Structural and biochemical approaches uncover multiple evolutionary trajectories of plant quinate dehydrogenases

Artyom Gritsunov[1,†], James Peek[1,†], Julio Diaz Caballero[1], David Guttman[1,2] and Dinesh Christendat[1,2,*]

[1]*Department of Cell and Systems Biology, University of Toronto, 25 Willcocks Street, Toronto, ON M5S 3B2, Canada, and*
[2]*Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, ON M5S 3B2, Canada*

SUMMARY

Quinate is produced and used by many plants in the biosynthesis of chlorogenic acids (CGAs). Chlorogenic acids are astringent and serve to deter herbivory. They also function as antifungal agents and have potent antioxidant properties. Quinate is produced at a branch point of shikimate biosynthesis by the enzyme quinate dehydrogenase (QDH). However, little information exists on the identity and biochemical properties of plant QDHs. In this study, we utilized structural and bioinformatics approaches to establish a QDH-specific primary sequence motif. Using this motif, we identified QDHs from diverse plants and confirmed their activity by recombinant protein production and kinetic assays. Through a detailed phylogenetic analysis, we show that plant QDHs arose directly from bifunctional dehydroquinate dehydratase–shikimate dehydrogenases (DHQD-SDHs) through different convergent evolutionary events, illustrated by our findings that eudicot and conifer QDHs arose early in vascular plant evolution whereas Brassicaceae QDHs emerged later. This process of recurrent evolution of QDH is further demonstrated by the fact that this family of proteins independently evolved NAD^+ and $NADP^+$ specificity in eudicots. The acquisition of QDH activity by these proteins was accompanied by the inactivation or functional evolution of the DHQD domain, as verified by enzyme activity assays and as reflected in the loss of key DHQD active site residues. The implications of QDH activity and evolution are discussed in terms of plant growth and development.

Keywords: central metabolism, quinate, shikimate, quinate dehydrogenase, protein evolution, Brassicaceae, *Solanum lycopersicum*, NAD^+ and $NADP^+$ cofactors.

INTRODUCTION

Quinate is a precursor for the biosynthesis of chlorogenic acids (CGAs) in plants (Cle *et al.*, 2008; Moglia *et al.*, 2014). Chlorogenic acids accumulate in leaves and fruits where they act as feeding deterrents and antifungal agents. They also have potent antioxidant properties which have been shown to protect crops from intense UV radiation damage (Mondolot *et al.*, 2006; Cle *et al.*, 2008; Hicks *et al.*, 2012). Chlorogenic acids are derived through conjugation of quinate and hydroxycinnamoyl moieties (Lallemand *et al.*, 2012; Escamilla-Treviño *et al.*, 2014). While CGA biosynthesis has been well studied, little information exists on the metabolism of quinate in plants.

Shikimate is produced from dehydroquinate through two reactions by the enzymes dehydroquinate dehydratase (DHQD) and shikimate dehydrogenase (SDH) (Herrmann and Weaver, 1999; Mir *et al.*, 2015). In microbes these two activities are performed by separate monofunctional

enzymes. In contrast, plants use a bifunctional DHQD-SDH enzyme to catalyze both reactions (Chaudhuri and Coggins, 1985; Chaudhuri *et al.*, 1986; Singh and Christendat, 2006). Quinate is produced from dehydroquinate at a branch point of the shikimate pathway via the activity of quinate dehydrogenases (QDHs). Although QDH activity has been detected in some plants, characterization of QDHs in plants has been limited (Gamborg, 1966; Guo *et al.*, 2014).

The crystal structures for the plant bifunctional DHQD-SDH and monofunctional bacterial DHQD and SDH have been determined and their biochemical properties have been thoroughly investigated (Moore *et al.*, 1993; Gourley *et al.*, 1994; Michel *et al.*, 2003; Singh and Christendat, 2006). Previously, we have determined the crystal structure of the DHQD-SDH complex from *Arabidopsis thaliana* (Singh and Christendat, 2006; Singh *et al.*,

2008). While *A. thaliana* possesses only a single *DHQD-SDH* gene other plants have multiple gene copies encoding homologous DHQD-SDH enzymes (Graziana *et al.*, 1980; Ossipov *et al.*, 2000). However, the biochemical functions of these plant *DHQD-SDH* homologs have not been studied in detail nor has their distribution across diverse plant species been investigated. Recently, it has been shown that two of the *DHQD-SDH* gene duplicates found in *Populus trichocarpa* encode enzymes with QDH activity (Guo *et al.*, 2014).

Quinate dehydrogenases have been more thoroughly studied in microbes, where they participate in the catabolism of quinate as a carbon source (Hoppner *et al.*, 2013). Microbial QDHs share a common structural fold with SDHs and catalyze the oxidation of both quinate and shikimate, with preference for quinate (Peek *et al.*, 2011; Hoppner *et al.*, 2013). In contrast, SDHs show specificity for shikimate. Previous studies have used structural and mutational analyses to understand the basis of the broader substrate specificity of QDHs (Lindner *et al.*, 2005; Hoppner *et al.*, 2013). However, this approach has proved challenging, and so far attempts to understand the molecular basis for the different substrate preferences of the SDH and QDH classes have not been fruitful.

A lack of information on primary sequence motifs that confer QDH activity has affected the identification of plant *QDH* genes. To date, only two *QDH* genes have been identified. These were found in *P. trichocarpa* by use of a brute force approach in which the multiple DHQD-SDH homologs from this plant were recombinantly produced and characterized (Guo *et al.*, 2014). Given the importance of quinate esters in crop protection, the ability to identify plant QDHs involved in quinate biosynthesis could have important agricultural applications (Niggeweg *et al.*, 2004; Mondolot *et al.*, 2006; Cle *et al.*, 2008; Wojciechowska *et al.*, 2014).

In this study, we demonstrate through structural, biochemical and phylogenetic studies that plant SDHs have undergone neofunctionalization to produce QDHs. Using structure-guided mutagenesis we have constructed a plant QDH using the *A. thaliana* SDH domain as a structural scaffold. The results of this analysis allow us to establish a substrate-specifying amino acid motif that facilitates the identification of QDHs in other plants. Detailed bioinformatics analyses reveal the distribution of QDH enzymes in multiple plant species, with some enzymes having a common origin and others undergoing recent recurrent convergent evolution. We further discover that QDHs that arose earlier have undergone a second round of duplication and have diverged in cofactor specificity while retaining their substrate preferences. We discuss the neofunctionalization of plant QDHs in the context of plant growth and development.

RESULTS

Exploring the substrate promiscuity of *A. thaliana* SDH via targeted mutagenesis of active site residues

In this study we set out to investigate the evolution and distribution of plant QDHs. Prior to this work, amino acid sequence motifs to differentiate QDHs from the closely related SDHs had not been established. We therefore began by exploring the sequence requirements for quinate usage. We have previously determined the three-dimensional crystal structure of the *A. thaliana* DHQD-SDH protein (gene ID: AT3G06350) with shikimate bound in the SDH active site (PDB ID: 2GPT). By superimposing quinate in place of the bound molecule of shikimate we determined that there are only four amino acid residues likely to contribute to specificity for one substrate instead of the other, namely S336, S338, T381 and Y550, all of which would be in the direct vicinity of the quinate C1-hydroxyl. Amino acid S336 had previously been shown by mutational analysis to be critical for shikimate binding (Singh and Christendat, 2006). Similarly, Y550 had been shown to be essential for maintaining the substrate in a catalytically competent position in the active site (Singh and Christendat, 2006). In comparison, mutation of S338 or T381 had a smaller impact on the kinetic properties of the enzyme. We therefore hypothesized that these residues could more easily be substituted to permit quinate usage.

Using site-directed mutagenesis and kinetic analyses we attempted to construct a functional quinate dehydrogenase using *A. thaliana* DHQD-SDH as the structural scaffold. We investigated whether S338 or T381 could be replaced with residues having a smaller R-group to permit quinate to be used as a substrate. The wild-type *A. thaliana* SDH domain shows a strict requirement for shikimate as a substrate and has no measurable activity using even very high concentrations of quinate (Table 1). An S338G mutant had kinetic properties similar to the wild-type enzyme using shikimate as a substrate, and likewise could not use quinate. In contrast, a T381G mutant catalyzed the oxidation of quinate with a turnover rate of 8.8 sec^{-1} and a K_M of 3.330 mm. A S338G/T381G double mutant did not show improved enzymatic activity with quinate compared with the T381G mutant. T381A and T381S mutants also accepted quinate as a substrate but were much less efficient than the T381G variant. These findings suggest that the size of the amino acid side chain at position 381 is a key determinant of substrate specificity.

To further explore this hypothesis, the T381G mutant was co-crystallized with quinate. The structure of the mutant enzyme clearly shows that absence of the T381 side chain creates sufficient space in the active site to accommodate the quinate C1-hydroxyl (Figure 1a). The

remaining interactions between the enzyme and the substrate are the same as in the shikimate-liganded wild-type structure (Figure 1b,c). For example, some of these interactions include, S336, S338, Y550 which form hydrogen bond interactions with the carboxylic group of shikimate. Similarly, the K385 and D423 catalytic dyad which interacts with C4-OH and participates in proton transfer during the reduction/oxidation of $NADP^+$/NADPH is retained in the T381G mutant.

Table 1 Michaelis–Menten kinetic parameters of wild-type (WT) and mutant *Arabidopsis thaliana* shikimate dehydrogenase

Enzyme	Substrate	k_{cat} (sec^{-1})	K_M (µM)
WT	Shikimate	516 ± 32	604 ± 93
WT	Quinate	No activity	No activity
S338G	Shikimate	427 ± 60	548 ± 19
S338G	Quinate	No activity	No activity
T381G	Shikimate	24.0 ± 1.5	1613 ± 226
T381G	Quinate	8.8 ± 0.7	3330 ± 491
T381S	Shikimate	172 ± 22	1539 ± 440
T381S	Quinate	0.062 ± 0.005	4485 ± 687
T381A	Shikimate	11.8 ± 1.2	2512 ± 501
T381A	Quinate	0.113 ± 0.007	5135 ± 551
S338G/T381G	Shikimate	22.0 ± 0.9	882 ± 100
S338G/T381G	Quinate	6.7 ± 0.5	4075 ± 529

Sequence analysis of putative QDHs to establish a sequence motif for their functional annotation

Quinate dehydrogenase activity has been reported in extracts from a small number of plants (Gamborg, 1966; Graziana *et al.*, 1980; Kang and Scheibe, 1993; Ossipov *et al.*, 2000; Guo *et al.*, 2014). Little is known about the identity of these enzymes or the distribution of QDHs in other plants. To identify potential QDHs from more diverse plant species we retrieved the amino acid sequences of homologs of *A. thaliana* DHQD-SDH by sequence similarity searches and analyzed variations in their SDH active site residues. This analysis revealed that the key catalytic dyad, K385-D423, is conserved in all sequences identified (Figure 1d) (Singh and Christendat, 2006). Similarly, residues that stabilize the carboxyl group of the substrate, especially S336 and Y550, show nearly complete conservation (Figure 1d). In contrast, we observed distinct variations at the position corresponding to T381 from the *A. thaliana* SDH domain (Figure 1d). It is important to note that every plant species analyzed contained at least one DHQD-SDH homolog with a threonine at this position, implying that they all contain a SDH that is likely to participate in shikimate biosynthesis. In a small number of mosses and monocots we observed a serine at position 381 (Figure 2). Based on our mutagenesis studies, the conservative substitution of a serine for a threonine is still likely to favor SDH activity
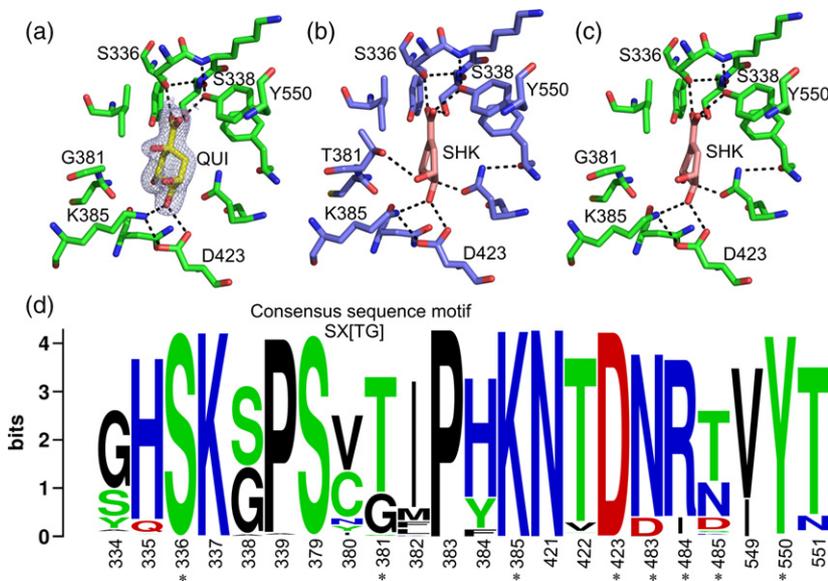

Figure 1. Structural and sequence analyses of *Arabidopsis thaliana* shikimate dehydrogenases (SDHs) for the determination of residues confirming substrate promiscuity.
(a–c) Active site representation of wild-type and T381G variant. (a) The T381G variant with quinate (PDB ID: 6BMQ). The electron density of quinate is adjusted to 1σ. (b) Wild-type SDH with shikimate (PDB ID: 2GPT). (c) T381G SDH variant in complex with shikimate (PDB ID: 6BMB).
(d) Sequence logo representation of selected SDH/quinate dehydrogenase (QDH) active site amino acids from a variety of plant species. Residues were numbered with respect to the *A. thaliana* SDH sequence (gene ID: AT3G06350). Significant residues are labeled by an asterisk. The key catalytic dyad K385 and D423 is shown. Residues that stabilize the carboxyl group of shikimate and quinate, Y550 and S336, are also represented. T or G found at position 381 are represented by the SX[TG] motif. DID or NRT motifs specifying NAD^+ or $NADP^+$ specificity are shown at the position 483–485. This figure was prepared with the WebLogo server (http://weblogo.berkeley.edu). [Colour figure can be viewed at wileyonlinelibrary.com].

over QDH activity. However, in most dicots, a second DHQD-SDH homolog was identified which contains a glycine at position 381. We predicted that these enzymes should possess QDH activity. In the *A. thaliana* SDH primary sequence, T381 is preceded by a variable residue and then by S379, which is strictly conserved in all sequences analyzed. We therefore propose that the consensus sequence motif, SX[TG] (Figure 1d), could be used to predict substrate specificity among plant SDH paralogs. In sequences with the SXG motif, we also observed variations in key cofactor binding residues. In particular, an asparagine–arginine–threonine (NRT) sequence motif (*A. thaliana* SDH residues 483–485) is typically associated with $NADP^+$ usage, while an aspartate–isoleucine–aspartate (DID) sequence motif is indicative of a preference for NAD^+ (Singh *et al.*, 2008; Peek and Christendat, 2015). Some of these putative $NADP^+$ QDHs also contain a predicted chloroplast target sequence.

Functional validation of substrate specificity predictions

To investigate the correlation between the SX[TG] motif and substrate specificity, we recombinantly expressed predicted QDHs and assayed the purified enzymes for activity with quinate and shikimate. The enzymes examined were derived from *Nicotiana tabacum* (SXG, DID motifs), *Solanum lycopersicum* (SXG, DID motifs), *Brassica rapa* (SXG, NRT motifs) and *Brassica napus* (SXG, NRT motifs). All the enzymes tested use quinate as a substrate, consistent with our hypothesis that the SXG motif is associated with quinate metabolism (Table 2, Figure S2 in the online Supporting Information). While the enzymatic efficiencies of the purified proteins were lower than that of the *A. thaliana* SDH domain, unlike the *A. thaliana* protein they exhibited broader substrate specificity and were capable of using both quinate and shikimate as substrates. This finding is in line with the results of our mutagenesis studies, which demonstrated that the introduction of a glycine at position 381 of the *A. thaliana* SDH domain confers dual substrate specificity. The *N. tabacum* and *S. lycopersicum* enzymes showed a moderate preference for quinate over shikimate, while the opposite trend was found for the enzymes from *B. rapa* and *B. napus*. Only low-level activity was detected for the *B. rapa* enzyme, probably as a result of the inherent instability of this particular protein when expressed recombinantly. As predicted by our analysis of cofactor-specifying motifs, the enzymes with the DID motif from *N. tabacum* and *S. lycopersicum* utilized NAD^+ while the enzymes with the NRT motif from *B. rapa* and *B. napus* used $NADP^+$.

Phylogenetic reconstruction of plant DHQD-SDH/QDHs to understand their duplication and diversification

The identified sequences were used to build a maximum likelihood phylogenetic tree to generate a global picture of the neofunctionalization of DHQD-SDH homologs in the plant kingdom (Figure 2). At the base of the tree, we find $NADP^+$-specific SDH sequences obtained from bryophyte (moss) genomes containing primarily SXT motifs with a few SXS motif sequences. Next, we observe a major branch (Branch A) that groups SXT SDH sequences from monocots, eudicots and conifers, and a few monocot and conifer sequences having the SXS motif (*Oryza sativa*, *Zea mays* and *Brachypodium distachyon*). In this branch, we also identify SXG (QDH $NADP^+$) sequences from eudicot Brassicaceae (*B. rapa*, *B. napus*, *Capsella grandiflora*, *Capsella rubella*), Rosaceae (*Fragraria vesca*, *Prunus persica*, *Malus domestica*, *Physocarpus opulifolius*) families and a small number of monocots (*Dioscorea villosa* and *Helonius bullata*). The rest of the eudicot SXG motif QDH sequences can be found embedded in the other major branch of the tree (Branch B). The eudicot QDH sequences have diverged into two distinct groups, with one having $NADP^+$ and the other having NAD^+ cofactor-specific motifs. At the base of Branch B, conifer QDH SXG motif sequences are found embedded in a monophyletic group. The conifer QDH in this branch is predicted, based on the presence of the NRT motif, to have $NADP^+$ cofactor specificity. In summary, with few exceptions, we observed that in early plant evolution QDH sequences emerged from SDH sequences and later on they evolved recurrently through independent gene duplication events in some eudicots.

Sequence analysis of DHQD domains

While the bifunctional *A. thaliana* DHQD-SDH catalyzes two sequential reactions in shikimate biosynthesis, the role of the DHQD domain in quinate metabolism is uncertain. We investigated the conservation of function of the DHQD domain of the newly identified DHQD-QDH proteins. Using structural analysis of the *A. thaliana* DHQD domain, we identified a number of conserved and functionally important DHQD active site residues. These include H214 and K241, which function as catalytic residues, and R279, an important substrate-binding residue (Singh and Christendat, 2006). Two of these three residues show significant variability in the DHQD domains of proteins with a predicted QDH domain (Figure 3). R279 shows the highest level of sequence variation with more than 75% of all predicted eudicot QDHs having a glutamine at this position (Table S2). H214 is also frequently substituted with another residue; in 50% of the eudicot QDH sequences a tyrosine is present at this position. When we analyzed the small number of QDH enzymes that are found within the clades that also contain a SDH homolog from the same plant, we found subtle or no variation in the DHQD active site residues described above. For example, the three residues are still retained in *M. domestica* and to a lesser extent in *B. rapa* and *B. napus* (a histidine replaces R279). Based on the conservation of key DHQD active site residues and the
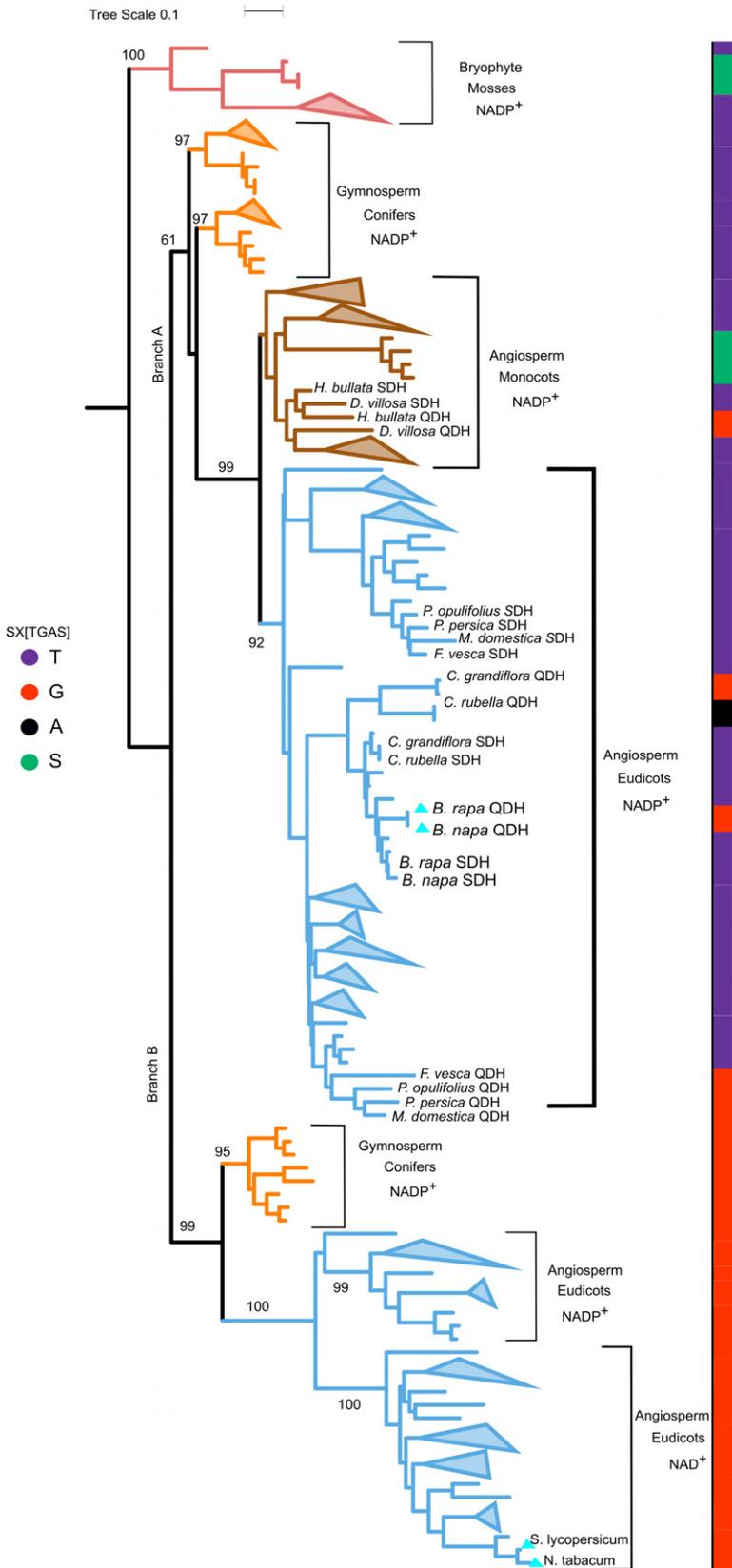
Figure 2. Phylogenetic tree representation of plant bifunctional dehydroquinate dehydratase-shikimate dehydrogenase (DHQD-SDH) and quinate dehydrogenase (QDH) protein sequences to illustrate their evolution and diversification.

SX[TGAS] sequence motifs are color coded. Bootstrap scores are labeled on the tree. The tree was rooted around an ancient bryophyte group represented by mosses. Several monocot and eudicot QDHs form monophyletic groups with their respective SDH paralogs, whereas the majority of QDHs from angiosperms form monophyletic NAD+ and NADP+ branches. Asterisks signify QDHs of undetermined cofactor preference. Cyan triangles specify sequences that were enzymatically evaluated in this study. [Colour figure can be viewed at wileyonlinelibrary.com].

phylogenetic proximity of these enzymes to DHQD-SDH paralogs from the same plants, we conclude that these enzymes emerged more recently in evolutionary history and are undergoing convergent evolution with respect to the DHQD-QDHs from Branch B.

Assessment of DHQD domain activity

To determine the functional impacts of the DHQD active site variations, we conducted activity assays using the recombinant proteins from *S. lycopersicum*, *N. tabacum* and *B. rapa*. Neither the *S. lycopersicum* nor the *N. tabacum* DHQD domains, which contain variations at positions H214 and R279, showed measurable DHQD activity, as determined by a coupled enzyme assay (Figure 3b, c). In comparison, the *B. rapa* protein had a low level of DHQD activity (Figure 3c). This finding is in agreement

with the retention of residues corresponding to the key catalytic groups (H214 and 241) in the *B. rapa* enzyme.

Analysis of *QDH* gene expression

Gene expression analysis was undertaken to better understand the physiological relevance of the newly identified *QDHs* in plants. Using publicly available expression data for *S. lycopersicum*, we observed that the NADP$^+$-specific enzyme shows higher expression levels in green tissues, including leaves, green stems and flowers and during the early stages of fruit development (Figure 4a). The gene for the NAD$^+$-dependent enzyme showed higher expression in the roots in the later stages of *S. lycopersicum* fruit development and especially during fruit ripening (Figure 4b). These unique expression profiles imply that *QDH* genes for NAD$^+$- and NADP$^+$-specific enzymes are likely to have

Table 2 Michaelis–Menten kinetic parameters of quinate dehydrogenases from *Solanum lycopersicum*, *Nicotiana tabacum*, *Brassica napus* and *Brassica rapa* plants

| Species	Substrate or cofactor		K_M (μM)	k_{cat} (sec^{-1})	k_{cat}/K_M [(sec^{-1} μM^{-1}) $\times$ 10^3]
	Fixed	Variable			
N. tabacum	NAD$^+$	Quinate	565 $\pm$ 23	20.8 $\pm$ 0.3	36.8
N. tabacum	NAD$^+$	Shikimate	314 $\pm$ 40	3.8 $\pm$ 0.1	12
N. tabacum	Quinate	NAD$^+$	107 $\pm$ 8	16.2 $\pm$ 0.3	151
S. lycopersicum	NAD$^+$	Quinate	635 $\pm$ 165	10.5 $\pm$ 0.2	16.5
S. lycopersicum	NAD$^+$	Shikimate	519 $\pm$ 138	0.74 $\pm$ 0.06	0.9
S. lycopersicum	Quinate	NAD$^+$	185 $\pm$ 18	9.4 $\pm$ 0.3	23.7
B. rapa	NADP$^+$	Quinate	219 $\pm$ 28	1.8 $\times$ 10^{-3} $\pm$ 6 $\times$ 10^{-5}	0.080
B. rapa	NADP$^+$	Shikimate	487 $\pm$ 37	0.335 $\pm$ 0.007	0.686
B. rapa	Shikimate	NADP$^+$	129 $\pm$ 5	0.493 $\pm$ 0.005	3.8
B. napus	NADP$^+$	Quinate	438 $\pm$ 43	7.0 $\pm$ 0.23	16
B. napus	NADP$^+$	Shikimate	271 $\pm$ 17	8.6 $\pm$ 0.16	31
B. napus	Quinate	NADP$^+$	189 $\pm$ 23	8.5 $\pm$ 0.3	45
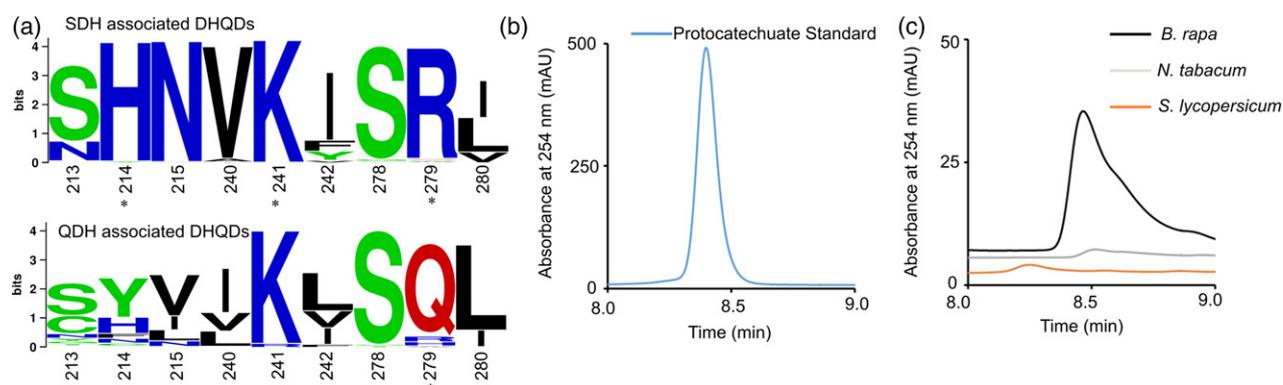
Figure 3. Bioinformatics and experimental evaluation of dehydroquinate dehydratase (DHQD) domain evolution.
(a) Sequence logo representation of select amino acid from the DHQD domain of plant bifunctional DHQD-shikimate dehydrogenase (SDH)/quinate dehydrogenase (QDH). Catalytic residues H214, K241 and substrate binding residue R279, labeled with an asterisk, were analyzed for the variability in DHQD domains of QDH- and SDH-associated proteins. Residues are numbered with respect to the *Arabidopsis thaliana* SDH sequence (gene ID: AT3G06350). This figure was prepared with the WebLogo server (http://weblogo.berkeley.edu).
(b, c) Qualitative HPLC analysis of DHQD domain activity from several plant species. The identity of the reaction end-product, protocatechuate, was further confirmed by LC-MS/MS analysis. [Colour figure can be viewed at wileyonlinelibrary.com].

distinct physiological roles (Winter *et al.*, 2007; Tomato Genome Consortium, 2012).

DISCUSSION

Detailed structural, biochemical and bioinformatics analyses were conducted to understand the distribution and biological role of QDH enzymes in plants. These analyses reveal that plant *SDH* gene duplicates retained most of their active site residues with a single residue substitution dictating substrate specificity (Figure 1d). These findings support the notion that plant SDHs have neofunctionalized to produce QDHs (Figure 2). In support of this hypothesis, structural analysis of *A. thaliana* SDH and of a T381G variant in complex with shikimate and quinate revealed that the glycine substitution is sufficient to produce an active site that permits quinate binding without drastically altering shikimate binding (Figure 1, Table 1). A similar substitution in *P. trichocarpa* SDH allowed quinate binding (Carrington *et al.*, 2018). Furthermore, a glycine is also present in *P. trichocarpa* QDH at the corresponding position of T381 (Guo *et al.*, 2014). These observations are consistent with our hypothesis that T381 confers selectivity to shikimate. Subsequently, we established a sequence motif, SXG, which we used to identify plant QDHs from diverse species (Figure 1d). In contrast to plant SDHs, which use NADP$^+$ as a cofactor, there are both NADP$^+$- and NAD$^+$-specific QDHs (Figures 1d and 3, Table 2). Interestingly, we observed both NADP$^+$- and NAD$^+$-specific QDHs in most eudicot plants investigated.

Our phylogenetic analysis has revealed the distribution of SDH and QDH enzymes in the plant kingdom and provided insight into the evolution of quinate metabolism in plants. Based on the distribution of the QDH enzymes through the phylogeny, we expect that the SDH gene duplicates underwent diversification for NADP$^+$- and NAD$^+$-specific utilization in addition to their evolution for QDH activity. The SDH enzymes, identified by their SXT motif, are found in all land plants and some contain a predicted chloroplast target sequence, which is consistent with the hypothesis that their shikimate pathway is of microbial origin derived through endosymbiosis with green algae (Peek *et al.*, 2011). In contrast, plant QDH sequences are restricted mainly to eudicots and conifers with a small number being identified in monocots. A detailed investigation into the evolution and distribution of QDH with emphasis on non-angiosperms is described elsewhere (Carrington *et al.*, 2018). The limited representation of QDHs in monocots is consistent with previous speculations that certain monocots do not have downstream quinate-processing enzymes such as hydroxycinnamoyl CoA quinate transferases (HQT) (Escamilla-Treviño *et al.*, 2014).

Plant QDHs probably evolved after the emergence of vascular plants because QDH sequences are not present in mosses. Although QDHs are also found in bacteria and fungi, these enzymes appear to have evolved independently of the plant enzymes via distinct gene duplication events. A small number of NADP$^+$-specific QDH sequences obtained from angiosperms appear to have distinct evolutionary histories (Figure 2). Moreover, these enzymes are in a monophyletic clade along with the SDH paralogs from their respective plants. This observation suggests that they underwent recent divergent evolution with respect to their SDH paralogs and convergent evolution with respect to Branch B, the major branch of QDHs. Biochemical analysis of representative QDHs from Brassicaceae supports this hypothesis; although these enzymes can catalyze the oxidation of quinate they still exhibit a preference for shikimate as a substrate, conversely NAD$^+$-utilizing QDHs are more active with quinate than with shikimate (Table 2, Figure S2). Further evidence for this hypothesis can be seen in their DHQD domain which still displays remnant DHQD activity. Although we found that the *B. rapa* protein
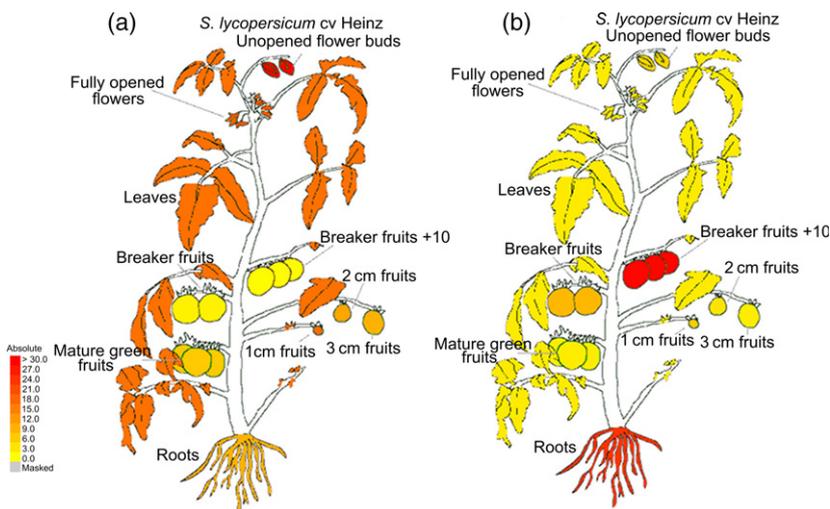

Figure 4. Tissue-specific gene expression profiles of *Solanum lycopersicum* quinate dehydrogenases (QDHs).
(a) *In silico* expression profiling of NADP$^+$ cofactor-specific *QDH* in whole *S. lycopersicum* (Solyc06g084460) and for (b) NAD$^+$ cofactor-specific *QDH* (Solyc10g038080). All expression values are expressed in units of reads per kilobase of transcript per million mapped reads (RPKM) with a 30-unit threshold. Colors on diagrams correlate with the scale bar. [Colour figure can be viewed at wileyonlinelibrary.com].

retained a low level of DHQD activity, this activity appears to have been selected against in other eudicot DHQD domains that arose earlier in their evolution (Figure 3a–c). The divergence of the function of DHQD domains associated with QDHs is likely to be biologically significant, as the presence of both an active DHQD domain and a QDH domain would result in competition for the substrate, dehydroquinate (Figure 5b).

Additional evidence to support our findings that QDHs are undergoing recurrent evolution is demonstrated by the fact that this family of proteins evolved NAD^+ and $NADP^+$ specificity independently in eudicots. While plant SDHs utilize $NADP^+$ as a cofactor, some of the QDHs we identified use $NADP^+$ while others use NAD^+. Further investigation of these QDHs revealed that some of the $NADP^+$ cofactor-utilizing proteins also contain a chloroplast-targeting peptide which is absent from the NAD^+-specific enzymes (Bischoff *et al.*, 2001; Ding *et al.*, 2007). Based on our expectation that $NADP^+$-utilizing enzymes are involved in anabolic pathways and NAD^+ ones in catabolic pathways, we proposed that the chloroplast-localized $NADP^+$-specific QDHs are involved in the generation of quinate from shikimate pathway intermediates for CGA production (Figure 5a). This is consistent with the fact that the plant shikimate pathway is localized in the chloroplast and is responsible for diverting about 20% of fixed carbon for the production

of secondary metabolites and aromatic amino acids (Herrmann and Weaver, 1999; Bischoff *et al.*, 2001; Maeda and Dudareva, 2012). The NAD^+-dependent QDHs, on the other hand, are most probably involved in diverting quinate away from CGA biosynthesis and in utilizing the quinate that is released from CGA during its breakdown (Figure 5a) (Ding *et al.*, 2007). Consistent with this model, silencing of a caffeoyl shikimate esterase in poplar results in buildup of caffeoyl/feruloyl quinate (Saleme *et al.*, 2017). The identification of CGA-specific esterase is the focus of future studies. Normally, the breakdown of CGA is expected to occur in the later stages of plant and fruit development when the demand for CGA is much lower (Mondolot *et al.*, 2006). This is supported by our gene expression analysis of *S. lycopersicum* QDHs (Figure 4). The $NADP^+$-specific *DHQD-QDH* from this plant, which contains a predicted chloroplast-targeting peptide, is highly expressed in green tissues and in the early stages of flower and fruit development (Figure 4a). In contrast, the NAD^+-specific *DHQD-QDH* is highly expressed in mature *S. lycopersicum* fruits and in the roots (Figure 4b). The NAD^+-specific protein does not have an identifiable chloroplast target peptide, which suggests that it may function in the cytosol (Ding *et al.*, 2007). Perfectly in line with these findings, the quinate levels in kiwi and citrus fruits are high during the early stages of fruit development and decrease markedly as the
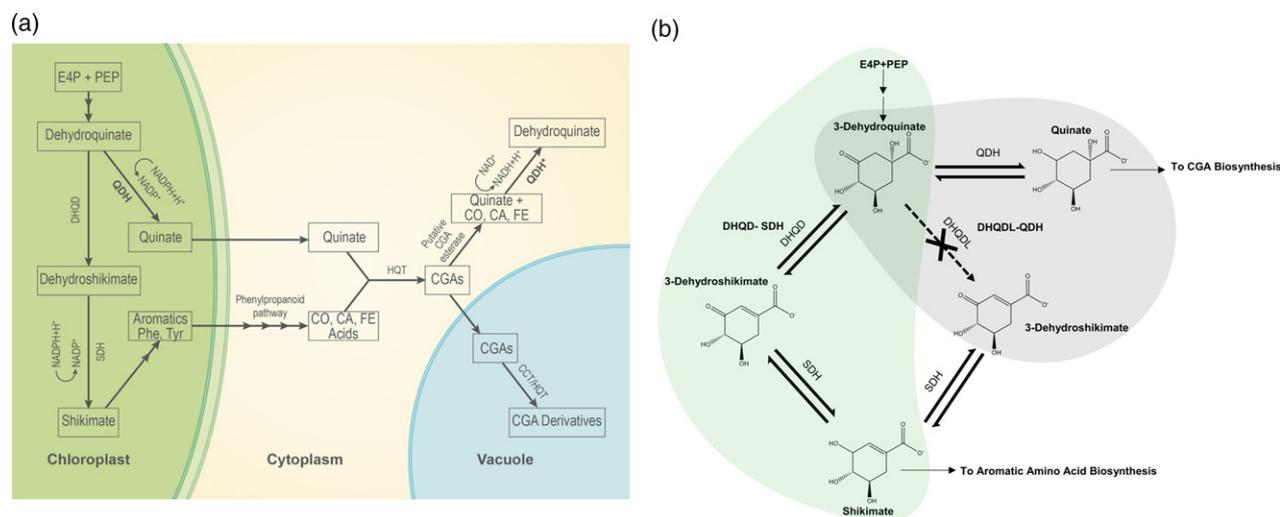

Figure 5. Models for the roles of dehydroquinate dehydratase (DHQD)-shikimate dehydrogenase (SDH)/quinate dehydrogenase (QDH) in quinate and shikimate metabolism in plants.

The shikimate pathway starts by the condensation of erythrose-4-phosphate (E4P) with phosphoenol pyruvate (PEP) which leads to the production of aromatic amino acids and central metabolites.

(a) In chloroplasts, 3-dehydroquinate is diverted away from the shikimate pathway via NADPH-dependent QDH reduction. In cytoplasm, quinate and *p*-coumaric (CO), caffeic (CA) and ferulic (FE) acids are incorporated into chlorogenic acid (CGA) biosynthesis by the action of hydroxycinnamoyl CoA quinate transferases (HQT). In vacuoles, additional hydroxycinnamoyl moieties can be crosslinked onto CGA to make di/trihydroxycinnamoyl-quinate esters by the action of chlorogenate transferases (CCT). CGAs serve as antioxidants and anti-feeding/-fungal agents. Upon breakdown of CGA, the released quinate is converted back into 3-dehydroquinate via NAD^+-dependent QDH* catalyzed oxidation with NAD^+ cofactor.

(b) Two consecutive reactions of shikimate pathway dehydroquinate dehydratase (DHQD) and shikimate dehydrogenase activities (SDH) are localized on the single polypeptide (DHQD-SDH). The duplication and neofunctionalization of bifunctional DHQD-SDHs resulted in the evolution of QDHs with an inactivated DHQD domain. The diversification/loss of function of the DHQD domain to DHQDL prevents competition for the metabolic intermediate 3-dehydroquinate. [Colour figure can be viewed at wileyonlinelibrary.com].

fruit approaches maturation (Albertini *et al.*, 2006; Marsh *et al.*, 2009).

The importance of the shikimate pathway in plants has been thoroughly studied (Davis, 1951; Weinstein *et al.*, 1962; Ding *et al.*, 2007). Similarly, the roles of CGA as an antioxidant and antifeeding and antifungal agent for crop protection are well established (Niggeweg *et al.*, 2004; Cle *et al.*, 2008; Sato *et al.*, 2011; Wojciechowska *et al.*, 2014). We therefore integrate our findings on the functional diversity of plants SDHs and QDHs to develop a model describing the interaction between shikimate and quinate metabolism in plants. The shikimate pathway enzymes are chloroplast localized and are involved in the production of dehydroquinate, a precursor for quinate biosynthesis. We envision that the $NADP^+$-dependent QDHs produce quinate in the chloroplast, which is then transported to the cytoplasm where it condenses with hydroxycinnamic acids to form CGA. It is established that the HQT involved in CGA biosynthesis is localized in the cytoplasm and vacuoles (Moglia *et al.*, 2014). We expect this process to be more actively occurring during the early stages of plant growth and development, because an actively growing plant requires a constant supply of CGA as an antioxidant and protection from UV light, herbivory and microbial pathogens (Mondolot *et al.*, 2006; Wojciechowska *et al.*, 2014). As the plant matures there is less demand for CGA, and as such quinate can be diverted for synthesis of other compounds (Ye *et al.*, 2015). We expect the NAD^+-specific QDH to divert the available quinate in the cytoplasm to dehydroquinate, which can then be redirected towards biosynthesis of aromatic compounds. While this model is based on our molecular characterization of the $NADP^+$- and NAD^+-dependent QDHs and their distinct gene expression profiles, detailed metabolomics and biological analyses will be required in the future for its validation.

EXPERIMENTAL PROCEDURES

Site-directed mutagenesis of *A. thaliana* SDH

A previously generated *A. thaliana* DHQD-SDH construct containing a C-terminal $6\times$ His/TEV cleavage site was used as a template for generating site-directed mutants (Singh and Christendat, 2006). Mutations were introduced into the wild-type *A. thaliana* SDH domain using the QuikChange kit (Agilent Genomics, https://www.agilent.com/). All mutations were verified by DNA sequencing. Mutant proteins were recombinantly expressed in *Escherichia coli* and purified according to established procedures (Christendat *et al.*, 2000; Singh and Christendat, 2006).

Crystallization of *A. thaliana* SDH mutant

Crystals of the *A. thaliana* DHQD-SDH mutant, T381G, were generated using the hanging drop vapor diffusion method in a crystallization condition consisting of 100 mM sodium citrate, pH 4.8, and 2 M ammonium sulfate. Each drop contained the protein (15 mg ml^{-1}) and crystallization solution in a 1:1 ratio (Benvenuti and Mangani, 2007). In addition, each drop contained either 20 mM quinate or 2 mM shikimate. Crystals were flash frozen in a liquid nitrogen stream using 15% glycerol as the cryoprotectant.

Diffraction data were collected using a HiFlux Homelab system (Rigaku, https://www.rigaku.com/) and processed using HKL2000 (Otwinowski and Minor, 1997). The structure of the T381G mutant in complex with quinate or shikimate was determined by molecular replacement using the wild-type *A. thaliana* DHQD-SDH–shikimate complex (PDB ID: 2GPT) as a search model. Molecular replacement and automated structure building were performed using the Phenix suite (Adams *et al.*, 2010). Additional structure building, including the placement of quinate, was performed with Coot (Emsley and Cowtan, 2004). All refinements were performed using Phenix Refine. Data collection and refinement statistics are listed in Table S1.

Phylogenetic analysis of plant DHQD-SDH/QDH peptides

Sequences for plant bifunctional DHQD-SDH/QDH proteins were identified in OneKP, Kiwifruit Genome Database, Phytozome and NCBI using the BLAST search engines (Altschul *et al.*, 1990; Cheng *et al.*, 2011; Goodstein *et al.*, 2012; Matasci *et al.*, 2014; O'Leary *et al.*, 2016). Retrieved sequences were aligned with the *A. thaliana* SDH protein (sequence ID: NP_187286.1). Sequences were aligned with MEGA7 software suite using the ClustalW method with default parameters (Appendix S1) (Kumar *et al.*, 2016). Gap opening penalties were set at 10 and gap extension penalties were set at 0.1 for pairwise alignments and 0.2 for multiple sequence alignments. A Gonnet protein weight matrix was used for the analysis of peptide divergence (Kumar *et al.*, 2016). This alignment of SDH/QDH protein sequences was then used to build a maximum-likelihood tree with 500 bootstrap iterations, a gamma distributed value of 5 and a Jones–Taylor–Thornton model. ITOL software was used for annotating the tree (Letunic and Bork, 2016).

The alignment of bifunctional DHQD-SDH/QDH peptide sequences described above was used for the analysis of active site variations in the DHQD domain. Amino acid residues corresponding to *A. thaliana* K241, H214 and R279 were examined for variations within the aligned DHQD-SDH/QDH sequences.

RNA isolation, cDNA generation, cloning

Brassica rapa, *B. napus*, *N. tabacum* leaves and *S. lycopersicum* ripening fruits were frozen in liquid nitrogen and ground to fine powder. RNA was extracted with Tri Reagent® (Sigma-Aldrich, http://www.sigmaaldrich.com/) according to the manufacturer's instructions. RNA pellets were suspended in RNase-free water. Complementary DNA was synthesized with an iScript™ cDNA synthesis kit (Bio-Rad, http://www.bio-rad.com/). The resulting cDNA was used as a template to amplify DHQD-QDH genes. The amplified genes were cloned into pET28a and pET28mod protein expression vectors and sequenced (Christendat *et al.*, 2000; Singh and Christendat, 2006).

Protein expression/purification

Recombinant protein expression and purification were carried out using an established protocol (Christendat *et al.*, 2000; Singh and Christendat, 2006; Peek *et al.*, 2017). Bacterial protein expression plasmids containing the genes for *S. lycopersicum* QDH, *N. tabacum* QDH, *B. rapa* QDH/SDH and *B. napus* QDH/SDH were transformed into *E. coli* BL21 CodonPlus (Agilent Technologies, https://www.agilent.com/). After lysis, cellular debris was pelleted by centrifugation and proteins were purified using nickel nitrilotriacetic acid affinity chromatography (NiNTA). The $6\times$ His tag on the *S. lycopersicum* QDH, *B. napus* and *B. rapa* QDH/SDH

constructs was removed by incubating with Tobacco Etch Virus (TEV) nuclear-inclusion-a endopeptidase. The N-terminal 6× His tag on the *N. tabacum* QDH was kept intact after purification of the protein. Purified proteins were dialyzed overnight at 4°C before being concentrated, flash frozen and stored at −80°C.

Kinetic analysis of the QDH/SDH enzyme activities

In order to test the ability of putative QDH/SDH enzymes to catalyze quinate oxidation, activities of purified enzymes were assayed spectrophotometrically by monitoring the reduction of NAD$^+$ or NADP$^+$ (ε = 6220 M^{-1} cm^{-1}) at 340 nm (Singh and Christendat, 2006). To determine the kinetic parameters for each substrate, quinate or shikimate concentrations were varied while the cofactor was kept at a saturating concentration of 2 mM. In cases where the cofactor concentration was varied, quinate or shikimate was kept at 3.2 mM. All enzyme reactions were performed in 100 μl of kinetics buffer (50 mM TRIS pH 8.8, 150 mM NaCl, 2.5 mM MgCl$_2$) at 25°C. Graphpad Prism 7® software was used for analysis of kinetic data.

Assessment of the DHQD domain activity via coupled enzymatic reaction and HPLC

DHQD activity was assessed using an enzyme-coupled assay. In this assay, we utilized a microbial dehydroshikimate dehydratase enzyme that converts the dehydroshikimate produced by DHQD to protocatechuate, which can readily be detected by HPLC (Peek *et al.*, 2017) (Figure S1). All reaction mixtures were set up with the respective cofactors, NAD$^+$ or NADP$^+$, quinate/shikimate and dehydroshikimate dehydratase enzyme. Reactions were incubated for 1 h at 22°C and then quenched with 5% formic acid, vortexed, centrifuged, flash-frozen and stored at −80°C. Reaction mixes were thawed and loaded onto an Agilent Zorbax SB-C18 Rapid Resolution 4.6 × 150 mm 3.5-μm HPLC column. The run parameters consisted of a constant flow rate of 0.75 ml min^{-1}, with a linear gradient of 95:5 to 5:95 (water:methanol with 0.1% formic acid) over 20 min. The elution profile was monitored at 254 nm.

Expression analysis

Expression analysis for the *S. lycopersicum* genes *Solyc10g038080.1* and *Solyc06g084460.2* (encoding DHQD-QDH enzymes requiring NAD$^+$ and NADP$^+$, respectively) was conducted using the Bio-Analytical Resource for Plant Biology (BAR server) Tomato Electronic Fluorescent Pictorgraph (eFP) feature (Winter *et al.*, 2007; Tomato Genome Consortium, 2012). All expression values are expressed as reads per kilobase of transcript per million mapped reads (RPKM) units with 30 units as the threshold.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Chemical reactions involved in coupled enzyme assay.

Figure S2. Saturation kinetic profiles for quinate dehydrogenases.

Table S1. Data collection and refinement statistics.

Table S2. Analysis of dehydroquinate dehydratase domain residue variation.

Appendix S1. Alignment of dehydroquinate dehydratase–quinate dehydrogenase/shikimate dehydrogenase truncated protein sequences.

REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221.

Albertini, M.V., Carcouet, E., Pailly, O., Gambotti, C., Luro, F. and Berti, L. (2006) Changes in organic acids and sugars during early stages of development of acidic and acidless citrus fruit. *J. Agric. Food Chem.* **54**, 8335–8339.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Benvenuti, M. and Mangani, S. (2007) Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nat. Protoc.* **2**, 1633–1651.

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.

Carrington, Y., Guo, J., Le, C.H., Fillo, A., Kwon, J., Tran, L. and Ehlting, J. (2018) Evolution of a secondary metabolic pathway from primary metabolism: Shikimate and quinate biosynthesis in plants. https://doi.org/10.1111/tpj.13990

Chaudhuri, S. and Coggins, J.R. (1985) The purification of shikimate dehydrogenase from *Escherichia coli*. *Biochem. J.* **226**, 217–223.

Chaudhuri, S., Lambert, J.M., McColl, L.A. and Coggins, J.R. (1986) Purification and characterization of 3-dehydroquinase from *Escherichia coli*. *Biochem. J.* **239**, 699–704.

Cheng, F., Liu, S., Wu, J., Fang, L., Sun, S., Liu, B., Li, P., Hua, W. and Wang, X. (2011) BRAD, the genetics and genomics database for Brassica plants. *BMC Plant Biol.* **11**, 136.

Christendat, D., Yee, A., Dharamsi, A. *et al.* (2000) Structural proteomics of an archaeon. *Nat. Struct. Biol.* **7**, 903–909.

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

Davis, B.D. (1951) Aromatic biosynthesis. I. The role of shikimic acid. *J. Biol. Chem.* **191**, 315–325.

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.

Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.

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.

Gamborg, O.L. (1966) Aromatic metabolism in plants III. Quinate dehydrogenase from mung bean cell suspension cultures. *Biochim Biophys Acta*, **128**, 483–491.

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

Gourley, D.G., Coggins, J.R., Isaacs, N.W., Moore, J.D., Charles, I.G. and Hawkins, A.R. (1994) Crystallization of a type II dehydroquinase from *Mycobacterium tuberculosis*. *J. Mol. Biol.* **241**, 488–491.

Graziana, A., Boudet, A. and Boudet, A.M. (1980) Association of the quinate : NAD⁺ oxidoreductase with one dehydroquinate hydro-lyase isoenzyme in corn seedlings. *Plant Cell Physiol.* **21**, 1163–1174.

Guo, J., Carrington, Y., Alber, A. and Ehlting, J. (2014) Molecular characterization of quinate and shikimate metabolism in *Populus trichocarpa*. *J. Biol. Chem.* **289**, 23846–23858.

Herrmann, K.M. and Weaver, L.M. (1999) THE SHIKIMATE PATHWAY. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50**, 473–503.

Hicks, J.M., Muhammad, A., Ferrier, J., Saleem, A., Cuerrier, A., Arnason, J.T. and Colson, K.L. (2012) Quantification of chlorogenic acid and hyperoside directly from crude blueberry (*Vaccinium angustifolium*) leaf extract by NMR spectroscopy analysis: single-laboratory validation. *J. AOAC Int.* **95**, 1406–1411.

Hoppner, A., Schomburg, D. and Niefind, K. (2013) Enzyme-substrate complexes of the quinate/shikimate dehydrogenase from *Corynebacterium glutamicum* enable new insights in substrate and cofactor binding, specificity, and discrimination. *Biol. Chem.* **394**, 1505–1516.

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

Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874.

Lallemand, L.A., Zubieta, C., Lee, S.G., Wang, Y., Acajjaoui, S., Timmins, J., McSweeney, S., Jez, J.M., McCarthy, J.G. and McCarthy, A.A. (2012) A structural basis for the biosynthesis of the major chlorogenic acids found in coffee. *Plant Physiol.* **160**, 249–260.

Letunic, I. and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245.

Lindner, H.A., Nadeau, G., Matte, A., Michel, G., Menard, R. and Cygler, M. (2005) Site-directed mutagenesis of the active site region in the quinate/shikimate 5-dehydrogenase YdiB of *Escherichia coli*. *J. Biol. Chem.* **280**, 7162–7169.

Maeda, H. and Dudareva, N. (2012) The shikimate pathway and aromatic amino Acid biosynthesis in plants. *Annu. Rev. Plant Biol.* **63**, 73–105.

Marsh, K.B., Boldingh, H.L., Shilton, R.S. and Laing, W.A. (2009) Changes in quinic acid metabolism during fruit development in three kiwifruit species. *Funct. Plant Biol.* **36**, 463–470.

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

Michel, G., Roszak, A.W., Sauve, V., Maclean, J., Matte, A., Coggins, J.R., Cygler, M. and Lapthorn, A.J. (2003) Structures of shikimate dehydrogenase AroE and its Paralog YdiB. A common structural framework for different activities. *J. Biol. Chem.* **278**, 19463–19472.

Mir, R., Jallu, S. and Singh, T.P. (2015) The shikimate pathway: review of amino acid sequence, function and three-dimensional structures of the enzymes. *Crit. Rev. Microbiol.* **41**, 172–189.

Moglia, A., Lanteri, S., Comino, C., Hill, L., Knevitt, D., Cagliero, C., Rubiolo, P., Bornemann, S. and 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 Physiol.* **166**, 1777–1787.

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

Moore, J.D., Hawkins, A.R., Charles, I.G., Deka, R., Coggins, J.R., Cooper, A., Kelly, S.M. and Price, N.C. (1993) Characterization of the type I dehydroquinase from *Salmonella typhi*. *Biochem. J.* **295**, 277–285.

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–D745.

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

Otwinowski, Z. and Minor, W. (1997) [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.

Peek, J. and Christendat, D. (2015) The shikimate dehydrogenase family: functional diversity within a conserved structural and mechanistic framework. *Arch. Biochem. Biophys.* **566**, 85–99.

Peek, J., Lee, J., Hu, S., Senisterra, G. and Christendat, D. (2011) Structural and mechanistic analysis of a novel class of shikimate dehydrogenases: evidence for a conserved catalytic mechanism in the shikimate dehydrogenase family. *Biochemistry*, **50**, 8616–8627.

Peek, J., Roman, J., Moran, G.R. and Christendat, D. (2017) Structurally diverse dehydroshikimate dehydratase variants participate in microbial quinate catabolism. *Mol. Microbiol.* **103**, 39–54.

Saleme, M.L.S., Cesarino, I., Vargas, L. et al. (2017) Silencing CAFFEOYL SHIKIMATE ESTERASE affects lignification and improves saccharification in poplar. *Plant Physiol.* **175**, 1040–1057.

Sato, Y., Itagaki, S., Kurokawa, T., Ogura, J., Kobayashi, M., Hirano, T., Sugawara, M. and Iseki, K. (2011) *In vitro* and *in vivo* antioxidant properties of chlorogenic acid and caffeic acid. *Int. J. Pharm.* **403**, 136–138.

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**, 7787–7796.

Singh, S., Stavrinides, J., Christendat, D. and Guttman, D.S. (2008) A phylogenomic analysis of the shikimate dehydrogenases reveals broadscale functional diversification and identifies one functionally distinct subclass. *Mol. Biol. Evol.* **25**, 2221–2232.

Tomato Genome Consortium. (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature*, **485**, 635–641.

Weinstein, L.H., Porter, C.A. and Laurencot, H.J. (1962) Role of the shikimic acid pathway in the formation of tryptophan in higher plants: evidence for an alternative pathway in the bean. *Nature*, **194**, 205–206.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. and Provart, N.J. (2007) An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PLoS ONE*, **2**, e718.

Wojciechowska, E., Weinert, C.H., Egert, B., Trierweiler, B., Schmidt-Heydt, M., Horneburg, B., Graeff-Hönninger, S., Kulling, S.E. and Geisen, R. (2014) Chlorogenic acid, a metabolite identified by untargeted metabolome analysis in resistant tomatoes, inhibits the colonization by Alternaria alternata by inhibiting alternariol biosynthesis. *Eur. J. Plant Pathol.* **139**, 735–747.

Ye, J., Hu, T., Yang, C., Li, H., Yang, M., Ijaz, R., Ye, Z. and Zhang, Y. (2015) Transcriptome profiling of tomato fruit development reveals transcription factors associated with ascorbic acid. Carotenoid and flavonoid biosynthesis. *PLoS ONE*, **10**, e0130885.