Evolution of a Secondary Metabolic Pathway from Primary Metabolism: Shikimate and quinate biosynthesis in plants

Yuriko Carrington, Jia Guo, Cuong H. Le, Alexander Fillo, Junsu Kwon, Lan T. Tran, Jürgen Ehlting

2018

https://doi.org/10.1111/tpj.13990

Original Research Article

Evolution of a Secondary Metabolic Pathway from Primary Metabolism: Shikimate and quinate biosynthesis in plants.

Yuriko Carrington[1], Jia Guo[1], Cuong H Le[2], Alexander Fillo[1], Junsu Kwon[1], Lan T Tran[3], Jürgen Ehlting[1*]

[1] Department of Biology & Centre for Forest Biology, University of Victoria, Victoria BC, Canada

[2] Department of Biochemistry and Microbiology & Centre for Forest Biology, University of Victoria, Victoria BC, Canada

[3] Department of Botany, University of British Columbia, Vancouver BC, Canada

*correspondence to Jürgen Ehlting, je@uvic.ca

Running head: Evolution of sikimate and quinate biosynthesis.

Keywords: molecular evolution, secondary metabolism, shikimate / quinate dehydrogenase, *Rhodopirellula baltica*, *Chlamydomonas reinhardtii*, *Physcomitrella patens*, *Selaginella moellendorfii*, *Pinus taeda*, *Populus trichocarpa*.

Summary

The shikimate pathway synthesizes aromatic amino acids essential for protein biosynthesis. Shikimate dehydrogenase (SDH) is a central enzyme of this primary metabolic pathway, producing shikimate. The structurally similar quinate is a secondary metabolite synthesized by quinate dehydrogenase (QDH). *SDH* and *QDH* belong to the same gene family, which diverged into two phylogenetic clades after a defining gene duplication just prior to the angiosperm/gymnosperm split. Non-seed plants that diverged before this duplication harbour only a single gene of this family. Extant representatives from the chlorophytes (*Chlamydomonas reinhardtii*), bryophytes (*Physcomitrella patens*), and lycophytes (*Selaginella moellendorfii*) encoded almost exclusively SDH activity *in vitro*. A reconstructed ancestral sequence representing the node just prior to the gene duplication also encoded SDH activity. QDH activity was gained only in seed plants following gene duplication. QDH enzymes of gymnosperms, represented here by *Pinus taeda*, may be reminiscent of an evolutionary intermediate since they encode equal SDH and QDH activities. The second copy in *P. taeda* maintained specificity for shikimate similar to the activity found in the angiosperm SDH sister clade. The codon for a tyrosine residue within the active site displayed a signature of positive selection at the node defining the QDH clade, where it changed to a glycine. Replacing the tyrosine to a glycine in a highly shikimate-specific angiosperm SDH was sufficient to gain some QDH function. Thus, very few mutations were necessary to facilitate the evolution of *QDH* genes.

Introduction

Gene duplication is a major mechanism providing the raw genetic materials for novel gene evolution and several hypotheses have been developed to explain how these gene copies manage to slip through silencing mutations (Moore and Purugganan, 2005; Weng, 2014). Plant genomes harbour a high number of duplicated genes, many connected to the expansion of specialized (often lineage specific) biochemical pathways in plants called secondary metabolism. Secondary metabolites or plant natural products fulfill diverse and important functions in chemical ecology by modifying precursors supplied by primary metabolism (Kroymann, 2011). One such primary metabolic pathway is the plant shikimate pathway, the biosynthetic route towards the aromatic amino acids (Phe, Tyr, and Trp) (Herrmann and Weaver, 1999). In plants, bacteria and fungi, the shikimate pathway is essential for protein biosynthesis and chemical or genetic interference with this pathway is lethal as demonstrated by the effectiveness of glyphosate as an herbicide and by inclusion of shikimate pathway biosynthetic genes in the catalog of embryo lethal genes in Arabidopsis (Singh and Shaner, 1998; Pagnussat *et al.*, 2005). The shikimate pathway must therefore be stringently maintained under strong purifying selection.

Both the end products and intermediates of the shikimate pathway are also used for synthesizing diverse secondary metabolites (Herrmann and Weaver, 1999). Phenylalanine gives rise to a multitude of phenylpropanoids (e.g. lignin, flavonoids, tannins, and hydroxycinnamic acid conjugates) that are found almost ubiquitously in plants (Vogt, 2010). Among them, the hydroxycinnamic ester chlorogenic acid (CGA) is particularly widespread. It is abundant for example in coffee, members of the *Solanaceae* and *Salicaceae* families (Niggeweg *et al.*, 2004), but is apparently absent in some species, such as the model plant *Arabidopsis thaliana* (Guo *et al.*, 2014). The biological functions of CGA in plants are diverse and range from acting as an insect feeding deterrent (Ikonen *et al.*, 2001; Leiss *et al.*, 2009) to protection against UV radiation (Grace *et al.*, 1998; Clé *et al.*, 2008). Quinate, a precursor of CGA and of other bioactive secondary metabolites is synthesized in a side-branch of the shikimate pathway via the reversible reduction of 3-dehydroquinate by quinate dehydrogenase (QDH) (Figure 1). The QDH from *Populus trichocarpa* shares extensive sequence similarity with the bifunctional enzyme dehydroquinate dehydratase / shikimate dehydrogenase (DQD/SDH) of the primary shikimate

pathway[i], and both enzymes catalyze similar types of reactions (Guo *et al.*, 2014). This points to a common ancestry between SDH (primary metabolism) and QDH (secondary metabolism). As shikimate is an essential intermediate for protein biosynthesis (Herrmann and Weaver, 1999) it shall be expected that a pre-duplication progenitor acted on shikimate. It remains unknown, however, whether the ancestor also encoded at least some QDH activity that was later augmented in one of its gene copies or if this copy gained QDH activity after duplication. Here we used a phylogenetic and comparative *in vitro* biochemical approach to test these alternate hypotheses. A defining gene duplication just prior to the angiosperm / gymnosperm split enabled evolution of QDH function from an ancestral SDH in seed plants, which likely had very little, if any QDH activity. However, this type of neofunctionalization may have been preceded by non-specific, "promiscuous" binding activities of SDHs encoded by single copy genes in earlier derived plant lineages.

Results

Evolutionary history of the SDH/QDH family

The phylogenetic history of the plant SDH/QDH family was examined across major taxonomic groups within the Viridiplantae including the green algae (Chlorophyta), mosses (Bryophyta), lycopods, gymnosperms, and angiosperms (Supplemental Table 1). The aquatic bacterial phylum Planctomycetes was also included as an out-group because this group was most similar in BLAST searches, and because it has previously been identified as the closest relatives of plant *SDH* genes (Richards *et al.*, 2006). Genes encoding for *SDH/QDH* appear to exist as a single copy in Plancotmycetes and the non-seed plants analyzed (chlorophytes, bryophytes, and lycopods), as our database searches revealed only a single full-length sequence per species. All of these taxonomic groups formed monophyletic clades in the phylogeny, generally with good bootstrap support, and follow their expected taxonomic relationships (Figure 2). This is consistent with a single-copy SDH/QDH gene family member being maintained throughout early plant evolution and a single gene being maintained in all three non-seed plant clades analyzed. In contrast, multiple SDH/QDH gene copies are found in most (but not all) seed plant species, both

[i] We here focus on the dehydrogenase domain and activity of the bifunctional enzyme and will therefore refer to the enzyme as SDH only for brevity reasons.

in gymnosperms and in angiosperms. These form two major clades within the seed plants each encompassing a gymnosperm and an angiosperm sister clade (Figure 2). One angiosperm clade was denoted the SDH clade because it contains 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 P. trichocarpa* (Guo *et al.*, 2014). None of the gymnosperm sequences forming the sister clade to the angiosperm SDHs have been characterized previously, but because of its phylogenetic position we denoted this clade as the gymnosperm SDH clade. In contrast, the previously characterized QDHs from *P. trichocarpa* (Poptr*QDH1 and PoptrQDH2*) (Guo *et al.*, 2014) are the only biochemically characterized members of the second clade (Figure 2), which we refer to as the angiosperm QDH clade. None of the gymnosperm sister clade members have been previously characterized, but again solely because of its phylogenetic position we denoted this clade as QDH. Both the QDH and the SDH clades underwent additional duplications at different times during the evolution of the respective lineages, giving rise to clearly separated subclades in each group (Figure 2).

SDH and QDH activity across the green plant lineage

All biochemically characterized SDH or QDH enzymes with sequence information available are from angiosperms (Figure 2). To follow enzymatic specificity throughout the plant lineage, we selected members representing each major clade for biochemical characterization (Figure 2). Species were selected based on available sequence information with a 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 (Figure 2). We included only the most likely reconstructed ancestral sequence for gene synthesis and biochemical characterization. Recombinant His_6-tagged proteins were heterologously expressed in *E. coli* and purified by affinity chromatography. Purified SDH or QDH enzymes had sizes consistent with the expectations based on the DNA constructs employed (Figure 3B, 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 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 3A). 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 (Supplemental Figure 1). SDH from the green alga *C. reinhardtii* and from the planctomycete *R. baltica* displayed lower maximal velocities towards shikimate than land plant SDHs (Table 1), but apparent affinities were similar across all SDHs with K_M values ranging from 100 μM to 280 μM (Table 1). Within seed plants, the pine protein representing the SDH clade (PintaSDH) displayed high activity and specificity for shikimate but no detectable activity with quinate (Table 1) comparable to angiosperm SDHs previously described. Representing the QDH clade from gymnosperms, PintaQDH reacted equally well with both shikimate and quinate (Figure 3A, Table 1). PintaQDH has similar apparent affinities and maximal velocities for both quinate and shikimate with a slightly (1.7 fold) higher specificity for quinate compared to shikimate (based on V_{max}/K_M, Table 1). Like all SDH enzymes tested here, PintaQDH is dependent on $NADP^+$ as a cofactor and showed negligible activities close to the detection limit (not exceeding 0.4 μmole 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^+ over $NADP^+$ as a cofactor with either shikimate or quinate as substrate as previously described (Guo et al., 2014).

Signatures of selection

In order to elaborate on the evolutionary forces that may have changed substrate preferences and therefore physiological functions of SDHs, ratios of nonsynonymous to synonymous substitutions were estimated. Initially using a branch model, signatures of positive selection

across the whole protein sequence were identified along branches leading to all plant SDHs as well as in the branches defining the SDH clade and QDH clade of seed plants (Supplemental Figure 2). We expected positive selection to act on only few active sites residues that define substrate specificity within a framework of a protein that evolved largely under purifying selection to maintain overall activity as a hydrolase. For this reason, we subsequently employed a branch-site model to identify episodic positive selection acting on specific sites over short evolutionary time spans (Yang and Nielsen, 2002; Guindon and Gascuel, 2003). The majority of sites showed no signatures of positive selection, consistent with the overall high degree of conservation of the protein family. However, a total of eight sites were found to have signatures of episodic positive selection along single branches at a false discovery level of $q < 0.05$ ($p < 0.001$). Among them two sites are located within the SDH active site: a Ser and a Thr corresponding to positions 338 and 381 respectively in the *A. thaliana* SDH sequence (Figure 4). Ser^{338} binds to the C1 carboxylate of shikimate (Singh and Christendat, 2006) and is conserved in most true SDHs of seed plants as well as in bacteria and green algae. However, Ser^{338} was substituted by Gly in the branch leading into the land plants and reverted back to Ser in the branch leading into the angiosperm SDH clade (Figure 4B). Both changes display signatures of positive selection albeit with low empirical Bayes factor support (< 5). Thr^{381} is conserved in most members across all SDH clades but was replaced under positive selection by Gly in the branch leading into the seed plant QDH clade. Subsequently this change was fixed in the QDH clade since only synonymous substitutions occurred at this position within the QDH clades and in consequence all extant gymnosperm and angiosperm QDH clade members analyzed encode a Gly at this position (Figure 4C).

Repeating evolutionary history: site directed mutagenesis

As a complement to the positive selection tests, the Ser^{338} to Gly mutation or the Thr^{381} to Gly mutation or both were introduced into the highly shikimate specific *Poptr*SDH1 background using site-directed mutagenesis. Ser^{338} and Thr^{381} in *A. thaliana* correspond to positions 275 and 318 in PoptrSDH1, respectively. The recombinant mutant enzymes were purified and analyzed using SDS-PAGE and Western Blotting, to determine successful purification (Figure 5A). Based on Michaelis Menten kinetic analyses (Figure 5C), the Ser275Gly mutant showed only slightly reduced maximum activity with shikimate (91 ± 5 μmoles NADPH mg^{-1} min^{-1}) compared to

wild-type *Poptr*SDH1 (103 ± 13 μmoles mg^{-1} min^{-1}). The K_M appeared to be relatively unaffected as well (Figure 5C). Notably, this mutant had no detectable activity with quinate. The $Ser^{275}Gly$ change is thus not sufficient to enable gain of quinate activity, consistent with enzymatic properties of non-seed plant SDHs that also contain a Gly at this position but lack detectable activity with quinate. The $Thr^{318}Gly$ mutant yielded only very little enzyme and in consequence the relative amounts of co-purified proteins from *E. coli* is high (Figure 5A). This could reflect a destabilizing effect of the mutation on the protein's active site or 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 $Thr^{318}Gly$ mutant at comparably high substrate concentration (0.6 mM – 5 mM) of shikimate and quinate. At these concentrations, $Thr^{318}Gly$ displayed low, but clearly detectable activities with both shikimate and quinate (Figure 5B, C). In contrast, the $Ser^{275}Gly/Thr^{318}Gly$ double mutant expressed well in *E. coli* and showed *bona fide* QDH activity besides its original SDH activity, which was severely reduced. Although the $Ser^{275}Gly/Thr^{318}Gly$ double mutant is clearly sufficient to confer gain of activity with quinate, its activity was lower than QDH activities of PintaQDH and *Poptr*QDH2 activity. This and the very high K_M value for quinate (2351 ± 1468 μM), suggests that other mutations were probably required to refine conversion of SDH, highly optimized for shikimate biosynthesis, to QDH.

Discussion

A defining gene duplication event just prior to the angiosperm / gymnosperm split (>300 Mya) gave rise to the SDH and QDH type enzymes in seed plants. This gene duplication facilitated the *de novo* evolution of a secondary metabolic pathway (quinate metabolism) from primary (shikimate) metabolism. Sequence analysis combined with functional and mutagenesis data revealed that very few positively selected amino acid changes were sufficient to set in motion functional diversification among *SDH* duplicates. This eventually led to the evolution of *QDH* genes via neofunctionalization.

We found signatures of positive selection (ω>1) in the branch subtending the green lineage suggesting optimization of SDH function in plants. Steps three and four 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). Strong purifying selection pressure to maintain the shikimate pathway is reflected by low ω values obtained for most branches across the phylogeny. In contrast, positive selection was detected in branches subtending the SDH and QDH clades of seed plants. While detection of positive selection in the branch subtending the QDH clade is expected if QDH evolved via neofunctionalization from SDH, it was puzzling to see positive selection acting on the branch leading into the seed plant SDH clade since SDH represents the ancestral function in this model. This left the possibility that dual shikimate and quinate activities were present in the ancestral sequence, which was then subfunctionalized after gene duplication. We rejected this hypothesis based on the enzymatic properties of the reconstructed pre-duplication ancestor and of the extant representatives from all lineages that diverged prior to duplication. All of these sequences showed nearly exclusive SDH activity.

Exploring episodic selection pressures at individual sites across the phylogeny, we found that a Ser and Thr required for binding shikimate (Singh and Christendat, 2006) both changed to Gly under signatures of positive selection. However, introducing a mutation of Ser^{275} (homologous to *Arabidopsis* Ser^{338}) to Gly in the PoptrSDH1 protein did not lead to detectable activity with quinate, and had only minor negative effects on V_{max} and K_m for shikimate. It was therefore surprising to see that this change may have occurred under positive selection (as suggested by the branch site model, albeit with low support) and was maintained for an extended period of time before reverting back to Ser in the angiosperm SDH clade. Indeed, this change coincided with a strong increase in SDH catalytic rates across land plant SDHs compared to green algal (*C. reinhardtii*) and bacterial (*R. baltica*) SDHs. QDH clade members contain an additional mutation, namely a change from Thr^{381} to Gly. The introduction of both changes into the framework of a highly shikimate specific enzyme (PoptrSDH1) is sufficient to confer clearly detectable QDH activity at the expense of notably reduced SDH activity. Unfortunately, the $T^{381}G$ single mutant yielded poor protein yields in the recombinant *E. coli* system we employed.

Nevertheless, our data suggest that this mutation alone is sufficient to gain QDH activity. Likewise, introduction of the equivalent Thr to Gly change into the *A. thaliana* SDH protein is also sufficient to gain QDH activity (Gritsunov *et al.*, 2018). In *Arabidopsis*, Lys^{385} and Asp^{423} 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 SDH/QDH protein family and showed no signs of positive selection acting on them, attesting to their functional relevance in mediating dehydrogenase catalytic activity.

Most functionally characterized SDHs are from angiosperms (Lourenco and Neves, 1984; Diaz and Merino, 1997; Singh and Christendat, 2006; Guo *et al.*, 2014). We here show that SDH sequences from a Planctomycete (*R. baltica*), a green alga (*C. reinhardtii*), a moss (*P. patens*) and a lycopod (*S. moellendorffii*) have comparable *in vitro* SDH properties and have no or negligible activity with quinate.

While *Poptr*QDH2 showed higher specificity towards quinate than shikimate (Guo *et al.*, 2014), *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. Our data suggest that *Pinta*QDH analyzed here corresponds to this isozyme purified from pine needle tissues. Both *Pinta*SDH and *Pinta*QDH preferred NADP over NAD as a coenzyme. Similar trends have been obtained for well-characterized angiosperm SDHs preferring NADP over NAD, fitting their description as NADPH-dependent dehydrogenases (Singh and Christendat, 2006; Muir *et al.*, 2011; Ding *et al.*, 2007). In contrast, and like other studied angiosperm QDHs (Minamikawa, 1977; Refeno *et al.*, 1982; Kang and Scheibe, 1993), poplar QDHs preferred NAD over NADP. *Pinta*QDH therefore appears to represent an intermediate that has the ability to act both on shikimate and quinate but still requires NADP as a coenzyme. Differential activities of *Poptr*QDH and *Pinta*QDHs might be caused by differences in evolutionary rates. While angiosperms underwent rapid evolution and species expansion during the Cretaceous period (125-100 MYA) (de Bodt *et al.*, 2005; Wang *et al.*, 2013), gymnosperms underwent much lower rates of speciation (Buschiazzo *et al.*, 2012; Pavy *et al.*, 2012) and have a seven-fold lower rate of molecular evolution (de la Torre *et al.*, 2017) consistent with the overall shorter branch lengths of both gymnosperm clades within the SDH/QDH phylogeny. Ostensibly, the path towards QDH was already initiated prior to the

angiosperm / gymnosperm split, but in angiosperms QDH activity was optimized faster than in gymnosperms owing to different rates of molecular evolution.

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 becomes optimized for its native function(s) while the other evolves a new one under positive selection subsequently after gene duplication (Moore and Purugganan, 2005). CGA, a major derivative of quinate, is thought to act as a storage and transport form of lignin precursors (Aerts and Baumann, 1994; Mondolot *et al.*, 2006; Singh and Christendat, 2007; Lallemand *et al.*, 2012). Thus, perhaps QDH activity was at least partially driven by its possible involvement in lignin biosynthesis in seed plants. However, apart from lignin biosynthesis, CGA has also been implicated in defense against pathogens (Sheppard and Peterson, 1976) and herbivores (Ikonen *et al.*, 2001; Leiss *et al.*, 2009), as an antioxidant (Grace *et al.*, 1998; Niggeweg *et al.*, 2004) and UV protectant (Clé *et al.*, 2008; Grace *et al.*, 1998) as well as in fruit browning (Weurman and Swain, 1953), so it is plausible that plants faced more than one selection pressure toward QDH function.

It is worth noting that both in the SDH and QDH clades, additional (more recent) gene duplications and losses took place shaping complex gene families in some seed plants clades. Within each clade, *P. trichocarpa* duplicates do not seem to be biochemically distinct, but show differential expression patterns during development and in response to environmental cues (Guo *et al.*, 2014) suggesting physiological divergence. Gritsunov *et al.*, (2018) identified a co-factor shift from NADP to NAD within the angiosperm QDH clade and propose an associated differentiation between catabolic and anabolic QDH functions within the angiosperm QDH clade. Furthermore, minute activity towards gallic acid biosynthesis has been detected in SDH enzymes from grape and walnut (Muir *et al.*, 2011; Bontpart *et al.*, 2016). Vice versa, some angiosperms have returned to a single-copy state maintaining only SDH activity, a likely consequence of gene loss after duplication. This is true for some eudicots (e.g. *A. thaliana*), but is particularly notable for monocots, which appear to generally lack QDH clade members. However, though rare, quinate and its derivatives, such as chlorogenic acid have been discovered in at least a few monocot species (Kweon *et al.*, 2001; Clifford *et al.*, 2006; Shen *et al.*, 2009) and an acyltransferase has been characterized from switchgrass (*Panicum virgatum*) that can

utilize quinate for the production of chlorogenic acid (Escamilla-Treviño *et al.*, 2014). In the absence of a QDH clade member it must be assumed that quinate is produced by an alternative enzyme in this monocot. Interestingly, within the angiosperm SDH clade, monocot sequences cluster into three distinct sub-groups indicating more recent duplications and diversification events. It appears plausible that following loss of *QDH* genes, some descending lineages faced selection pressures to regain QDH activity. This has indeed been found by Gritsunov *et al.*, (2018), who identified a recent SDH duplicate within the *Brassicaceae* that gained QDH independently.

In summary, we found that adaptive mutations in an *SDH* gene duplicate led to the origination of a quinate biosynthetic pathway in seed plants largely via neofunctionalization. We thereby established clear molecular evolutionary and biochemical links between plant primary and secondary metabolism, helping shed further light on the mechanisms driving adaptive plant biochemical diversity. Although unlikely to be of physiological relevance, it is still interesting to note that we observed minute QDH activities for the single-copy SDH from *S. moellendorfii* and the reconstructed pre-duplication ancestral protein. Weng (2014) refers to non-specific catalytic activities as "metabolic noise", which can produce novel metabolites albeit at an initially low efficiency. Such inadvertent activities might become beneficial to a population and driven to fixation by selection (Khersonsky *et al.*, 2006; Weng, 2014). It appears plausible that "metabolic noise" of SDH provided a takeoff point for the evolution of *QDH* genes by neofunctionalization.

Experimental Procedures

Sequence data collection and phylogenetic reconstruction. To find homologs, 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) was performed 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. A total of 238 sequences from 135 plant species covering major lineages of the Viridiplantae were included as well as an additional nine sequences from bacteria belonging to the Planctomycetes to root the phylogeny. Species and sequence information is provided in Supplemental Table 1. A

multiple sequence alignment of amino acid sequences was generated using Dialign v2.2.1 (Subramanian *et al.*, 2005) on the Mobyle@Pasteur platform (Neron *et al.*, 2009). The resulting alignment was trimmed to include only alignment positions with at least 10% diagonal similarity using BioEdit v7.2.5 (Ibis Biosciences). The trimmed alignment was used to estimate the best suited maximum likelihood model for phylogenetic reconstruction with ProtTest v3.4 (Darriba *et al.*, 2011) and the resulting model (Le and Gascuel model with a proportion of invariant sites and with rate variation among sites [LG +I +G]) was used for reconstructing a maximum likelihood phylogeny and for bootstrap analyses (1,024 replicates) using PhyML v3.2 mpi (Guindon and Gascuel, 2003) on the WestGrid computer cluster (Compute Canada). Trees were visualized and rooted using FigTree v1.4.2 (Rambaut, 2014).

Positive selection tests. A maximum likelihood phylogeny derived from amino acid alignments (generated as described above) was used as a guide tree for estimating selection forces. Ratios of nonsynonymous to synonymous substitution rates (dN/dS) were estimated across a codon alignment corresponding to the respective protein alignment. Codons were aligned using PAL2NAL (Suyama *et al.*, 2006). First, a branch model was implemented using Codeml within the PAML v4.5 package(Yang, 2007) with the following settings: seqtype = 1, codonfreq = 2, clock = 0, model = 1, Nsites = 0, icode = 0, fix_kappa = 0, fix_omega = 0, fix_alpha = 1, alpha = 0. Next, the Mixed Effects Model of Evolution (MEME) implemented in DataMonkey (Pond and Frost, 2005) was used to test for episodic positive selection acting on individual sites (Murrell *et al.*, 2012). We used the automatic model selection tool available on the DataMonkey server to first find the most likely pattern of nucleotide substitutions. A substitution map of sites with episodic signatures of positive selection that were also located within the SDH active site were obtained using the Single Likelihood Ancestor Counting (SLAC) method (Kosakovsky Pond and Frost, 2005) also implemented in DataMonkey.

Ancestral Reconstruction. Ancestral SDH/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).

Gene cloning and recombinant protein purification. Plasmid constructs with either *PoptrSDH1* or *PoptrQDH2 from P. trichocarpa* in the pQE30 expression vector (Qiagen) 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 (sequences provided in Supplemental Table 1) and chemically synthesized by Genescript and obtained in the pUC57 vector. These sequences were sub-cloned 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* DH5α. Sanger sequencing was performed by Sequetech 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 (OD_{600}) of 0.4 to 0.6. Protein expression was induced with 0.06 mM isopropyl-1-1-thio-β-D-galactoside (IPTG) for 24 hrs at 19˚C. Cells were harvested by centrifuging and stored at -80˚C for at least one hour. Recombinant proteins were purified by Ni-NTA affinity chromatography. Frozen cell pellets were re-suspended 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 hr with gentle rocking prior to sonication (5 x 10 s 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% Nickel-NTA agarose beads (Qiagen) for 1 hr on ice with gentle rocking. The Ni-NTA lysate was washed 2 times with 10 mL of wash buffer 1 (50 mM NaH_2PO_3 and 300 mM NaCl) followed by 3 washes with 4 mL of wash buffer 2 (50 mM NaH_2PO_3, 300 mM NaCl and 20 mM imidazole) and eluted 4 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.

SDS PAGE and Western Blotting. A fraction of each protein elution was analyzed either by SDS-PAGE or Western Blotting to assess their purities. For SDS-PAGE, elutions were boiled for 20 min in 2 X crack buffer prior to gel electrophoresis. Protein samples were separated on a 10%

polyacrylamide and visualized by staining with GelCode Blue Stain Reagent (Thermo Fisher Scientific) 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™ Kit (Thermo Fisher Scientific) following the manufacturer's recommended methods.

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) 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 to 20 µg) was used per reaction depending on the protein sample and velocity of the observed reaction. Reactions were carried out for 90 s at room temperature. Enzyme activities initially obtained in units of Abs/s, were converted to concentration units (µmoles NADPH s^{-1} mg^{-1}) using the extinction coefficient of NADPH at 340nm ($6.22X10^3$L mol^{-1} cm^{-1}) and normalized for the amount of enzyme used. Kinetic properties were determined by testing multiple (typically ten) shikimate or quinate concentrations with the appropriate cofactor. The apparent K_M value and maximum velocities (Vmax) of three replicates (independent protein purifications) were modelled to the Michaelis-Menten equation using the 'drm' package implemented in R.

Site directed mutagenesis. Two codon sites found to be under positive selection in QDH proteins were introduced into the shikimate-specific SDH1 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) with wild-type *Poptr*SDH1 cloned into pQE30(Guo *et al.*, 2014) as a template. Following the validation of desired mutations by Sanger sequencing (Sequetech), the mutant *Poptr*SDH1 plasmids were electro-transformed into *E.coli* M15 cells. These were cultured for expression of recombinant His_6-tagged wild-type and mutant *Poptr*SDH1. Induction, purification as well as both protein characteristics (i.e. SDS PAGE and Western Blotting) and kinetic analyses of recombinant proteins were carried out as described above.

Author contributions

Conceptualization: JE, JG, and YC; Methodology: JG and YC; Investigation: YC, JG, AF, JK, and LT; Formal Analysis: CHL and JE; Writing – Original Draft: YC; Writing – Review & Editing: JE and YC; Supervision: JE; Funding Acquisition: JE.

Acknowledgments

This work has been supported by a Discovery Grant (to JE) from the Natural Sciences and Engineering Research Council of Canada (NSERC) and YC was supported by a stipend from the NSERC Collaborative Research and Training Experience (CREATE) Program in Forests and Climate Change. AL and JK received undergraduate research project awards from the Centre for Forest Biology at the University of Victoria. We appreciate early access to the 1KP transcriptome database.

Conflict of interest

The authors declare no conflicts of interest.

Supporting Information

Supplemental Table 1. Sequences included in phylogenetic analyses.

Supplemental Figure 1. Michaelis-Menten kinetics of SDH and QDH enzymes.

Supplemental Figure 2. Detection of positive selection across branches.

References

Aerts, R.J. and Baumann, T.W. (1994) Distribution and utilization of chlorogenic acid in Coffea seedlings. *J. Exp. Bot.*, **45**, 497–503.

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

Bodt, S. de, Maere, S. and Peer, Y. van de (2005) Genome duplication and the origin of angiosperms. *Trends Ecol. Evol.*, **20**, 591–597.

Bonner, C.A. and Jensen, R.A. (1994) Cloning of cDNA encoding the bifunctional dehydroquinase.shikimate dehydrogenase of aromatic-amino-acid biosynthesis in Nicotiana tabacum. *Biochem. J.*, **302 (Pt 1**, 11–4.

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

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

Clé, C., Hill, L.M., Niggeweg, R., Martin, C.R., Guisez, Y., Prinsen, E. and 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, M.N., Wu, W. and Kuhnert, N. (2006) The chlorogenic acids of Hemerocallis. *Food Chem.*, **95**, 574–578.

Darriba, D., Taboada, G.L., Doallo, R. and Posada, D. (2011) ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, **27**, 1164–1165.

Degnan, S.M. (2014) Think laterally: horizontal gene transfer from symbiotic microbes may extend the phenotype of marine sessile hosts. *Front. Microbiol.*, **5**, 638.

Diaz, J. and Merino, F. (1997) Shikimate dehydrogenase from pepper (Capsicum annuum) seedlings. Purification and properties. *Physiol. Plant.*, **100**, 147–152.

Ding, L., Hofius, D., Hajirezaei, M.R., Fernie, A.R., Börnke, F. and Sonnewald, U. (2007) Functional analysis of the essential bifunctional tobacco enzyme 3-dehydroquinate dehydratase/shikimate dehydrogenase in transgenic tobacco plants. *J. Exp. Bot.*, **58**, 2053–2067.

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

Goodstein, D.M., Shu, S., Howson, R., et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.*, **40**, D1178–D1186.

Grace, S.C., Logan, B.A. and Adams, W.W. (1998) Seasonal differences in foliar content of chlorogenic acid, a phenylpropanoid antioxidant, in Mahonia repens. *Plant, Cell Environ.*, **21**, 513–521.

Gritsunov, A., Peek, J., Caballero, J.D., Guttman, D. and Christendat, D. (2018) Structural and biochemical approaches uncover multiple evolutionary trajectories of plant quinate dehydrogenases. *Plant J.*, **this issue**.

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

Guo, J., Carrington, Y., Alber, A. and Ehlting, J. (2014) Molecular characterization of quinate and shikimate metabolism in Populus trichocarpa. *J. Biol. Chem.*, **289**, 23846–58.

Herrmann, K.M. and Weaver, L.M. (1999) The shikimate pathway. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **50**, 473–503.

Ikonen, A., Tahvanainen, J. and Roininen, H. (2001) Chlorogenic acid as an antiherbivore defence of willows against leaf beetles. *Entomol. Exp. Appl.*, **99**, 47–54.

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

Khersonsky, O., Roodveldt, C. and Tawfik, D. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.*, **10**, 498–508.

Kosakovsky Pond, S.L. and Frost, S.D.W. (2005) Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.*, **22**, 1208–1222.

Kroymann, J. (2011) Natural diversity and adaptation in plant secondary metabolism. *Curr. Opin. Plant Biol.*, **14**, 246–251.

Kweon, M.-H., Hwang, H.-J. and Sung, H.-C. (2001) Identification and antioxidant activity of novel chlorogenic acid derivatives from bamboo (Phyllostachys edulis). *J. Agric. Food Chem.*, **49**, 4646–4655.

la Torre, A.R. de, Li, Z., Peer, Y. van de and Ingvarsson, P.K. (2017) Contrasting Rates of Molecular Evolution and Patterns of Selection among Gymnosperms and Flowering Plants. *Mol. Biol. Evol.*, **34**, 1363–1377.

Lallemand, L.A., Zubieta, C., Lee, S.G., et al. (2012) A structural basis for the biosynthesis of the major chlorogenic acids found in coffee. *Plant Physiol.*, **160**, 249–60.

Leiss, K.A., Maltese, F., Choi, Y.H., Verpoorte, R. and Klinkhamer, P.G.L. (2009) Identification of chlorogenic acid as a resistance factor for thrips in chrysanthemum. *Plant Physiol.*, **150**, 1567–75.

Lourenco, E.J. and Neves, V.A. (1984) Partial purification and some properties of shikimate

dehydrogenase from tomatoes. *Phytochemistry*, **23**, 497–499.

Matasci, N., Hung, L.-H., Yan, Z., et al. (2014) Data access for the 1,000 Plants (1KP) project. *Gigascience*, **3**, 17.

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

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

Moore, R.C. and Purugganan, M.D. (2005) The evolutionary dynamics of plant duplicate genes. *Curr. Opin. Plant Biol.*, **8**, 122–128.

Muir, R.M., Ibáñez, A.M., Uratsu, S.L., et al. (2011) Mechanism of gallic acid biosynthesis in bacteria (Escherichia coli) and walnut (Juglans regia). *Plant Mol. Biol.*, **75**, 555–565.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K. and Kosakovsky Pond, S.L. (2012) Detecting individual sites subject to episodic diversifying selection H. S. Malik, ed. *PLoS Genet.*, **8**, e1002764.

Neron, B., Menager, H., Maufrais, C., Joly, N., Maupetit, J., Letort, S., Carrere, S., Tuffery, P. and Letondal, C. (2009) Mobyle: a new full web bioinformatics framework. *Bioinformatics*, **25**, 3005–3011.

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

O'Leary, N.A., Wright, M.W., Brister, J.R., et al. (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733-45.

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

Pagnussat, G.C., Yu, H.-J., Ngo, Q.A., et al. (2005) Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. *Development*, **132**, 603–614.

Pavy, N., Pelgas, B., Laroche, J., Rigault, P., Isabel, N. and Bousquet, J. (2012) A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. *BMC Biol.*, **10**, 84.

Pond, S.L.K. and Frost, S.D.W. (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*, **21**, 2531–3.

Rambaut, A. (2014) FigTree. *FigTree*. Available at: http://tree.bio.ed.ac.uk/software/figtree/.

Refeno, G., Ranjeva, R. and Boudet, A.M. (1982) Modulation of quinate: NAD+ oxidoreductase activity through reversible phosphorylation in carrot cell suspensions.

Planta, **154**, 193–198.

Richards, T.A., Dacks, J.B., Campbell, S.A., Blanchard, J.L., Foster, P.G., McLeod, R. and Roberts, C.W. (2006) Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal gene transfer, and endosymbiotic replacements. *Eukaryot. Cell*, **5**, 1517–31.

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

Sheppard, J.W. and Peterson, J.F. (1976) Chlorogenic acid and verticillium wilt of tobacco. *Can. J. Plant Sci.*, **56**, 157–160.

Singh, B.K. and Shaner, D.L. (1998) Rapid determination of glyphosate injury to plants and identification of glyphosate-resistant plants. *Weed Technol.*, **12**, 527–530.

Singh, S.A. and Christendat, D. (2006) Structure of Arabidopsis dehydroquinate dehydratase-shikimate dehydrogenase and implications for metabolic channeling in the shikimate pathway. *Biochemistry*, **45**, 7787–96.

Singh, S.A. and Christendat, D. (2007) The DHQ-dehydroshikimate-SDH-shikimate-NADP(H) complex: Insights into metabolite transfer in the shikimate pathway. *Cryst. Growth Des.*, **7**, 2153–2160.

Subramanian, A.R., Weyer-Menkhoff, J., Kaufmann, M. and Morgenstern, B. (2005) DIALIGN-T: an improved algorithm for segment-based multiple sequence alignment. *BMC Bioinformatics*, **6**, 66.

Suyama, M., Torrents, D. and Bork, P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.*, **34**, W609–W612.

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

Wang, B., Zhang, H. and Jarzembowski, E.A. (2013) Early Cretaceous angiosperms and beetle evolution. *Front. Plant Sci.*, **4**, 360.

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

Weurman, C. and Swain, T. (1953) Chlorogenic acid and the enzymic browning of apples and pears. *Nature*, **172**, 678–678.

Yang, Z. (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, **24**, 1586–1591.

Yang, Z. and Nielsen, R. (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.*, **19**, 908–17.

Figures

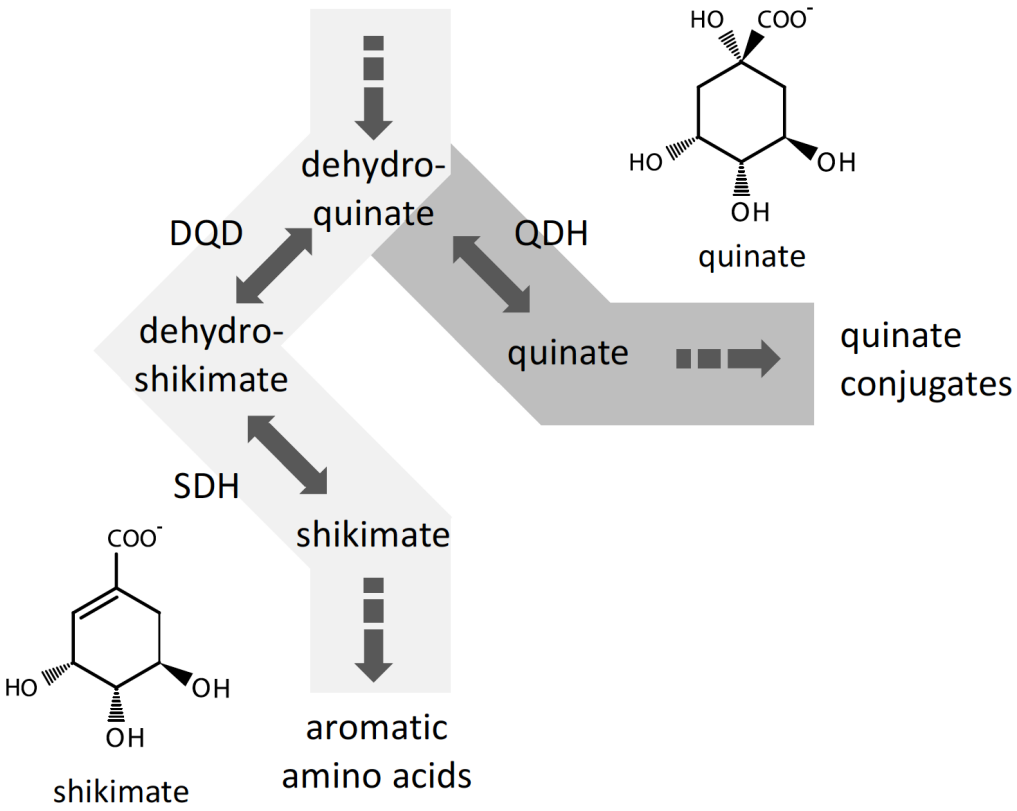

Figure 1: Schematic representation of the proposed link between reactions catalyzed by the shikimate pathway enzyme DQD/SDH and QDH involved in quinate metabolism. Shikimate is produced from dehydroquinate via a two-step reaction and subsequently channelled to downstream reactions in the pathway. Quinate is reversibly formed from a side-branch of the shikimate pathway from dehydoquinate and may be converted to more structurally complex secondary metabolites or to dehydroquinate to fuel the shikimate pathway. Structural similarity between shikimate and quinate is shown.

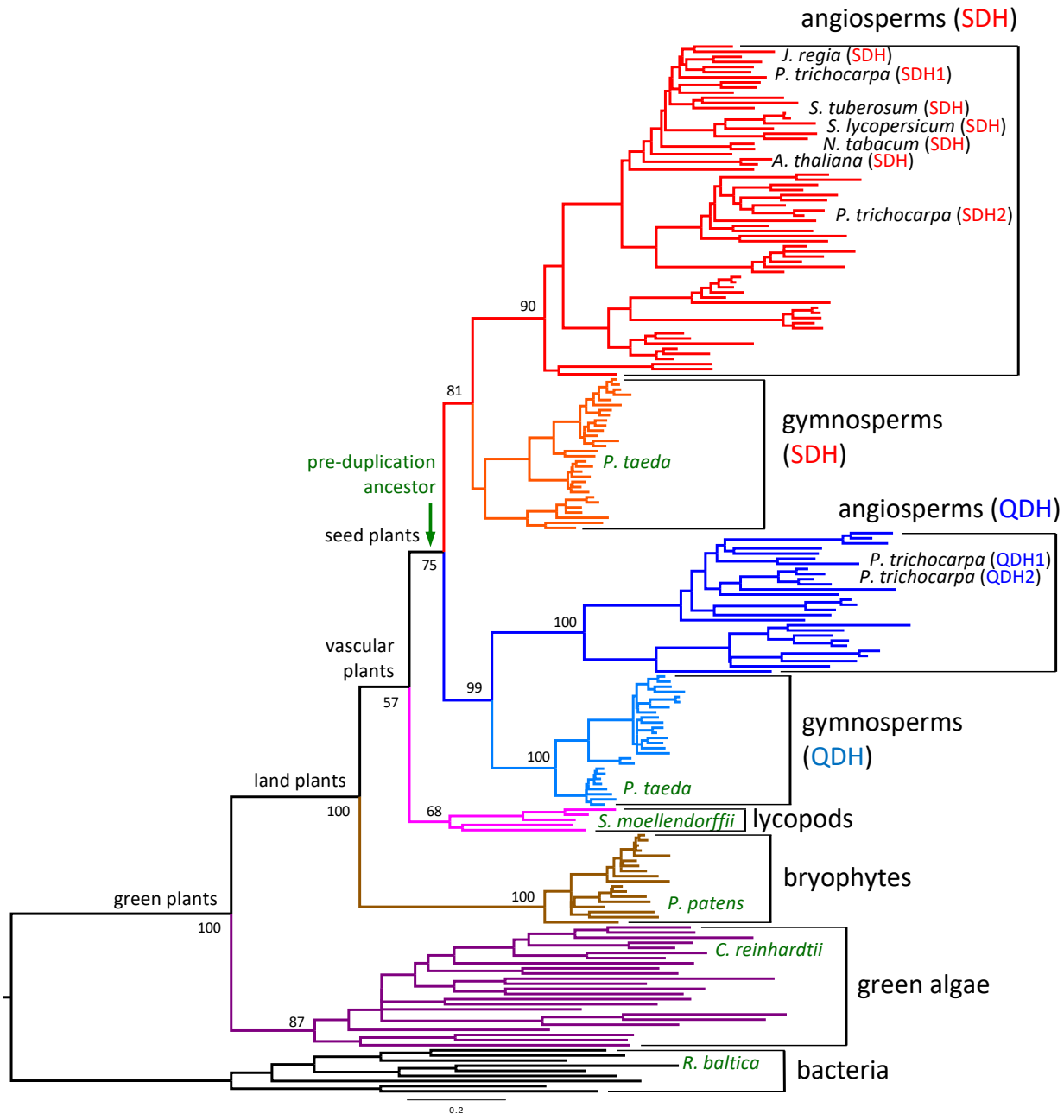

Figure 2: Phylogenetic reconstruction of the *SDH/QDH* gene family across land plants. A trimmed protein alignment was used for a maximum likelihood reconstruction. Bootstrap values (from 1,024 replicates) are given in percent for branches leading into the major clades only. Clades depicting taxonomic 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. Branch length are drawn to scale based on estimated amino acid substitutions per site as indicated.

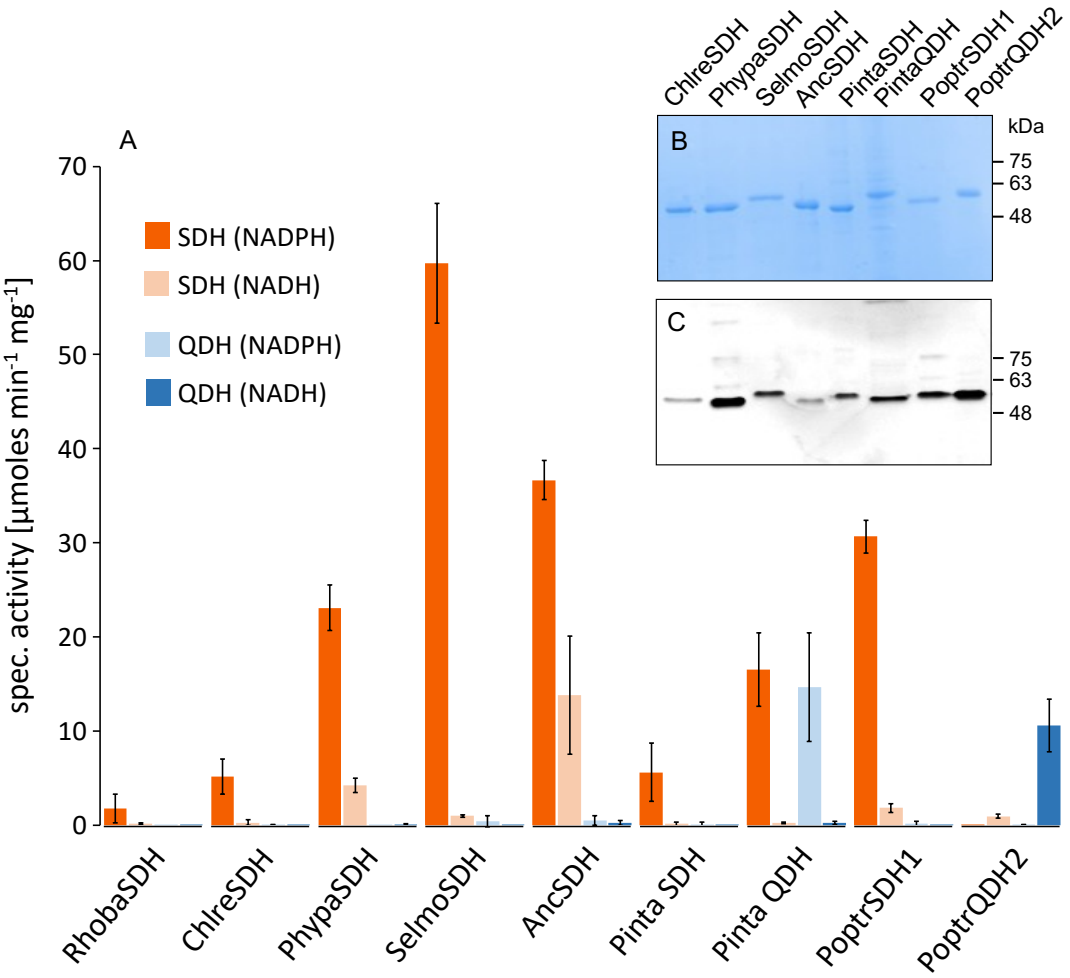

Figure 3 Enzyme activities with shikimate and quinate.

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). Enzymes 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 protein amount used. Shown is the mean of three replicates (independent protein purifications) and error bars denote standard deviation. Species abbreviations: *Pinus taeda* (Pinta), *Selaginella moellendorfii* (Selmo), *Physcomitrella patens* (Phypa), *Chlamydomonas reinhardtii* (Chlre). Anc122 represents the pre-duplication ancestor sequence reconstructed from the phylogeny. SDH1 and QDH2 from *Populus trichocarpa* (Poptr) (Guo *et al.*, 2014) were included as controls for comparison.

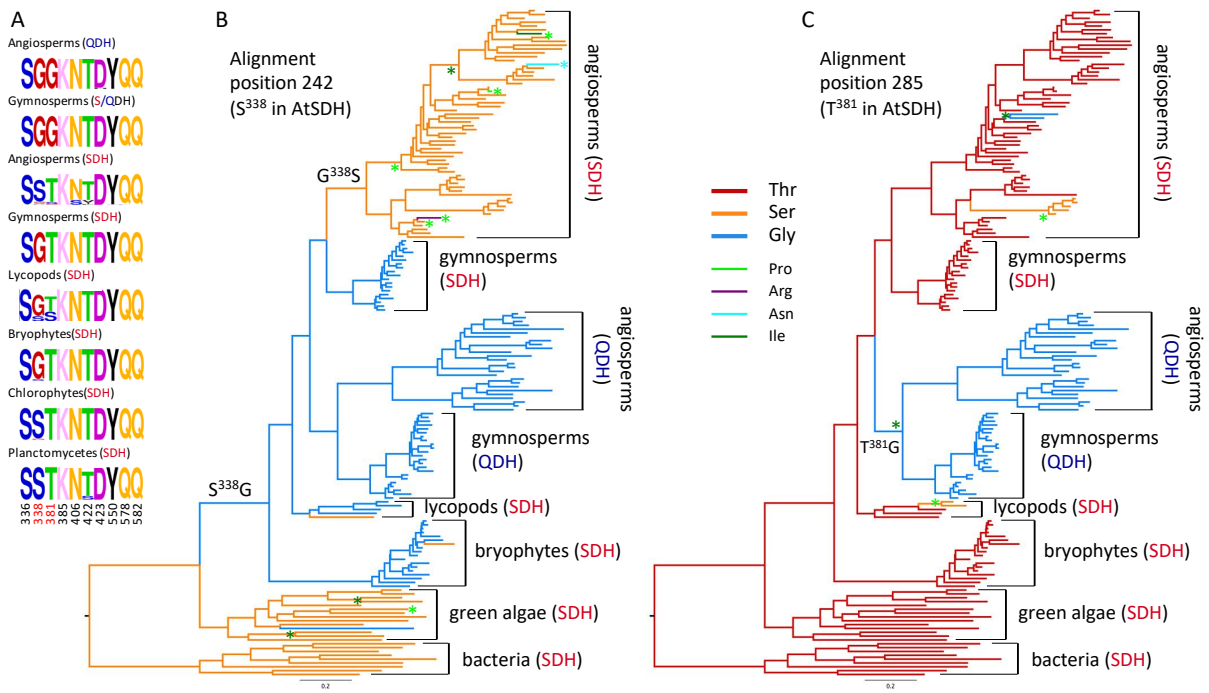

Figure 4: Selection test and amino acid substitution map.

A: Sequence logo representing active site residues extracted from an alignment of 174 protein sequences. The number below shows the respective amino acid position in the structurally characterized SDH from *Arabidopsis thaliana* (AtSDH). A maximum likelihood phylogeny based on the protein alignment was generated and the corresponding codon alignment was used to identify signatures of episodic positive selection using MEME. Only two active site residues, highlighted in red in A), showed signatures of selection. For these two sites, amino acid substitutions were mapped onto the phylogeny using SLAC and are shown for alignment codon position 242 (corresponding to S^{338} in AtSDH) (B) and position 285 (corresponding to T^{381} in AtSDH) (B). Nodes with an Empirical Bayes Factor for dN/dS larger than 10 and 100 are indicated by light green, and dark green asterisks, respectively. Amino acids are colour coded as indicated.

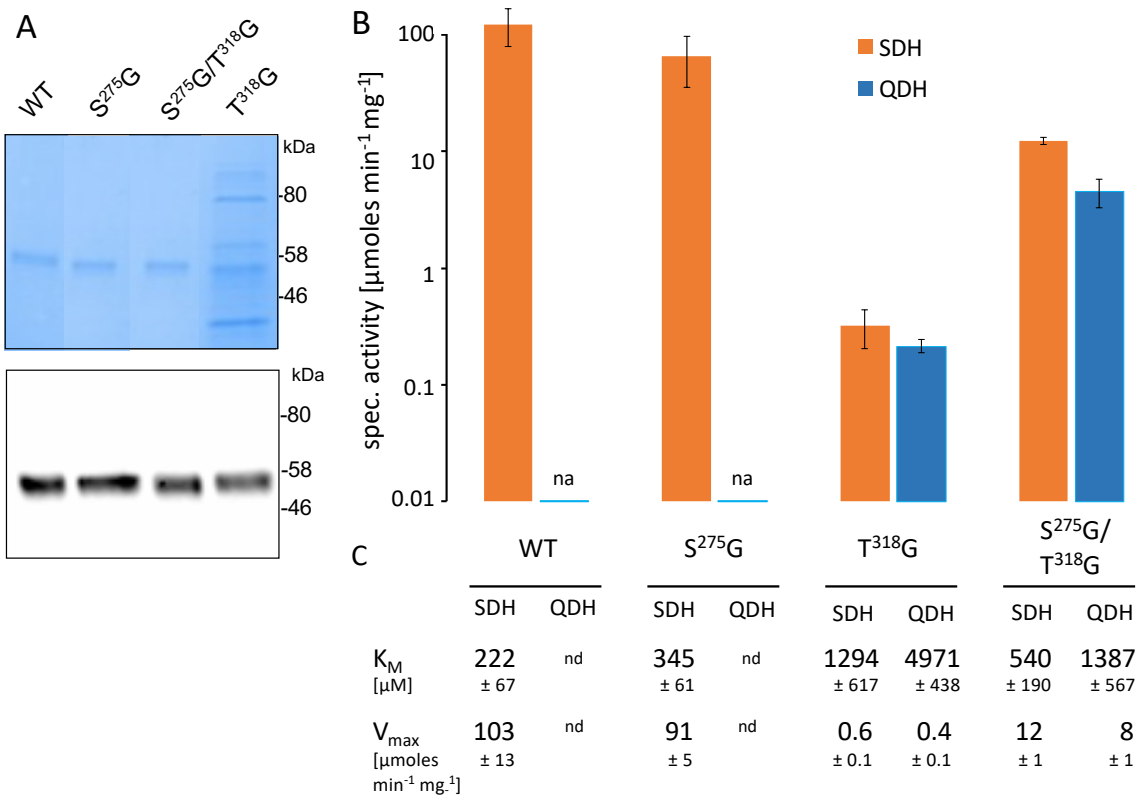

Figure 5 Activities of mutant PoptrSDH1 with shikimate and quinate.

Amino acid changes at position 275 (Ser to Gly) or 318 (Thr to Gly) or both where introduced through site-directed mutagenesis into the PoptrSDH1 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). Enzyme was incubated with shikimate or quinate in the presence of $NADP^+$ to determine SDH or QDH activity, respectively. Specific activity at 5 mM substrate concentration are shown as bar graphs (B). Kinetic constants were determined from three replicate purifications using at least nine substrate concentrations ranging from 0.05 mM to 5 mM (except for T318G, where substrate concentrations ranged from 0.6 mM to 5 mM) (C). Kinetic properties of mutant enzymes in comparison to wild type (WT) PoptrSDH1 enzyme; na: no activity detectable; nd: activity too low to be determined.

Table 1 Enzymatic properties based on Michaelis Menten kinetic analysis

Enzyme	Substrate	Co-factor	V_{max}* [µmole min^{-1} mg^{-1}]	K_M* [µM]	V_{max} / K_M
RhobaSDH	Shikimate	NADPH	5.1 +/- 0.3	101 +/- 24	51
ChlreSDH	Shikimate	NADPH	4.8 +/- 0.2	120 +/- 24	42
PhypaSDH	Shikimate	NADPH	51.8 +/- 2.9	239 +/- 47	217
SelmoSDH	Shikimate	NADPH	36.2 +/- 2.0	279 +/- 51	130
Anc122SDH	Shikimate	NADPH	90.5 +/- 4.2	243 +/- 40	372
PintaSDH	Shikimate	NADPH	32.4 +/- 2.6	218 +/- 63	149
PintaQDH	Shikimate	NADPH	12.8 +/- 1.0	820 +/- 162	16
PintaQDH	Quinate	NADPH	17.7 +/- 2.3	677 +/- 229	26

*based on three replicate experiments from independent protein purifications; for each replicate ten substrate concentrations ranging from 0.05 mM to 5 mM were used; kinetic constants were modelled using non-linear regression to the Michaelis-Menten equation; standard error of model provided