019). Altogether this suggests that a combination of *QDH*, *HQT*, CSE and a potential chlorogenic acid esterase is involved in the metabolism of quinate and chlorogenic acid, and possibly redirects these compounds towards lignin biosynthesis.

5.1.3 Localization of QDH?

If QDH plays a role in chlorogenic acid and lignin biosynthesis, then this scenario conforms neatly to its predicted localization in the cell. The oxidative polymerization of monolignols during lignin biosynthesis is a cytoplasmic process (Eudes *et al*., 2016) and localization studies on HQT from tomato (Moglia *et al*., 2014) and *Lonicera japonicum* (Li *et al*., 2019) suggest that chlorogenic

acid biosynthesis also occurs in the cytoplasm, although data pertaining to such is somewhat at odds (as will be discussed later). Nevertheless, this would lend support to the notion that chlorogenic acid serves as a storage form of lignin precursors. *N. tabacum* has two SDH isoforms, NtDHD/SHD-1 and Nt-DHD/SHD-2 (note their sequences correspond to *Nicta*SDH and *Nicta*QDH of the S/QDH family tree), the first of which behaves like typical plant SDH, displaying activity towards shikimate and $NADP^+$ and possessing a chloroplast signal peptide targeting it to the chloroplast. On the other hand, NtDHD/SHD-2, shows activity with shikimate but lacks the signal peptide, thereby localizing to the cytosol. The cytoplasmic localization of *QDH* and *HQT* points to the biosynthesis of quinate in the cytoplasm followed by its conversion to chlorogenic acid. The question then arises about how dehydroquinate, which is an intermediate of the chloroplastic shikimate pathway, is made available to QDH in the cytoplasm. On the one hand, the discovery of cytosolic isoforms of DAHP synthase, EPSP synthase and chorismite mutase points to an alternative shikimate pathway in the cytosol (Ding *et al*., 2007) that could supply dehydroquinate to quinate in a branch pathway. In support of this idea, RNAi knock-down of cytosolic chorismate mutase in petunia leads to a reduction in phenylalanine and phenolic volatiles in flowers (Qian *et al*., 2019).

On the other hand, there is disagreement surrounding the cytosolic localization of QDH and HQT, An $NADP^+$-dependent QDH has been isolated from both the cytosol and chloroplasts of larch needles (Ossipov *et al*., 1995) and HQT from *L. japonicum* dually localizes in the chloroplast and cytoplasm (Li *et al*., 2019). Histochemical staining points to the presence of caffeoyl-quinic esters in the chloroplasts of developing *Coffea canephora*, where they may possibly protect against UV-B damage (Mondolot *et al*., 2006) and this finding is shared by the presence of chlorogenic acid in the cytosol and chloroplasts of young *L. japonicum* (Li *et al*., 2019). If this were the case, then it is expected that chloroplastic and cytoplasmic QDH and HQT participate in chlorogenic acid biosynthesis in their corresponding compartments. Interestingly, both NAD^+ and $NADP^+$-dependent QDH classes are found among angiosperms. The former comprises the QDH clade displaying lower sequence similarity to well-characterized SDH proteins and shows activity with NAD^+ and QDH *in vitro*. The latter include sequences from Brassicaceae, Rosaceae and some monocots that phylogenetically cluster within the angiosperm SDH clade, but show activity towards shikimate and quinate using $NADP^+$ *in vitro* (Gritsunov *et al*., 2018). Given some of these enzymes possess a transit peptide that the NAD^+-dependent QDH's lack, it is

suggested the former plays a role in the reduction of dehydroquinate to quinate using NADPH, while cytosolic QDH catalyzes the NAD^+-dependent oxidation of quinate released from the degradation of chlorogenic acid to dehydroquinate (Gritsunov *et al*., 2018). Several independent groups point to the transport of quinate from the chloroplast to the cytosol *(*Bonner and Jensen, 1998; Gritsunov *et al*., 2018; Ossipov *et al.*, 1995). The fate of quinate in the cytoplasm, and whether it is converted to chlorogenic acid, lignin, or hydroxybenzoic acids (Ossipov *et al*., 1995) or if it re-enters the chloroplast to provide a carbon source for the shikimate pathway (Bonner and Jensen, 1998) is yet to be determined.

5.1.4 Developing a method for analyzing QDH products

As mentioned earlier, the objectives of this experiment were to understand the *in vivo* functions of *P. trichocarpa QDH* genes through their overexpression in *Arabidopsis*. It was predicted that novel production of quinate and its derivatives would be detected in transgenic plants using LC-MS. So far, several approaches have been developed for analyzing phenolics (e.g. chlorogenic acid) and sugars in plants, but mostly independently of each other. Since a large portion of the methods and results of this chapter describes the development of an LC-MS capable of analyzing both organic acids (e.g. quinate) and phenolic compounds in *Arabidopsis* leaves, the availability of different analyzation techniques to study plant compounds will be briefly outlined in this section. As mentioned above RPHPLC-UV-MS using a C18 column is routinely used in the laboratory for analyzing slightly polar compounds (Meyer, 2010; Jorge *et al*., 2016). Particularly beneficial for this study, a reliable and proven method for extracting plant phenolics has been developed by neighbouring colleagues (Constabel and Yip, personal communications 2014). On the other hand, HILIC-MS is better suited for analyzing sugars and organic acids that are poorly retained on hydrophobic columns (Meyer, 2010; Jorge *et al*., 2016). Earlier works have established methods for simultaneously analysing chlorogenic acid and quinic acid by coupling HPLC with tandem mass spectrometry (MS/MS) (e.g. Fang *et al*., 2002). A major advantage of LC-MS/MS is that it provides structural information about fragmented ions in addition to their molecular masses (Jorge *et al*., 2016). However, some drawbacks are its costly equipment that may not be user friendly or accessible. Alternatively, both quinic acid and its derivatives have been quantified using direct infusion mass spectrometry (DIMS) (Mendonça *et al*., 2008; Ng *et*

al., 2004; Ouni *et al*., 2011). This approach is fast, requires little sample and provides as much information as LC-MS. However, because it bypasses the initial separation step, selectivity may be lowered, which complicates spectra analyses (Kirwan *et al*., 2014). Overall, the versatility of these approaches have undoubtedly proven useful for mining through the rich layers of plant secondary metabolomes (Piasecka *et al*., 2019): however the challenge then becomes to customize a method addressing a particular biological question (Jorge *et al*., 2016), the identification of quinate and its derivatives in transgenic *Arabidopsis* leaves.

5.1.5 Experimental objectives: search for novel production of quinate and quinate derivatives in transgenic *Arabidopsis* overexpressing *P. trichocarpa* QDH

The objective of this experiment was to examine the *in vivo* roles of *QDH* genes, for which no information is known. Here *P. trichocarpa QDH* genes (*PoptrQDH* and *PoptrQDH2*) were constitutively expressed under the CaMV35S promoter in *Arabidopsis*. These two isoforms were chosen out of the tree QDH enzymes found in *P. trichocarpa* because their activities *in vitro* have already been established (Carrington *et al*., 2018; Guo *et al*., 2014). *Arabidopsis* is ideal for studying exogenous *QDH* gene functions because it is not known to accumulate quinic acid or its major derivative, chlorogenic acid (Niggeweg *et al*., 2004). However, it possesses an HCT enzymes that is capable of, *in vitro*, utilizing quinate to form chlorogenic acid (Hoffmann *et al*., 2004). It was hypothesized that overexpression of *QDH* in *Arabidopsis* will lead to novel production of quinate and its derivatives in transgenics if the precursor is available in the subcellular compartment the QDH is targeted to. While the sample size of this experiment was not statistically large enough, metabolite analyses suggests quinic acid was present in at least one transgenic individual compared to wildtype. In addition, novel production of chlorogenic acid was observed in another *PoptrQDH* overexpressing mutant. The results imply a feasibility of using this method (albeit with modifications to the extraction procedure) to test *QDH* functions *in vivo* and that can be applied to other plant systems in the future. As a whole, this study provides a first glimpse into the biological activities *P. trichocarpa* QDH proteins.

5.2 Methods

Credit for the cloning of *PoptrQDH* and *PoptrQDH2* goes to Jia Guo and is described in Guo *et al*. (2014)

5.2.1 Plant growth conditions

Arabidopsis plants were transformed by pCAMBIA 2300-*PoptrQDH* or *PoptrQDH*2 by *Agrobacterium* vacuum infiltration. The seeds were harvested and grown subsequently as the F1 generation. Seeds were surface sterilized in a solution of 1% bleach and 0.0.1% Tween for 5 minutes followed by 70% ethanol (v/v) for 5 minutes and rinsed (5X) in sterile water in a total volume of 1 mL. Transformed seeds were selected on 1/2 MS media containing kanamycin (50µg/mL) or without added antibiotics in the case of wildtype seeds. Both seed types were cold stratified at 4˚C for two days and germinated in a growth chamber set 21˚C and a 16 hour photoperiod with a light intensity of 120µmol m^{-2} s^{-1}. Ten-day old seedlings were transferred to soil and grown in a greenhouse growth chamber set to the same conditions as described earlier. F1 plants were grown until maturity and dried seeds were harvested as the F2 generation. The F2 generation plants were either grown using the same conditions as above or under a short-day cycle [8/16 hours (day/light)].

5.2.2 Control plant growth conditions

Z. mays seeds were soaked overnight in water and transferred to soil pots; *N. tabacum* seeds were sown directly in soil and both seeds were cold stratified at 4˚C for two days. *Z. mays* and *N. tabacum* were grown in the same growth chambers as *Arabidopsis* plants. Hybrid (717) poplars were propagated on woody plant media (WPM) and grown in a greenhouse growth chamber set to a long day (16 hr light 8 hr dark) cycle.

5.2.3 RT-PCR

RNA was extracted from young, ~two week old *Arabidopsis* using TRIzol (Invitrogen) and DNASe treated with Ambion Turbo DNAse (Invitrogen). The cDNA was synthesized using SuperScript III (Thermo Fisher Scientific) and its integrity was checked using PCR with actin house-keeping gene primers (Forward sequence 5’ GCGACAATGGAACTGGAA 3’ and Reverse 5’ GGATAGCATGTGGAAGTGCATACC 3’). Expression of *PoptrQDH* and *PoptrQDH2* was determined by PCR using primers specific for each gene (Forward *QDH* 5’

GGCTTAAUATGGGGCGTGCTGGATC 3' and Reverse 5' GGTTTAAUTCAGAATTTGGCTAGAACAATCTCCC: Forward *QDH*2 5' GGCTTAAUATGGGGAGTGTTGGAGTCCTGAC 3' and Reverse 5' GGTTTAAUTCAGAATTTGGCTAAAACAATCTCCC 3').

5.2.4 Extraction of phenolic acids and organic acids

From Constabel and Yip (~~2014,~~ personal communications, 2014)

Mature plant leaves were harvested and frozen immediately in liquid nitrogen. Frozen tissues were ground using an ice cold mortar and pestle and freeze-dried. Freeze-dried leaves (50-100 mg) were disrupted in a Precellys bead-beater and sonicated for 10 minutes. Homogenous mixtures were centrifuged (15, 000 rpm for 15 minutes) and the supernatant was collected in a borosilicate tube. The extraction was repeated twice more and the pooled supernatants were speed-vacced overnight to dryness. The extract pellet was resuspended in 70% methanol (v/v) and loaded onto a Phenomenex Strata X-C18 cartridge activated with methanol. The filtrate was rinsed (3X) with water and eluted (3X) in methanol in a total volume of 3 mL. The pooled eluents were dried overnight in a speed vac and subsequently resuspended in 10% methanol (v/v) to a final concentration of 10 mg/mL. Extraction of organic acids was performed as described above except using either methanol or acidified water (HCl, pH2) as the extraction solvent and without concentrating on a SPE cartridge.

5.2.5 LC-UV/CAD analyses for phenolic compounds and organic acids

Extracts were filtered and centrifuged prior to LC analyses. Phenolic compounds were analyzed using a Thermo Dionex Ultimate 3000 HPLC system consisting of a HPG-3400 pump, an ACC-3000T autosampler and a Dionex 3000 PDA detector. The mobile phase consisted of water with 1% (v/v) formic acid (A) and acetonitrile with 0.4% (v/v) formic acid (B). Phenolic compounds were separated on a Phenomenex Kinetex C18 column (150 x 4.6 mm, 2.6 μm) using the following gradient conditions: 0-5 min, 3% B; 49-50 min, 100% B and 58-60 min, 3% B, flow rate 0.6 mL/min. The separation of organic acids was carried out on a TSKgel Amide-80 column (25 cm X 4.6 mm, 5μm) and compounds were eluted isocratically with 95% A [water with 5% formic acid (v/v)] and 5% B [acetonitrile with 5% formic acid (v/v)] at a flow rate of 1.2 mL/min and detected

using a Corona CAD. Three technical replicates were run per sample and a solvent blank was run every three samples.

5.2.6 UPLC-analyses of chlorogenic acid

Phenolic compounds were also analyzed using a Waters UPLC-PDA-MS system paired with an Acquity BEH C18 column (2.1 mm X 50 mm, 1.7μm). The mobile phases consisted of water with 0.4% formic acid (v/v) (A) and acetonitrile with 0.4% formic acid (v/v) (B). The following gradient conditions were used: 0-5 min, 3% A; 5-9 min, 100% B, 9-15 min, 3% B at a flow rate of 0.4 mL/min. Three technical replicates were run per sample and a solvent blank was run every three samples.

5.2.7 Orbitrap-analysis of chlorogenic acid and quinic acid

High resolution mass spectrometry (HRMS) was performed using a Thermo Dionex 3000 coupled to a PDA and an Orbitrap Q Exactive Plus mass spectrometer (Thermo Fisher). The following isocratic method was developed using water with 1% formic acid (v/v) A and acetonitrile with 1% formic acid (v/v) B and the amide column as before: 0-10 min, 10% B. An extended 30-minute wash step was run in between every sample to ensure there was no carry over. Standard curves of quinic acid and chlorogenic acid in 10% (v/v) methanol were prepared using concentrations ranging from 0.001 to 0.2 ppm and 10 to 250 ppm. Ions were quantified by peak area integration using the Thermo Xcalibur software.

Results

5.1 Characterization of PoptrQDH and PoptrQDH 2 OX Arabidopsis

To investigate the biological functions of *QDH* genes in plants, *QDH* and *QDH 2* from *P. trichocarpa* were transformed into *Arabidopsis*. Transgene expression was confirmed among F1 and F2 plants although it decreased among most F3 individuals and was absent in F4 plants (Figure 5.1-2**)**.

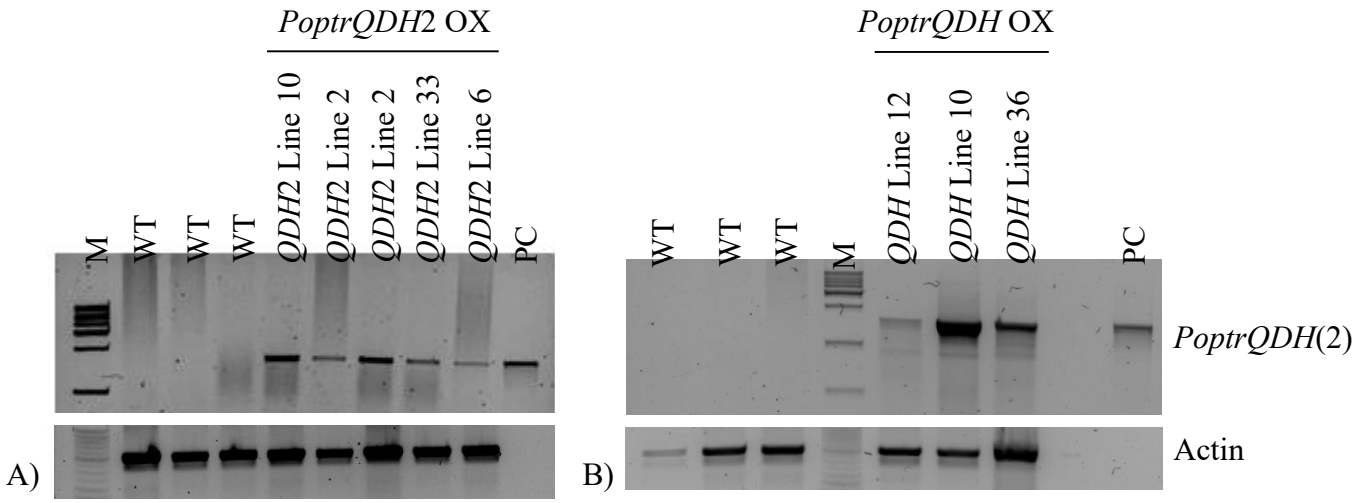


Figure 5.1: Expression of PoptrQDH and PoptrQDH2 in transgenic Arabidopsis. *PoptrQDH2* (A) and *PoptrQDH* (B) gene expression in different transgenic lines of F2 generation plants is shown. Wildt-type *Arabidopsis* are not shown to expression either transgene. RT-PCR was also performed using house-keeping gene, actin-subunit, primers. PC; positive control (either pQE30::*PoptrQDH* or pQE30::*PoptrQDH2*, M; molecular ladder [either M101R-1 (Biobasic) for gene-specific RT-PCR or B7025 (NEB) for house-keeping RT-PCR].

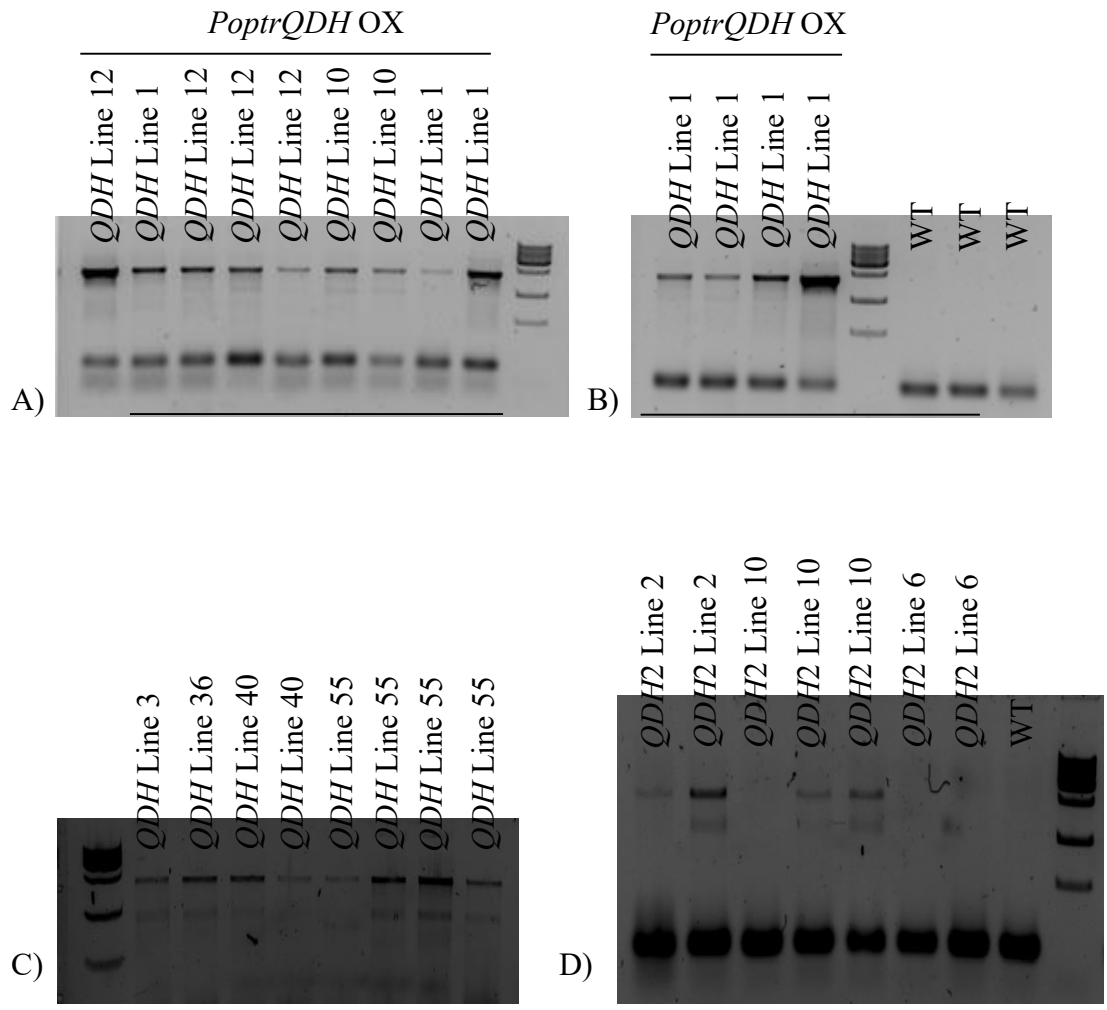


Figure 5.2: *PoptrQDH* and QDH2 expression in transgenic *Arabidopsis* and transgene silencing. Expression of *PoptrQDH* (A, B and C) and *Poptr* in different transgenic lines of the F2 generation and in siblings of the same transgenic line (note each lane represents DNA extracted from a different plant). Transgene expression is consistent among different siblings (represented by different lanes). In contrast, F3 generation plants show signs of silencing in both *PoptrQDH* (not shown) and *PoptrQDH*2 OX lines (D).

None of the transgenics showed outward phenotypic differences to wildtype *Arabidopsis*. Similarly, initial screening of phenolic acids and organic acids of wildtype and transgenic *Arabidopsis* using HILIC-UV-CAD showed no differences between them (Figure 5.3 and 5.4-6).

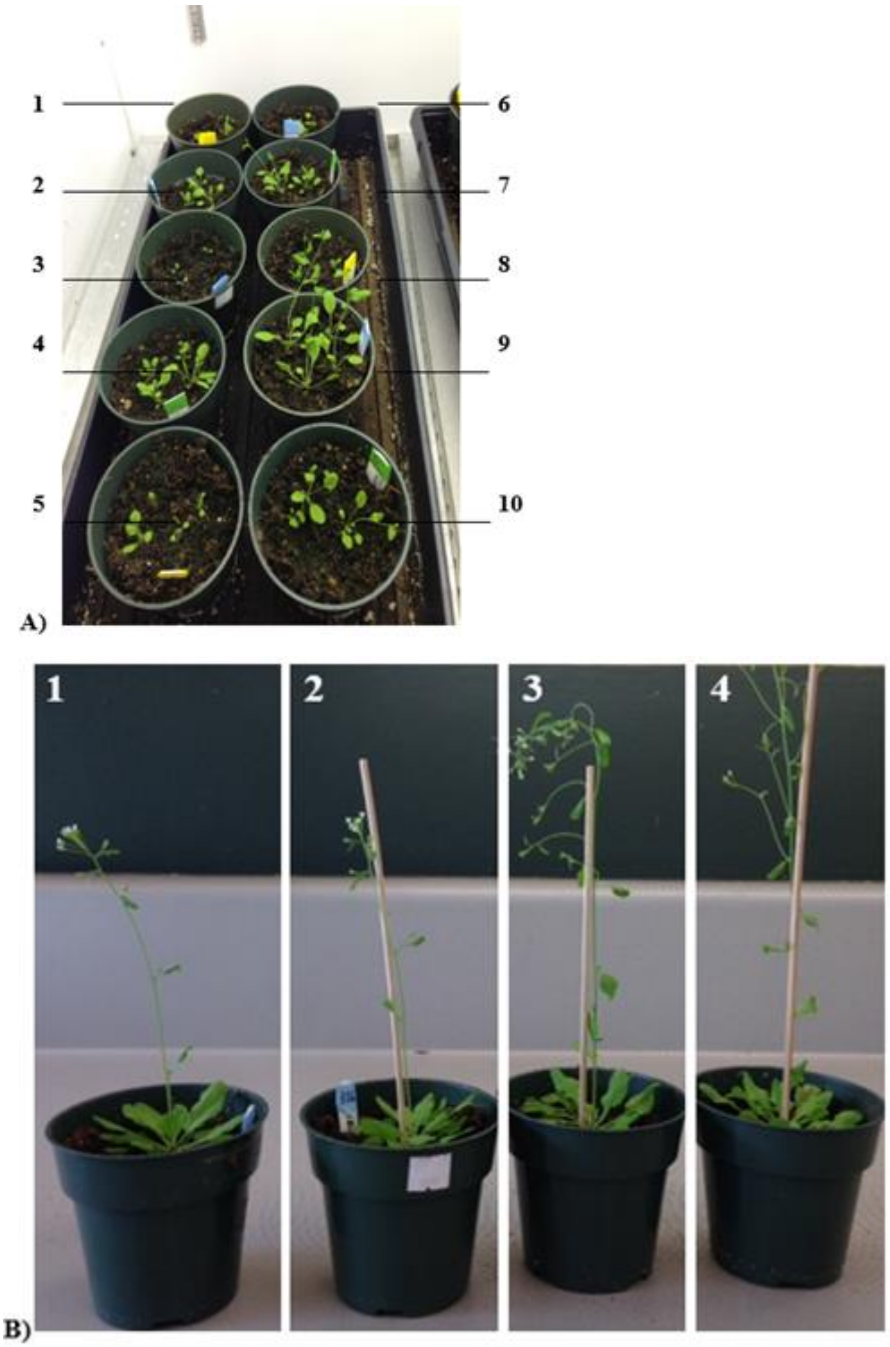


Figure 5.3: Wildtype and transgenic *Arabidopsis* plants: (A) Two week old wildtype (4, 7, 10) and transgenic *Arabidopsis* overexpressing *Poptr*QDH (1, 5, 8) and *Poptr* QDH2 (3, 6, 9) grown

in a greenhouse chamber under a long day cycle. Transgenic *Arabidopsis* occasionally didn't not grow past the initial germination stage (3). (B) Three months old wiltype (3,5) and *Poptr*QDH2 overexpressing *Arabidopsis* (1,2) grown under a short-day cycle. Physical differences were not observed between plants regardless of growth conditions used.

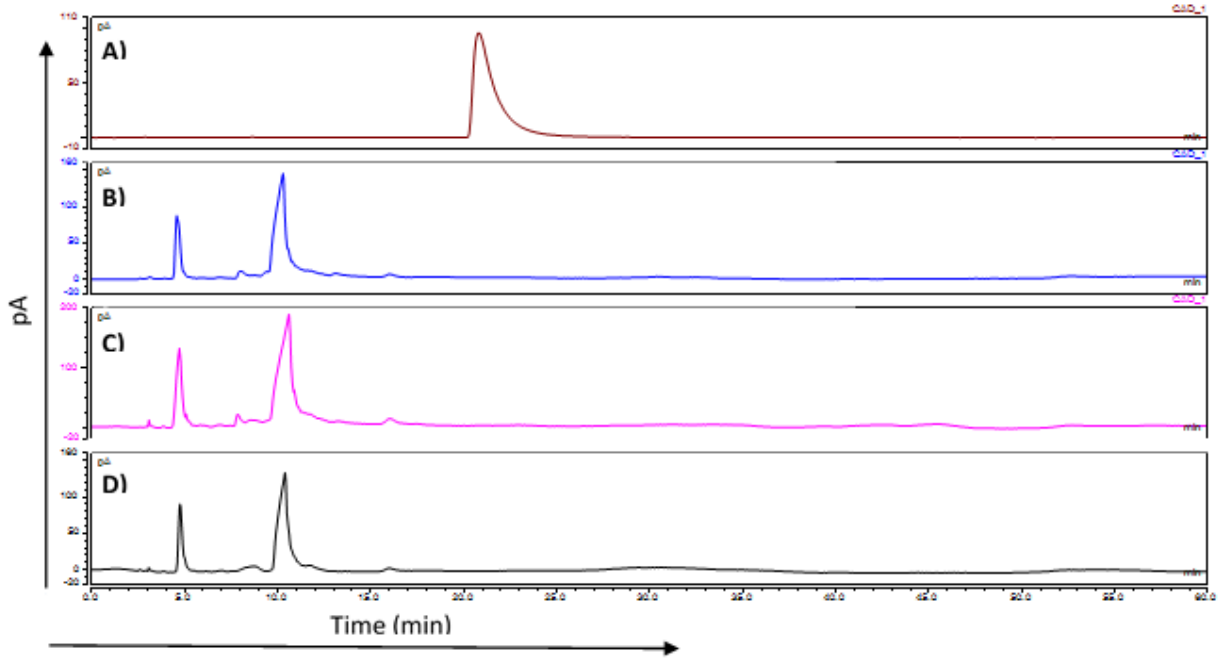


Figure 5.4: LC-CAD analysis of quinic acid standard and organic extracts of Arabidopsis. A 60 ppm quinic acid standard (A) and the acidic extracts of wildtype *Arabiodpsis* (B-D) were analyzed by HILIC-CAD. No quinic acid was observed in the wildtypes. pA; picoAmps

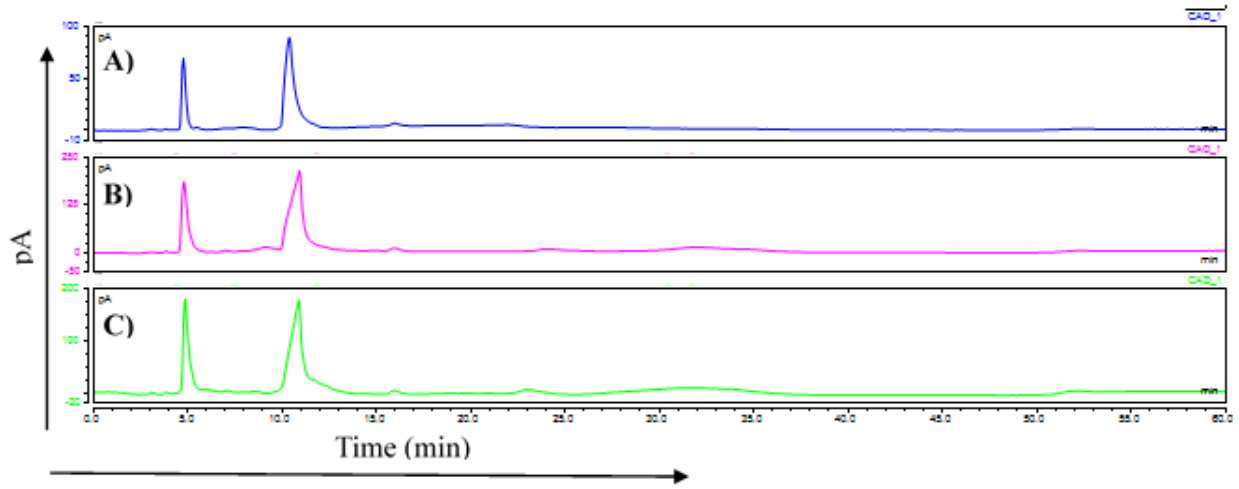


Figure 5.5: LC-CAD organic acid analyses of wildtype and transgenic Arabidopsis. Acidic extracts of three transgenic *Arabiodpsis* (all from the same line, *PoptrQDH* OX Line 1) (A-C) were analyzed by HILIC-CAD. No differences were observed between them or compared to wildtype extracts (above).

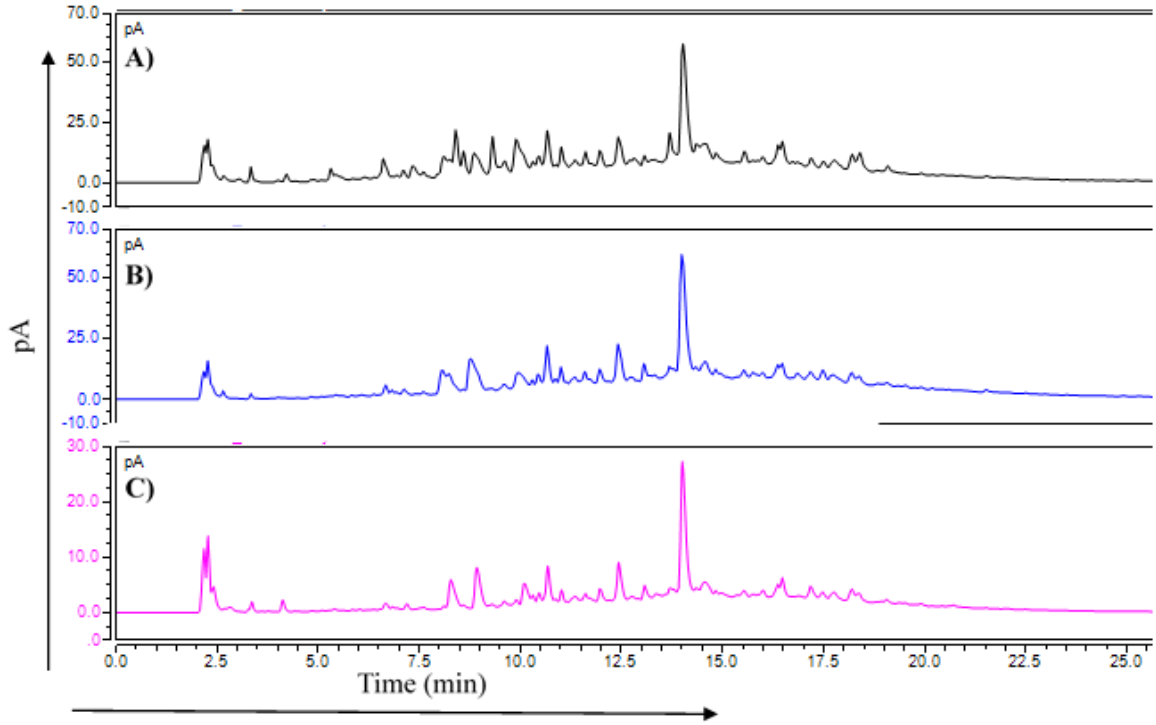


Figure 5.6: LC-UV phenolic analysis of wildtype and transgenic Arabidopsis. Methanol extracts of (A) wildtype and transgenic *Arabidopsis* from *PoptrQDH* OX Line 1 (B) and Line 12 (C). Separation of phenolic compounds was done using RP-HPLC-UV. Note the total run time (60 minutes) is not shown, but no differences were observed between the individuals.

Sections 5.5 to 5.4 summarizes results during method development. While these results were not used for the final analysis, I have decided to keep them in this thesis in case it helps any other students who attempt to develop their own LC-MS methods.

5.2 Metabolite analyses/method development using the HILIC/RP—UPLC-MS

Phenolic screening of wildtype and transgenic *Arabidopsis* was confirmed using UHPLC-MS and a BEH C18 column with a modified gradient adapted from Tang (Personal communications, 2014). The former is more informative (providing mass data on top of UV data) and is more sensitive than regular HPLC-UV. The UHPLC is also optimal for quick surveying of metabolites because of its high-pressure tolerance, enabling shorter run-times. In addition, the autosampler equipped on the UPLC-MS system used here required smaller injection volumes ($\underline{\geq}$ 2μL) than other available instruments ($\underline{\geq}$ 80μL). This allowed fine tuning of the gradient methodology without using up too much plant sample. Methanol extractions from greenhouse grown hybrid poplars, which are known to accumulate the quinate-derived chlorogenic acid as well as wildtype *Arabidopsis* spiked with chlorogenic acid, were performed as technical controls. These ensured that chlorogenic acid was not lost during extractions and that the newly made UHPLC gradient was suitable for its separation and detection. While the positive controls appeared to show no technical errors (Figure 5.7-8) chlorogenic acid could not be found in wildtype or transgenic plants (data not shown).

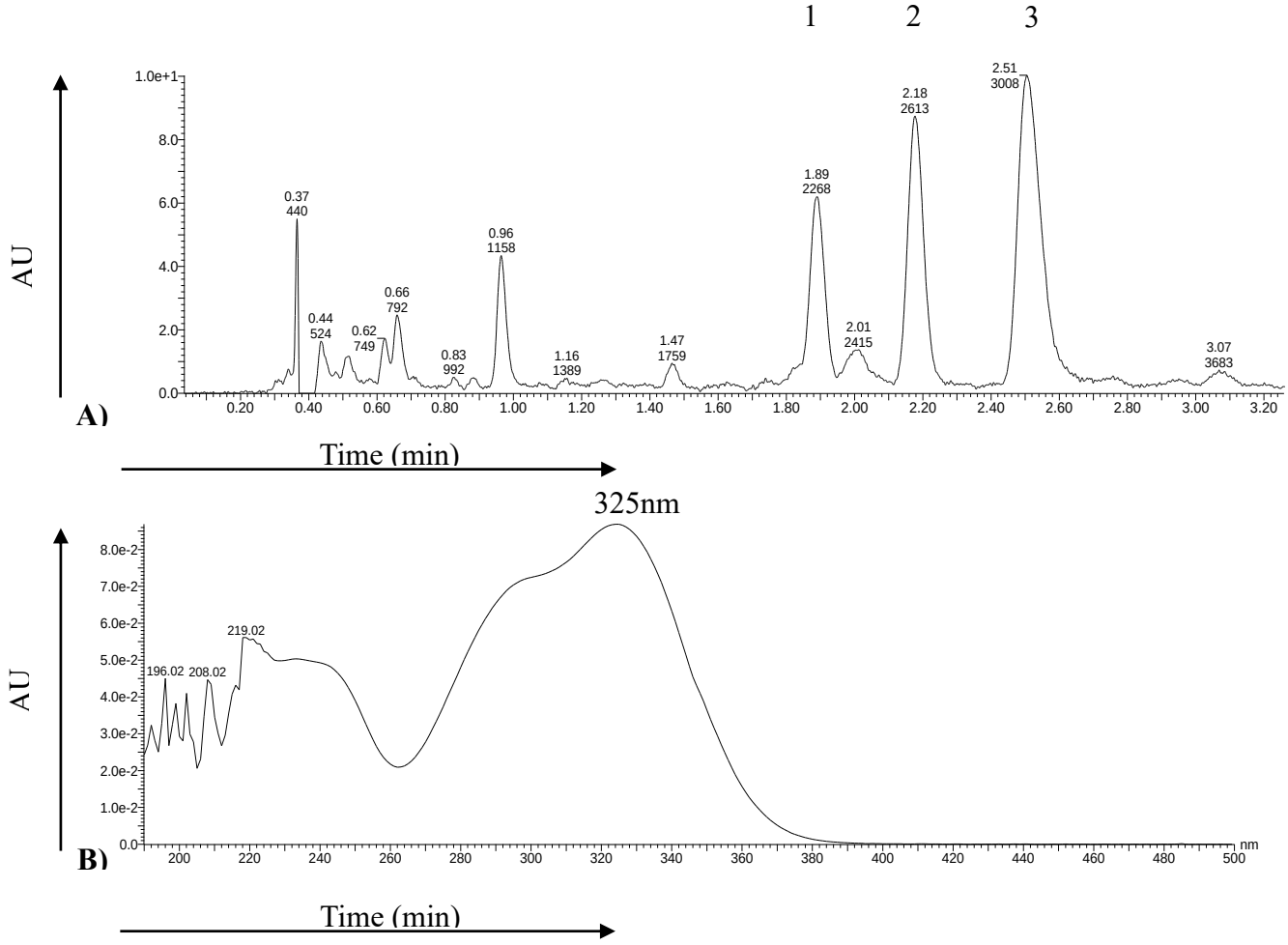


Figure 5.7: Methanol extract of hybrid young 717 analysed using RP-UHPLC-UV-MS. Three peaks (1,2,3) were observed to have the same mass as chlorogenic acid based on single ion monitoring scans (not shown) and UV spectra. A) The diode array chromatogram (190-500nm) of the entire run and (B) at ~1.9 minutes, corresponding to peak 1, is shown. All peaks three peaks showed an absorption maximum at ~325 nm (2 and 3 not shown), which is characteristic of chlorogenic acid. AU; absorbance units

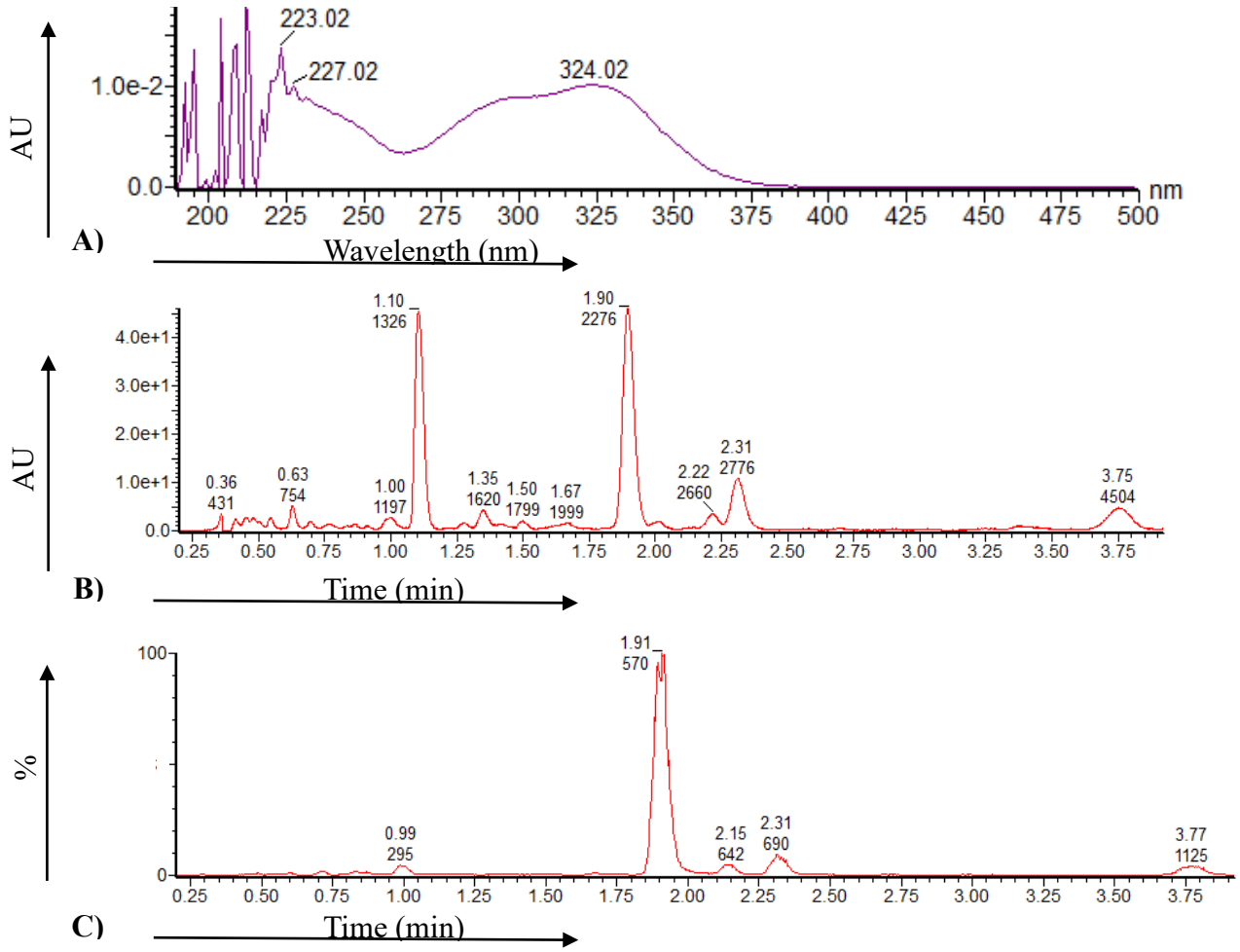


Figure 5.8: UHPLC-MS analysis of the methanol extract of wildtype Arabidopsis spiked with chlorogenic acid. The diode array (A) at 1.9 minutes and the entire run (B) is shown. Based on A), a peak at ~1.9 minutes has a UV maximum at ~325 nm, characteristic of chlorogenic acid. This peak at ~1.9 minutes corresponds to a large peak shown in the chromatogram (C) set to record a mass of 353, or the negative ion of chlorogenic acid (minus H^+).

Although organic acids are poorly retained on C18 columns, some authors have used them to separate quinic acid from plant matrices, albeit using MS/MS for detection. While MS/MS was not available here, it was worth investigating whether the BEH C18 column previously used for phenolic analyses on the UHPLC could also be used to study quinic acid. Several attempts were made to analyze organic acids using RP-UHPLC-MS and gradients were developed based on the methods of Fang *et al*. (2002) who used C18 columns (albeit of different models) and MS/MS to analyze phenolic acids and their precursors in plants. A Kinetic-C18 column, more suitable for detecting polar compounds, and a combination of solvents (mostly H_2O, MeOH and ACN acidified with 0.1-0.4% formic acid) were also tested by modifying a gradient from the column manufacturer's recommendations. However, these did not produce desirable results and quinic acid was observed to have a retention time of < 1 minute under the tested conditions.

5.3 Metabolite analyses/method development using HILIC-LC-Orbitrap

On the Orbitrap, high pH conditions (>pKa of quinic acid) were initially tested using H_2O/ACN and 2.5mM ammonium acetate (pH5.8) as an additive. This method was based on a protocol by Moldoveanu and Davis (2014), who used the same amide column and HILIC-MS/MS to analyze quinic acid and sugar alcohols in tobacco. Retention of quinic acid was favoured under these conditions (Rt~7-11 minutes, depending on flow rate): however, shifts in retention time were observed after ~30 minutes, possibly due to changing pH. Peak shift occurred regardless of the buffering system used (either ammonium bicarbonate/carbonic acid, ammonium formate/formic acid or ammonium acetate/acetic acid). In addition to inconsistent retention times, a second problem with this method was a build up of salt and plant matrix materials in the ion source and capillary tubes probably due to their low solubility in the solvent conditions selected, thereby increasing cleaning times between runs. Thus, an alternative method using low pH conditions—which in the past, did not show shifting peaks—was investigated.

Because the acidic conditions previously used on the LC-CAD system (H_2O/ACN with 4% formic acid) are not MS friendly, this method could not be directly applied to the Orbitrap even though the same HILIC amide column was installed as before. Thus, a modified version of this method, H_2O/ACN acidified with 0.4% formic acid, was tested instead. The peak shape of a quinic acid standard was good, but retention time was further optimized by increasing the concentration of

acid (1% formic acid). Switching to combination of H_2O/MeOH seemed to improve peak resolution over H_2O/ACN. The final method, H_2O/MeOH (90/10, v/v) run isocratically at 0.6 mL/min for 14 minutes enabled analysis of quinic acid and chlorogenic acid.

5.4 Quinic acid and chlorogenic acid analyses with HILIC-Orbitrap-ESI-MS

Quinic acid was shown to have a retention time of ~5.3 minutes and a mass of 191.0546 $\pm$ 0.0019 (Figure 5.9) in negative ion mode based on three technical replicates of eight standards.

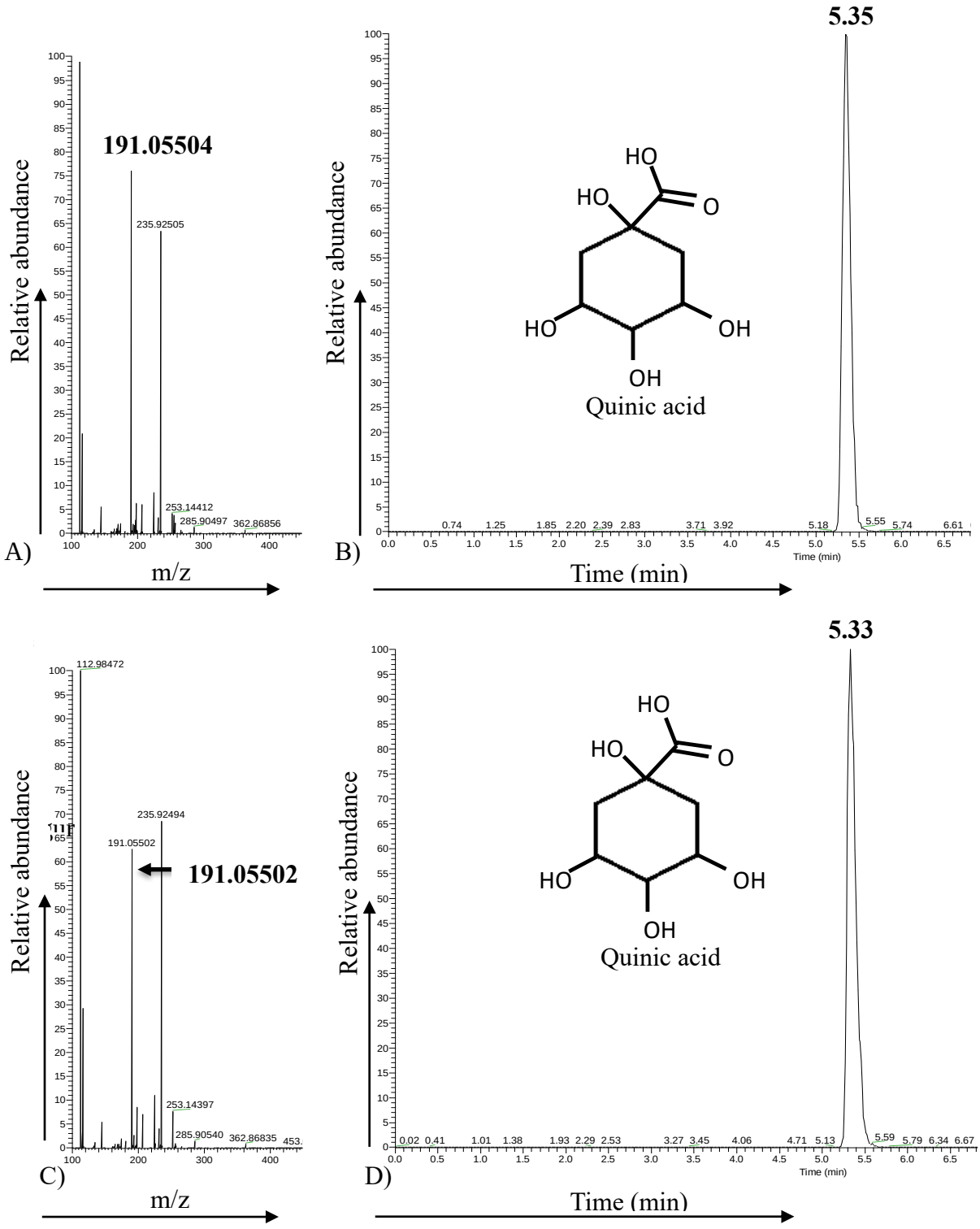


Figure 5.9: LC-Orbitrap analysis of quinic acid. Quinic acid standards 0.1 (A) and 0.008 ppm (B) run ~1 hour apart using HILIC-Orbitrap-ESI-MS shown to elute at ~5.3 minutes. Optimal peak resolution and consistency in retention times was observed.

Quinic acid was not observed in the majority of wildtype and transgenic *Arabidopsis* plants (Figure 5.10). Only two individuals, from *PoptrQDH*2 OX Line 33 and *PoptrQDH* OX Line 36, showed

quinate content similar to those in positive controls (Figure 5.11). Clear peaks were observed in the positive controls [mature *N. benthamiana* and *Z. mays* (Figure 5.12) and 717 *in vitro* plantlets (Figure 5.13)]. A "physiologically relevant amount" of quinic acid was searched for in transgenic plants based on comparisons to amounts found in control plants. However, observed levels of quinic acid ~~observed~~ in the leaves of young 717, and mature *N. benthamiana* and *Z. mays* plants were lower (< 5 ppm) than the accuracy range of a calibrated Orbitrap (Perry *et al*., 2008): thus, quantification of quinic acid could not be performed. These samples also showed considerable variation, likely as a result of the extraction procedure used.

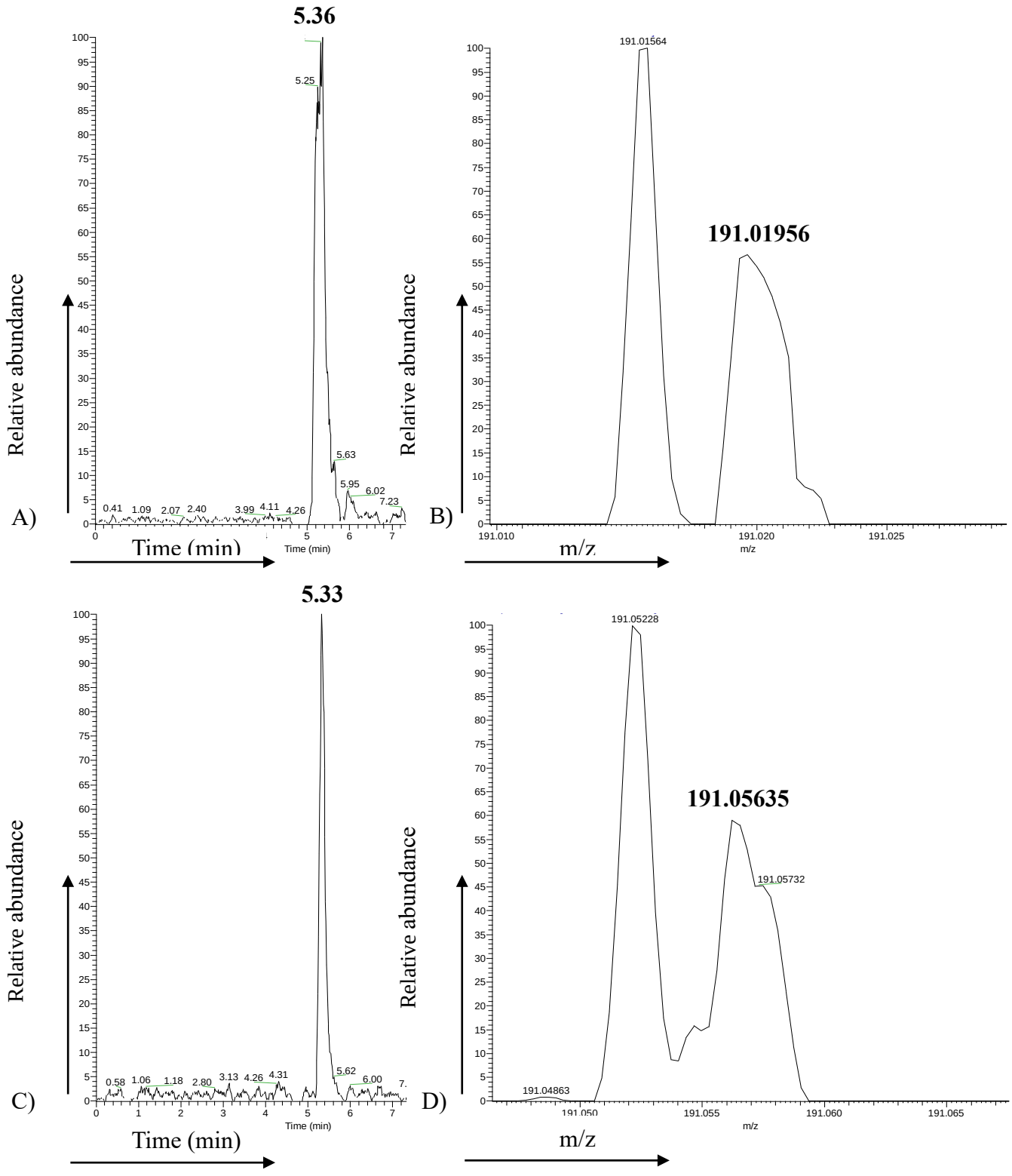


Figure 5.10: LC-Orbitrap search for quinic acid in wildtype Arabidopsis. The selective ion chromatogram (A and C) and spectra (B and D) are shown. In (A) and (C) the software parameters

were set to scan a mass range of 191.04-191.06 m/z, corresponding to the expected mass range of a quinic acid ion in negative MS mode. Within this window, peaks were observed to elute at ~5.3 minutes which is the expected elution time of quinic acid. The corresponding mass spectra of these peaks are shown to the left (B and D): however, the peaks are poorly resolved as a result of low concentration. Note the spectra are normalized to the highest peak shown and smoothing was applied to chromatograms (A and C) to enhance visualization.

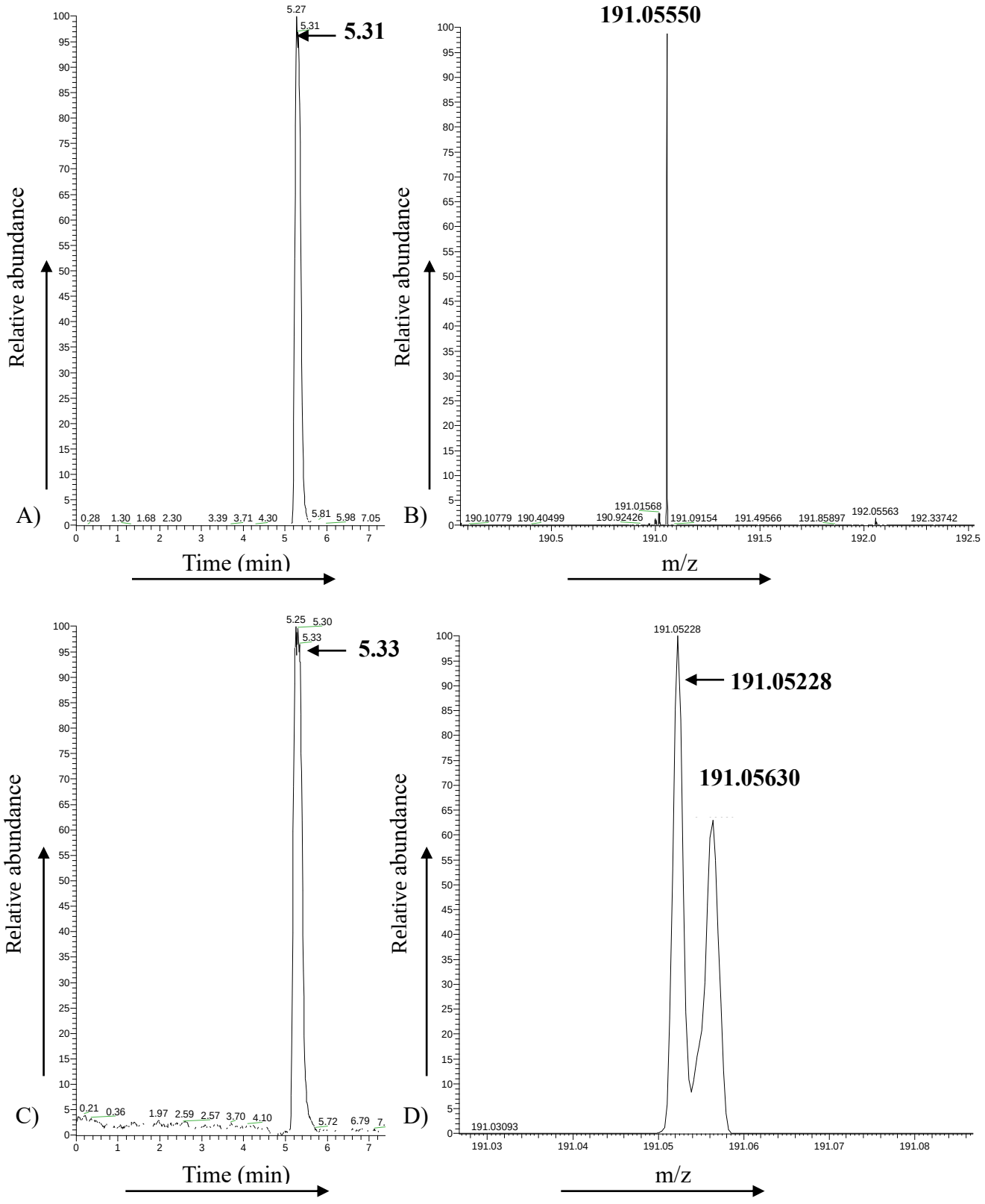

A)
Relative abundance
Time (min)
5.27
5.31
5.31
0.28 1.30 1.68 2.30 3.39 3.71 4.30 5.81 5.98 7.05
B)
Relative abundance
m/z
191.05550
190.10779 190.40499 190.92426 191.01568 191.09154 191.49566 191.85897 192.05563 192.33742
C)
Relative abundance
Time (min)
5.25 5.30
5.33
5.33
0.21 0.36 1.97 2.59 2.57 3.70 4.10 5.72 6.79
D)
Relative abundance
m/z
191.05228
191.05228
191.05630
191.03093

Figure 5.11: LC-Orbitrap search for quinic acid in transgenic Arabidopsis. The selective ion chromatograms of an individual from *PoptrQDH2* Line 33 (A) and *PoptrQDH* Line 36 (C) which had the highest quinic acid content among the analysed *Arabidopsis* samples and consequently, whose spectra (B and D) were relatively sharper than the rest, are shown. The spectra are normalized to the highest peak shown and smoothing was applied to chromatograms (A and C) to enhance visualization.

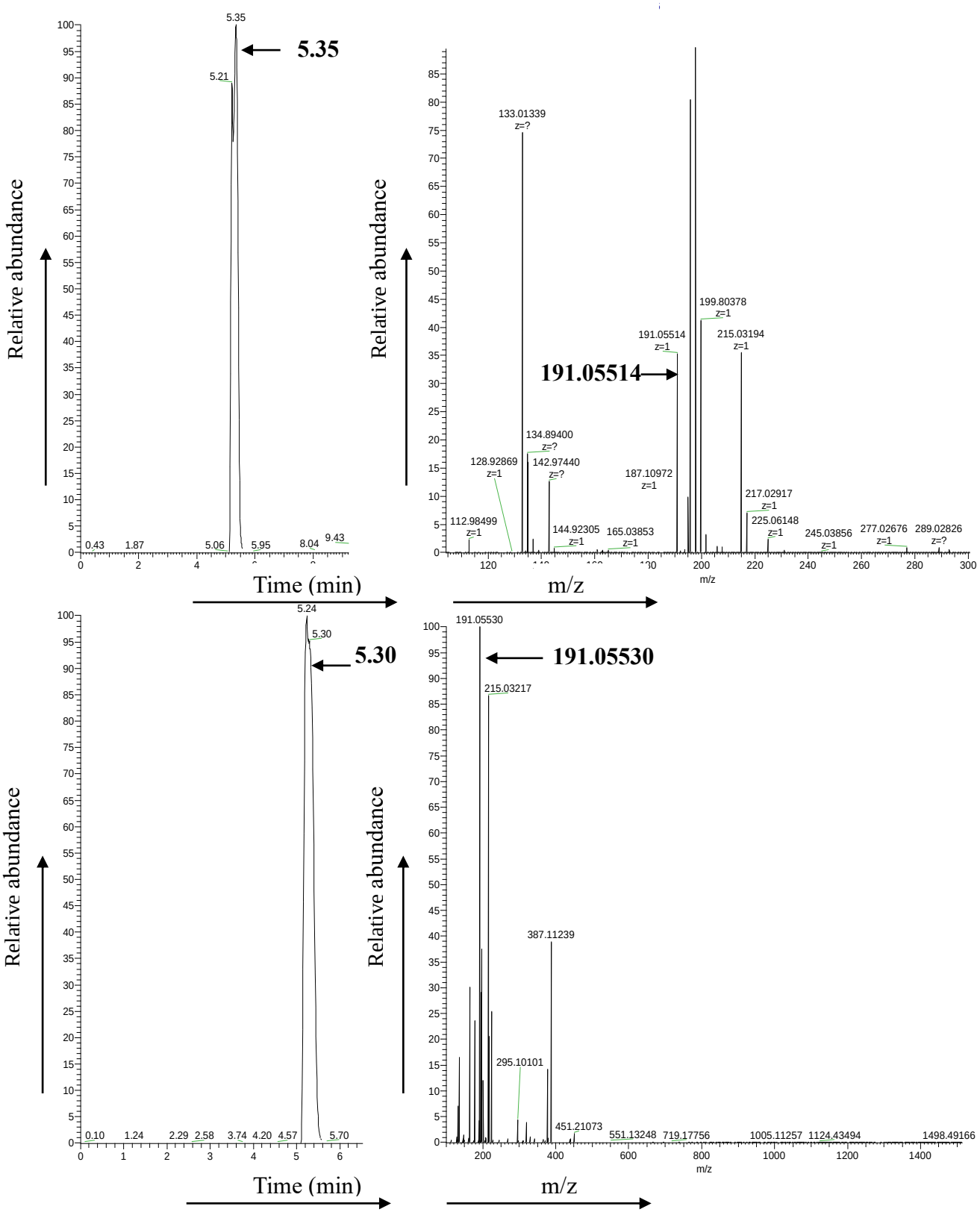


Figure 5.12: LC-Orbitrap analysis of quinic acid in positive control plants. Chromatograms

of *N. benthamiana* (A) and *Z.mays* (B) serving as positive extraction controls and their corresponding spectra (C and D) are shown. A sharp peak corresponding to quinic acid (~191.055 m/z, Rt. ~5.3 min) can be observed without normalization. Smoothing was applied to the chromatograms

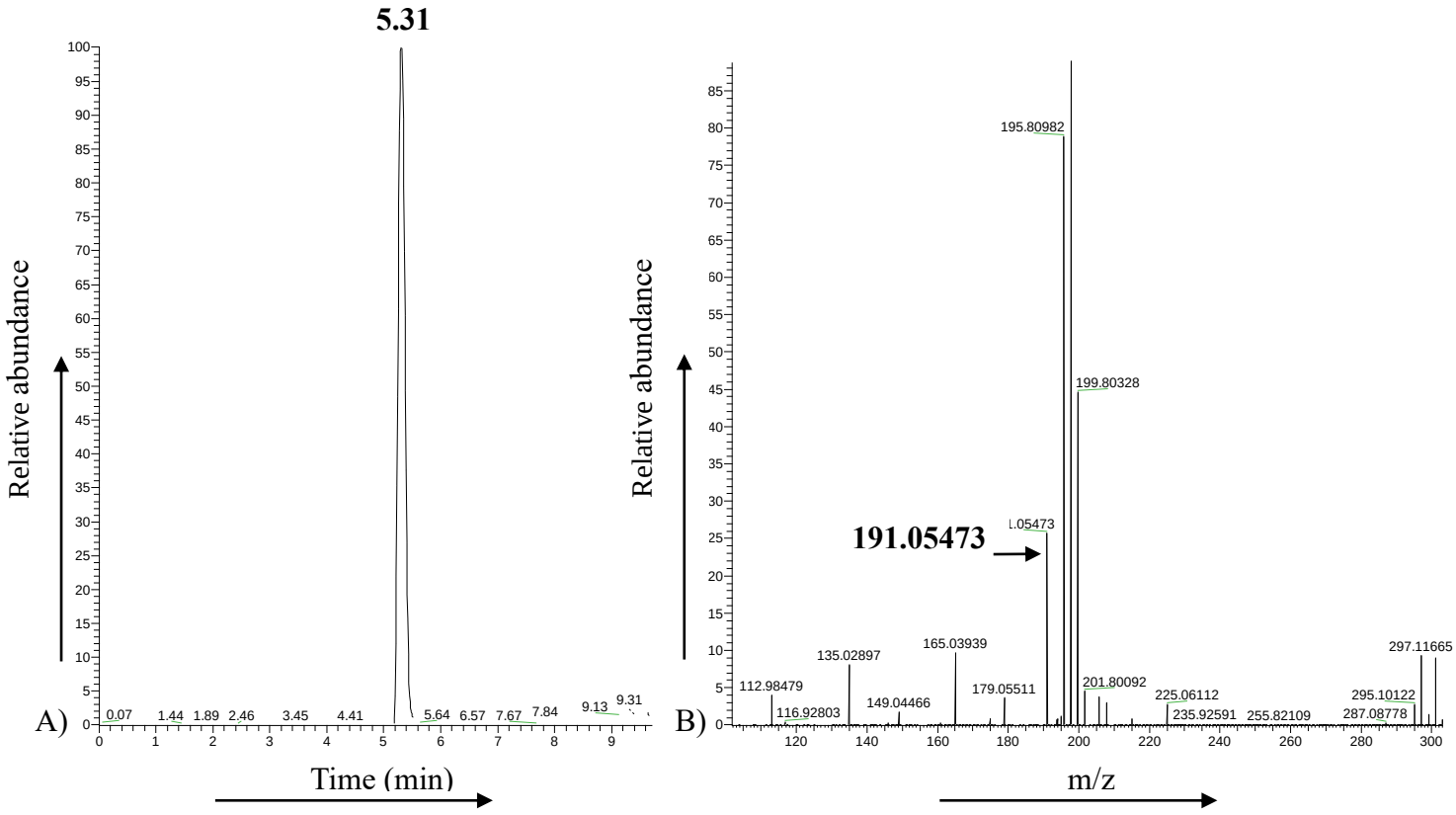


Figure 5.13: LC-Orbitrap analysis of quinic acid in a positive control plant. The chromatogram (A) and spectra (B) of the methanol extract of a 717 *in vitro* plantlet used as a positive control is shown. A peak eluting at 5.31 minutes was observed in the mass range of quinic acid. Smoothing was applied to the chromatogram.

Both a chlorogenic standard (Figure 5.14), and a peak corresponding to it in a 717 positive control

plant (Figure 5.15) was shown to elute at ~11.6 minutes and have a mass of ~353.08. Chlorogenic acid was observed at higher concentrations (determined by interpolating from a standard curve) in control plants than quinic acid (Table 5.1). The concentrations of the most abundant chlorogenic isomer in the leaves of an *in vitro* 717 plantlet and two four months old *N. benthamiana* plants were observed to be 269 and 951 and 1054 ppm respectively. This peak eluted at ~11.6 minutes. Two other major peaks at about 9.4 minutes and 10 minutes were also observed with the same expected mass and spectral properties as chlorogenic acid. Most of the *Arabidopsis* plants tested did not appear to contain chlorogenic acid (Figure 5.16). However, this may likely be an underestimate of true values since each sequence run was initially 10 minutes long. Earlier it was believed this was enough to detect chlorogenic acid until the third isomer at ~12 minutes was later observed, and the run time was appropriately extended. During the extended runs, one transgenic plant (*PoptrQDH* OX Line 10) showed a peak with a similar retention time, expected mass and UV spectra properties as the chlorogenic acid isomer eluting at ~12 minutes (Figure 5.17).

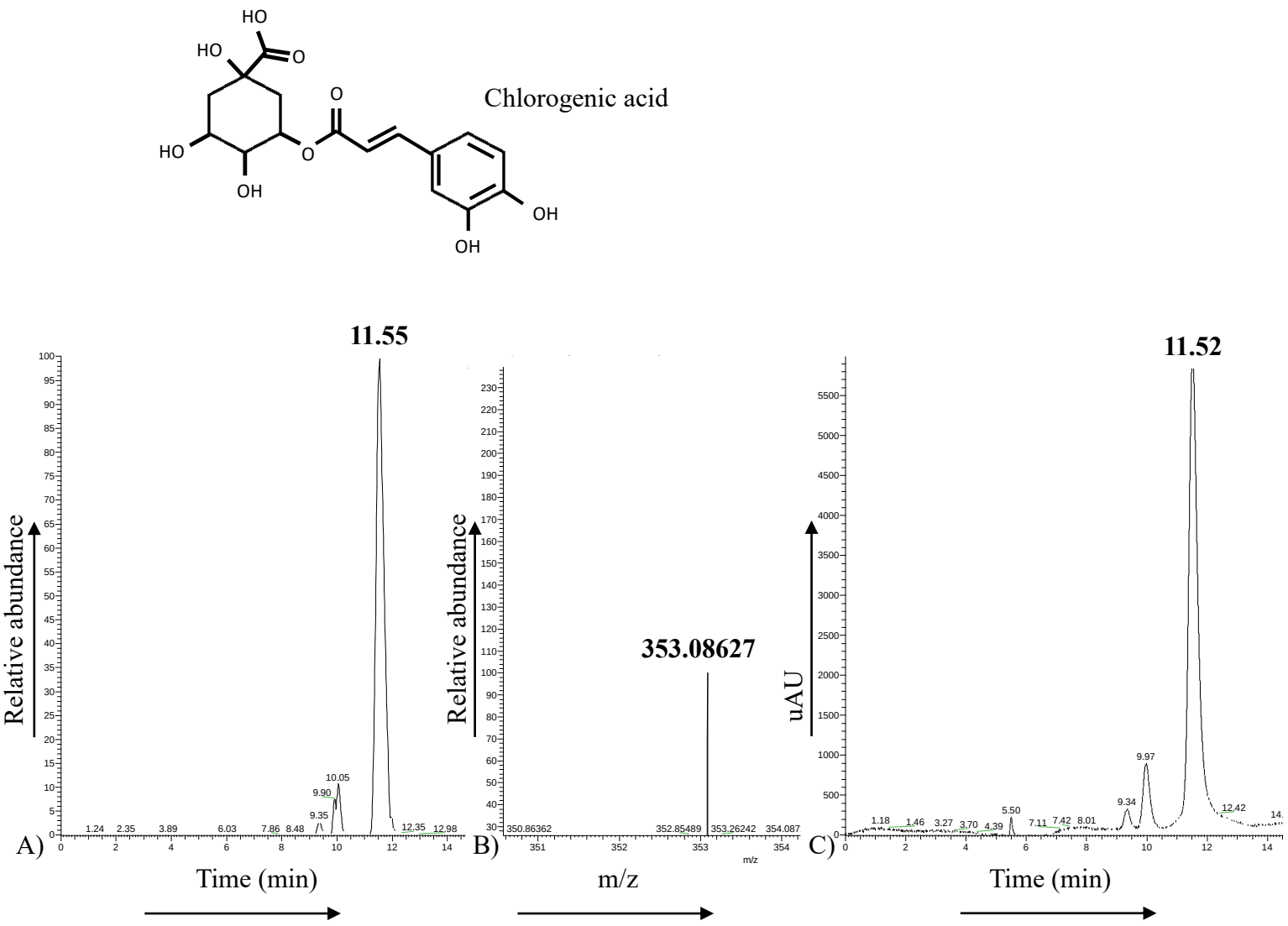


Figure 5.14: LC-Orbitrap analysis of a chlorogenic acid standard. (A) A chromatogram scanning at a mass range of 353.08-353.1 is shown. The chromatogram showed a peak eluting at ~11.6 min. The spectrum at approximately this time frame shows one peak with the expected mass of chlorogenic acid (B). This peak also showed absorbance at 325 nm (C), which is a yet another characteristic of chlorogenic acid. Smaller peaks eluting at ~9.4 min and 10.0 min are also shown to have a mass of ~353.08 (not shown) and absorb at 325 nm and are therefore likely isomers of chlorogenic acid.

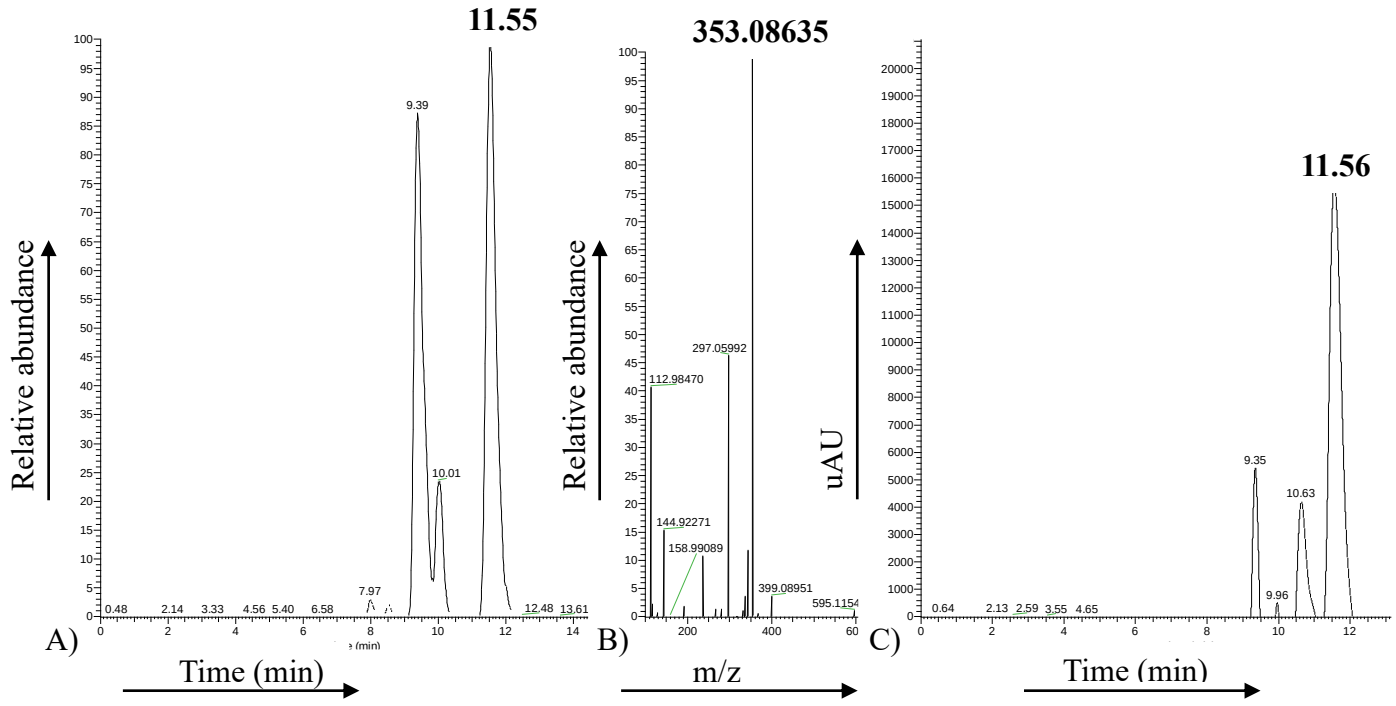


Figure 5.15: LC-Orbitrap analysis of chlorogenic acid in young 717. (A) The chromatogram showed a peak eluting at ~11.6 min, similar to the elution time of a chlorogenic standard. The spectrum at approximately this time frame shows one peak with the expected mass of chlorogenic acid (B). This peak also showed absorbance at 325 nm (C), which is another characteristic of chlorogenic acid.

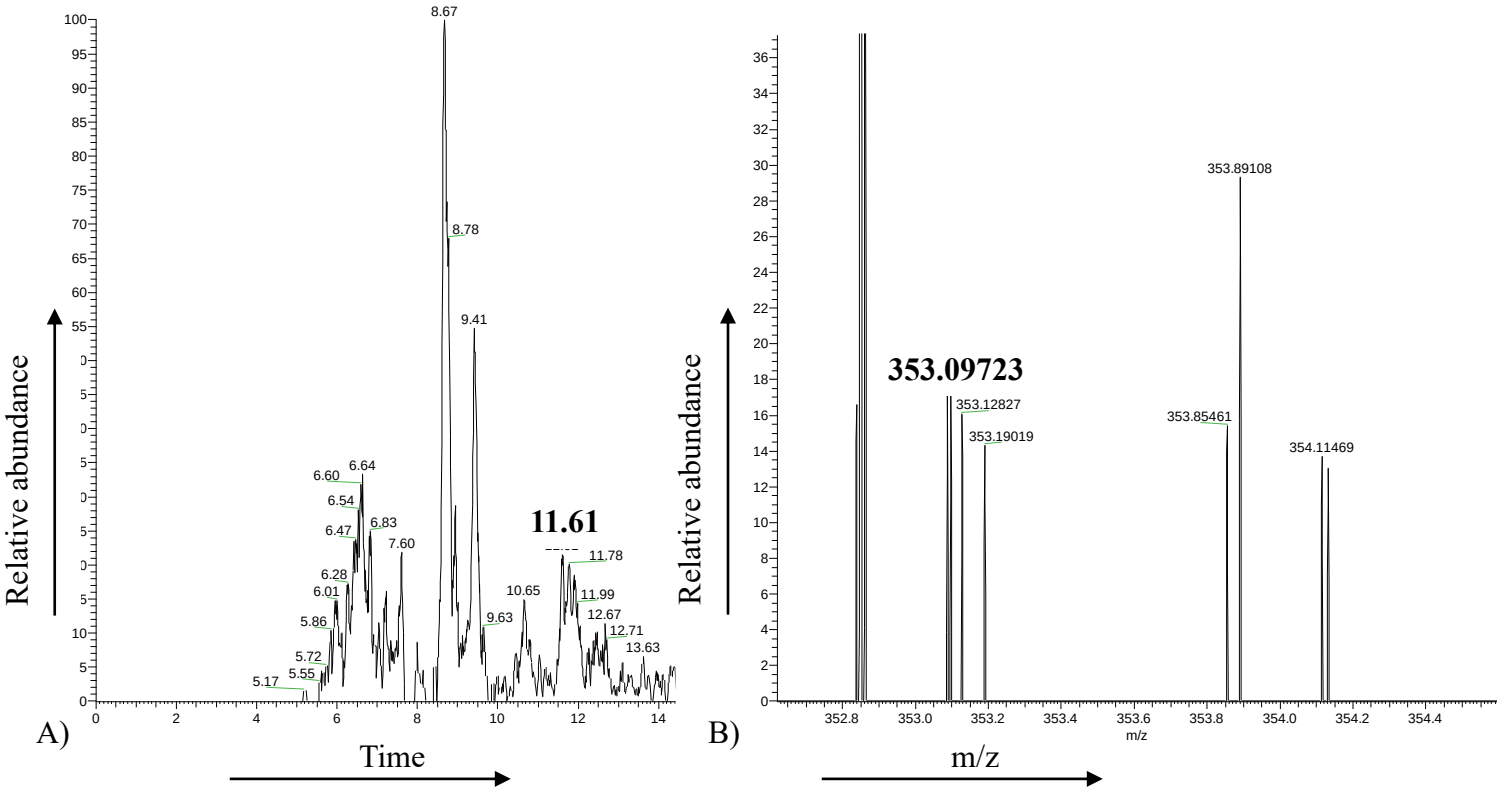


Figure 5.16: LC-Orbitrap search for chlorogenic acid in wildtype Arabidopsis. (A) A chromatogram scanning at a mass range of 353.08-353.1 is shown. The chromatogram showed a peak eluting at ~11.6 min. Here the spectrum (B) at this time does not show a mass corresponding to chlorogenic acid (~354.08 min) as the closest mass is 353.09723 (however, a magnified version of this spectrum is shown in the appendix, where peaks more similar in mass to chlorogenic acid are shown). This peak was not observed to absorb at 325 nm (not shown) and therefore does not appear to be chlorogenic acid. Similar results were obtained for other wildtype and most transgenic plants.

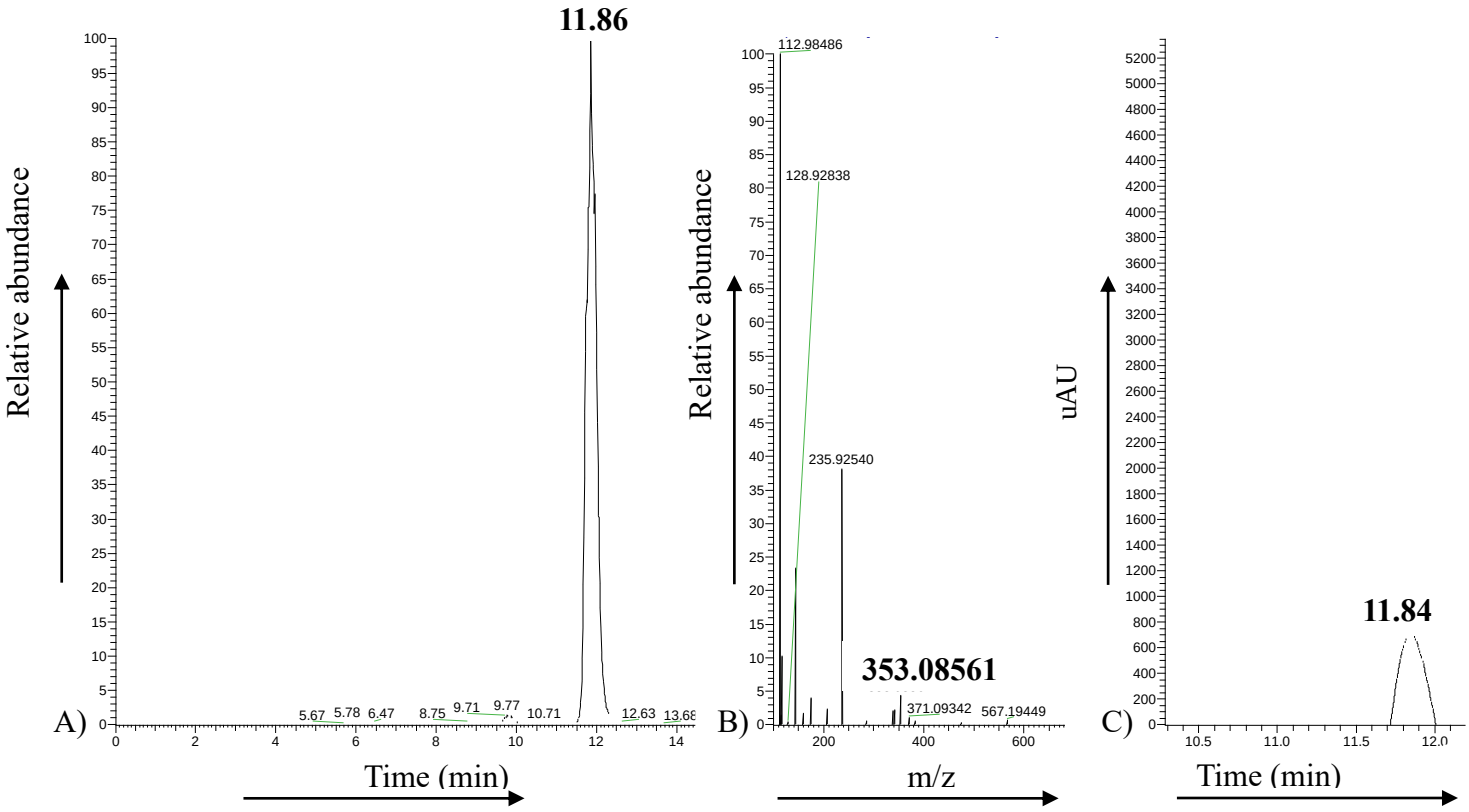


Figure 5.17: LC-Orbitrap analysis of chlorogenic acid in transgenic Arabidopsis. (A) A chromatogram of the leaf methanol extract of a transgenic *Arabidopsis* (*PoptrQDH* Line 10) scanning at a mass range of 353.08-353.1. The chromatogram showed a peak eluting at ~the expected elution time as one of the three chlorogenic acid isomers. The spectrum at approximately this time frame shows one peak with the expected mass of chlorogenic acid (B). This peak also showed absorbance at 325 nm (C), which is characteristic of chlorogenic acid.

Table 5.1: LC-Orbitrap analysis of chlorogenic acid in leaf methanol extract of wildtype and transgenic Arabidopsis. Plants including two months old *in vitro* 717's and four months old *N. benthamiana* served as extraction controls and to depict "physiologically relevant" amounts of chlorogenic acid. Methanol extracts were concentrated overnight and respuspended in 10% methanol to a final concentration of 10mg/mL which were then injected onto a HILIC column coupled to an Orbitrap-MS. At low concentrations (<5 ppm), poor peak resolution made it difficult to determine exact retention times and mases. Estimated concentrations were determined by interpolating from a standard curve. A standard of chlorogenic acid was observed to have a mass of ~353.08627 in negative ion mode, an elution time of ~11.6 minutes and to absorb at 353 nm

Sample	Rt (min)	*m/z* (negative ion mode)	Concentration (ppm)	Absorbs@325nm?
WT *Arabidopsis*	11.6	353.08097, 353.08408	0	No
PoptrQDH2 Line 2	11.7	353.07984, 353.08462	0	No
PoptrQDH2 Line 6	11.8	353.08058, 353.08498	0	No
PoptrQDH Line 10	10.5-12 (indistinct)	353.08133	0	No
PoptrQDH Line 1	11.8	353.08457	0	No
PoptrQDH Line 10	11.9	353.08561	13	Yes
N. benthamiana	11.6	353.08643	951	Yes
N. benthamiana	11.5	353.08657	1054	Yes
717	11.6	353.08635	259	Yes

5.4 Discussion

5.4.1 Summary

In spite of the growing interest in the biochemical diversity of plant *SDH* gene copies, few works so far have focused on their biological roles. In this study, *P. trichocarpa QDH genes* were introduced into *Arabidopsis* via *Agrobacterium*-mediated transformation. The model plant lacks its own *QDH* genes and is not known to produce quinate or its derivatives. All recent works on plant *QDH* genes have involved their *in vitro* characterization (e.g. Carrington *et al*., 2018; Gritsunov *et al*., 2018; Guo *et al*., 2014). These studies suggest *QDH* genes endow quinate biosynthetic activities to plants. Whilst their actual sequences are unknown, earlier investigations since the 1970's have also characterized QDH proteins isolated from crude extracts of mungo beans (Minamikawa, 1977), carrots (Refeno *et al*., 1982) and peas (Leuschner *et al*., 1995) capable of catalyzing the NAD(P)(H)-dependent interconversion of quinate to dehydroquinate. For this reason it made sense to search for novel quinate biosynthetic activities in transgenic *Arabidopsis* heterologously expressing *QDH*. Endogenous quinate exists primarily conjugated to caffeic acid—as chlorogenic acid (also mentioned in chapter one)—rather than in its free form in plants where it occurs. To understand the complete effects of introducing *QDH* genes, both quinate in addition to chlorogenic acid and any other changes to the normal metabolomes of wildtype *Arabidopsis* were therefore searched. This study is the first to report a prototype method for the simultaneous determination of quinic acid and chlorogenic acid in transgenic *Arabidopsis* using HILIC-MS.

5.4.2 Transgene silencing in *PoptrQDH* and *PoptrQDH2* lines

Ectopic expression of *PoptrQDH* and *PoptrQDH2* was successfully confirmed in F1 and F2 *Arabidopsis* plants indicating successful delivery and stable integration of transgenes. In contrast, nearly complete silencing of these transgenes were observed in later generations. Both variation in the expression of, and silencing of exogenous genes is the common bane of plant researchers (Matzke and Matzke, 1995; Schubert *et al*., 2004). Although data pertaining to the latter is somewhat at odds, transgene silencing appears to occur epigenetically rather than by direct alterations to the DNA sequence. In general, transgene silencing results either from stopped transcription (transcriptional gene silencing) or transcript degradation (posttranscriptional gene silencing). Silencing can result from positional effects of DNA integration. For example, when

transgenes are arranged as inverted repeats, the encoded transcripts can potentially form hairpins that are recognized and targeted by small interfering RNA (siRNA) silencing machinery (Schubert *et al*., 2004).

Another speculation is that transgene silencing is determined by promoter strength and gene copy number. One study conducted on the expression of various reporter gene constructs introduced into *Arabidopsis* using the T-DNA delivery system, found that homozygous transgenics showed higher transgene expression than hemizygotes. However, the correlation between transcript abundance and gene expression became negative once a certain threshold of transcripts was reached and exceedance of this limit activated transcript degradation pathways. So far it is unknown what determines this cut-off value since the accumulation of some transgene transcripts appear to be more tolerable than others. It may be sequence dependent: however, overall this implies the risk of dosage-dependent transgene silencing may be higher when using strong (e.g. CaMV353) rather than weak promoters (Schubert *et al*., 2004).

On the other hand, *cis* transcriptional gene silencing can result when DNA integrates near highly methylated loci or transcriptionally inactive heterochromatin (Fagard and Vaucheret, 2000; Rajeevkumar *et al*., 2015) . However, in some cases methylation occurs regardless of the methylated states of neighbouring genes, which likely involves an inherent defense mechanism of plants to find and deactivate foreign invading DNA (Matzke and Matzke, 1995). Thus, one short coming of genetic transformation is the unpredictability of transgene silencing. Not only this, but methylation patterns and the deactivated states of genes are stably passed from parent to offspring. One study showed the expression of GUS: promoter constructs in transgenic rice gradually decreases over generations. While parent (T_0) lines actively expressed GUS in vascular and epidermal tissues (albeit at varying levels), subsequent generations showed progressive silencing, first in the vasculature followed by epidermal tissues. Loss of GUS expression was correlated with a steady increase in CG methylation of the promoter (Klöti *et al*., 2002).

5.4.3 Quinic acid is present in positive control plants but not in the majority of wildtype and transgenic *Arabidopsis*

Strikingly, peaks with a retention time and m/z value matching that of quinic acid were detectable in the leaf methanol extracts of both wildtype and transgenic *Arabidopsis*. This result was unexpected due to the absence of *QDH* genes in the *Arabidopsis* genome. However, it should be emphasized their levels were very low and beyond the linear range of instrument calibration. Outside this range, mass accuracy is compromised such that it becomes difficult to accurately predict ions. In this study, *Nicotiana benthamiana*, *Zea mays* and hybrid poplar served as positive extraction controls because they do possess *QDH* genes and are known to accumulate quinate-derivatives. Comparisons to levels of quinate in these plants suggest the observed amounts in wild-type and transgenic *Arabidopsis* are lower than physiological concentrations. Yet in one individual (*PoptrQDH2* OX Line 33) the level of observed quinate was similar to those extracted from positive control plants. A similar but slightly lower amount of quinic acid was observed in another individual of a different overexpressing line (*PoptrQDH* Line 2 36) which again resembled quinic acid levels in positive controls.

Given that the analyzed F2 plants all appeared to express either transgene, it is unclear why quinic acid was not observed in most of them. Lack of quinic acid may be attributable to post-translational problems. The degradation of recombinant proteins is a commonly encountered issue in transgenic plants. Plants are abundant in proteases (*Arabidopsis* for example has 1900 protease encoding genes) that are present across cellular compartments. Factors determining hydrolysis of recombinant proteins is shrouded in mystery but appears to be dependent on their amino acid sequences. Some may be improperly folded in which case they become automatically targeted for degradation by the ubiquitin proteasome system (Pillay *et al*., 2013). Protein degradation can be detected using Western Blot analyses, specifically by looking for protein bands that are smaller than expected (Doran, 2006).

5.4.4 Chlorogenic acid is present in positive control plants and possibly in one transgenic *Arabidopsis*

A surprising finding of this study was the determination of chlorogenic acid in one transgenic plant (*PoptrQDH* 10). Unlike its precursor, chlorogenic acid has an absorption maximum of 325 nm

such that it can be identified with greater confidence than quinic acid using a combination of retention time, *m/z* value and UV spectra data. In this individual, and in control plants, levels of chlorogenic acid were higher than quinic acid and could be quantified using a standard curve. Previous reports have determined chlorogenic acid as the major peak in the phenolic profiles of plants including coffee (Clifford *et al*., 2006), plums (Fang *et al*., 2002) and *N. benthamiana* (Ncube *et al*., 2014). In Sorbus species and *N. benthamiana*, levels of chlorogenic acid can reach as much as 21.91 and 28.4mg/g dry weight respectively (Gaivelyte *et al*., 2013; Ncube *et al*., 2014). Less work has focused on quantifying quinic acid, which seems to be present at lower levels. The quinic acid content of angiosperms ranges from 0-~8mg/g fresh weight following the general trend that its levels are higher in woody plants than herbaceous plants (Yoshida *et al*., 1975). The results of this study follow these trends: however, it should be noted that the extraction procedure used was more suitable for phenolic compounds than organic acids. For this reason, some quinic acid is expected to have been lost and the data obtained can be regarded as more qualitative than quantitative.

5.4.5 Biosynthetic pathways for chlorogenic acid

The biosynthesis of chlorogenic acid has been a matter of debate and is thought to occur either through the esterification of caffeoyl CoA with quinate or through the esterification of *p*-coumaroyl CoA with quinate followed by hydroxylation. In solanaceous plants (e.g. tomato and tobacco) for example, it is thought to occur via the former pathway catalyzed by hydroxycinnamoyl-coenzyme A quinate hydroxycinnamoyl transferase (HQT). Supporting this idea, tomato plants overexpressing HQT produce higher levels of chlorogenic acid than wildtype while conversely silencing of HQT leads to its decreased concentrations (Niggeweg *et al*., 2004). The alternative route involves two steps: esterification of p-coumaroyl CoA with quinate by HQT followed by hydroxylation of the product (p-coumaroyl quinate) to caffeoylquinate by a p-coumaroyl quinate/shikimate 3'-hydroxylase (C3'H) (Jeszka-Skowron *et al*., 2016). The fact that an HQT is involved in both pathways creates a bit confusion. However, this discrepancy is somewhat resolved by the presence of two HQTs in artichoke, which might allow for the separate production of caffeoylquinic acid and p-coumaroyl quinic acid using two different enzymes (Sonnante *et al*., 2010).

Arabidopsis is not known to endogenously produce quinic acid or its derivatives and lacks genes encoding HQT (Niggeweg *et al*., 2004) yet other members of the Brassicaceae family (e.g. kale, Brussel sprouts) appear to synthesize chlorogenic acid. In addition, chlorogenic acid has been determined in *B. napus* which was recently reported to encode an SDH duplicate encoding quinate biosynthetic activity *in vitro* (Gritsunov *et al*., 2018). This points to *Arabidopsis* being unique in the family, having likely lost the ability to synthesize quinate and its derivatives. However, *Arabidopsis* does possess a homolog of HQT: hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase (HCT). Both *Arabidopsis* HCT and C3H reportedly accepts quinate and shikimate (Hoffmann *et al.*, 2004) and p-coumaroyl quinate and p-coumaroyl shikimate (Schoch *et al*., 2001) respectively as substrates *in vitro*. Importantly, this provides a potential route towards chlorogenic acid biosynthesis (Niggeweg et al., 2004) and opened the possibility that chlorogenic acid might accumulate in *QDH* overexpressing *Arabidopsis* (Figure 5.18). However, with one exception, no differences were observed between the phenolic profiles of wildtype and transgenic *Arabidopsis*. Given that *in vitro* assays are performed using saturating substrate conditions, it is unsurprising that product profiles *in vitro* do not perfectly match those *in vivo*. C3H from *Lonicera japonica* has about a twofold higher affinity for p-coumaroyl shikimate (K_M 9 $\pm$2 μM) than p-coumaroyl quinate (20$\pm$2 μM) (Pu *et al*., 2013). Assuming that *Arabidopsis* HCT and C3H also prefer to accept shikimate and p-coumaroyl shikimate as substrates, it is unlikely that quinate derivatives will accumulate to detectable levels in transgenic plants. However, since an extended run was not performed for the majority of plant samples, it is possible these simply went unnoticed.

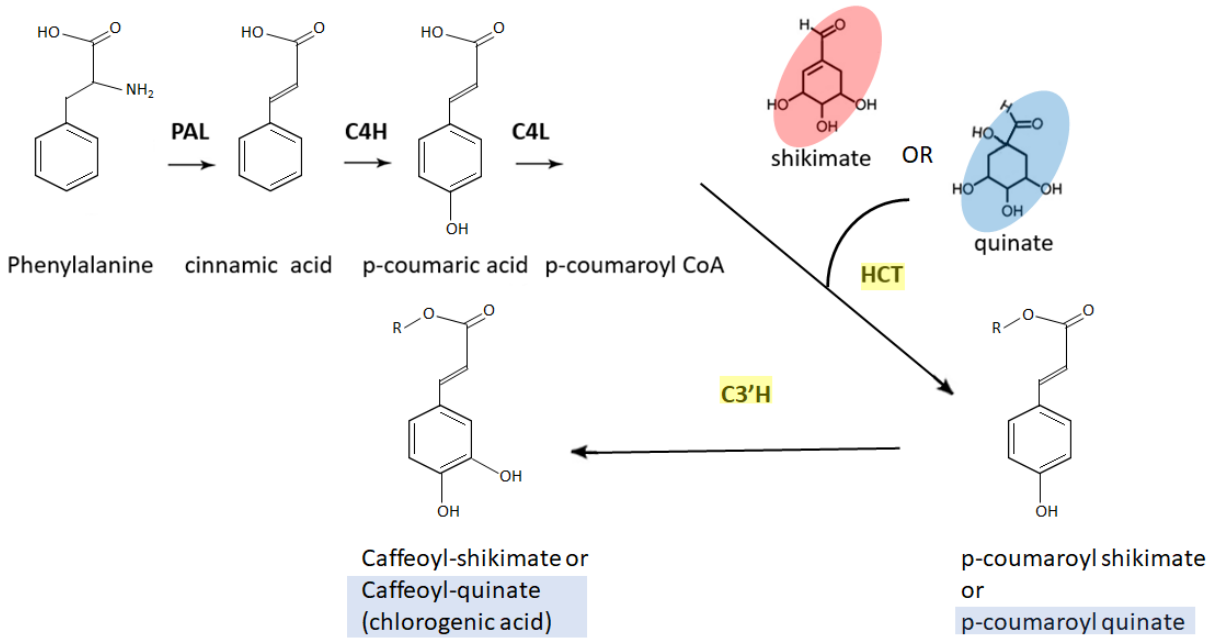


Figure 5.18: Proposed pathway of chlorogenic acid biosynthesis in transgenic *Arabidopsis*.

5.4.6 Conclusion

A major limitation of this study was small sample size due to unexpected gene silencing (some F3 generation and all of the F4 generation offspring showed almost complete silencing of *PoptrQDH* and *PoptrQDH2* either due to segregation or (post) transcriptional modifications) and because most plant materials were used up during method development. One drawback of the Orbitrap (or at least the injection needle used here) compared to the UHPLC, is that it requires a higher injection volume (~100µL versus 2µL). All the leaves from three *Arabidopsis* plants from the same line produced just enough sample to analyze either quinic acid or phenolic compounds in a single technical run. Therefore, the extraction of both compounds had to be performed simultaneously. Isolation of phenolic acids includes a solid phase extraction (SPE) step that helps to separate other compounds that might also absorb in the UV range, and that can potentially complicate data analysis. Solvents (e.g. methanol) used for extracting phenolic compounds also works well for extracting quinic acid as reported previously (Mendonça *et al.*, 2008; Ng *et al.*, 2004; Ouni *et al.*, 2011): however, earlier tests showed that some quinate is lost during SPE. Thus, it should be taken into consideration that quinic acid extraction could have been further optimized had it not been desirable to simultaneously analyze phenolic compounds. If sample limitation is not an issue in the future, this method can be optimized by performing separate extractions for quinic acid and

chlorogenic acid. To isolate quinic acid, methanol extractions can be performed without SPE or alternatively using acidified water and/or an anion exchange column. Since chlorogenic acid could be detected in positive control plants, it is possible that an inability to quantify quinic acid was due to the fact it was produced as quickly as it was used up in plants. A similar instance has been reported for E4P, the starting intermediate of the shikimate pathway. While E4P is presumably found in all plants because it is an essential metabolite for growth and development, its levels were consistently found to be close to or below the detection limit of instruments (Bonner and Jensen, 1998). Although statistical tests could not be performed, the detection of chlorogenic acid in one transgenic plant is of particular interest. If repetition of these results can be obtained, then this will provide exciting new evidence for evolutionary 'leaps.' In an argument against sudden changes during evolution, Darwin numerously cited "Natura non facit saltum" ("nature makes no leap") in his pinnacle work, "The origin of species" (Padian, 2009) . However, substrate promiscuity in combination with complex interconnected metabolic networks may provide a mechanism by which organisms (at least plants) can rapidly diversify.

Chapter 6

6.1 Evolution of S/QDH genes and a brief history of land plants

In chapter three, *Poptr* SDH, representing the seed plant SDH clade showed the greatest catalytic activity with shikimate and $NADP^+$ and the results from both chapters three and four suggest that angiosperm sequences in general show greater substrate versatility than non-seed plants and conifers, given their ability to wield a combination of shikimate, quinate, NAD^+ and $NADP^+$. Greater catalytic activities and functional diversity of seed plant sequences compared to non-seed plants makes sense in light of land plant evolution. Extant land plants (embryophytes) are comprised of two major groups: the bryophytes (e.g. mosses) and vascular plants with the latter further split into the lycophytes, monilophytes (ferns) and seed plants (angiosperms and gymnosperms). The~~ir~~ exact origin of land plants remains a conundrum because of gaps in the fossil record (Kenrick and Crane, 1997; Weng and Chapple, 2010): however, ultrastructural and biochemical data support the charophycean green algae as their sister group (Kranz *et al*., 1995; Waters, 2003) and the earliest land plants likely evolved from a related group of freshwater algae (Weng and Chapple, 2010). As they embarked on land, pioneering plants were presumably threatened by damaging UV-B radiation, a lack of support against gravity, temperature fluctuations, desiccation, bacteria and fungi (de Vries and Archibald, 2018; Kenrick and Crane, 1997; Weng and Chapple, 2010). To cope and survive, they evolved a series of anatomical and physiological adaptations (Bennici, 2008; de Vries and Archibald, 2018; Kenrick and Crane, 1997) including the use of chemical defense compounds. Arguably, the phenylpropanoid pathway which shields against the sun's harmful rays as well as pests and pathogens and provides support against gravity was one of the first and most critical of metabolic pathways to have evolved (de Vries and Archibald, 2018; Kenrick and Crane, 1997; Weng and Chapple, 2010). Lignification in particular marks a turning point in the history of vascular plants (Peter and Neale, 2004; Weng and Chapple, 2010). Deposition of this complex heteropolymer into the secondary cell walls of xylem provided both the mechanical stability and compressional resistance necessary to conduct water and nutrients over long distances in order to promote upright growth of giant trees (Weng and Chapple, 2010; Xu *et al*., 2009). Forest ecosystems derived from early vascular plant lineages became common by the mid-Devonian, and increased sizes of plants were probably driven by a race for sunlight (de Vries and Archibald, 2018; Kenrick and Crane, 1997; Weng and Chapple, 2010).

Increased growth and lignification in turn, was likely propelled by greater carbon flux from the shikimate pathway.

6.2 Further blooming of S/QDH's in angiosperms

A possible outcome of gene duplication is the alteration of gene expression patterns between parent and daughter genes as in the case of vertebrate δ-crystallin (Piatigorsky, 1991; Piatigorsky *et al*., 1988; Wistow, 1993) and plant CYP98A3 (Matsuno, *et al*., 2009). Subfunctionalization in the form of expression divergence is believed to provide duplicated genes an initial foothold in the genome before natural selection can act on them (Gu *et al.*, 2002; Zhang, 2003). In light of this observation, subfunctionalization may have also played a role in the retention and subsequent diversification of plant SDH duplicates. For example, previous *in silico* experiments have revealed that expression of *P. trichocarpa SDH* and *QDH* genes differ spatially and temporally. *Poptr*SDH analyzed in this study for example, is constitutively expressed in all tissues and especially in shoot meristems, *i.e.* sites of growth whereas *Poptr*QDH (also studied here) is preferentially expressed in vascular tissues. Another QDH isoform (not examined in this study) shows predominantly foliage-specific expression which is increased in response to wounding and fungal infection (Guo *et al*., 2014) similar to lignin biosynthetic genes (Dixon and Paiva, 1995; Liu, *et al*., 2018). Thus, these expression data may imply that QDH activities were favoured at least by angiosperms, for supporting lignification, photoprotection as well as defending against pathogens. So far, expression profiling of pine SDH and QDH enzymes has not been undertaken but could merit information about their functional role(s) in conifers.

Flowering plant *S/QDH* genes not only differ in their expression profiles, but also in their cofactor preferences. The ability to use either NAD^+ or $NADP^+$ is thought to help tease apart the production of essential aromatic amino acids from metabolism of secondary compounds such as chlorogenic acid (Gritsunov *et al*., 2018). Divergence in regulation (either at the gene or protein level) of seed plant SDH and QDH clade members reflect additional (more recent) rounds of gene duplication within the SDH/QDH phylogeny. Whether or not these extra SDH and QDH copies possess novel gene functions has recently come to light. SDH copies of grapevine (*Vitis vinifera*) and walnut (*Juglans regia*), for example, show minute activities towards gallic acid biosynthesis *in vitro* (Bontpart *et al*., 2016; Muir *et al*., 2011). Still other angiosperms lack QDH clade members likely

due to gene loss. This is true for some eudicots (e.g. *A. thaliana*), and especially monocots. For this reason, the discovery of chlorogenic acid in at least a few monocot species creates an element of unexpected surprise (Clifford *et al*., 2006; Kweon *et al*.,, 2001; Shen *et al*., 2009). Monocot SDH sequences are further unique in that they cluster into three distinct sub-groups indicating more recent duplication and diversification events. As mentioned before and remarkably, some SDH duplicates in *Brassicaceae* have acquired quinate biosynthetic activities independently of other angiosperm QDH clade members following initial loss of *QDH* genes (Gritsunov *et al*., 2018). This points to a sort-of craftsmanship character to evolution, which tinkers through trial and error. The dynamic cycling of birth, death and re-birth of duplicated genes may have also taken place in other gene families molding lineage-specific characteristics.

6.3 Significance and future outlooks

As a whole, this work provides clear biochemical links between primary and secondary metabolism in plants and helps to shed light on the molecular forces underlying plant biochemical diversity. While the availability of whole-genome sequences for a variety of model species have undoubtedly expedited genetic and biochemical research in the field of plant science, attention is mostly paid to flowering plants, which has the risk of leaving other, non-angiosperm species untouched and in the dust (Song and Mitchell-Olds, 2011). Yet the characterization of early derived lineages occupying pivotal nodes of the green family tree is particularly important for understanding evolution. Studies on the evolution of plant terpene synthases (reviewed in Chen *et al*., 2011) and CYP98's (Renault *et al.*, 2017*)* represent the few published works so far that have performed a comprehensive analysis of secondary metabolic gene evolution. In both works, a gene family including members in angiosperms and mosses were investigated. The present work's evolutionary analysis extends further down the family tree by also including green algae.

To the best of my knowledge, chapter five of this study is the first to describe a potential method for investigating the *in vivo* activities of *QDH* genes, which in turn, facilitates opportunities for studying their biological roles. This can be done by testing the resistance of *QDH* over-expressing (or knock-out) plants against various biological and abiotic stresses. Despite the handful of reports that have helped elucidate the functions of chlorogenic acid in plants, much less is known about its precursor, quinic acid. Due to a scarcity of data, only a patchwork of speculations can be

provided. Given that it is an organic acid, it is assumed that quinic acid functions like other organic acids. Yet this simple idea does not greatly narrow the possibilities as organic acids play diverse roles in plants; their presence and concentration varies among taxa and are regulated by internal and external cues. As previously discussed, perhaps the most significant contribution of quinic acid to plant growth is to provide lignin precursors in the form of chlorogenic acid (Mondolot *et al*., 2006). Alongside a possible function in growth and like chlorogenic acid, quinic acid may also be involved in plant defense: however, there is insufficient evidence supporting this claim. The antioxidant properties of chlorogenic acid, for example, are likely due to the effects of caffeic acid rather than quinic acid, which is unable to scavenge free radicals *in vitro* (Grace *et al*., 1998). Interestingly, both chlorogenic acid and feruloylquinic are more abundant in chrysanthemums resistant to western flower thrips than in susceptible ones (Leiss *et al*., 2009). It is unclear if insecticidal properties of these compounds are due to caffeic, ferulic or quinic acid, or all the above.

On the other hand, unexpected clues towards understanding *QDH* functions may lie in looking at primary, instead of secondary metabolites. Notably, other polyols (e.g. mannitol, sorbitol, and myo-inositol) are observed to accumulate alongside sugars (e.g. sucrose and trehalose) and nitrogen containing compounds (e.g. proline) in water-stressed plants. Such osmolytes are useful in lowering the osmotic potential of cells without interfering with metabolism (Bohnert *et al*., 1995; Dichio *et al*., 2009; Williams *et al*., 2015). The common ice plant (*Mesembryanthemum crystallinum*), for example, accumulates D-pinitol in response to salinity stress. D-pinitol is used by the salt-loving succulent to sequester and store incredibly high concentrations of sodium (>1 M) for use in regulating osmotic homeostasis (Bohnert *et al*., 1995). Due to their hydroxyl groups, polyols as a whole may function in hydrating metabolites and preventing undesirable metabolic reactions (Williamson *et al*., 2002). On that note, others (e.g. mannitol) may also provide cellular protection by scavenging ROS induced by abiotic stress (Bianco and Avellone, 2014; Conde *et al*., 2015). Apart from their seeds, there are only a few plant species capable of surviving prolonged periods of drought without suffering permanent damage. The vegetative tissues of such remarkable species, known as resurrection plants, can withstand 90 to 100% water loss for months to years. Upon rehydration, the plants spring back to life from their inanimate state as suggested by their nickname, without a change in normal metabolism (Gechev*et et al*., 2005). Interestingly

the most abundant metabolite in one such resurrection plant, *Myrothamnus flabellifolius* is the quinic acid derivative, 3,4,5 tri-*O*-galloylquinic acid (Gechev *et al*., 2012; Moore *et al*., 2005). This compound is observed to interact with the plasma membrane and may help in maintaining its fluidity much like cholesterol. Like other gallotannins, it may also scavenge ROS, thereby preventing lipid peroxidation (Moore *et al*., 2005). However, its antioxidant effects are likely conferred by gallic acid, rather than quinic acid (Rice-Evans *et al*., 1997). On that note, an investigation into whether *QDH* gain-of-function (or loss-of-function) mutants are resistant (or susceptible thereof) to drought conditions should make for an interesting scientific endeavour. Importantly, understanding the biological significance of QDH proteins will help reveal the invisible forces shaping diversification of the S/QDH family.

Bibliography

Amborella Genome Project (2013) The Amborella Genome and the Evolution of Flowering Plants. Science 342: 1241089–1241089

Augusto L, Davies TJ, Delzon S, De Schrijver A (2014) The enigma of the rise of angiosperms: can we untie the knot? Ecology Letters 17: 1326–1338

Baker PJ, Britton KL, Rice DW, Rob A, Stillman TJ (1992) Structural Consequences of Sequence Patterns in the Fingerprint Region of the Nucleotide Binding Fold Implications for Nucleotide Specificity. Journal of Molecular Biology 228: 662-671

Barbehenn RV, Peter Constabel C (2011) Tannins in plant–herbivore interactions. Phytochemistry 72: 1551–1565

Barkman T, Zhang J (2009) Evidence for escape from adaptive conflict? Nature 462: E1–E1

Baruch A, Jeffery DA, Bogyo M (2004) Enzyme activity - It's all about image. Trends in Cell Biology 14: 29–35

Becker P (2000) Competition in the regeneration niche between conifers and angiosperms: Bond's slow seedling hypothesis. Functional Ecology 14: 401–412

Begley TP (2006) Cofactor biosynthesis: An organic chemist's treasure trove. Natural Product Reports 23: 15–25

Bennici A (2008) Origin and early evolution of land plants. Communicative and Integrative Biology 1: 212–218

Bianco RL, Avellone G (2014) Diurnal Regulation of Leaf Water Status in High- and Low-Mannitol Olive Cultivars. Plants 3: 196–208

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

Bloom JD, Arnold FH (2009) In the light of directed evolution: pathways of adaptive protein evolution. Proceedings of the National Academy of Sciences of the United States of America 106: 9995–10000

Bohnert HJ, Nelson DE, Jensen RG (1995) Adaptations to Environmental Stresses. The Plant Cell 7: 1099–1111

Bond WJ, Lee WG, Craine JM (2004). Plant structural defences against browsing birds: a legacy of New Zealand's extinct moas. Oikos 104: 500–508

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

Bonner CA, Jensen RA (1998) Upstream Metabolic Segments That Support Lignin Biosynthesis. ACS Symposium Series 697: 29–41

Bontpart T, Marlin T, Vialet, S, Guiraud J-L, Pinasseau L, Meudec E, Sommerer N, Cheynier V, Terrier N (2016) Two shikimate dehydrogenases, VvSDH3 and VvSDH4, are involved in gallic acid biosynthesis in grapevine. Journal of Experimental Botany 67: 3537–3550

Boudet A-M (2007) Evolution and current status of research in phenolic compounds. Phytochemistry 68: 2722–2735

Boudet A-M (2012) Polyphenols: From Plant Adaptation to Useful Chemical Resources. In Recent Advances in Polyphenol Research, Halbwirth H, Stich K, Cheynier V, and Quideau S (eds.), Wiley-Blackwell Publishing, pp 41–70

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

Bradford MM (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry 7: 248-254

Brinkmann-Chen S, Flock T, Cahn JKB, Snow CD, Brustad EM, McIntosh JA, Meinhold P, Zhang L, Arnold FH (2013) General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. Proceedings of the National Academy of Sciences of the United States of America 110: 10946–10951

Brochmann C, Brysting AK, Alsos IG, Borgen L, Grundt HH, Scheen A-C, Elven R (2004) Polyploidy in arctic plants. Biological Journal of the Linnean Society 82: 521–536

Broderick JB (2001) Coenzymes and Cofactors. In *Encyclopedia of Life Sciences*. Nature Publishing Group, pp. 1-11

Burns E, Ifrach I, Carmeli S, Pawlik J, Ilan M (2003) Comparison of anti-predatory defenses of Red Sea and Caribbean sponges. I. Chemical defense. Marine Ecology Progress Series 252: 105–114

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

Cahn JKB, Werlang, CA, Baumschlager A, Brinkmann-Chen S, Mayo SL, Arnold FH (2017) A General Tool for Engineering the NAD/NADP Cofactor Preference of Oxidoreductases.

ACS Synthetic Biology 6: 326–333

Cai W, Pei J, Grishin NV (2004) Reconstruction of ancestral protein sequences and its applications. BMC Evolutionary Biology 4: 33

Carrington Y, Guo J, Le CH, Fillo A, Kwon J, Tran LT, Ehlting J (2018) Evolution of a secondary metabolic pathway from primary metabolism: shikimate and quinate biosynthesis in plants. The Plant Journal 95: 823–833

Carugo O, Argos P (1997) NADP-Dependent enzymes. I: Conserved stereochemistry of cofactor binding. Proteins: Structure, Function, and Genetics 28: 10–28

Chánique AM, Parra LP (2018) Protein engineering for nicotinamide coenzyme specificity in oxidoreductases: Attempts and challenges. Frontiers in Microbiology 9: 194

Charles-Dominique T, Barczi J-F, Le Roux E, Chamaillé-Jammes S (2017) The architectural design of trees protects them against large herbivores. Functional Ecology 31: 1710–1717

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. The Plant Journal 66: 212–229

Chen S, Krinsky BH, Long M (2013) New genes as drivers of phenotypic evolution. Nature Reviews Genetics 14: 645–660

Chen Z, Lee WR, Chang SH (1991) Role of aspartic acid 38 in the cofactor specificity of Drosophila alcohol dehydrogenase. European Journal of Biochemistry 202: 263–267

Clé C, Hill LM, Niggeweg R, Martin CR, Guisez Y, Prinsen E, Jansen, M. A. K. (2008). Modulation of chlorogenic acid biosynthesis in *Solanum lycopersicum*; consequences for phenolic accumulation and UV-tolerance. Phytochemistry 69: 2149–2156

Clifford MN, Knight S, Surucu B, Kuhnert N (2006) Characterization by LC-MS^n of Four New Classes of Chlorogenic Acids in Green Coffee Beans: Dimethoxycinnamoylquinic Acids, Diferuloylquinic Acids, Caffeoyl-dimethoxycinnamoylquinic Acids, and Feruloyl-dimethoxycinnamoylquinic Acids. Journal of Agricultural and Food Chemistry 54: 1957–1969

Clifford MN, Wu W, Kuhnert N (2006) The chlorogenic acids of Hemerocallis. Food Chemistry 95: 574–578

Colin PY, Kintses B, Gielen F, Miton CM, Fischer G, Mohamed MF, Hyvönen M, Morgavi DP, Janssen DB, Hollfelder F (2015) Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. Nature Communications 6: 10008

Conant GC, Wolfe KH (2008) Turning a hobby into a job: How duplicated genes find new

functions. Nature Reviews Genetics 9: 938–950

Conde A, Regalado A, Rodrigues D, Costa JM, Blumwald E, Chaves MM, Gerós H (2015) Polyols in grape berry: transport and metabolic adjustments as a physiological strategy for water-deficit stress tolerance in grapevine. Journal of Experimental Botany 66: 889–906

Constabel CP and Yip L (2014) Personal Communications

Copley SD (2015) An evolutionary biochemist's perspective on promiscuity. Trends in Biochemical Sciences 40: 72–78

Crisp MD, Cook LG (2011). Cenozoic extinctions account for the low diversity of extant gymnosperms compared with angiosperms. New Phytologist 192: 997–1009

Croteau R, Kutchan TM, Lewis NG (2000) Natural products (plant secondary metabolites). In Biochemistry and molecular biology of plants, Buchanan B, Gruissem W and Jones RL (eds.), American Society of Plant Physiologists, pp 1250–1318

Darwin C (1859) On the Origin of Species, John Murray

De La Torre AR, Li Z, Van de Peer Y, Ingvarsson PK (2017) Contrasting Rates of Molecular Evolution and Patterns of Selection among Gymnosperms and Flowering Plants. Molecular Biology and Evolution 34: 1363–1377

de Vries J, Archibald JM (2018) Plant evolution: landmarks on the path to terrestrial life. New Phytologist 217: 1428–1434

Degnan SM (2014) Think laterally: horizontal gene transfer from symbiotic microbes may extend the phenotype of marine sessile hosts. Frontiers in Microbiology 5: 638

Dempsey DA, Vlot AC, Wildermuth MC, Klessig DF (2011) Salicylic Acid biosynthesis and metabolism. The Arabidopsis Book 9: e0156

Deng C, Cheng C-H C, Ye H, He X, Chen L (2010) Evolution of an antifreeze protein by neofunctionalization under escape from adaptive conflict. Proceedings of the National Academy of Sciences of the United States of America 107: 21593–21598

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

Dichio B, Margiotta G, Xiloyannis C, Bufo SA, Sofo A, Cataldi TRI (2009) Changes in water status and osmolyte contents in leaves and roots of olive plants (Olea europaea L.) subjected to water deficit. Trees 23: 247–256

Ding L, Hofius D, Hajirezaei MR, Fernie AR, Börnke 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-lnduced Phenylpropanoid Metabolism. The Plant Cell 7: 1085-1097

Doran PM (2006) Foreign protein degradation and instability in plants and plant tissue cultures. Trends in Biotechnology 24: 426–432

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

Eudes A, Pereira JH, Yogiswara S, Wang G, Benites VT, Baidoo EEK, Lee TS, Adams PD, Keasling JD, Loqué D (2016) Exploiting the Substrate Promiscuity of Hydroxycinnamoyl-CoA:Shikimate Hydroxycinnamoyl Transferase to Reduce Lignin. Plant and Cell Physiology 57: 568.

Fagard M, Vaucheret H (2000) (Trans)gene silencing in plants : How Many Mechanisms? Annual Review of Plant Physiology and Plant Molecular Biology 51: 167–194

Fang N, Yu S, Prior RL (2002) LC/MS/MS characterization of phenolic constituents in dried plums. Journal of Agricultural and Food Chemistry 50: 3579–3585

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

Force A, Lynch M, Pickett B, Amores A, Yan, Y-L, Postlethwait J (1999) Preservation of Duplicate Genes by Complementary, Degenerative Mutations. Genetics 151: 1531-1545

Frick KM, Kamphuis LG, Siddique KHM, Singh KB, Foley RC (2017) Quinolizidine alkaloid biosynthesis in lupins and prospects for grain quality improvement. Frontiers in Plant Science 8: 87

Fritz RS, Hochwender CG, Lewkiewicz DA, Bothwell S, Orians CM (2001) Seedling herbivory by slugs in a willow hybrid system: developmental changes in damage, chemical defense, and plant performance. Oecologia 129: 87–97

Funke T, Han H, Healy-Fried ML, Fischer M, Schönbrunn E (2006) Molecular basis for the herbicide resistance of Roundup Ready crops. Proceedings of the National Academy of Sciences of the United States of America 103: 13010–13015

Gaeta RT, Pires JC (2010) Research review Homoeologous recombination in allopolyploids: the polyploid ratchet. New Phytologist 186: 18–28

Gaivelyte K, Jakstas V, Razukas A, Janulis V (2013) Variation in the contents of neochlorogenic

acid, chlorogenic acid and three quercetin glycosides in leaves and fruits of rowan (*Sorbus*) species and varieties from collections in lithuania. Natural Product Communications 8: 1105–1110

Gakière B, Hao J, de Bont L, Pétriacq P, Nunes-Nesi A, Fernie AR (2018) NAD $^{+}$ Biosynthesis and Signaling in Plants. Critical Reviews in Plant Sciences 37: 259–307

Galperin MY, Koonin EV (2012) Divergence and convergence in enzyme evolution. Journal of Biological Chemistry 287: 21–28

Gamborg OL (1966) Aromatic metabolism in plants III. Quinate dehydrogenase from mung bean cell suspension cultures. BBA - Enzymology AndBiological Oxidation 128: 483–491

Gechev TS, Dinakar C, Benina M, Toneva V, Bartels D (2012) Molecular mechanisms of desiccation tolerance in resurrection plants. Cellular and Molecular Life Sciences 69: 3175–3186

Geiselman JA, McConnell OJ (1981) Polyphenols in brown algae *Fucus vesiculosus* andAscophyllum nodosum: Chemical defenses against the marine herbivorous snail, *Littorina littorea*. Journal of Chemical Ecology 7: 1115–1133

Glover NM, Redestig H, Dessimoz C (2016) Homoeologs: What Are They and How Do We Infer Them? Trends in Plant Science 21: 609–621

Goodstein DM, Shu S. Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research 40: D1178–D1186

Grace SC, Logan BA, Adams WW (1998) Seasonal differences in foliar content of chlorogenic acid, a phenylpropanoid antioxidant, in Mahonia repens. Plant, Cell and Environment 21: 513–521

Gritsunov A, Peek J, Diaz Caballero J, Guttman D, Christendat D (2018) Structural and biochemical approaches uncover multiple evolutionary trajectories of plant quinate dehydrogenases. The Plant Journal 95: 812–822

Gu Z, Nicolae D, Lu HH.-S, Li WH (2002) Rapid divergence in expression between duplicate genes inferred from microarray data. Trends in Genetics: 18: 609–613

Guo J, Carrington Y, Alber A, Ehlting J (2014) Molecular characterization of quinate and shikimate metabolism in *Populus trichocarpa*. The Journal of Biological Chemistry 289: 23846–23858

Guo, W, Grewe F, Fan W, Young GJ, Knoop V, Palmer JD, Mower JP (2016) *Ginkgo* and *Welwitschia* Mitogenomes Reveal Extreme Contrasts in Gymnosperm Mitochondrial

Evolution. Molecular Biology and Evolution 33: 1448–1460

Hagerman, A. E. (2002) Hydrolyzable Tannin Structural Chemistry. In Tannin Handbook, Hagerman, AE, Miami University Press, pp. 1-5

Hanley ME, Lamont BB, Fairbanks MM, Rafferty CM (2007) Plant structural traits and their role in anti-herbivore defence. Perspectives in Plant Ecology, Evolution and Systematics, 8: 157–178

Hanukoglu I (2015) Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. Biochemistry and Molecular Biology Education 43: 206–209

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

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

Hashida S, Takahashi H, Uchimiya H (2009) The role of NAD biosynthesis in plant development and stress responses. Annals of Botany 103: 819–824

Hay ME, Fenical W (1988) Marine Plant-Herbivore Interactions: The Ecology of Chemical Defense. Annual Review of Ecology and Systematics 19: 111–145

Hegarty M, Coate J, Sherman-Broyles S, Abbott R, Hiscock S, Doyle J (2013) Lessons from Natural and Artificial Polyploids in Higher Plants. Cytogenetic and Genome Research 140: 204–225

Herrmann KM (1995) The Shikimate Pathway: Early Steps in the Biosynthesis of Aromatic Compounds. The Plant Cell 7: 907–919

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

Hoffmann L, Besseau S, Geoffroy P, Ritzenthaler C, Meyer D, Lapierre C, Pollet B, Legrand M (2004) Silencing of hydroxycinnamoyl-coenzyme A shikimate/quinate hydroxycinnamoyltransferase affects phenylpropanoid biosynthesis. Plant Cell 16: 1446–1465

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 Sciences 109: 2966–2971

Huang K, Li M, Liu Y, Zhu M, Zhao G, Zhou Y, Zhang L, Wu Y, Dai X, Xia T, Gao L (2019). Functional analysis of 3-dehydroquinate dehydratase/shikiamte dehydrogenase involved in shikimate pathway in *Camellia sinensis*. Frontiers in Plant Science 10: 1268

Hughes AL (2005) Gene duplication and the origin of novel proteins. Proceedings of the National Academy of Sciences of the United States of America 102: 8791–8792

Ishimaru K, Nonaka G-I, Nishioka I (1987) Gallic acid esters of proto-quercitol, quinic acid and (−)-shikimic acid from quercus mongolica and q. myrsin aefolia. Phytochemistry 26: 1501–1504

Jeszka-Skowron M, Sentkowska A, Pyrzyńska K, De Peña MP (2016) Chlorogenic acids, caffeine content and antioxidant properties of green coffee extracts: influence of green coffee bean preparation. European Food Research and Technology 242: 1403–1409

Jorge TF, Mata AT, António C (2016) Mass spectrometry as a quantitative tool in plant metabolomics. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374: 20150370

Kallberg Y, Persson B (2006) Prediction of coenzyme specificity in dehydrogenases/ reductases. A hidden Markov model-based method and its application on complete genomes. FEBS Journal 273: 1177–1184

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

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

Khersonsky O, Roodveldt C, Tawfik D (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. Current Opinion in Chemical Biology 10: 498–508

Kirwan JA, Weber RJM, Broadhurst DI, Viant MR (2014) Direct infusion mass spectrometry metabolomics dataset: A benchmark for data processing and quality control. Scientific Data 1: 1–13

Klöti A, He X, Potrykus I, Hohn T, Fütterer J (2002) Tissue-specific silencing of a transgene in rice. Proceedings of the National Academy of Sciences of the United States of America, 99: 10881–10886

Kranz HD, Mikš D, Siegler M-L, Capesius I, Sensen CW, Huss VAR (1995) The origin of land plants: Phylogenetic relationships among charophytes, bryophytes, and vascular plants inferred from complete small-subunit ribosomal RNA gene sequences. Journal of Molecular Evolution 41: 74–84

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

Kweon M-H, Hwang H-J, Sung H-C (2001) Identification and antioxidant activity of novel chlorogenic acid derivatives from bamboo (*Phyllostachys edulis*). Journal of Agricultural

and Food Chemistry 49: 4646–4655

Laurino P, Tóth-Petróczy Á, Meana-Pañeda R, Lin W, Truhlar DG, Tawfik DS (2016) An Ancient Fingerprint Indicates the Common Ancestry of Rossmann-Fold Enzymes Utilizing Different Ribose-Based Cofactors. PLOS Biology 14: e1002396

Lauvergeat V, Kennedy K, Feuillet C, McKie JH, Gorrichon L, Baltas M, Boudet AM, Grima-Pettenatic J, Douglas KT, Sabatier P (1995) Site-Directed Mutagenesis of a Serine Residue in Cinnamyl Alcohol Dehydrogenase, a Plant NADPH-Dependent Dehydrogenase, Affects the Specificity for the Coenzyme. Biochemistry 34: 12426-12434

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: 1567–1575

Lesk AM (1995) NAD-binding domains of dehydrogenases. Current Opinion in Structural Biology 5: 775–783

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

Li X, Bonawitz N D, Weng JK, Chapple C (2010) The growth reduction associated with repressed lignin biosynthesis in Arabidopsis thaliana is independent of flavonoids. Plant Cell 22: 1620–1632

Li Y, Kong D, Bai M, He H, Wang H, Wu H (2019) Correlation of the temporal and spatial expression patterns of HQT with the biosynthesis and accumulation of chlorogenic acid in Lonicera japonica flowers. Horticulture Research 6: 73

Liesche J, Windt C, Bohr T, Schulz A, Jensen KH (2015) Slower phloem transport in gymnosperm trees can be attributed to higher sieve element resistance. Tree Physiology 35: 376–386

Liu Q, Luo L, Zheng L (2018) Lignins: Biosynthesis and Biological Functions in Plants International Journal of Molecular Sciences 19: 335

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J (2000) The Molecules of Life. In Molecular Cell biology 4th ed., Freeman WH (ed.), WH Freeman and Company, section 1.2

Lunzer, M., Miller, S. P., Felsheim, R., and Dean, AM (2005). The Biochemical Architecture of an Ancient Adaptive Landscape. New Series: 310: 499-501

Lynch M (2007) The origins of genome architecture, Sinauer Associates Inc.

Lynch M, Marinov GK (2015) The bioenergetic costs of a gene. Proceedings of the National Academy of Sciences of the United States of America 112: 15690–15695

Mahesh V, Million-Rousseau R, Ullmann P, Chabrillange N, Bustamante J, Mondolot L, Morant M, Noirot M, Hamon S, de Kochko A, Werck-Reichhart D, Campa C (2007) Functional characterization of two p-coumaroyl ester 3′-hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis. Plant Molecular Biology 64: 145–159

Martínez Cuesta S, Rahman SA, Furnham N, Thornton JM (2015) The Classification and Evolution of Enzyme Function. Biophysical Journal 109: 1082–1086

Maruta T, Ogawa T, Tsujimura M, Ikemoto K, Yoshida T, Takahashi H, Yoshimura K, Shigeoka S (2016) Loss-of-function of an Arabidopsis NADPH pyrophosphohydrolase, AtNUDX19, impacts on the pyridine nucleotides status and confers photooxidative stress tolerance. Scientific Reports 6: 7432

Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M, Burleigh JG, Gitzendanner MA, Wafula E, Der JP, dePamphilis CW, Roure B, Philippe H, Ruhfel BR, Miles NW, Graham SW, Mathews S, Surek B, Melkonian M, Soltis DE, Soltis PS, Rothfels C, Pokorny L, Shaw JA, DeGironimo L, Stevenson DW, Villarreal JC, Chen T, Kutchan TM, Rolf M, Baucom RS, Deyholos MK, Samudrala R, Tian Z, Wu X, Xiao S, Zhang Y, Wang J, Leebens-Mack J, Wong GK-S (2014) Data access for the 1,000 Plants (1KP) project. GigaScience 3: 17

Matsuno M, Compagnon V, Schoch GA, Schmitt M, Debayle D, Bassard, J-E, 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

Matsuura HN, Fett-Neto AG (2015) Plant Alkaloids: Main Features, Toxicity, and Mechanisms of Action. In Plant Toxins, Gopalakrishnakone P, Carlini C, Ligabue-Braun R (eds.), Springer Netherlands, pp. 1-15

Matzke MA, Matzke AJ (1995) How and why do plants inactivate homologous (trans)genes? Plant Physiology 107: 679–685)

Mendonça JCF, Franca AS, Oliveira LS, Nunes M (2008) Chemical characterisation of non-defective and defective green arabica and robusta coffees by electrospray ionization-mass spectrometry (ESI-MS). Food Chemistry 111: 490–497

Meyer VR (2010) Practical High-Performance Liquid Chromatography: 5th Edition. John Wiley and Sons, Ltd

Michel G, Tonon T, Scornet D, Cock JM, Kloareg B (2010) Central and storage carbon metabolism of the brown alga Ectocarpus siliculosus: insights into the origin and evolution of storage carbohydrates in Eukaryotes. New Phytologist 188: 67–81

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

Mizutani M (2012) Impacts of diversification of cytochrome P450 on plant metabolism. Biological and Pharmaceutical Bulletin 35: 824–832

Moglia A, Lanteri S, Comino C, Hill L, Knevitt D, Cagliero C, Rubiolo P, Bornemann S, Martin C (2014) Dual catalytic activity of hydroxycinnamoyl-coenzyme a quinate transferase from tomato allows it to moonlight in the synthesis of both mono- and dicaffeoylquinic acids. Plant Physiology 166: 1777–1787

Moloveanu SC and Davis MF (2014) Analysis of quinic acid and of myo-Inositol in tobacco. Beiträge zur Tabakforschung International/Contributions to Tobacco Resarch 25: 498-506

Mondolot L, La Fisca P, Buatois B, Talansier E, de Kochko A, Campa C (2006) Evolution in caffeoylquinic acid content and histolocalization during Coffea canephora leaf development. Annals of Botany 98: 33–40

Moore JP, Westall KL, Ravenscroft N, Farrant JM, Lindsey GG, Brandt WF (2005) The predominant polyphenol in the leaves of the resurrection plant Myrothamnus flabellifolius, 3,4,5 tri-O-galloylquinic acid, protects membranes against desiccation and free radical-induced oxidation. The Biochemical Journal 385: 301–308

Moore R C, Purugganan MD (2005) The evolutionary dynamics of plant duplicate genes. Current Opinion in Plant Biology 8: 122–128

Muir R M, Ibáñez 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

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

Ncube EN, Mhlongo MI, Piater LA, Steenkamp PA, Dubery IA, Madala NE (2014) Analyses of chlorogenic acids and related cinnamic acid derivatives from Nicotiana tabacum tissues with the aid of UPLC-QTOF-MS/MS based on the in-source collision-induced dissociation method. Chemistry Central Journal 8: 66

Ng LK, Lafontaine P, Vanier M (2004) Characterization of cigarette tobacco by direct electrospray ionization-ion trap mass spectrometry (ESI-ITMS) analysis of the aqueous extract - A novel and simple approach. Journal of Agricultural and Food Chemistry 52: 7251–7257

Niggeweg R, Michael AJ, Martin C (2004). Engineering plants with increased levels of the

antioxidant chlorogenic acid. Nature Biotechnology 22: 746–754

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pufar S, Rangwala SH, Rausch, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tosltoy I, Tully RE, Vatsan AR, Wallin C, WEbb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD Pruitt KD (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research 44: D733-745

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

Osipov VI, Shein IV (1987) Study of the causes of the inactivation of quinate dehydrogenase from *Pinus sylvestris* l. needles. Biokhimiya 52: 194–197

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

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

Ossipov V, Salminen J-P, Ossipova S, Haukioja E, Pihlaja K (2003) Gallic acid and hydrolysable tannins are formed in birch leaves from an intermediate compound of the shikimate pathway. Biochemical Systematics and Ecology 31: 3-16

Otto SP (2007) The Evolutionary Consequences of Polyploidy. Cell 131: 452–462

Ouni Y, Taamalli A, Gómez-Caravaca AM, Segura-Carretero A, Fernández-Gutiérrez A, Zarrouk M (2011) Characterisation and quantification of phenolic compounds of extra-virgin olive oils according to their geographical origin by a rapid and resolutive LC-ESI-TOF MS method. Food Chemistry 127: 1263–1267

Padian K (2009) Ten myths about Charles Darwin. BioScience 59: 800-804

Palm GJ, Reisky L, Böttcher D, Müller H, Michels EAP, Walczak MC, Berndt L, Weiss MS, Bornscheuer UT, Weber G (2019) Structure of the plastic-degrading *Ideonella sakaiensis* MHETase bound to a substrate. Nature Communications 10(1): 1717

Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K (2000) Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. Proceedings of the National Academy of Sciences of the United States of America 97: 6960–6966

Panchy N, Lehti-Shiu M, Shiu S-H (2016) Evolution of Gene Duplication in Plants. Plant

Physiology 171: 2294–2316

Peisajovich SG, Tawfik DS (2007) Protein engineers turned evolutionists. Nature Methods 4: 991–994

Perozich J, Kuo I, Wang BC, Boesch JS, Lindahl R, Hempel J (2000) Shifting the NAD/NADP preference in class 3 aldehyde dehydrogenase. European Journal of Biochemistry 267: 6197–6203

Perry RH, Cooks RG, Noll RJ (2008) Orbitrap mass spectrometry: Instrumentation, ion motion and applications. Mass Spectrometry Reviews 27: 661–699

Peter G, Neale D (2004) Molecular basis for the evolution of xylem lignification. Current Opinion in Plant Biology 7: 737–742

Petsko GA, Ringe D (2014) Protein Structure and Function, Lawrence and Robertson M (eds.), Sinauer Associates Inc.

Piasecka A, Kachlicki P, Stobiecki M (2019) Analytical methods for detection of plant metabolomes changes in response to biotic and abiotic stresses. International Journal of Molecular Sciences 20: 379

Piatigorsky J (1991) The recruitment of crystallins: new functions precede gene duplication. Science 252: 1078–1079

Piatigorsky J, O'Brien WE, Norman BL, Kalumuck K, Wistow GJ, Borras T, Nickerson JM, Wawrousek EF (1988) Gene sharing by delta-crystallin and argininosuccinate lyase. Proceedings of the National Academy of Sciences of the United States of America 85: 3479–3483

Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends in Plant Science 5: 39–445

Pillay P, Schlüter U, Van Wyk S, Kunert KJ, Vorster BJ (2013) Proteolysis of recombinant proteins in bioengineered plant cells. Bioengineered 5: 15-20

Popper ZA, Tuohy MG (2010) Beyond the green: Understanding the evolutionary puzzle of plant and algal cell walls. Plant Physiology 153: 373–383

Pu G, Wang P, Zhou B, Liu Z, Xiang F (2013) Cloning and characterization of *Lonicera japonica* p-coumaroyl ester 3-hydroxylase which is involved in the biosynthesis of chlorogenic acid. Bioscience, Biotechnology and Biochemistry 77: 1403–1409

Qian Y, Lynch JH, Guo L, Rhodes D, Morgan JA, Dudareva N (2019) Completion of the cytosolic post-chorismate phenylalanine biosynthetic pathway in plants. Nature Communications 10: 1–15

Radwan MA (1972) Occurrence and genotypic differences of chlorogenic acid in Douglas-fir foliage. USDA For. Serv. Res. Note PNW-RN-173. Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Forest and Range Experiment Station, 6 P

Radwan MA (1975) Genotype and Season Influence Chlorogenic Acid Content in Douglas-fir Foliage. Canadian Journal of Forest Research 5: 281–284

Rajeevkumar S, Anunanthini P, Sathishkumar R (2015) Epigenetic silencing in transgenic plants. Frontiers in Plant Science 6: 693

Ramsey J (2011) Polyploidy and ecological adaptation in wild yarrow. Proceedings of the National Academy of Sciences of the United States of America 108: 7096–7101

Raven JA, Giordano M (2014) Algae. Current Biology 24: R590–R595

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

Renault H, Alber A, Horst NA, Basilio Lopes A, Fich EA, Kriegshauser L, Wiedemann G, Ullmann P, Herrgott L, Erhardt M, Pineau E, Ehlting J, Schmitt M, Rose JKC, Reski R, Werck-Reichhart D (2017) A phenol-enriched cuticle is ancestral to lignin evolution in land plants. Nature Communications 8

Ribeiro AJM, Tyzack JD, Borkakoti N, Holliday GL, Thornton JM (2020) A global analysis of function and conservation of catalytic residues in enzymes. Journal of Biological Chemistry 295: 314–324

Rice-Evans C, Miller N, Paganga G (1997) Antioxidant properties of phenolic compounds. Trends in Plant Science 2: 152–159

Richards TA, Dacks JB, Campbell SA, Blanchard JL, Foster PG, McLeod R, Roberts CW (2006) Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. Eukaryotic Cell 5: 1517–1531

Robinson PK (2015) Enzymes: principles and biotechnological applications. Essays in Biochemistry 59: 1–41

Roden JS, Canny MJ, Huang CX, Ball MC (2009) Frost tolerance and ice formation in *Pinus radiata* needles: ice management by the endodermis and transfusion tissues. Functional Plant Biology 36: 180

Romani A, Pinelli P, Galardi C, Mulinacci N, Tattini M (2002) Identification and quantification of galloyl derivatives, flavonoid glycosides and anthocyanins in leaves of *Pistacia lentiscus* L. Phytochemical Analysis 13: 79–86

Ronel M, Lev-Yadun S (2012) The spiny, thorny and prickly plants in the flora of Israel. Botanical Journal of the Linnean Society 168: 344–352

Rosell A, Valencia E, Ochoa WF, Fita I, Parés X, Farrés J (2003) Complete Reversal of Coenzyme Specificity by Concerted Mutation of Three Consecutive Residues in Alcohol Dehydrogenase. Journal of Biological Chemistry 278: 40573–40580

Rosell A, Valencia E, Parés X, Fita I, Farrés J, Ochoa WF (2003) Crystal structure of the vertebrate NADP(H)-dependent alcohol dehydrogenase (ADH8). Journal of Molecular Biology 330: 75–85

Sankoff D, Zheng C (2018) Whole Genome Duplication in Plants: Implications for Evolutionary Analysis. Methods in molecular biology 1704: 291–315

Sattler MC, Carvalho CR, Clarindo WR (2016) The polyploidy and its key role in plant breeding. Planta 243: 281–296

Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, Werck-Reichhart D (2001) CYP98A3 from *Arabidopsis thaliana* is a 3′-Hydroxylase of Phenolic Esters, a Missing Link in the Phenylpropanoid Pathway. Journal of Biological Chemistry 276: 36566–36574

Schubert D, Lechtenberg B, Forsbach A, Gils M, Bahadur S, Schmidt R (2004) Silencing in *Arabidopsis* T-DNA transformants: The predominant role of a gene-specific RNA sensing mechanism versus position effects. Plant Cell 16: 2561–2572

Shen H, Fu C, Xiao X, Ray T, Tang Y, Wang Z, Chen F (2009) Developmental control of lignification in stems of lowland switchgrass variety Alamo and the effects on saccharification efficiency. BioEnergy Research 2: 233–245

Shibata T, Miyake H, Tanaka R, Miyasaki T, Kawaguchi S (2014) The Influence of Phlorotannins and Bromophenols on the Feeding Behavior of Marine Herbivorous Gastropod Turbo cornutus. American Journal of Plant Sciences 5: 387–392

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 Sciences of the United States of America 109: 14888–14893

Silva N, Mazzafera P, Cesarino I (2019) Should I stay or should I go: are chlorogenic acids mobilized towards lignin biosynthesis? Phytochemistry 166: 112063

Singh SA, Christendat D (2006) Structure of Arabidopsis dehydroquinate dehydratase-shikimate 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 and

Design 7: 2153–2160

Solecka D (1997). Role of phenylpropanoid compounds in plant responses to different stress factors. Acta Physiologiae Plantarum 19: 257–268

Solymosi K (2013) Plastid Structure, Diversification and Interconversions I. Algae. Current Chemical Biology 6: 167–186

Song BH, Mitchell-Olds T (2011) Evolutionary and Ecological Genomics of Non-Model Plants. Journal of Systematics and Evolution 49: 17-24

Sonnante G, D'Amore R, Blanco E, Pierri CL, de Palma M, Luo J, Tucci M, Martin C (2010) Novel hydroxycinnamoyl-coenzyme a quinate transferase genes from artichoke are involved in the synthesis of chlorogenic acid. Plant Physiology 153: 1224–1238

Steinberg PD, Estestt JA, Winter FC (1995) Evolutionary consequences of food chain length in kelp forest communities. Ecology 92: 8145–8148

Stival Sena J, Giguère I, Boyle B, Rigault P, Birol I, Zuccolo A, Ritland K, Ritland C, Bohlmann J, Jones S, Bousquet J, Mackay J (2014) Evolution of gene structure in the conifer *Picea glauca*: a comparative analysis of the impact of intron size. BMC Plant Biology 14: 95

Tang H (2014) Personal Communications

Tawfik OK, DS (2010) Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. Annual Review of Biochemistry 79: 471–505

Theis N, Lerdau M (2003) The Evolution of Function in Plant Secondary Metabolites International Journal of Plant Sciences 164: S93–S102

Thomas BC, Pedersen B, Freeling M (2006) Following tetraploidy in an Arabidopsis ancestor genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Research 16: 934

Trudeau DL, Tawfik DS (2019) Protein engineers turned evolutionists—the quest for the optimal starting point. Current Opinion in Biotechnology 60: 46–52

Tzin V, Galili G (2010) The Biosynthetic Pathways for Shikimate and Aromatic Amino Acids in *Arabidopsis thaliana*. The Arabidopsis Book 8: e0132

Verpoorte R, Alfermann AW (2000). Metabolic Engineering of Plant Secondary Metabolism, Verpoorte R and Alfermann R (eds.), Springer Science+Business Media, pp. 1-18

Vogt T (2009) Phenylpropanoid Biosynthesis Molecular Plant 3: 2–20

Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, Verstrepen KJ

(2012) Reconstruction of Ancestral Metabolic Enzymes Reveals Molecular Mechanisms Underlying Evolutionary Innovation through Gene Duplication. PLoS Biology 10: e1001446

War AR, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, Sharma HC (2012) Mechanisms of plant defense against insect herbivores. Plant Signaling and Behavior 7: 1306–1320

Waters ER (2003) Molecular adaptation and the origin of land plants. Molecular Phylogenetics and Evolution 29: 456–463

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

Weng J-K, Noel JP (2012). The Remarkable Pliability and Promiscuity of Specialized Metabolism. Cold Spring Harbor Symposia on Quantitative Biology 77: 309–320

Weng J-K (2014) The evolutionary paths towards complexity: a metabolic perspective. New Phytologist 201: 1141–1149

Wertheim B, Beukeboom LW, van de Zande L (2013) Polyploidy in Animals: Effects of Gene Expression on Sex Determination, Evolution and Ecology. Cytogenetic and Genome Research 140: 256–269

Wijnker E, Schnittger A (2013) Control of the meiotic cell division program in plants. Plant Reproduction 26: 143–158

Williams B, Njaci I, Moghaddam L, Long H, Dickman MB, Zhang X, Mundree S (2015) Trehalose Accumulation Triggers Autophagy during Plant Desiccation. PLOS Genetics, 11: e1005705

Williamson DJ, Jennings BD, Guo W-W, Pharr DM, Ehrenshaft M (2002) Sugar alcohols, salt stress, and fungal resistance: polyols--multifunctional plant protection. American Society for Horticultural Science 127: 467–473

Wink M (1999). Functions of plant secondary metabolites and their exploitation in biotechnology. In Biochemistry of plant secondary metabolism. Annual plant reviews, Vol. 2, Wink M (ed.), Sheffield Academic Press, pp.208–209

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

Wistow G (1993) Lens crystallins: gene recruitment and evolutionary dynamism. Trends in Biochemical Sciences 18: 301–306

Xu Z, Zhang D, Hu J, Zhou X, Ye X, Reichel KL, Stewart NR, Syrenne RD, Yang X, Gao P, Shi

W, Doeppke C, Sykes RW, Burris JN, Bozell JJ, Cheng (M) Z-M, Hayes DG, Labbe N, Davis M, Stewart Jr CN, Yuan JS (2009) Comparative genome analysis of lignin biosynthesis gene families across the plant kingdom. BMC Bioinformatics 10: S3

Yogendra KN, Pushpa D, Mosa K A, Kushalappa AC, Murphy A, Mosquera T (2014) Quantitative resistance in potato leaves to late blight associated with induced hydroxycinnamic acid amides. Functional and Integrative Genomics 14: 285–298

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

Zhang J (2003) Evolution by gene duplication: an update. Trends in Ecology and Evolution, 18: 292–298

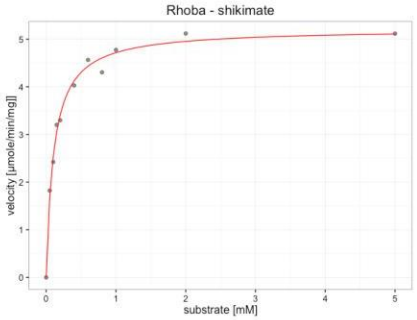
Zimmer CT, Garrood WT, Singh KS, Randall E, Lueke B, Gutbrod O, Matthiesen S, Kohler M, Nauen R, Davies TGE, Bass C (2018) Neofunctionalization of Duplicated P450 Genes Drives the Evolution of Insecticide Resistance in the Brown Planthopper. Current Biology 28: 268-274

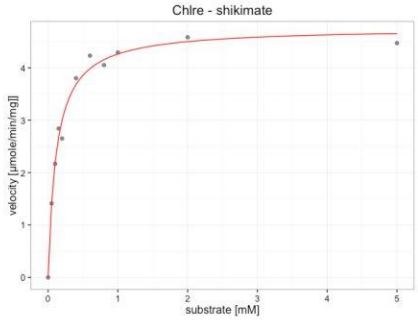
Appendix

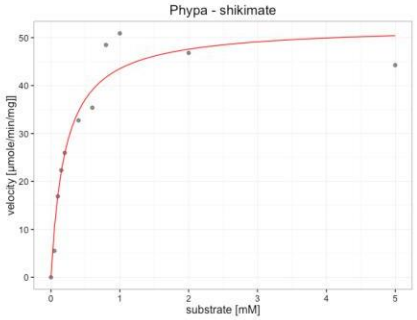
Appendix A Michaelis Menten kinetics of purified enzymes

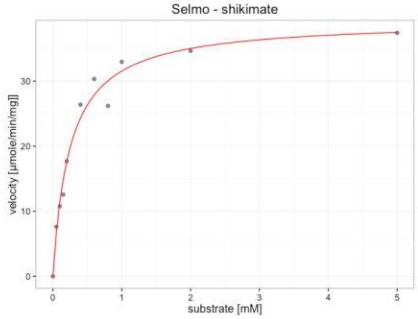
The following figures below have been partly published in Carrington *et al*. (2018)

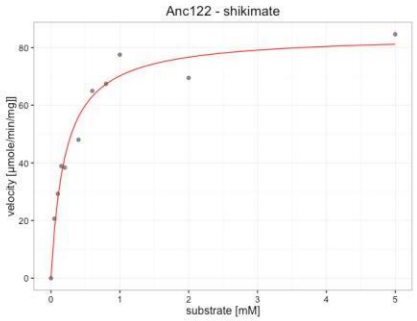
(Note: PintaSDH with NAD^+ has only one replicate)

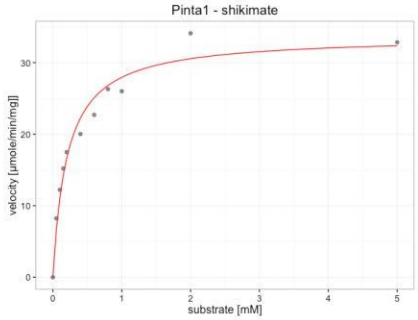


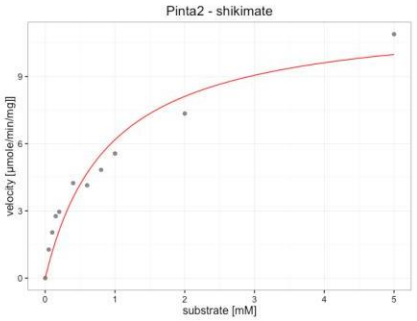


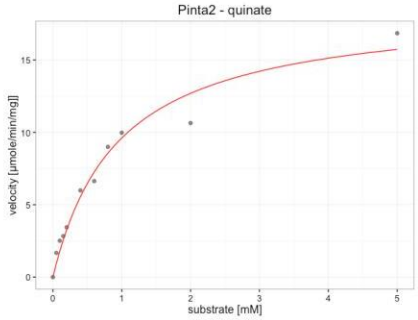


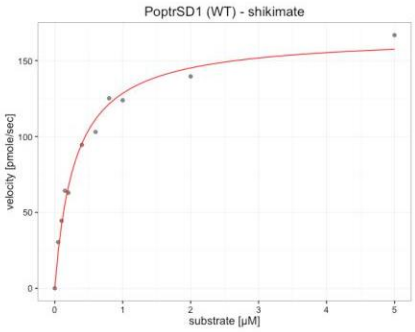


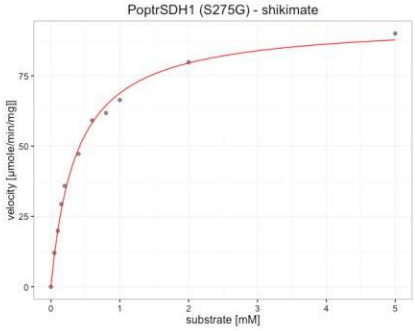


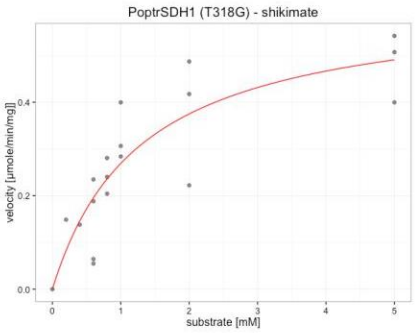


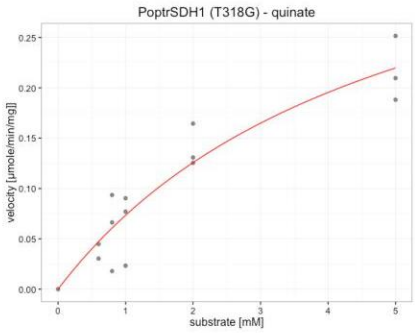


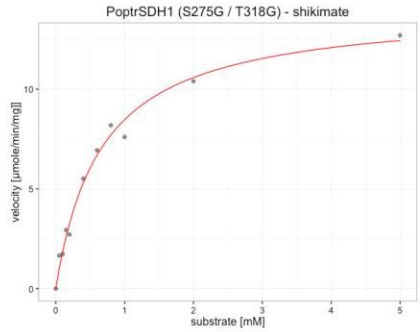


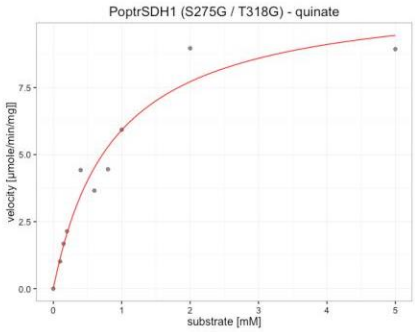

PoptrSD1 (WT) - shikimate
velocity [pmole/sec]
substrate [µM]

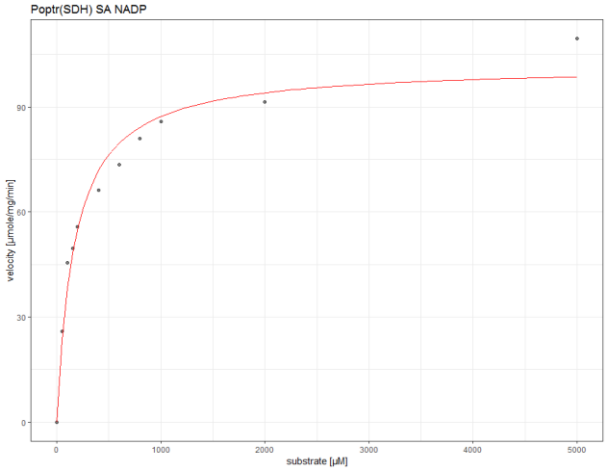

PoptrSDH1 (S275G) - shikimate
velocity [µmole/min/mg]
substrate [mM]

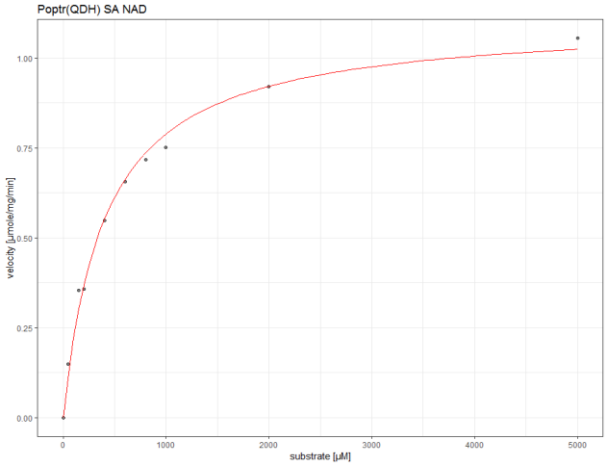

PoptrSDH1 (T318G) - shikimate
velocity [µmole/min/mg]
substrate [mM]

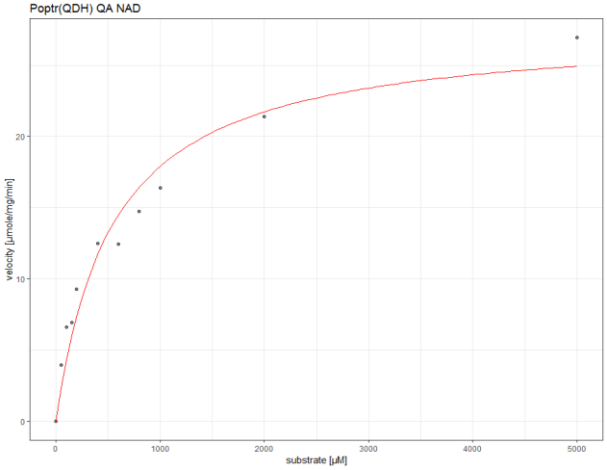

PoptrSDH1 (T318G) - quinate
velocity [µmole/min/mg]
substrate [mM]

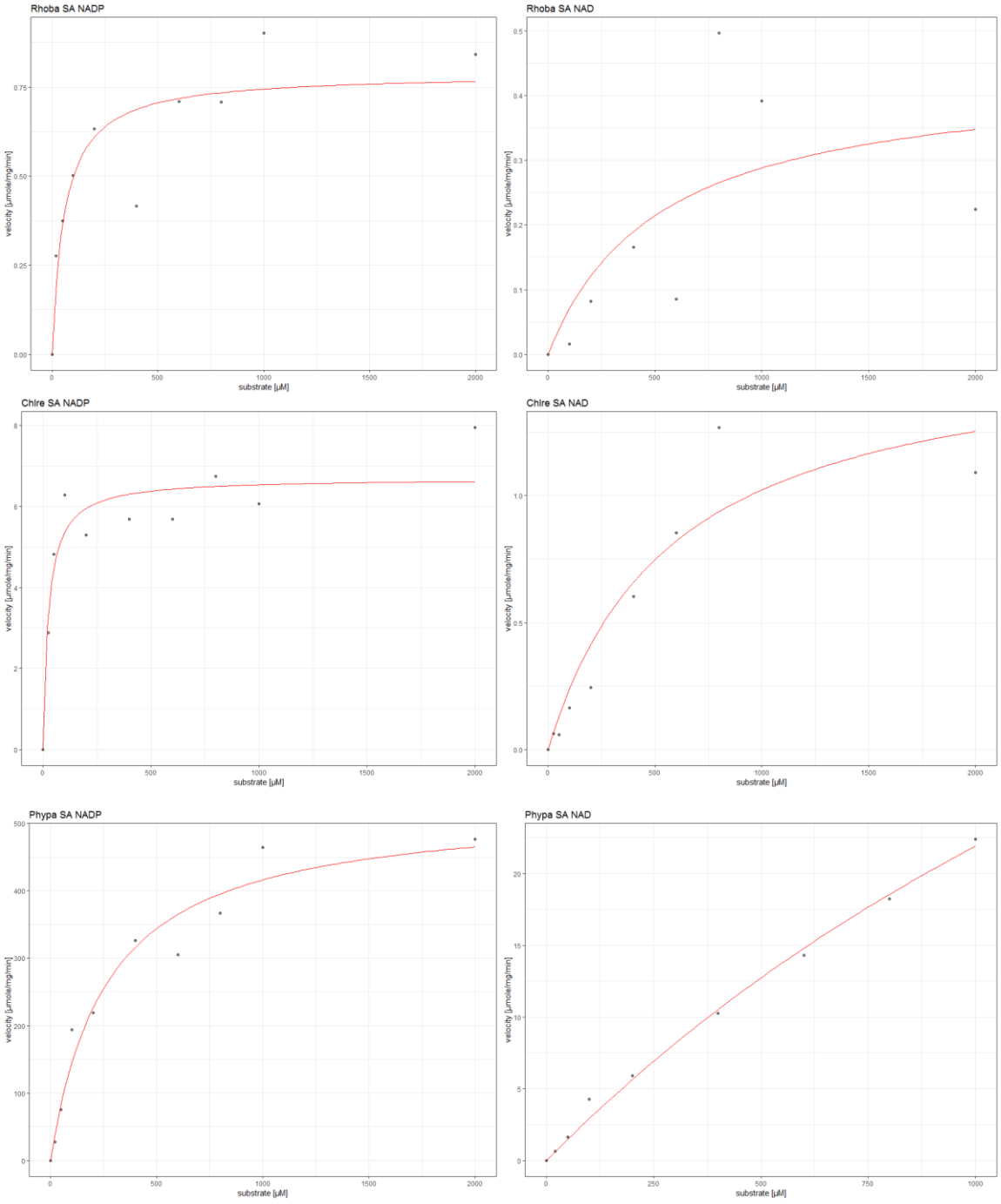

PoptrSDH1 (S275G / T318G) - shikimate
velocity [µmole/min/mg]
substrate [mM]

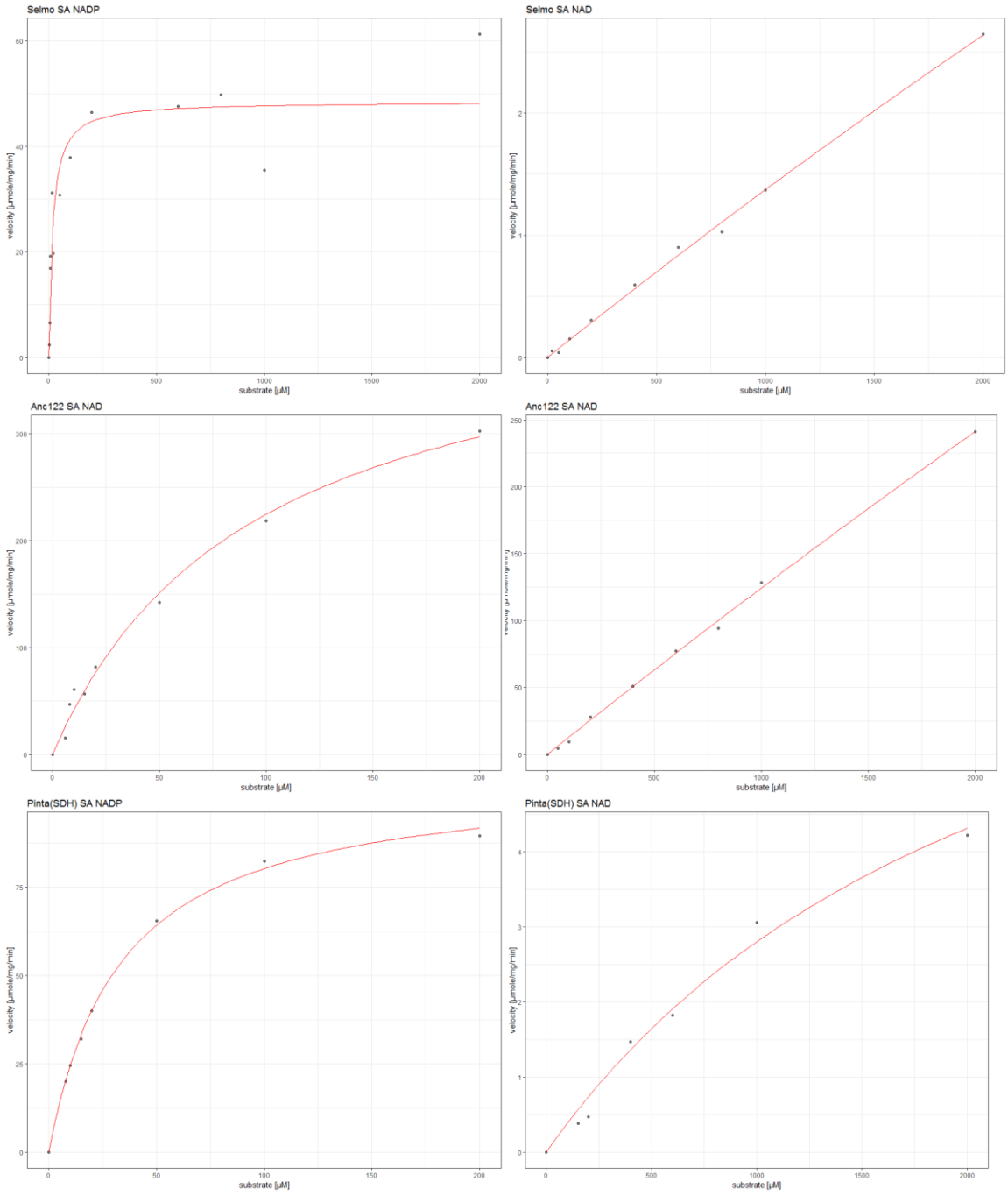

PoptrSDH1 (S275G / T318G) - quinate
velocity [µmole/min/mg]
substrate [mM]

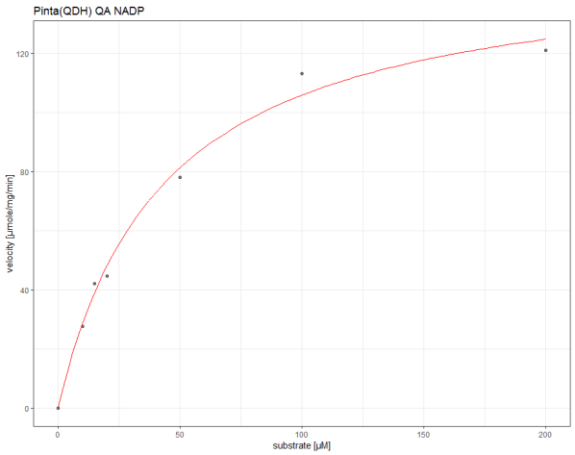

Poptr(SDH) SA NADP
velocity [µmole/mg/min]
0
30
60
90
0
1000
2000
3000
4000
5000
substrate [µM]

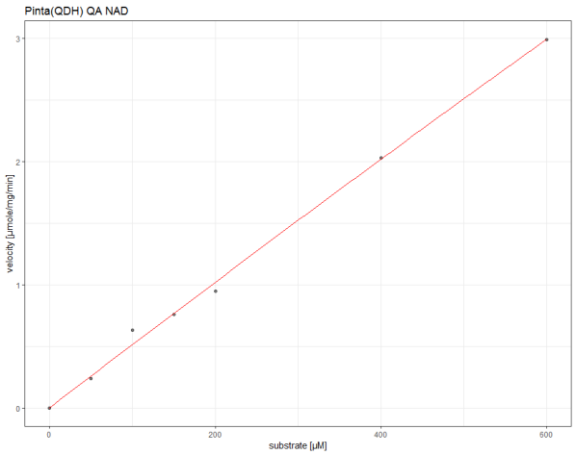

Poptr(QDH) SA NAD
velocity [µmole/mg/min]
0.00
0.25
0.50
0.75
1.00
0
1000
2000
3000
4000
5000
substrate [µM]

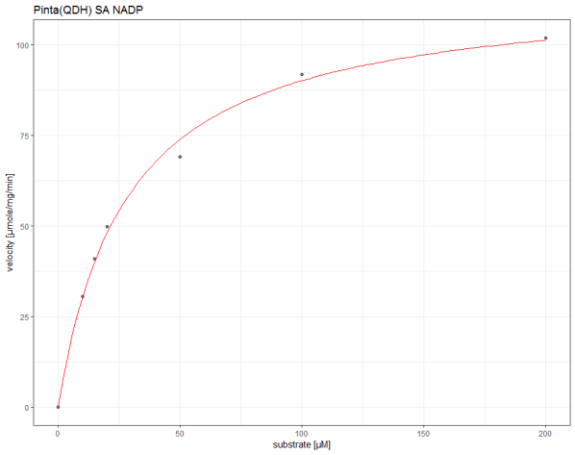

Poptr(QDH) QA NAD
velocity [µmole/mg/min]
0
10
20
0
1000
2000
3000
4000
5000
substrate [µM]

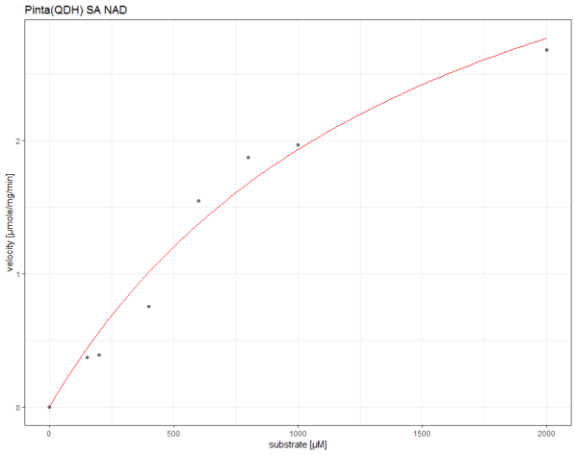

Rhoba SA NADP
velocity [µmole/mg/min]
substrate [µM]
Rhoba SA NAD
velocity [µmole/mg/min]
substrate [µM]
Chlre SA NADP
velocity [µmole/mg/min]
substrate [µM]
Chlre SA NAD
velocity [µmole/mg/min]
substrate [µM]
Phypa SA NADP
velocity [µmole/mg/min]
substrate [µM]
Phypa SA NAD
velocity [µmole/mg/min]
substrate [µM]

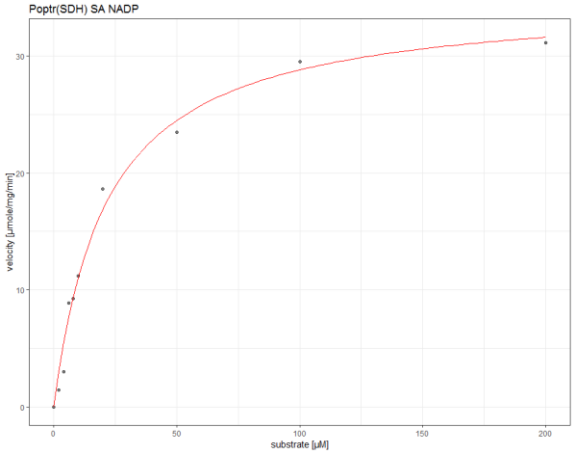

Selmo SA NADP
velocity [µmole/mg/min]
substrate [µM]
Selmo SA NAD
velocity [µmole/mg/min]
substrate [µM]
Anc122 SA NAD
velocity [µmole/mg/min]
substrate [µM]
Anc122 SA NAD
substrate [µM]
Pinta(SDH) SA NADP
velocity [µmole/mg/min]
substrate [µM]
Pinta(SDH) SA NAD
velocity [µmole/mg/min]
substrate [µM]

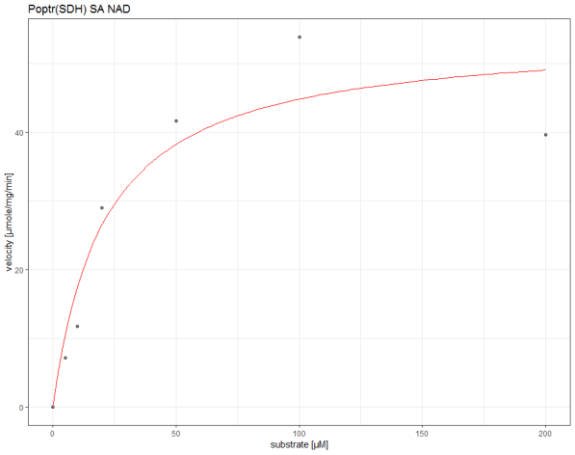

Pinta(QDH) QA NADP
velocity [µmole/mg/min]
substrate [µM]

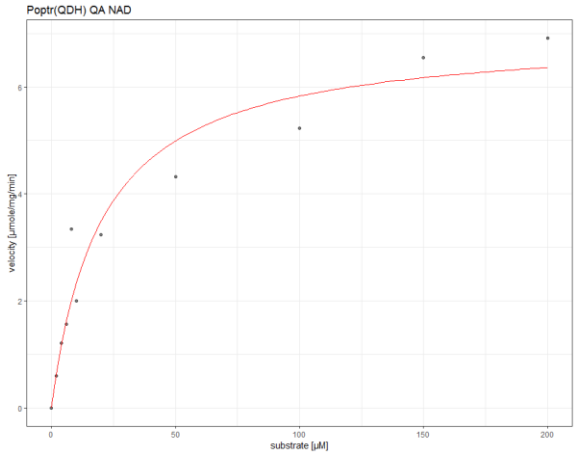

Pinta(QDH) QA NAD
velocity [µmole/mg/min]
substrate [µM]

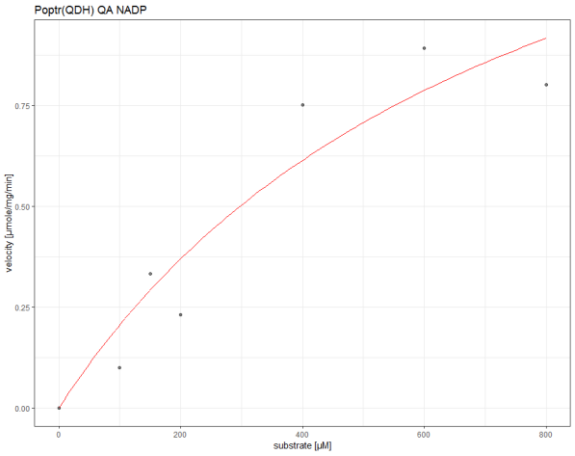

Pinta(QDH) SA NADP
velocity [µmole/mg/min]
substrate [µM]

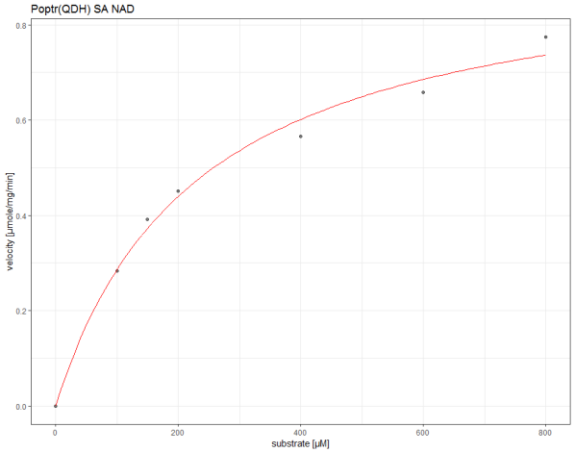

Pinta(QDH) SA NAD
velocity [µmole/mg/min]
substrate [µM]


Poptr(SDH) SA NADP
velocity [µmole/mg/min]
substrate [µM]


Poptr(SDH) SA NAD
velocity [µmole/mg/min]
substrate [µM]

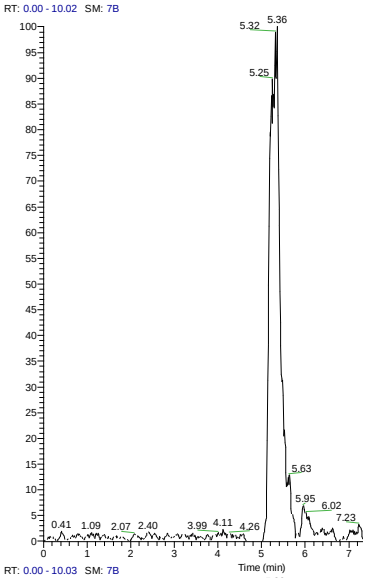

Poptr(QDH) QA NAD
velocity [µmole/mg/min]
substrate [µM]
0
2
4
6
0
50
100
150
200

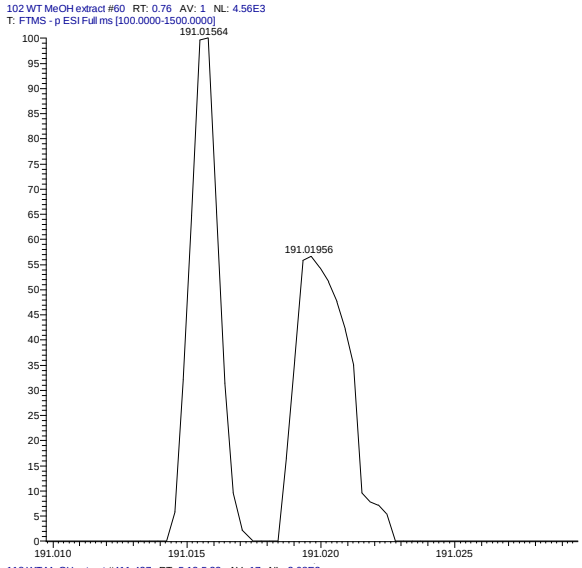

Poptr(QDH) QA NADP
velocity [µmole/mg/min]
substrate [µM]
0.00
0.25
0.50
0.75
0
200
400
600
800

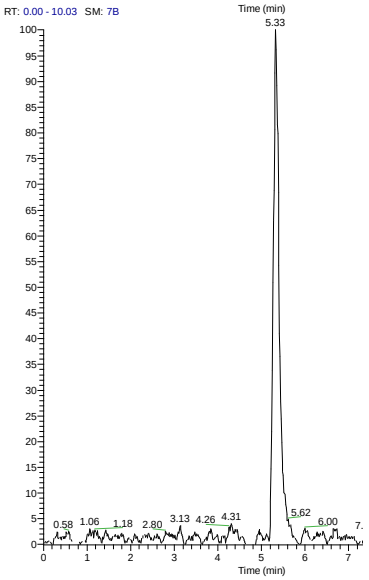

Poptr(QDH) SA NAD
velocity [µmole/mg/min]
substrate [µM]
0.0
0.2
0.4
0.6
0.8
0
200
400
600
800

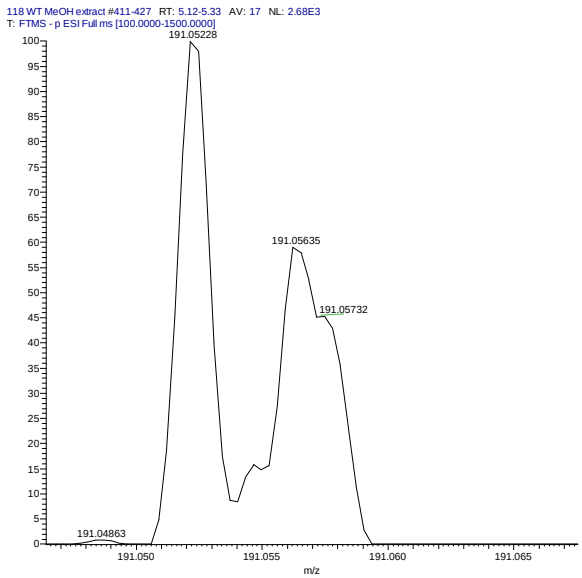
Appendix B Organic acid analysis on the HILIC-Orbitrap-MS

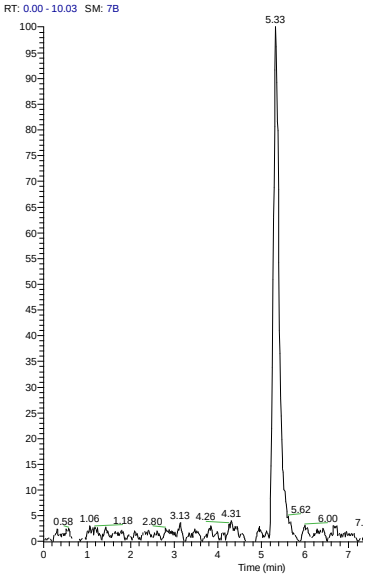
Wildtype *Arabidopsis*

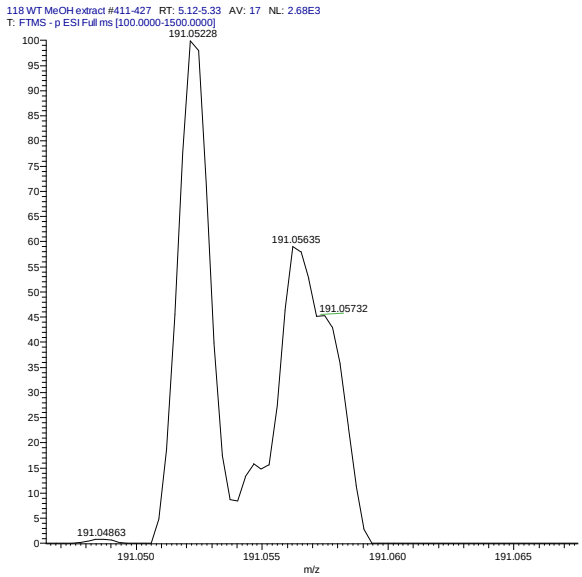


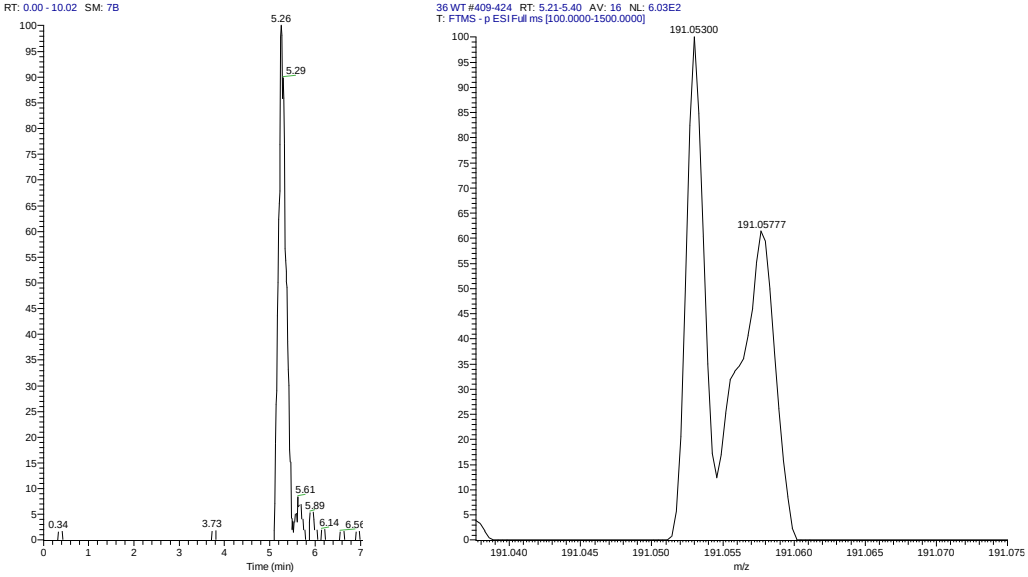


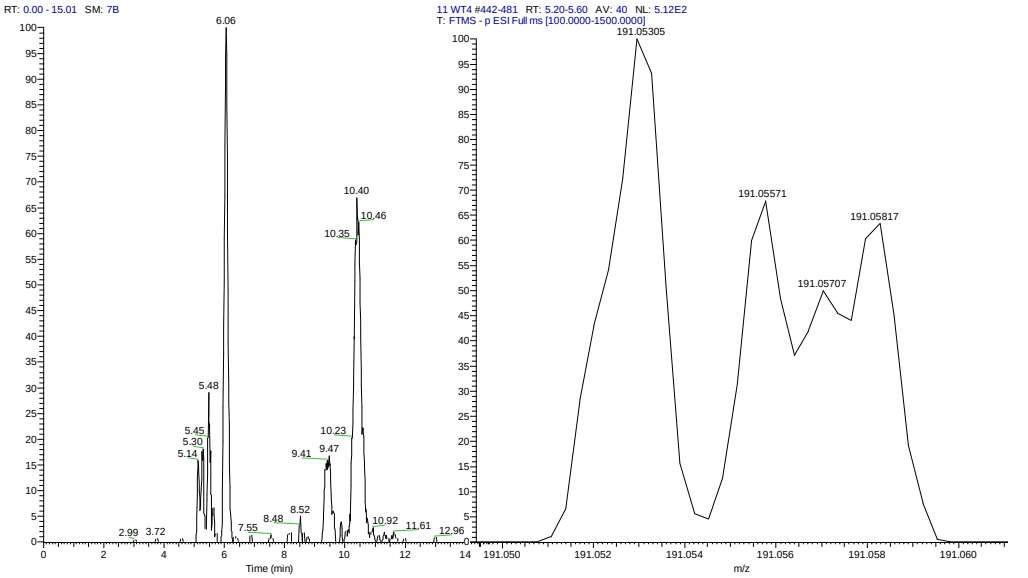


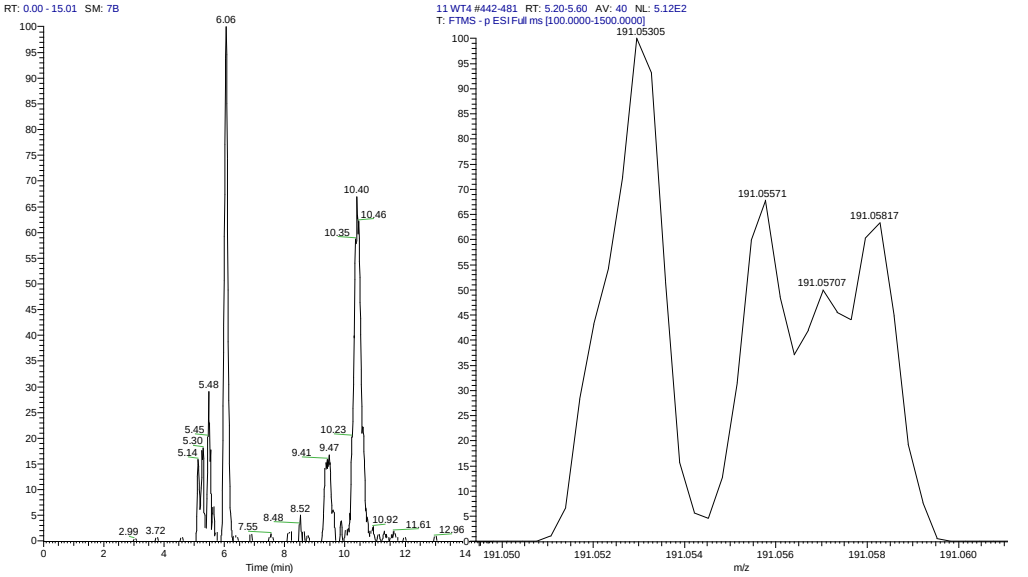




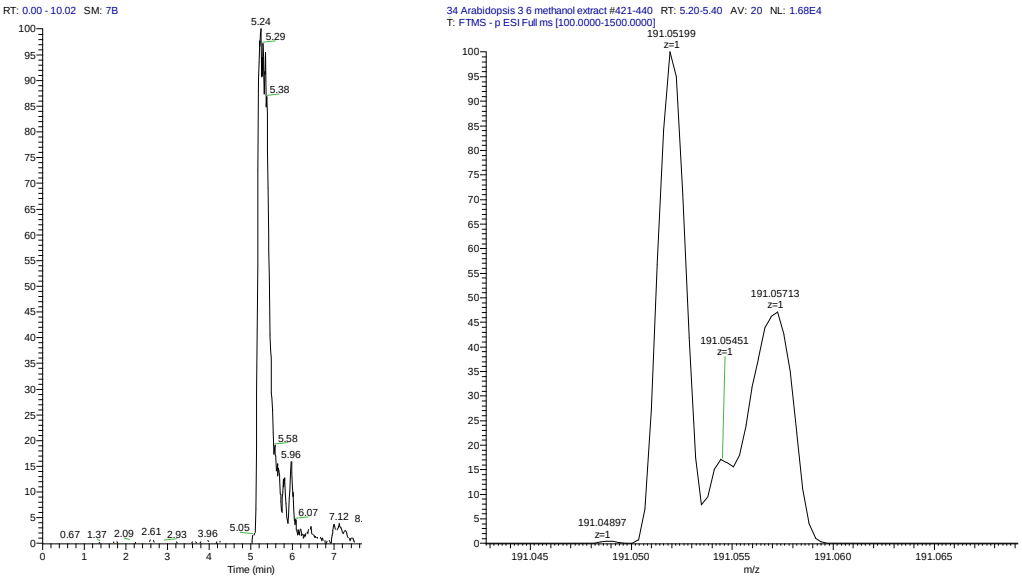


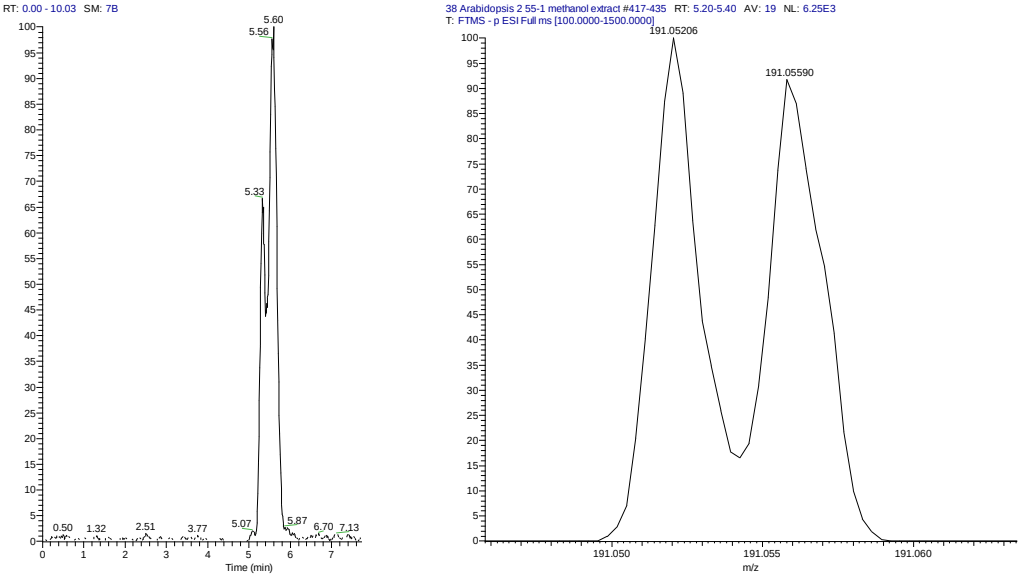

RT: 0.00 - 10.02 SM: 7B
5.26
5.29
5.61
5.89
6.14
0.34
3.73
Time (min)
36 WT #409-424 RT: 5.21-5.40 AV: 16 NL: 6.03E2
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05300
191.05777
m/z

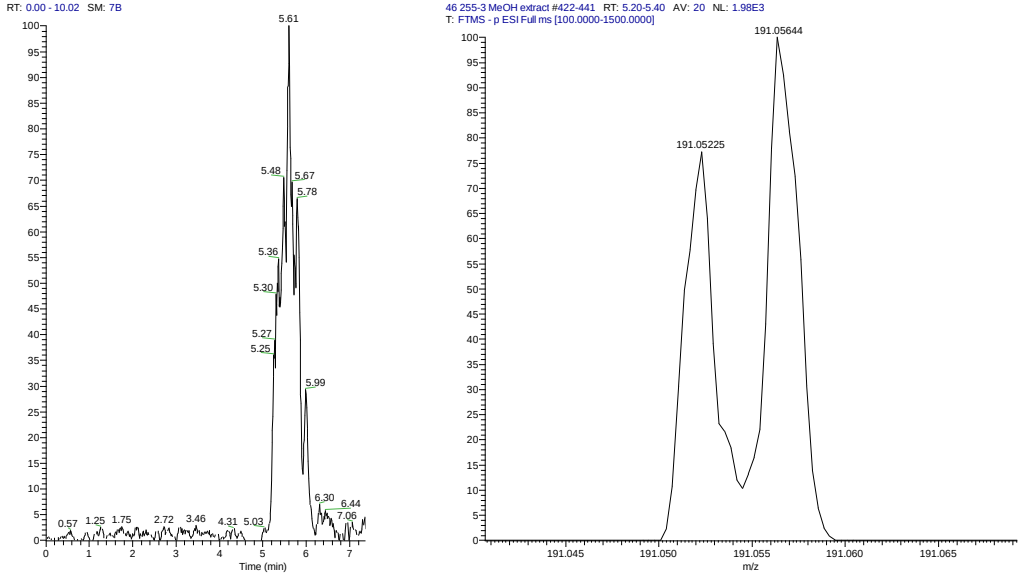

RT: 0.00 - 15.01 SM: 7B
6.06
10.40
10.46
10.35
5.48
5.45
5.30
5.14
10.23
9.41
9.47
8.48
8.52
7.55
10.92
11.61
12.96
2.99
3.72
Time (min)
11 WT4 #442-481 RT: 5.20-5.60 AV: 40 NL: 5.12E2
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05305
191.05571
191.05817
191.05707
m/z

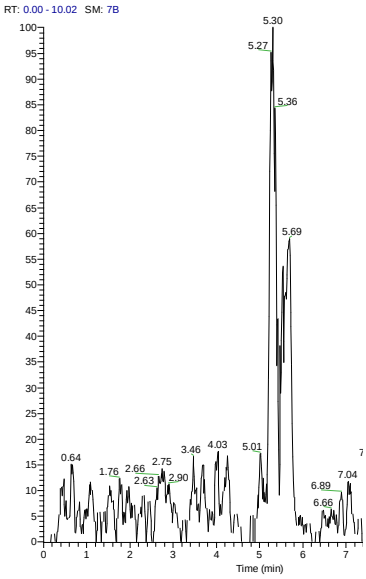

RT: 0.00 - 15.01 SM: 7B
6.06
10.40
10.46
10.35
5.48
5.45
5.30
5.14
10.23
9.41
9.47
8.48
8.52
7.55
10.92
11.61
12.96
2.99
3.72
Time (min)
11 WT4 #442-481 RT: 5.20-5.60 AV: 40 NL: 5.12E2
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05305
191.05571
191.05817
191.05707
m/z

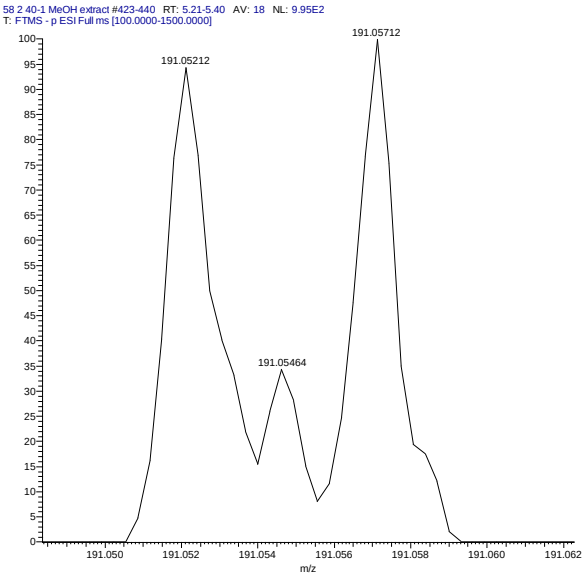
Transgenic *Arabidopsis*

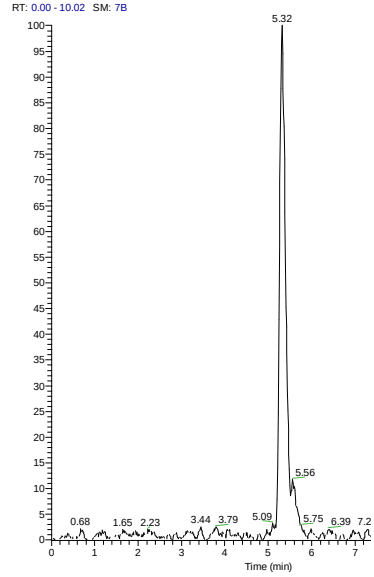


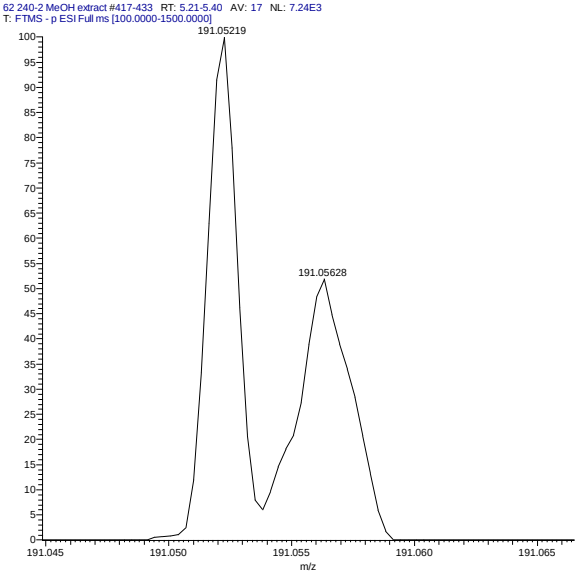


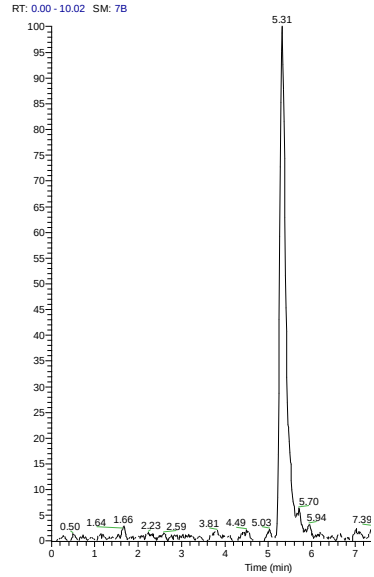


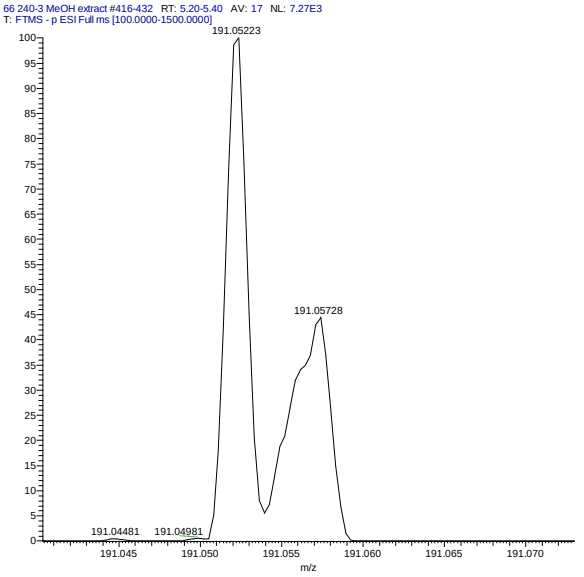

RT: 0.00 - 10.02 SM: 7B
5.30
5.27
5.36
5.69
0.64
1.76
2.66
2.63
2.75
2.90
3.46
4.03
5.01
6.89
7.04
6.66
Time (min)

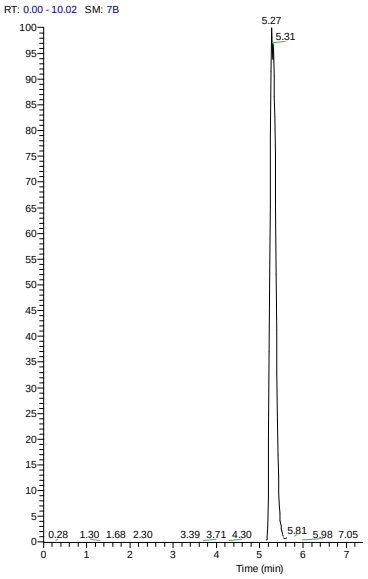

58 2 40-1 MeOH extract #423-440 RT: 5.21-5.40 AV: 18 NL: 9.95E2
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05712
191.05212
191.05464
m/z

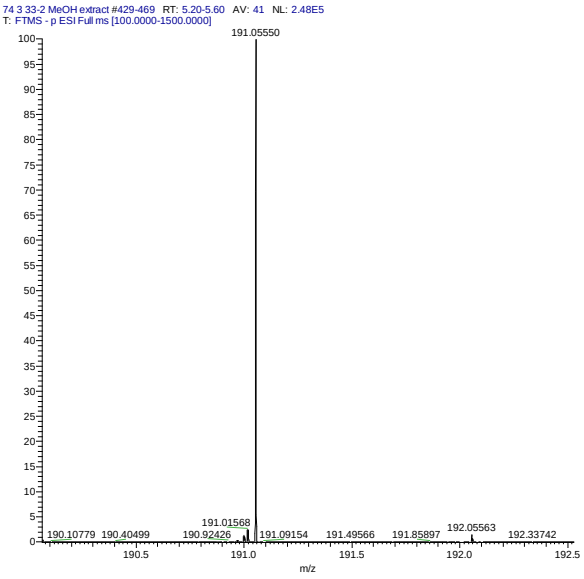

RT: 0.00 - 10.02 SM: 7B
5.32
5.56
0.68
1.65
2.23
3.44
3.79
5.09
5.75
6.39
Time (min)

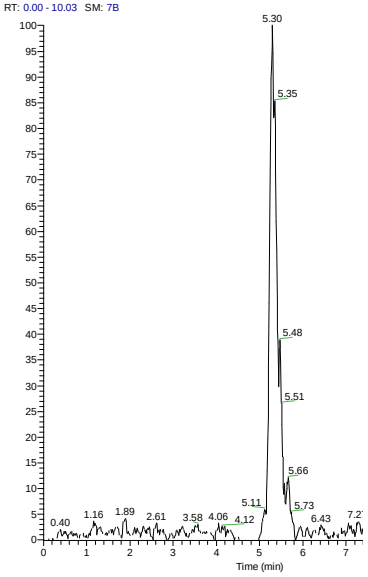

62 240-2 MeOH extract #417-433 RT: 5.21-5.40 AV: 17 NL: 7.24E3
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05219
191.05628
m/z

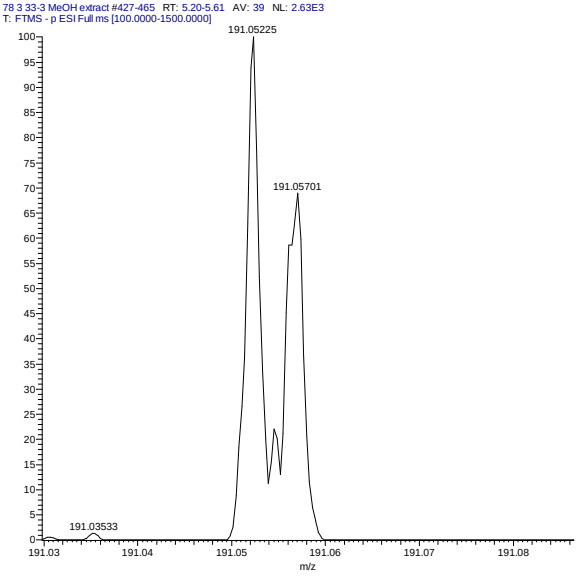

RT: 0.00 - 10.02 SM: 7B
5.31
5.70
5.94
0.50
1.64
1.66
2.23
2.59
3.81
4.49
5.03
Time (min)

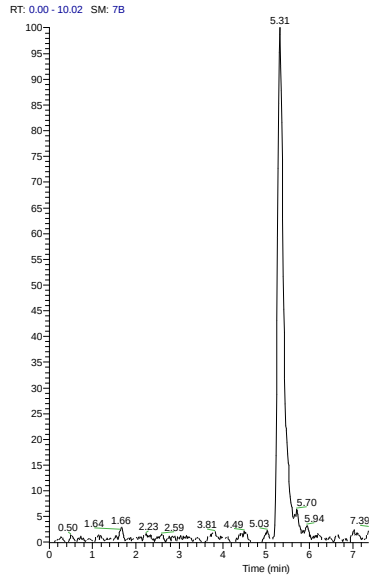

66 240-3 MeOH extract #416-432 RT: 5.20-5.40 AV: 17 NL: 7.27E3
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05223
191.05728
191.04481
191.04981
m/z

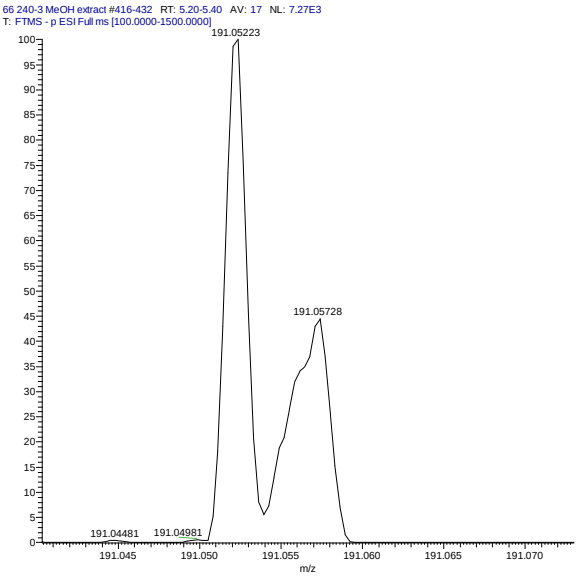

RT: 0.00 - 10.02 SM: 7B
5.27
5.31
0.28 1.30 1.68 2.30 3.39 3.71 4.30 5.81 5.98 7.05
Time (min)

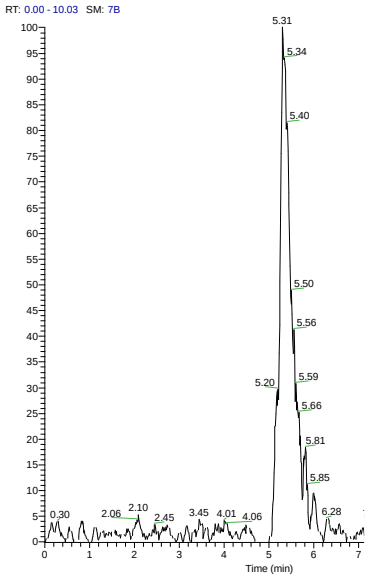

74 3 33-2 MeOH extract #429-469 RT: 5.20-5.60 AV: 41 NL: 2.48E5
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05550
191.01568
190.10779 190.40499 190.92426 191.09154 191.49566 191.85897 192.05563 192.33742
m/z

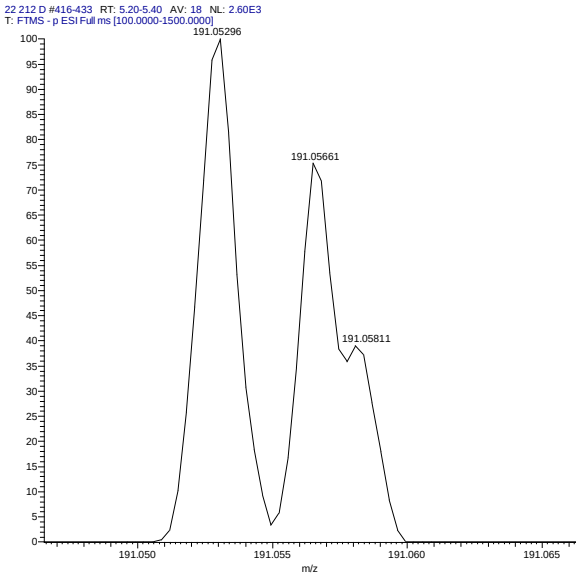

RT: 0.00 - 10.03 SM: 7B
5.30
5.35
5.48
5.51
5.66
5.73
5.11
0.40 1.16 1.89 2.61 3.58 4.06 4.12 6.43
Time (min)

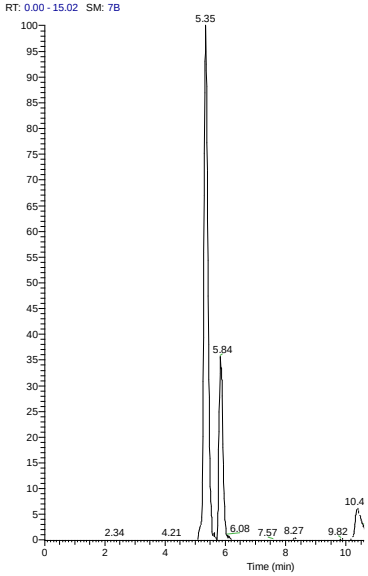

78 3 33-3 MeOH extract #427-465 RT: 5.20-5.61 AV: 39 NL: 2.63E3
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05225
191.05701
191.03533
m/z

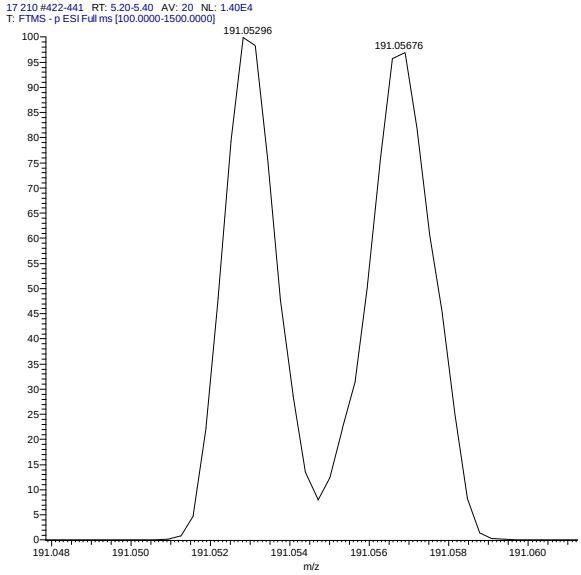

RT: 0.00 - 10.02 SM: 7B
5.31
5.70
5.94
0.50 1.64 1.66 2.23 2.59 3.81 4.49 5.03
Time (min)

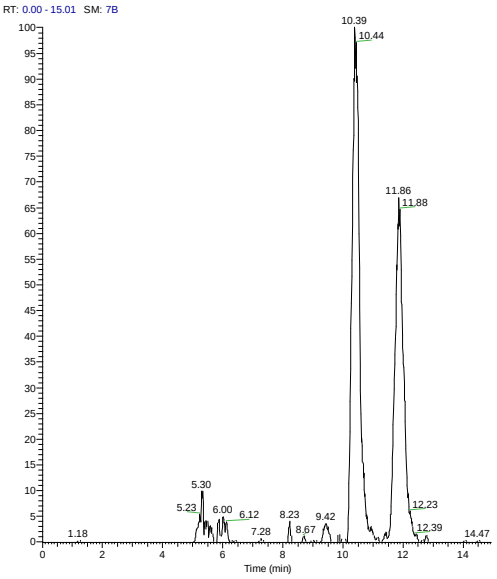

66 240-3 MeOH extract #416-432 RT: 5.20-5.40 AV: 17 NL: 7.27E3
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05223
191.05728
191.04481 191.04981
m/z

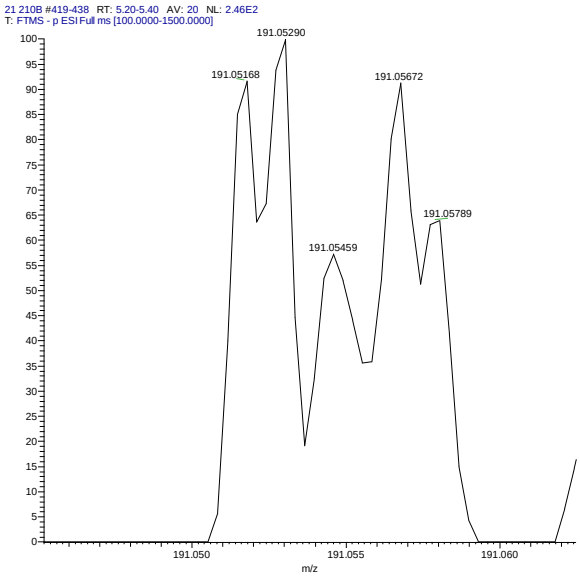

RT: 0.00 - 10.03 SM: 7B
5.31
5.34
5.40
5.50
5.56
5.20
5.59
5.66
5.81
5.85
6.28
0.30
2.06
2.10
2.45
3.45
4.01
4.06
Time (min)

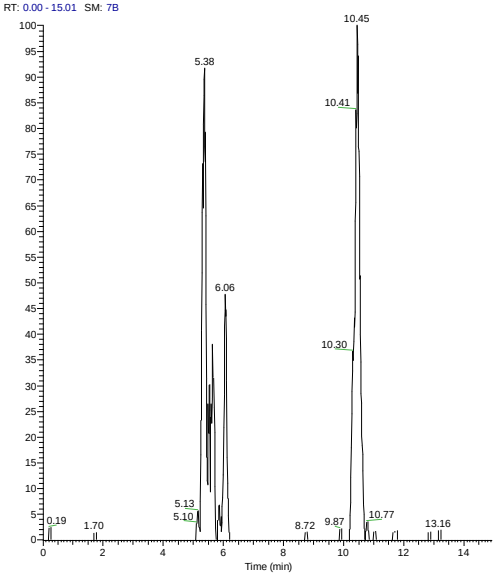

22 212 D #416-433 RT: 5.20-5.40 AV: 18 NL: 2.60E3
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05296
191.05661
191.05811
m/z

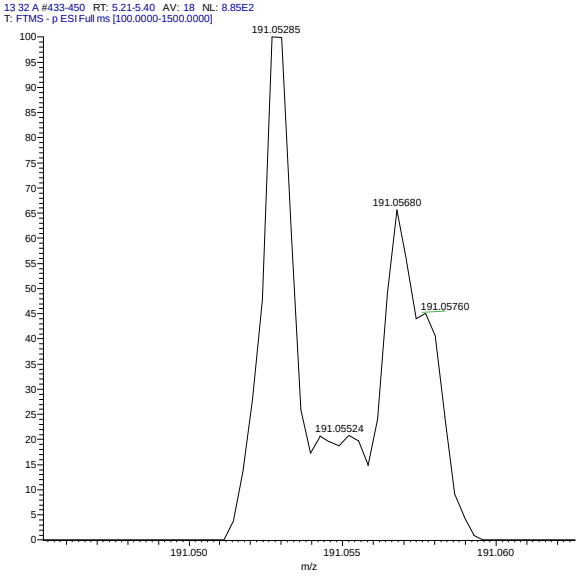

RT: 0.00 - 15.02 SM: 7B
5.35
5.84
10.4
2.34
4.21
6.08
7.57
8.27
9.82
Time (min)

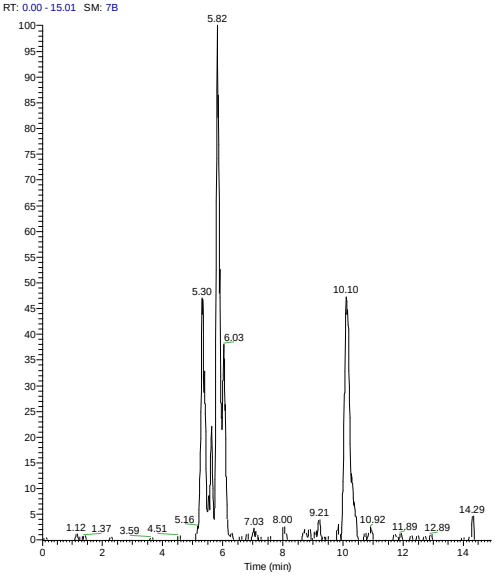

17 210 #422-441 RT: 5.20-5.40 AV: 20 NL: 1.40E4
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05296
191.05676
m/z

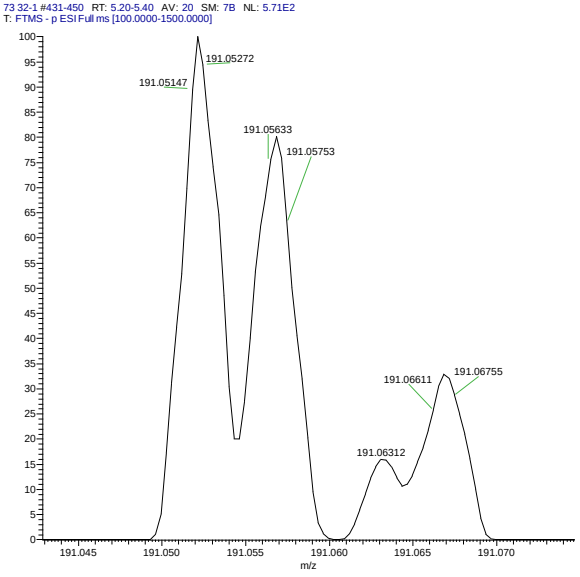

RT: 0.00 - 15.01 SM: 7B
10.39
10.44
11.86
11.88
5.30
5.23
6.00
6.12
8.23
9.42
8.67
7.28
12.23
12.39
1.18
14.47
Time (min)

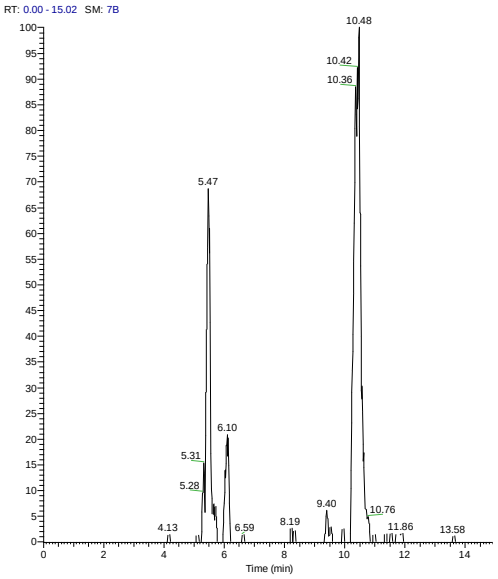

21 210B #419-438 RT: 5.20-5.40 AV: 20 NL: 2.46E2
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05290
191.05168
191.05672
191.05789
191.05459
m/z

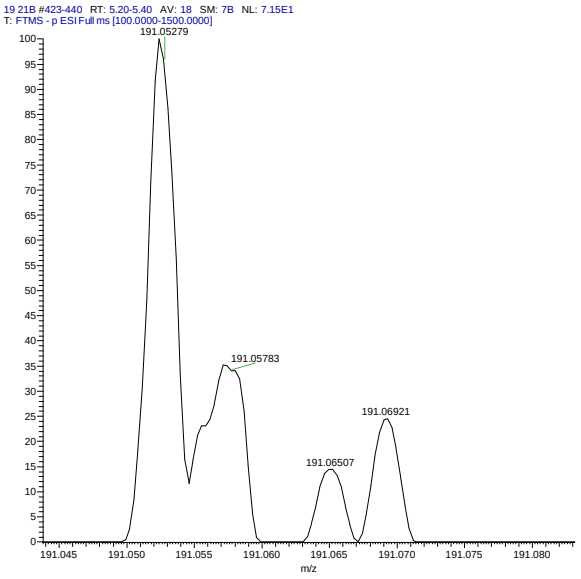

RT: 0.00 - 15.01 SM: 7B
10.45
5.38
10.41
6.06
10.30
5.13
5.10
0.19
1.70
8.72
9.87
10.77
13.16
Time (min)

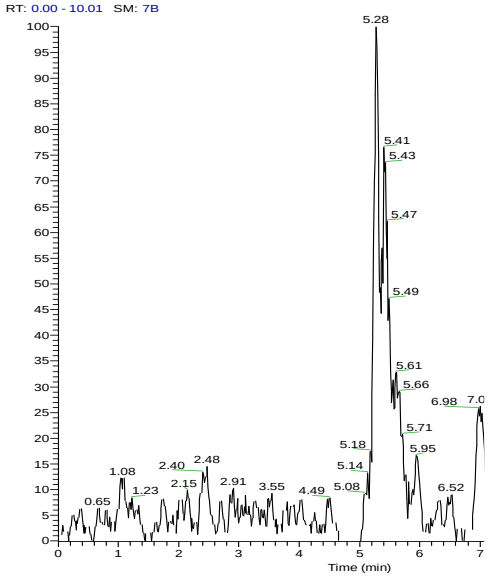

13 32 A #433-450 RT: 5.21-5.40 AV: 18 NL: 8.85E2
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05285
191.05680
191.05760
191.05524
m/z

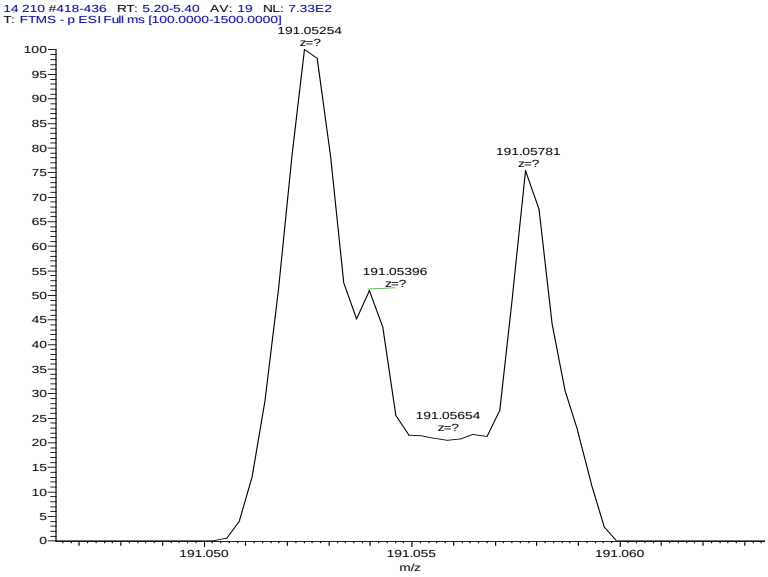

RT: 0.00 - 15.01 SM: 7B
5.82
5.30
6.03
10.10
1.12
1.37
3.59
4.51
5.16
7.03
8.00
9.21
10.92
11.89
12.89
14.29
Time (min)


73 32-1 #431-450 RT: 5.20-5.40 AV: 20 SM: 7B NL: 5.71E2
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05272
191.05147
191.05633
191.05753
191.06611
191.06755
191.06312
m/z

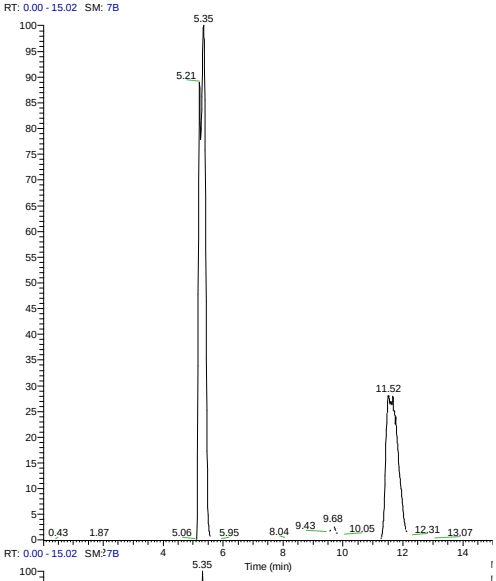

RT: 0.00 - 15.02 SM: 7B
10.48
10.42
10.36
5.47
6.10
5.31
5.28
4.13
6.59
8.19
9.40
10.76
11.86
13.58
Time (min)

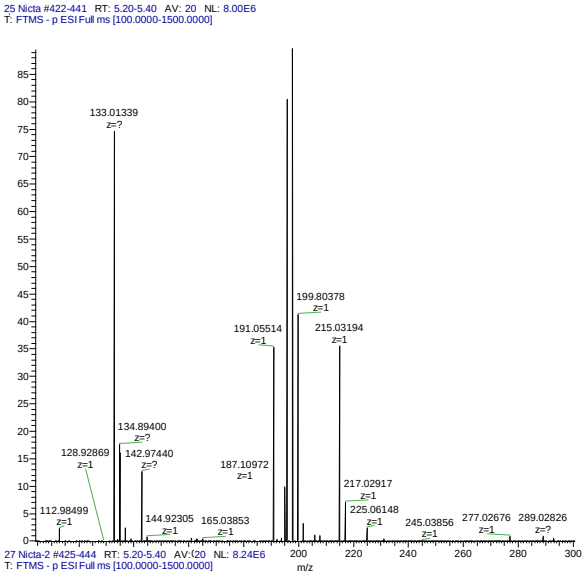

19 21B #423-440 RT: 5.20-5.40 AV: 18 SM: 7B NL: 7.15E1
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05279
191.05783
191.06921
191.06507
m/z

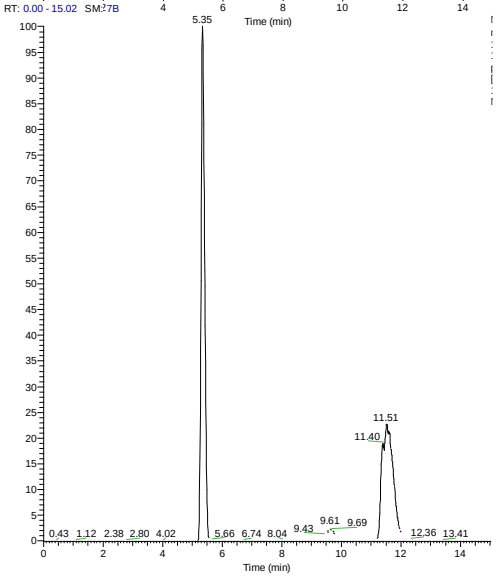

RT: 0.00 - 10.01 SM: 7B
5.28
5.41
5.43
5.47
5.49
5.61
5.66
6.98
7.0
5.71
5.95
5.18
5.14
5.08
1.08
1.23
2.40
2.48
2.15
2.91
3.55
4.49
0.65
6.52
Time (min)

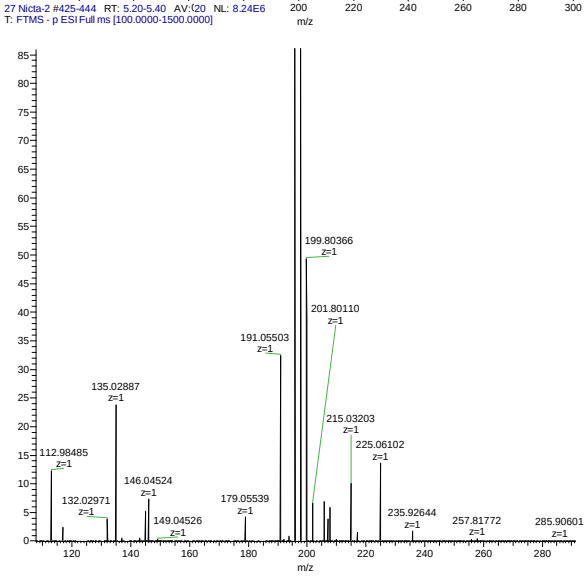

14 210 #418-436 RT: 5.20-5.40 AV: 19 NL: 7.33E2
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05254
z=?
191.05781
z=?
191.05396
z=?
191.05654
z=?
m/z

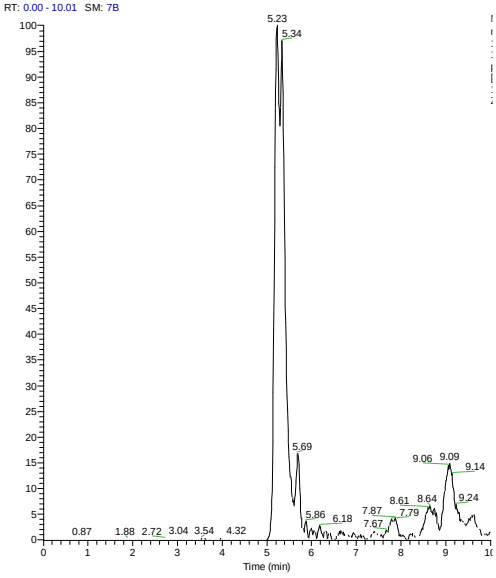
Positive control plants

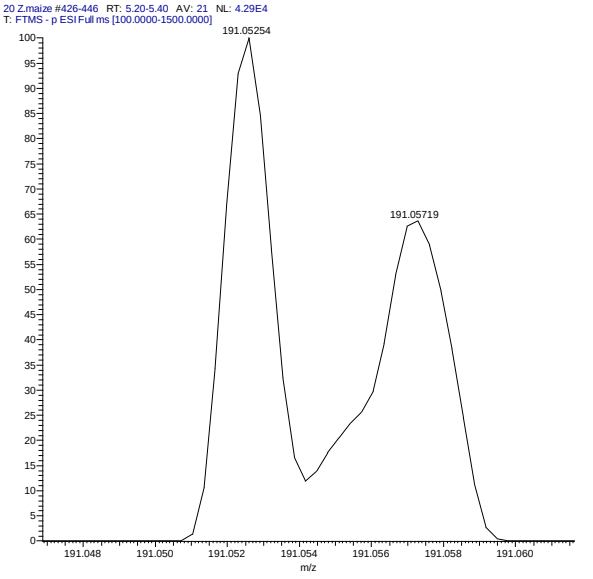


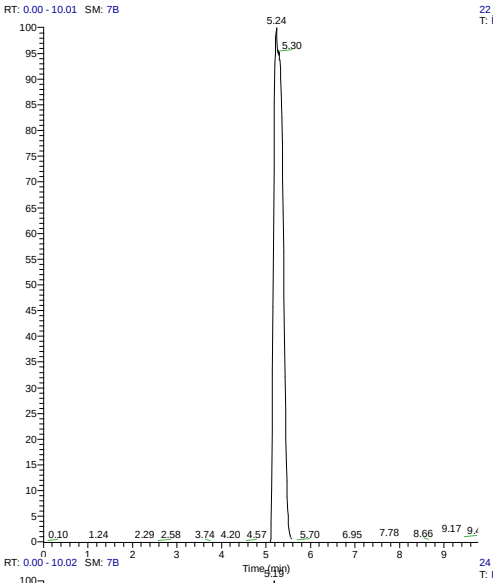


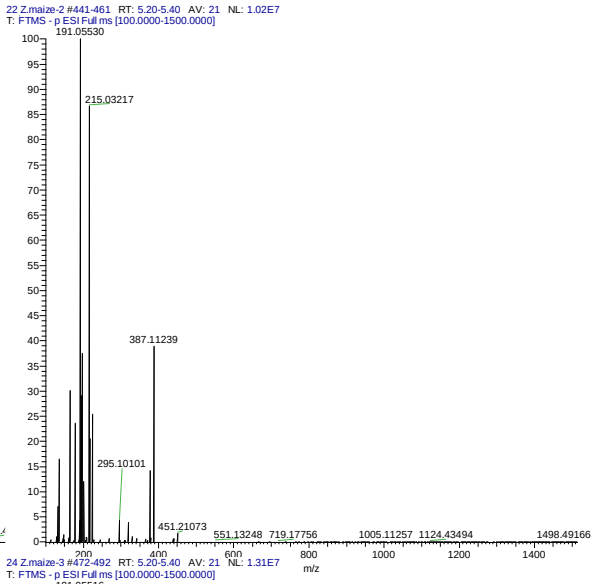


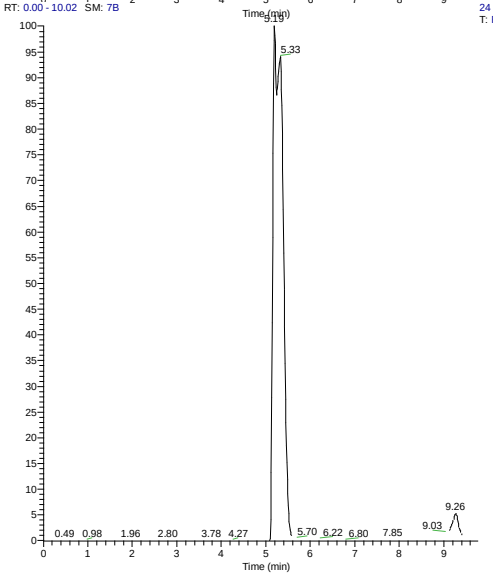


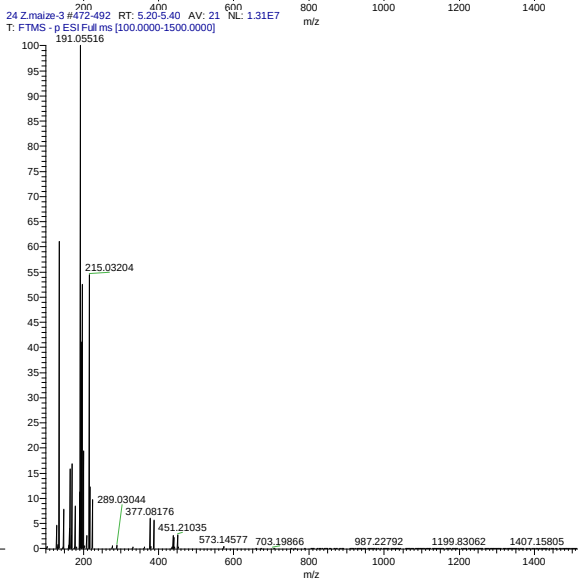


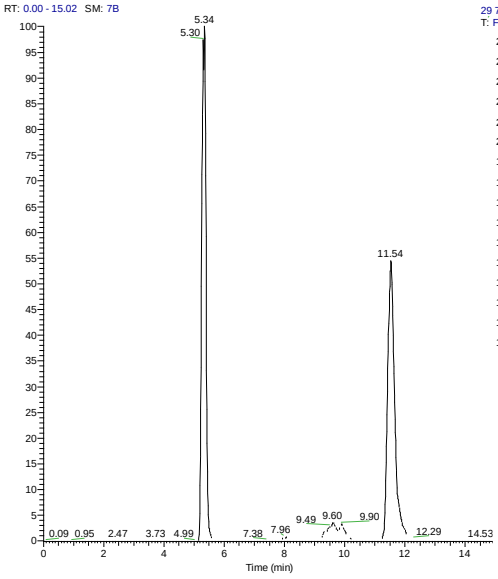


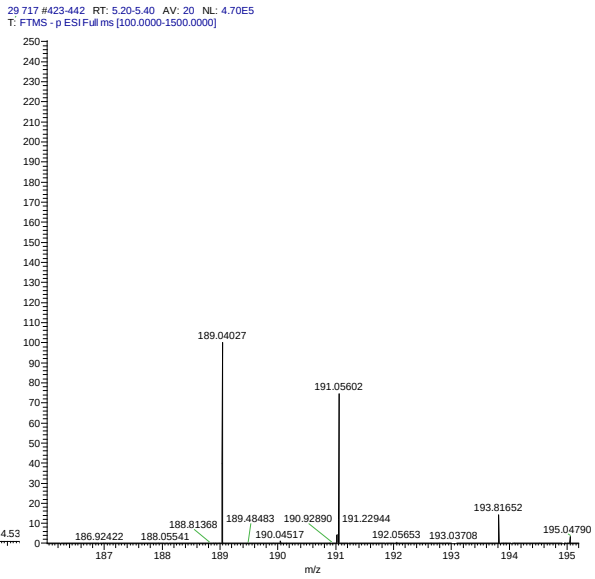

RT: 0.00 - 10.01 SM: 7B
5.24
5.30
Time (min)

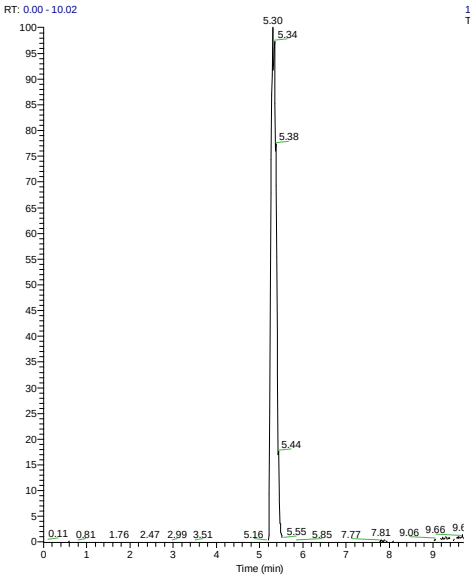

22 Z.maize-2 #441-461 RT: 5.20-5.40 AV: 21 NL: 1.02E7
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05530
215.03217
387.11239
295.10101
451.21073
m/z

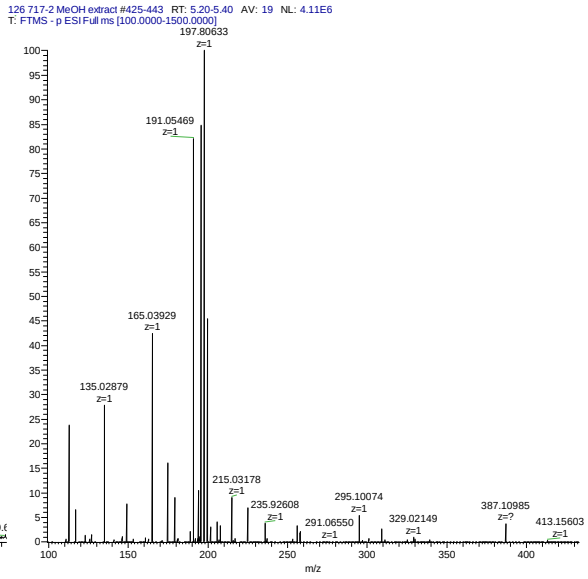

RT: 0.00 - 10.02 SM: 7B
5.19
5.33
9.26
Time (min)

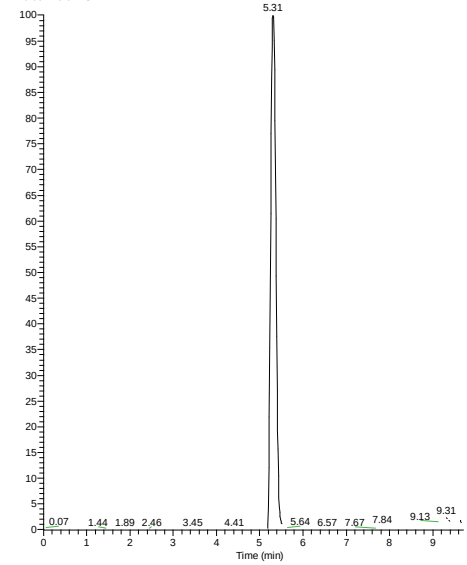

24 Z.maize-3 #472-492 RT: 5.20-5.40 AV: 21 NL: 1.31E7
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05516
215.03204
289.03044
377.08176
451.21035
m/z

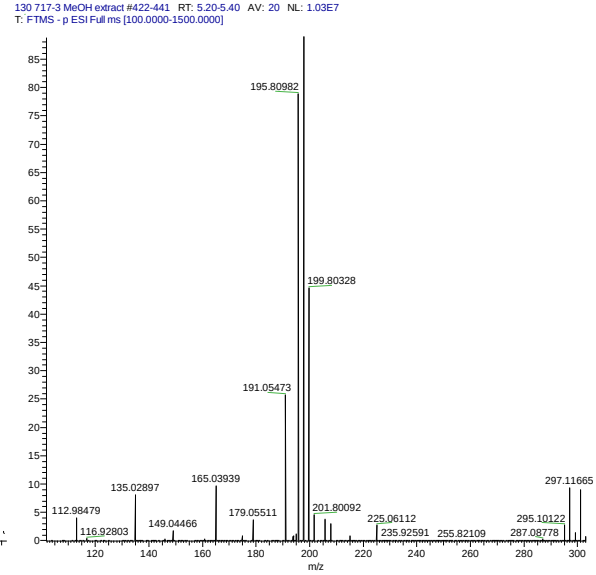

RT: 0.00 - 15.02 SM: 7B
5.34
5.30
11.54
Time (min)

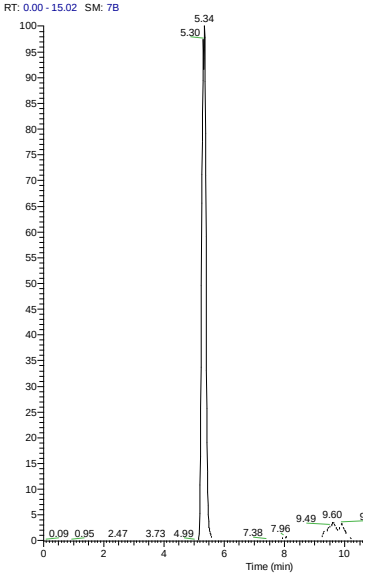

29 717 #423-442 RT: 5.20-5.40 AV: 20 NL: 4.70E5
T: FTMS - p ESI Full ms [100.0000-1500.0000]
189.04027
191.05602
193.81652
m/z

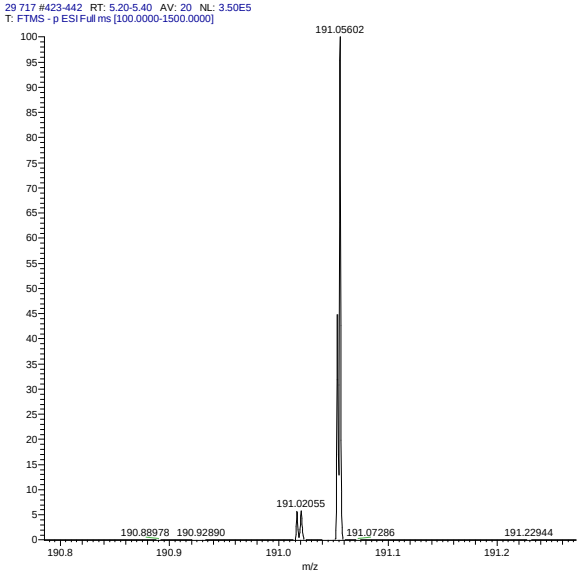

RT: 0.00 - 10.02
5.30
5.34
5.38
5.44
0.11 0.81 1.76 2.47 2.99 3.51 5.16 5.55 5.85 7.77 7.81 9.06 9.66
Time (min)

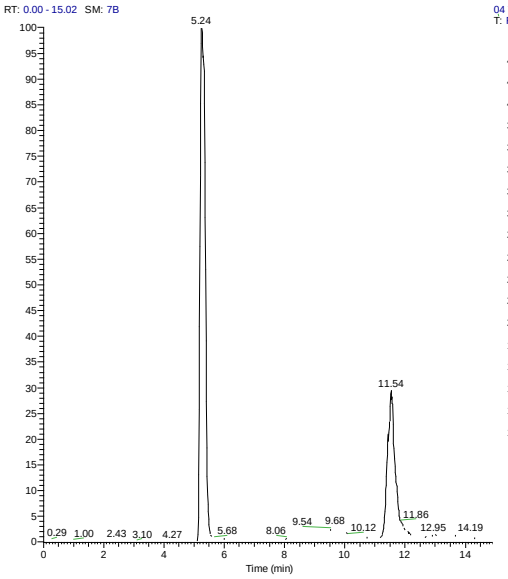

126 717-2 MeOH extract #425-443 RT: 5.20-5.40 AV: 19 NL: 4.11E6
T: FTMS - p ESI Full ms [100.0000-1500.0000]
197.80633 z=1
191.05469 z=1
165.03929 z=1
135.02879 z=1
215.03178 z=1
235.92608 z=1
291.06550 z=1
295.10074 z=1
329.02149 z=1
387.10985 z=?
413.15603 z=1
m/z

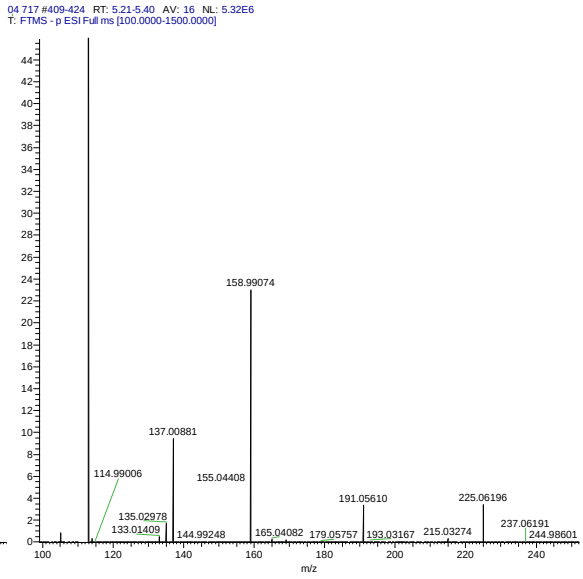

RT: 0.00 - 10.02 SM: 7B
5.31
0.07 1.44 1.89 2.46 3.45 4.41 5.64 6.57 7.67 7.84 9.13 9.31
Time (min)


130 717-3 MeOH extract #422-441 RT: 5.20-5.40 AV: 20 NL: 1.03E7
T: FTMS - p ESI Full ms [100.0000-1500.0000]
195.80982
199.80328
191.05473
135.02897
165.03939
112.98479
116.92803
149.04466
179.05511
201.80092
225.06112
235.92591
255.82109
287.08778
295.10122
297.11665
m/z


RT: 0.00 - 15.02 SM: 7B
5.34
5.30
0.09 0.95 2.47 3.73 4.99 7.38 7.96 9.49 9.60
Time (min)

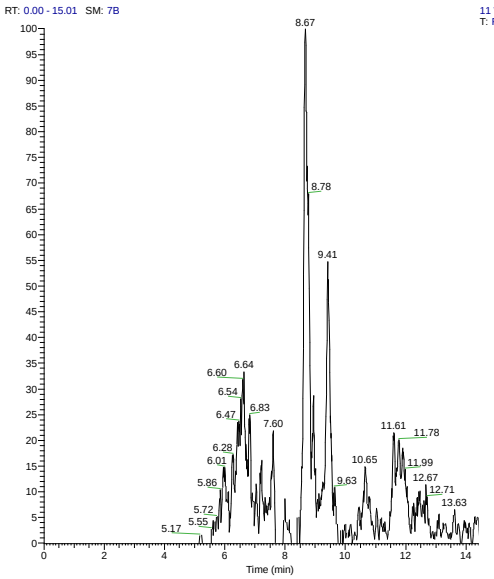

29 717 #423-442 RT: 5.20-5.40 AV: 20 NL: 3.50E5
T: FTMS - p ESI Full ms [100.0000-1500.0000]
191.05602
191.02055
190.88978 190.92890 191.07286 191.22944
m/z

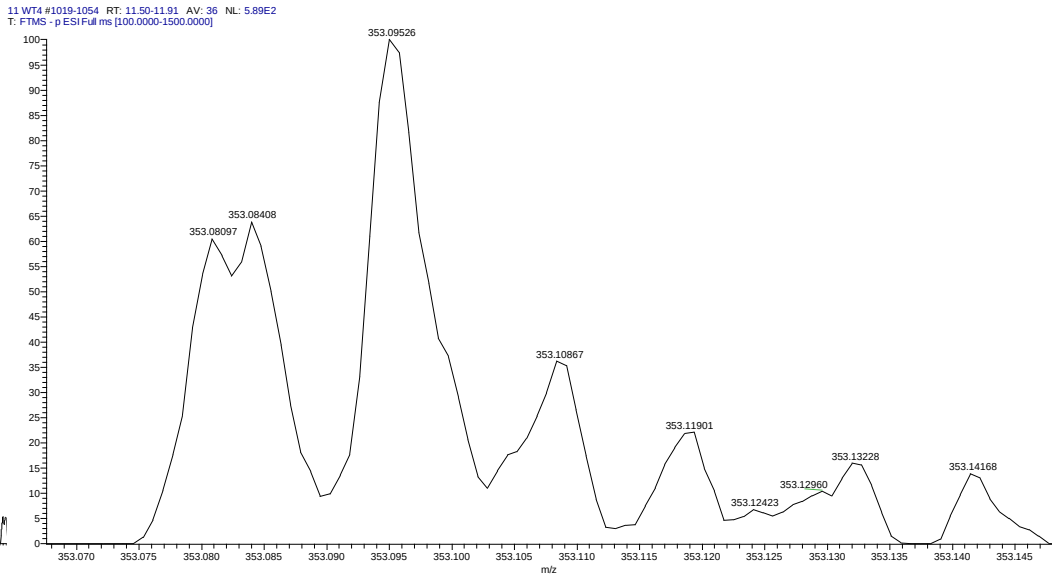

RT: 0.00 - 15.02 SM: 7B
5.24
11.54
0.29
1.00
2.43
3.10
4.27
5.68
8.06
9.54
9.68
10.12
11.86
12.95
14.19
Time (min)


04 717 #409-424 RT: 5.21-5.40 AV: 16 NL: 5.32E6
T: FTMS - p ESI Full ms [100.0000-1500.0000]
158.99074
137.00881
114.99006
155.04408
191.05610
225.06196
135.02978
133.01409
144.99248
165.04082
179.05757
193.03167
215.03274
237.06191
244.98601
m/z

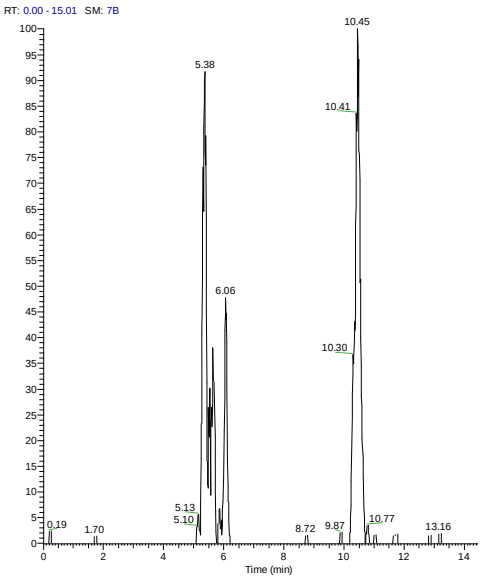
Appendix C Phenolic acid analysis on the HILIC-Orbitrap-MS

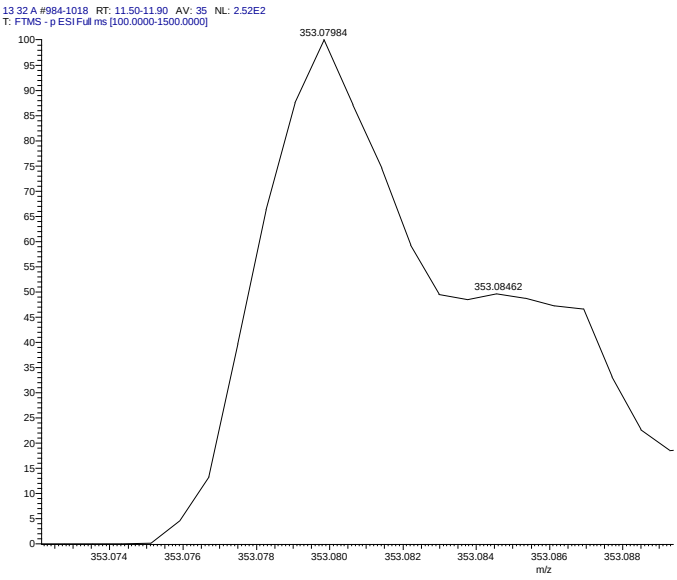
Wildtype *Arabidopsis*

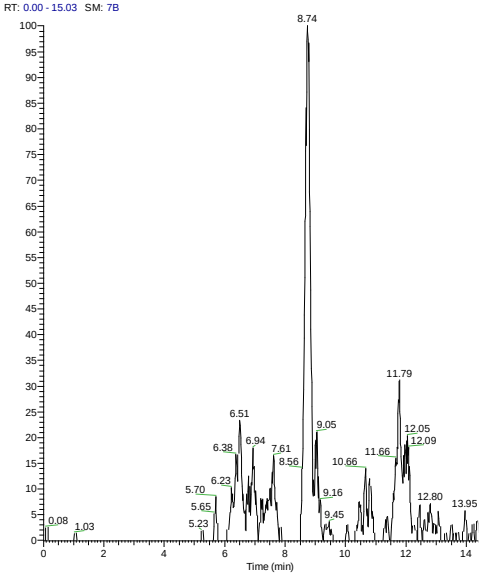


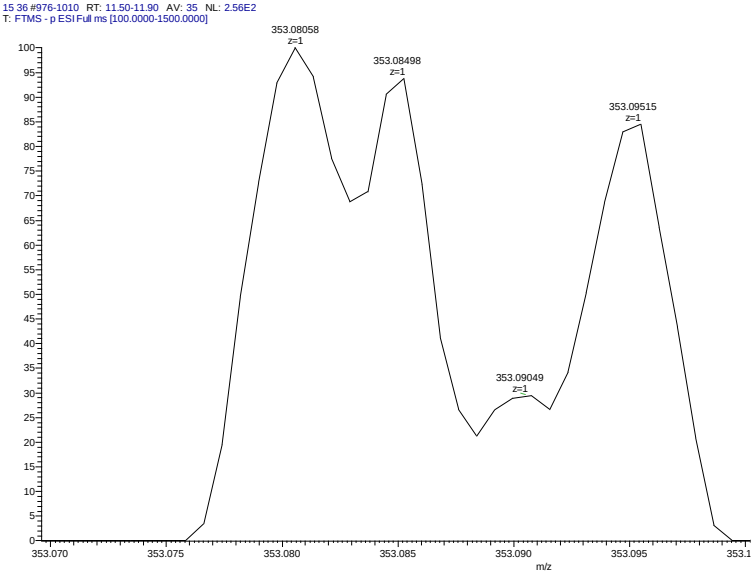


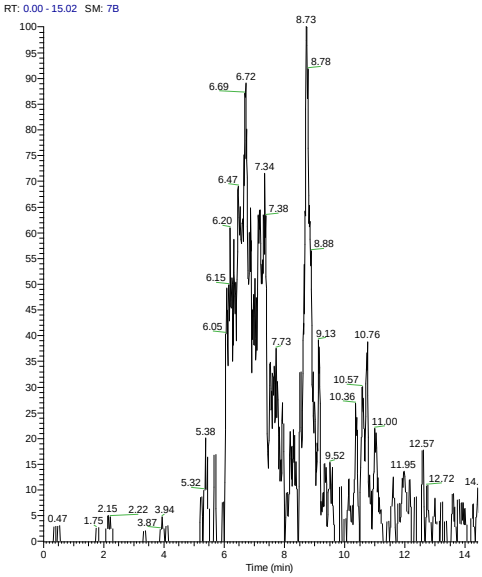
Transgenic *Arabidopsis*

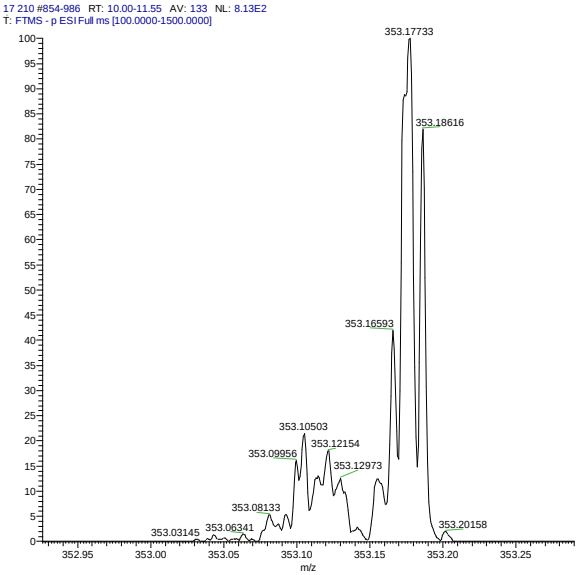


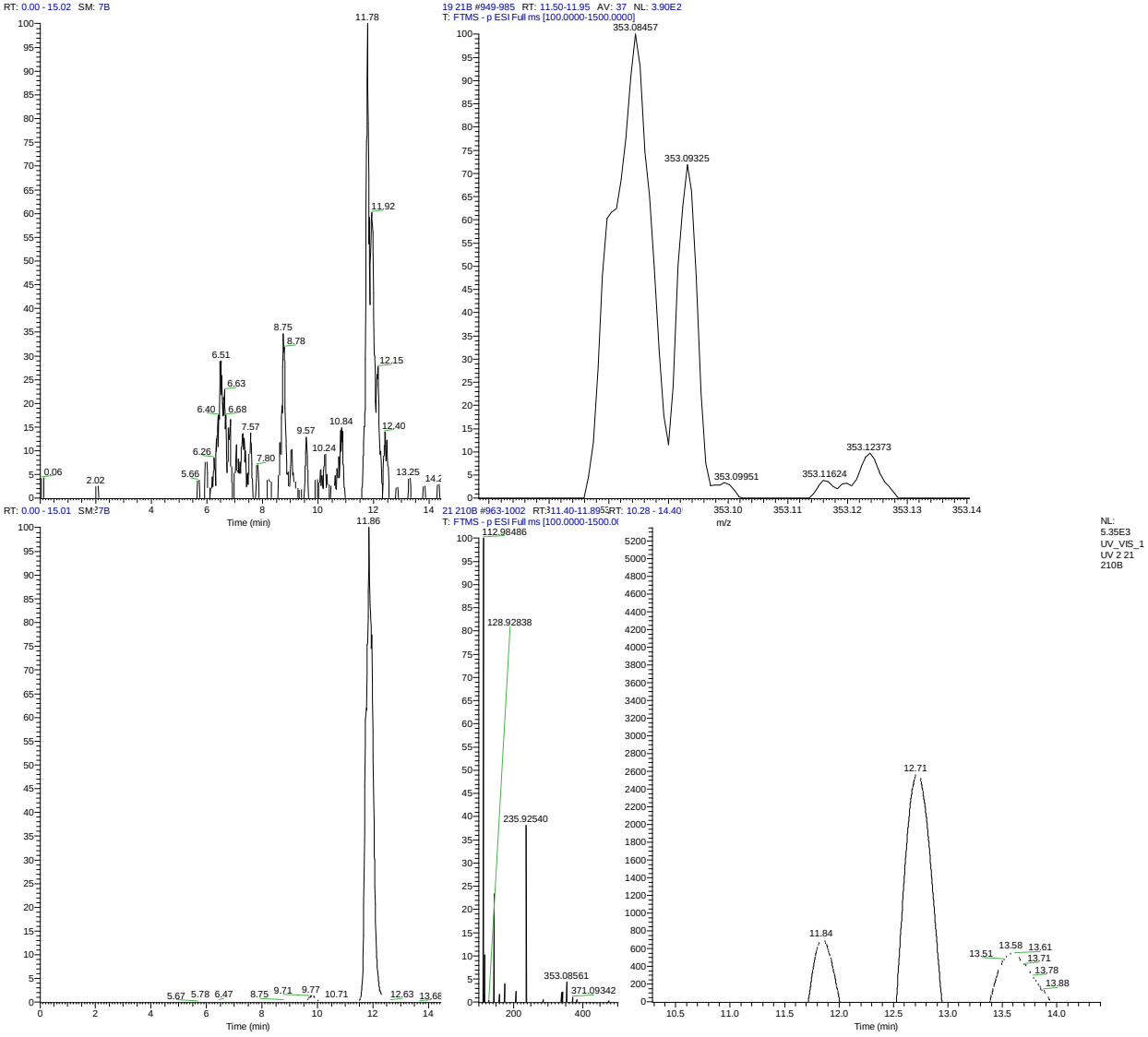




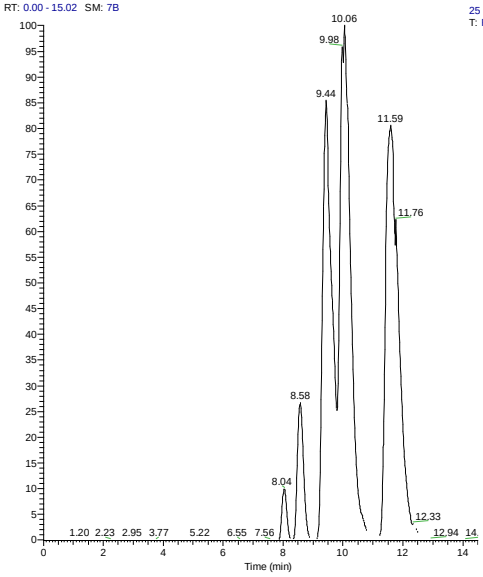


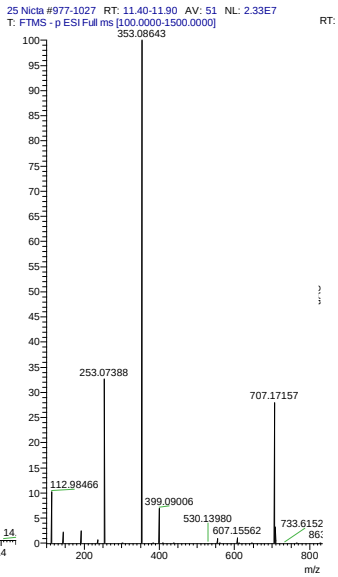


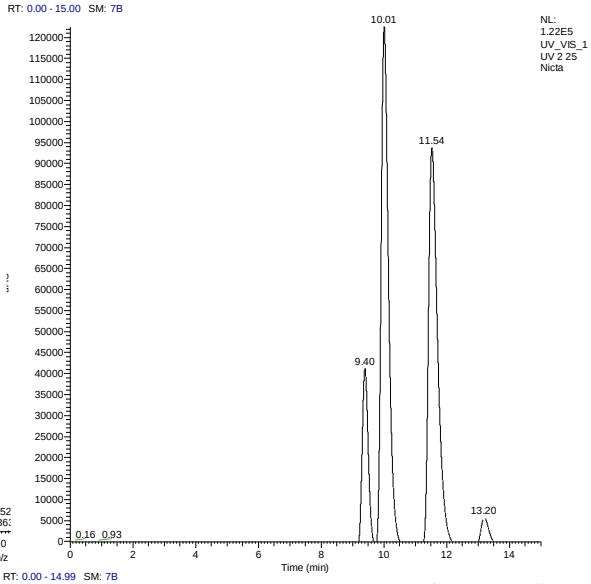


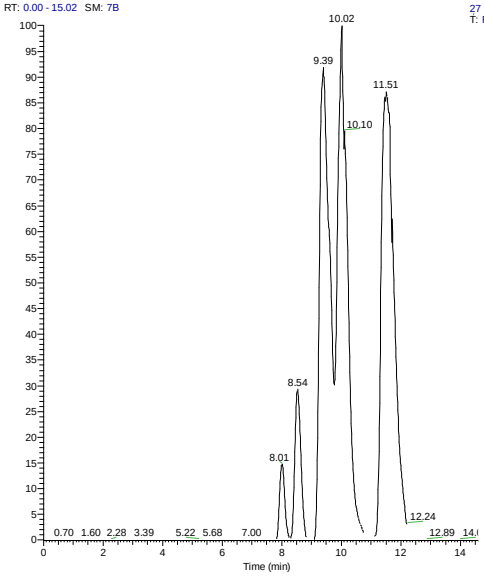

RT: 0.00 - 15.02 SM: 7B
11.78
11.92
8.75
8.78
6.51
6.63
12.15
6.40
6.68
7.57
9.57
10.84
12.40
6.26
10.24
7.80
5.66
13.25
0.06
2.02
Time (min)
19 21B #949-985 RT: 11.50-11.95 AV: 37 NL: 3.90E2
T: FTMS - p ESI Full ms [100.0000-1500.0000]
353.08457
353.09325
353.12373
353.09951
353.11624
m/z
RT: 0.00 - 15.01 SM: 7B
11.86
5.67 5.78 6.47 8.75 9.71 9.77 10.71 12.63
Time (min)
21 210B #963-1002 RT: 11.40-11.895 RT: 10.28 - 14.40
T: FTMS - p ESI Full ms [100.0000-1500.00
112.98486
128.92838
235.92540
353.08561
371.09342
NL:
5.35E3
UV_VIS_1
UV 2 21
210B
12.71
11.84
13.51
13.58
13.61
13.71
13.78
13.88
Time (min)

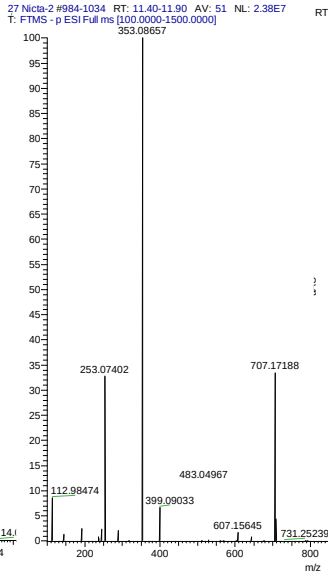
Positive control plants

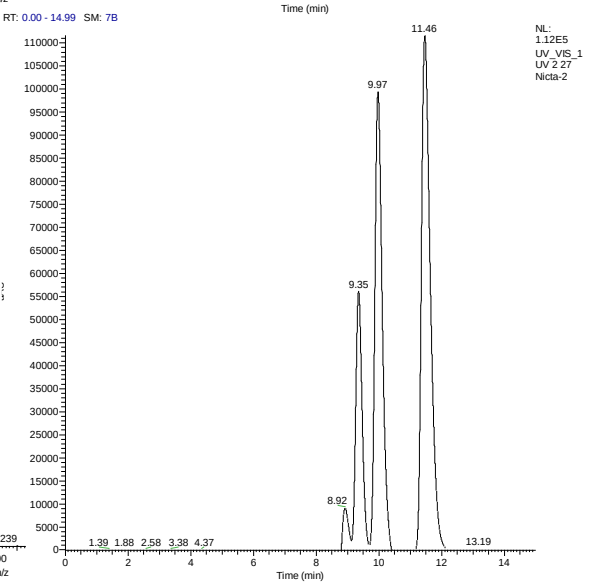


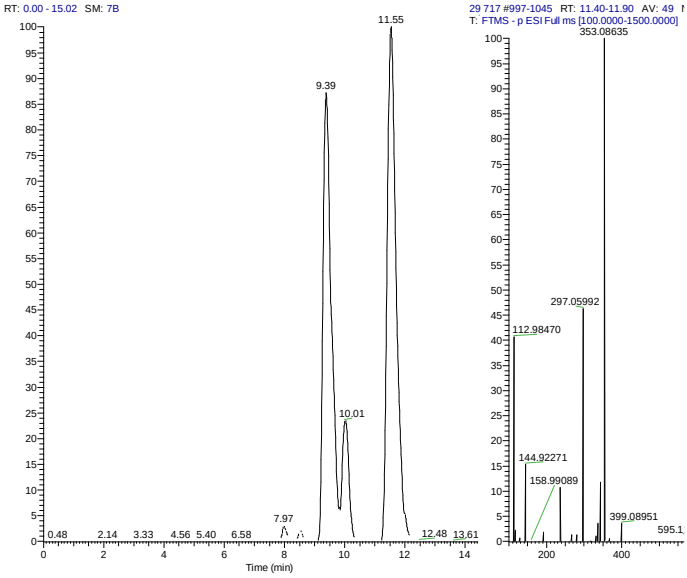


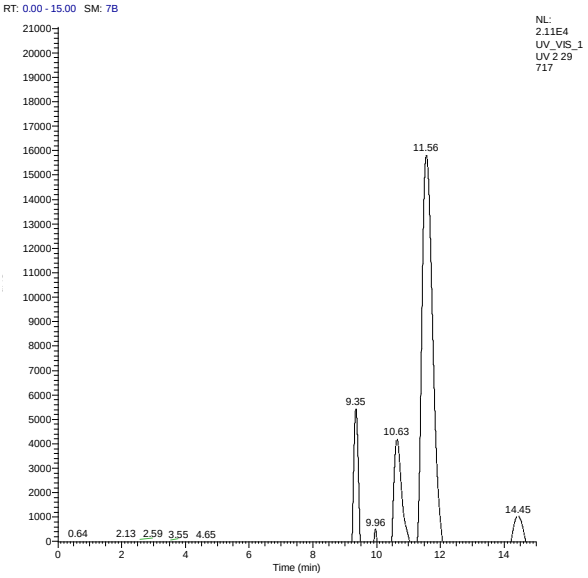












End. Thank you for reading