THE UNIVERSITY of ADELAIDE

Molecular and structural characterization of the candidate enzymes responsible for tartaric acid synthesis in the grapevine

By

Yong Jia

Thesis by a combination of conventional and publication formats submitted to The University of Adelaide for the degree of Doctor of Philosophy

The University of Adelaide

School of Agriculture Food and Wine

November 2015

Table of Contents

Abstract

L-tartaric acid (TA) is accumulated to be the dominant organic acid in the grape (*Vitis vinifera*) berries and remains metabolically stable during berry ripening and the wine making process. It confers a low pH and a "sharp" flavor to wine, affecting many quality aspects such as colour, taste, microbial stability and aging potential. Exogenous TA is widely used as a flavorant and an antioxidant additive in the food and wine industry. The metabolic intermediates involved in the TA synthesis pathway in grapevine have been well characterized. The primary pathway utilizes L-ascorbic acid (Vitamin C) as the precursor which is sequentially converted to 2-keto-L-gulonic acid (2KLG), L-idonic acid (IA), 5-keto-D-gluconic acid (5KGA) and TA by several biochemical reactions. Only one candidate enzyme, *V. vinifera* L-idonate-5-dehydrogenase (LIDH, EC 1.1.1.264), involved in this pathway has been characterized so far. It catalyzes the inter-conversion of IA and 5KGA, the rate-limiting step of the TA biosynthesis pathway. In addition, another enzyme, a putative 2-keto-L-gulonate reductase (VV2KR), which is assumed to be responsible for the reduction of 2KLG to IA, has also been reported. However, further enzymatic characterization of this enzyme is still needed to validate its metabolic function in grapevine. The present study aims to investigate the molecular and structural characteristics of these two proteins, thereby improving our understanding on TA biosynthesis in grapevine.

LIDH is highly homologous to sorbitol dehydrogenase (SDH, EC 1.1.1.14). We employed a computational approach to show that LIDH originates from the functional divergence of SDH in grapevine. We demonstrated that core eudicot (the strongly supported monophyletic group following the early-diverging eudicot lineages (Soltis, Senters et al. 2003)) SDHs have evolved into two distinctive phylogenetic lineages (Class I and Class II) due to positive natural selection after tandem gene duplication in the common ancestor of core eudicot plants. LIDH was identified as a Class II SDH. While the Class I SDH gene is universally conserved, Class II SDH genes are retained only in some plants including *V. vinifera*, *Solanum tuberosum*, *Theobroma cacao*. The

distribution of SDH genes among plant genomes showed a positive correlation between the occurrence of Class II SDH genes and the capacity for TA synthesis. These results provided new insights into the TA biosynthesis pathway from the evolutionary perspective. Protein modeling analyses revealed three amino acid substitutions at the catalytic sites between Class I and Class II SDHs. These three amino acid changes are suspected to be responsible for the unique enzymatic profile of LIDH. To confirm this hypothesis, Class I and Class II SDH genes from *V. vinifera*, *T. cacao* and *S. tuberosum* were cloned in this study. Site-mutations of the three amino acids were performed to assess their impact on enzyme substrate specificity. Recombinant protein of both wild-type and mutant SDHs from the above plants were expressed and purified. Enzyme kinetic tests confirmed that LIDH has a preference for L-idonate over D-sorbitol as its substrate. The Class I SDH from *T. cacao* demonstrated the highest activity with D-sorbitol but could hardly utilize L-idonate. The recombinant protein of the other SDHs could not be purified in this study due to troubles in protein induction and purification. Despite this, the preliminary results showed a significant enzymatic divergence between Class I and Class II SDHs, supporting the putative role of Class II SDHs in the TA biosynthesis pathway. In addition, a complete enzymatic characterization of the recombinant VV2KR was performed. VV2KR has the highest substrate specificity with D-glyoxylate, followed by hydroxypyruvate and 2KLG. We showed that VV2KR could catalyze the reduction of 2KLG to L-idonic acid effectively using NADPH as the preferred coenzyme. We went further to crystallize the recombinant VV2KR to determine its 3-dimmentional structure. The ligand-free crystal structure of VV2KR was solved to a resolution of 2.1 Å. VV2KR has the highest similarity with *Coleus blumei* hydroxypyruvate/glyoxylate reductase (CbHPR) and *Arabidopsis thaliana* hydroxypyruvate/glyoxylate reductase (AtHPR) isoform 2 (~75% amino acid sequence identity). The VV2KR monomer structure adopts the typical folding of the D-isomer 2-hydroxyacid dehydrogenase (2KDH) proteins and has the highest structural homology with the CbHPR structure (main-chain atom Root Mean Square Deviation of 0.76 Å; main-chain atom include all atoms in the peptide chain except

those from the R-group, i.e. the side-chain). It represents the second plant HPR structure being determined to date. Structural comparison of VV2KR and CbHPR with other HPR structures revealed some unique structural features for plant HPRs. The commonly accepted substrates for HPRs (D-glyoxylate, hydroxypyruvate, pyruvate) and the 2KLG intermediate from TA pathway were docked to the VV2KR structure by a computational method. The potential biological function of VV2KR in the TA biosynthesis pathway in grapevine was discussed.

In summary, we focused on the molecular and structural characterization of the candidate enzymes responsible for TA biosynthesis in grapevine. The results of the present study highlighted the effects of enzyme structural variations on their in-vivo biological functions and significantly extended our understanding on the molecular mechanism of the TA biosynthesis pathway in grapevine.

Declaration

This work contains no material previously submitted for a degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, no material which has been published or written by any other person except where due reference is made in the text.

I acknowledge that the copyright of published works contained within this thesis (as listed below-p9) resides with the copyright holder of those works.

Date 03-April-2016

Yong Jia

Statement of the contributions of jointly authored papers and manuscripts

1. Yong Jia., Darren CJ Wong, Crystal Sweetman, John B Bruning, Christopher M Ford. (2015). "New insights into the evolutionary history of plant sorbitol dehydrogenase." BMC Plant Biology 15(1):101.

Author contributions: YJ conceived the research. YJ and DCJW did sequence retrieval, curation and gene duplication characterization. YJ performed phylogenetic, synteny, natural selection and ancestral sequence analyses. YJ drafted the manuscript. JBB and YJ carried out protein modeling analyses. DCJW performed the transcript expression and gene co-expression analysis. CS and DCJW assisted with the drafting of the manuscript. CMF and JBB supervised the project.

2. **Yong Jia, Crista A Burbidge, Crystal Sweetman, John B Bruning, Christopher M Ford.** "Identification and structural characterization of a putative 2-keto-L-gulonic acid reductase from *Vitis vinifera*." Prepared for the submission to Acta Crystallographica Section D.

Author contributions: YJ, JBB and CMF conceived the research. CAB performed gene cloning and transformation. YJ and JBB purified the protein for enzymatic tests and crystallization. JBB collected the diffraction data. YJ and JBB solved the structure together. YJ, CS and CAB performed the enzymatic tests. YJ performed kinetic data and protein structure analyses. YJ drafted the manuscript. JBB and CMF contributed to the editing of the manuscript. CMF and JBB supervised the project.

The above papers and manuscripts are displayed in this thesis in either their published forms or submission forms according to the author guidelines of the specific journal.

The following authors agree with the Statement of the contributions of jointly authored papers and manuscripts and give consent to their inclusion in this thesis.

Yong Jia

John B Bruning

Crista A Burbidge

Darren CJ Wong

Crystal Sweetman

Christopher M Ford

Acknowledgement

I wish to acknowledge my supervisors Associate Professor Chris Ford and Dr. John Bruning. I would like to thank Chris for his guidance and support throughout my PhD and beyond, for introducing me into a research area that has changed my life. I feel extremely grateful for his persistent encouragement and invaluable suggestions regarding my experiments, without which I wouldn't have been able to go through those depressing PhD days. I will never forget the very enjoyable, relaxing, sometimes "aimless", chats that we had in office, which always successfully solve my problems the other way around. I would like also to thank my co-supervisor John for revealing me the novel world of crystals and protein structures, for making me feel that doing research could also be fun and cool. I appreciate John's patience and trust for me to start learning protein crystallization over half-way into my PhD, for teaching me all the relevant skills hand by hand.

I would like to thank members from the Ford's lab, including Dr. Crystal Sweetman, Dr. Damian Drew, Emma Drew, Dr. Jake Dunlevy, Dr. Darren Wong, David Contreras Pezoa, Emily Higginson, Denise Ong, Huihui Chong, Karen Francis. Also Ass/Prof. Colin Jenkins, Ass/Prof. Kathleen Soole from Flinders University. Special thanks to Dr. Crystal Sweetman, for teaching and guiding me since the first day of my PhD, the first person I would come to when I encounter any problem in the lab, and also for taking me to a lot of activities and parties and allowing me the opportunities to get in touch with local cultures. To Dr. Darren Wong, for the valuable suggestions and motivating talks regarding my research project. I miss those days working side by side with all you guys in the lab. How I wish sometime in the future we could still go back to the lab and run SDS-PAGE together. Also to members from John's lab: Alice Kroker, Andrew C. Marshall, Long Tan and Jia Truong, thanks for all the valuable advices and help regarding my experiment. It's been really lucky to meet them. I would also like to acknowledge all of the colleagues on level 4 of WIC building. These members include Dr. Sandra M. O. Mantilla, Dr. Susan Bastian, Yaelle Saltman, Dr. Ancheng Huang, Hannah Wang, Mariola Kwiatkowski, Renata Ristic, Dr. David Jeffery, Jame Wang, Joanna Gambetta, Sijing Li, Chen Liang. My thanks also go to Ee Lin Tek, Dr. Jin Zhang, Jiao Jiang from the microbiology lab level 3. They make my PhD life in Australia funny and interesting. Special thanks to Prof. Dabing Zhang, Dr. Gang Li and Xiujuan Yang from PGC. Thanks for their great help and advices with my research and beyond. I would also like to acknowledge all my CSC friends, especially Xiaoyu Chen, Cuicui Zhao, Huihui Bi etc. And also to the numerous friends I met in Highgate lodge, it's been so lucky to know them.

I wish to acknowledge my family: my parents, my sister, my young niece and nephew, and all my relatives. Thanks for always being within touch whenever I need support and encouragement. Whatever I do and wherever I am, they are always the most important people in my heart and make who I am. Finally, a special thank you to Miss Zhuolin Pan. She is the most beautiful girl that I have ever met.

Chapter 1. Introduction

1.1 General Introduction

Vitis vinifera is one of the most economically important fruit species. According to the most recent report by the International Organization of Vine and Wine (OIV), the world production of grape and wine in 2014 reached 75.1 million tons and 27.1 billion liters, respectively (OIV 2014). Wine, the major industrial product from the grape berries, is well known for its delicacy during the production process and may be affected by many factors. One of these factors corresponds to the berry acidity, which is mainly contributed by the content of tartaric acid (Castineira, Pena et al. 2002) and also the counterions such as potassium (Boulton 1980). L-tartaric acid (TA) is a four-carbon strong organic acid that accumulates to be the dominant organic acid in grape berries (Hale 1962). It is rapidly synthesized during the early stage of berry development and remains metabolically stable throughout the fruit ripening and wine-making processes (Iland and Coombe 1988, Coombe and McCarthy 2000). The content of TA is critical for wine-making. TA lowers the pH of grape juice, must and wine and significantly reduces the risk of microbial spoilage and oxidation during the fermentation and aging processes (Jackson 2000). The low pH of wine helps to maintain a desired colour which is an important quality element for red wine. TA also gives wine a sensory "sharp" flavor which balances the sweetness derived from sugar and alcohol, thereby significantly affecting the palatability of wine. As such, exogenous TA has become the primary candidate added to raise the acidity of wine and is widely used also as a flavorant and antioxidant additive in the food and pharmacy industry. The world production of TA was expected to reach 58,000 tons in 2012 (Jose 2008) and was predicted to continue rising due to the increasing market demand. Currently, industrial production of TA is mainly via a chemical method using naturally sourced potassium bi-tartrate, a by-product from the wine-making industry (Jose 2008).

The distribution of TA among higher plant displays a typical species-dependent characteristic. In contrast to malic acid, citric acid and ascorbic acid, which are the major organic acids commonly occurring in most plants, TA was found to accumulate

to a substantial level in only three plant families, i.e. *Vitaceae*, *Geraniaceae* and *Leguminoceae* (Stafford 1959). Among these, the TA content in *Vitaceae* species is significantly higher than any other plants (Stafford 1959, Burbidge 2011). The reason why *Vitaceae* plants accumulate so much TA remains a mystery to biologists, with some suggesting that it might be due to co-evolution with birds (van der Pijl 1982, Hardie 2000, Gerós, Chaves et al. 2012). The tart taste as a result of the presence of TA may prevent the immature grape berries from predation by birds or other aggressors.

Previous studies have shown that three distinct pathways of TA biosynthesis are present in higher plants. Two of them employ ascorbic acid (Vitamin C) as the precursor and involve the cleavage of the six-carbon chain at C4/C5 and C2/C3, with the four carbon fragment converted to TA in the end (Burbidge 2011). Ascorbic acid is an important antioxidant and also functions as an enzyme cofactor in plants. It has been shown to participate in numerous biological processes such as cell division, cell expansion, photosynthesis, stress responses etc (Smirnoff and Wheeler 2000, Hancock and Viola 2005). The identification of ascorbic acid as the starting compound for TA production in plants has added great biological significance to the TA pathway, which therefore serves as an alternative degradation route for ascorbic acid (Green and Fry 2005, DeBolt, Melino et al. 2007). The metabolic intermediates of TA biosynthesis from ascorbic acid have been well defined using the radioisotope labeling method. The six-carbon ascorbic acid is consecutively converted to 2KLG, L-idonic acid and 5KGA (Saito and Kasai 1984, Malipiero, Ruffner et al. 1987). Then 5KGA is further cleaved at either C4/C5, yielding TA and a two-carbon fragment (Saito and Kasai 1969, Wagner, Yang et al. 1975, Saito and Kasai 1984, DeBolt, Hardie et al. 2004) or at C2/C3, producing TA and oxalic acid (Hough and Jones 1956, Wagner and Loewus 1973). In *V. vinifera*, TA is mainly produced via the first pathway (cleavage at C4/C5) (Saito and Kasai 1969, Wagner, Yang et al. 1975, Saito and Kasai 1984, Saito and Loewus 1989, DeBolt, Hardie et al. 2004). Recent studies showed that the second pathway is also functional in grapevine, although to a much lesser degree (DeBolt, Hardie et al. 2004). In G*eraniaceae* plants, TA production proceeds via the cleavage

at C2/C3 (Wagner and Loewus 1973, Wagner, Yang et al. 1975). The third TA pathway exists in some *Leguminoceae* plants including *Phaseolus vulgaris* (Saito and Loewus 1989). In these plants, TA is derived from D-glucose which is also converted to 5KGA using D-gluconic acid as the intermediate (Saito and Loewus 1989). This ascorbic acid non-inclusive pathway has also been indicated to contribute to a minor proportion of the TA produced in *Vitis labrusca* cv Delaware and *Parthenocissus quinquefolia* L. plants (Saito and Loewus 1989). As such, all three TA biosynthesis pathways are now found to be functional in *Vitaceae* plants. The reasons for the presence of multiple TA biosynthesis pathways in plants and the intriguing relationship between the TA pathway and ascorbic acid metabolism are still poorly understood.

Compared to the well-established chemical intermediates, the genetic basis for the corresponding biochemical reactions in TA pathway is largely unknown. A number of previous studies have implied that the key steps in TA pathway are enzymatically catalyzed (Saito 1984, Malipiero, Ruffner et al. 1987, Salusjarvi, Povelainen et al. 2004). *Vitaceae* plants accumulate TA to be the dominant organic acid in young berries over a relatively short period (Iland and Coombe 1988, Coombe and McCarthy 2000), suggesting a highly active and biologically regulated metabolic process. At present, *V. vinifera* L-idonate-5-dehydrogenase (LIDH) is the only enzyme that has been characterized to be responsible for TA biosynthesis. It shares a close homology with plant sorbitol dehydrogenase and catalyzes the inter-conversion of L-idonic acid and 5KGA, the rate-limiting step in TA synthesis in grapevine (DeBolt, Cook et al. 2006). The second step of the TA pathway corresponds to the conversion of 2KLG to L-idonic acid (Saito and Kasai 1984, DeBolt, Cook et al. 2006). This step has been suspected to be catalyzed by a ketogluconate reductase (Burbidge 2011). Prokaryotic homologs of this type of protein performing the same reaction have been reported in a number of bacterial organisms including *Erwinia herbicola* (Truesdell, Sims et al. 1991), *Brevibacterium ketosoreductum* (Yum, Bae et al. 1998) and *Escherichia coli* (Yum, Lee et al. 1998). However, the corresponding enzyme in grapevine has not been identified. The enzymatic mechanism for the cleavage of

5KGA into TA also remains unclear. In the primary TA synthesis pathway in grapevine, 5KGA was suggested to be firstly converted to a four-carbon TA semialdehyde by a transketolase, then by a semialdehyde dehydrogenase to TA (Salusjarvi, Povelainen et al. 2004). However, none of these enzymes have been characterized.

While TA accumulates to a substantial level in a limited number of species (Stafford 1959, Burbidge 2011), most other plants have been shown to produce no TA or only in a trace amount. The unusual distribution of TA among higher plants exemplifies the significant phenotypic diversity of the plant kingdom. With the revelation of an increasing amount of genome sequence data, current comparative genomic studies showed that the majority of the gene families in plant genomes are largely conserved, even within those phylogenetically alienated species (Zhang, Zhang et al. 2002). The diverse phenotypes of different plants have been partly attributed to the variations of the gene family size and the functional divergence after gene duplication (Hughes 2002, Zhang, Zhang et al. 2002, Taylor and Raes 2004, Adams and Wendel 2005, Flagel and Wendel 2009). Gene duplication is widespread in plant genomes. It provides the genetic materials that natural selection could act on, thereby contributing to the phenotypic diversification of plants (Hughes 2002, Zhang, Zhang et al. 2002, Flagel and Wendel 2009). Notably, LIDH, the enzyme identified to be responsible for the rate-limiting step of TA biosynthesis in grapevine, shares a close homology with plant sorbitol dehydrogenase (DeBolt, Cook et al. 2006, Jia, Wong et al. 2015). LIDH displays principle activity on the oxidation of L-idonic acid to 5KGA, which is in contrast to plant SDHs. It has been suggested that the unusual enzymatic activity of LIDH was due to amino acid changes between proteins encoded by paralogous genes in grapevine (DeBolt, Cook et al. 2006). Phylogenetic and genetic structural analyses are needed to elaborate the evolutionary origin of LIDH gene, which would allow us better understand how TA is synthesized in grapevine. In addition, protein structural analyses are important to understand the *in vivo* function of candidate enzymes (Orengo, Todd et al. 1999, Lavery and Sacquin-Mora 2007, Worth, Gong et al. 2009). The determination of the 3-dimensional structure of protein gives us deep insights into

the structural basis of the biological functions associated with the candidate enzymes. Amino acid substitutions, especially those at the catalytic site, may change the enzyme substrate specificity (Lavery and Sacquin-Mora 2007). This is the molecular basis of gene functional evolution (Flagel and Wendel 2009; Worth, Gong et al. 2009). The present thesis aims to characterize the candidate enzymes involved in the TA biosynthesis pathway in grapevine, with a focus on the effects of gene evolution and protein structure on their biological functions. With the combination of computational, enzyme kinetic characterization and protein crystallization methods, we wish to better understand how and why TA is accumulated in grapevine and other higher plants.

1.2 The chemistry of tartaric acid

L-tartaric acid, (2R, 3R)-2,3-dihydroxybutanedioic acid, is a four-carbon dicarboxylic acid commonly found in many plants (Stafford 1959; Mattick, Plane et al. 1980). It has a first dissociation constant (pKa_1) of 2.98 and a second dissociation constant (pKa_2) of 4.34 (**Table 1**) (Dawson 1959). As shown in **Figure 1**, the extent of the dissociation of TA depends on the pH of the aqueous solutions in which it is dissolved (Mattick, Plane et al. 1980). Comparable to other common organic acids, such as malic acid (pKa_1=3.4, pKa_2=5.2) and citric acid (pKa_1=3.13, pKa_2=4.76, pKa_3=6.39) (Dawson 1959), TA is one of the strongest organic acids in higher plants. In addition to the *levo* form, another two isoforms of TA (*dextro* and *meso*) exist in nature (**Figure 2**). The *levo* and *dextro* forms of TA are able to rotate the polar light to the opposite direction while the *meso* form is a mixture of equal amount of the *levo* and *dextro* forms, and is optically inactive. TA has played an important role in the history of chiral chemistry. The successful separation of the optically active forms of TA (*levo* and *dextro*) by Louis Pasteur in 1860 led to the discovery of molecular chirality. The method that Pasteur used to resolve TA into its enantiomers laid the foundation for modern diastereomeric salt formation. At present, TA is still one of the cheapest and most readily available initial compounds for many chiral synthesis reactions. The majority of the TA occurring in plants is in the *levo* form (Stafford 1959; Wagner, Yang et al. 1975), however, minor amounts of the *dextro* and *meso* forms have also been identified in *Vitaceae*, *Geraniaceae* and spinach plants. Little information on the

molecular mechanism of the production of the dextro form is available. The present study focuses on the biosynthesis of the levo form of TA.

Table 1. Dissociation pK values for L-(+)-tartaric acid, malic acid and citric acid.

Acids	Step 1	Step 2	Step 3
Tartaric acid	$H_2Ta \rightleftarrows H^+ + HTa^-$ $K_1 = 1.04 \times 10^{-3}$ $pKa_1 = 2.98$	$HTa^- \rightleftarrows H^+ + Ta^{2-}$ $K_2 = 4.55 \times 10^{-5}$ $pKa_2 = 4.34$	NA
Malic acid	$H_2Ma \rightleftarrows H^+ + HMa^-$ $pKa_1 = 3.4$	$HMa^- \rightleftarrows H^+ + Ma^{2-}$ $pKa_2 = 5.2$	NA
Citric acid	$H_3Cia \rightleftarrows H^+ + H_2Cia^-$ $pKa_1 = 3.13$	$H_2Cia^- \rightleftarrows H^+ + HCia^{2-}$ $pKa_2 = 4.78$	$HCia^{2-} \rightleftarrows H^+ + Cia^{3-}$ $pKa_3 = 6.39$

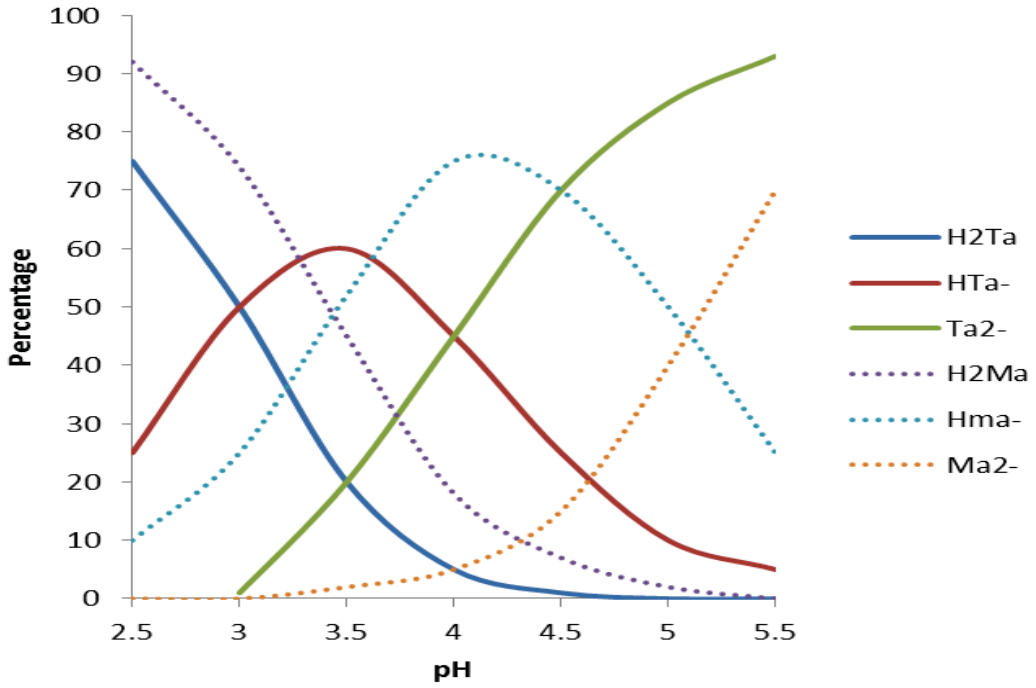


Figure 1. Relative concentration of tartrate and malate ionic forms in water at various pH conditions. Ta refers tartaric acid. Ma refers malic acid. (Mattick, Plane et al. 1980).

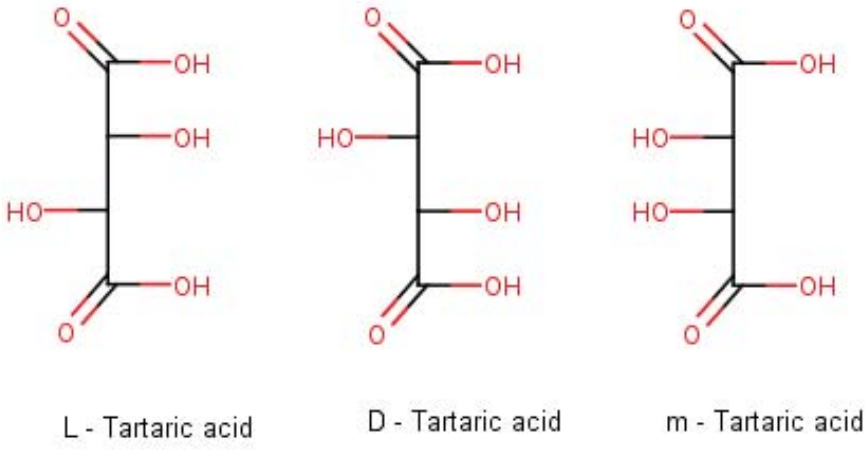


Figure 2. **Molecular configurations of tartaric acid isomers.** L-Tartaric acid, D-Tartaric acid and m-Tartaric acid represent the *levo*, *dextro* and *meso* forms, respectively. The L and D forms are optically active while the m form is optically inactive.

1.3 Unequal distribution in higher plants

Pioneering work in the 1960s by Stafford shed some light on the unusual occurrence of TA in higher plants. TA was present in most of the 49 angiosperm species screened. However, only 9 species (*V. vinifera*, *V. labruscanna*, *V. valifornica*, *Parthenocissus tricuspidata*, *P. quinquefolia*, *Pelargonium hortorum*, *Bauhinia malabarica*, *Phaseolus vulgaris* and *Coleus blumei*) were shown to accumulate TA to a substantial level. Most of these species are from the *Vitaceae*, *Geraniaceae* and *Leguminosae* families (Stafford 1959). A comprehensive profile of the TA content in various angiosperm plants was summarized by Burbidge (Burbidge 2011) (**Table 2**). In general, *Vitaceae* plants exhibit a dramatically higher level of TA than plants from other families. While some *Leguminosae* plants also accumulate a comparable amount of TA as *Vitaceae* plants, most species from this family have relatively lower levels of TA (Stafford 1959). The similar observation was made with the *Geraniaceae* plants (Stafford 1959). In addition to these three plant families, a few species such as *Coleus blumei* (*Labiatae*) also retain a large amount of TA (Stafford 1959). Within those prominent TA-accumulating families, the level of TA also was found to vary greatly among different species and varieties (Stafford 1959; Saito and Loewus 1989). In

addition, a similar pattern of TA accumulation was found among different varieties of *Solanum tuberosum*, with some also accumulating a significant amount of tartaric acid (Galdon, Mesa et al. 2010). This suggests that the TA biosynthesis may not only be regulated at the genetic level, but could also be affected significantly by the environmental factors. The fact that some potato varieties also produce a large amount of TA expands the TA-accumulating species to the *Solanaceae* family as well. However, *Solanum lycopersicum* (tomato), another important species from this family, has been shown to produce no TA (Suarez, Rodriguez et al. 2008). Additionally, TA has also been detected in some fruit species such as citrus (Nour, Trandafir et al. 2010) and pear (Hudina and Stampar 2000, Sha, Li et al. 2011), although to a significantly lower level. No TA or trace amount of TA is present in apple, a *Rosaceae* plant (Fuleki, Pelayo et al. 1995). Compared to other prevalent organic acids such as malic acid and citric acid, our knowledge on TA's distribution profile among higher plants is still very fragmentary. This is partially due to a lack of research interest on TA, which has been shown to be metabolically stable after synthesis and doesn't seem to have any biological function. More effort should be made to reveal a complete picture of the distribution of TA among higher plants.

Table 2. **Content of tartaric acid in different angiosperm plants**. Fresh weight (FW). Adapted from Burbidge (Burbidge 2011) based on data collected by DeBolt (DeBolt 2006), Melino (Melino 2009) and Stafford (Stafford 1959).

Species	Family	µM/g FW leaf tissue	mg/g FW berry or tuber tissue
Vitis vinfera	*Vitaceae*	182	12.8
Vitis riparia	*Vitaceae*		17.7
Vitis acerifolia	*Vitaceae*		16.9
Vitis bloodworthiana	*Vitaceae*		16.5
Vitis amerensis	*Vitaceae*		11.9
Vitis jaquomontii	*Vitaceae*		11.3
Vitis californica	*Vitaceae*		9.6
Vitis champinii	*Vitaceae*		9.5
Parthenocissus henryana	*Vitaceae*		8
Bauhinia malabarica	*Leguminosae*	154	
Vitis labruscana	*Vitaceae*	121	
Parthenocissus tricuspidata	*Vitaceae*	57	
Parthenocissus quinquefolia	*Vitaceae*	54	
Pelargonium hortorum	*Geraniaceae*	29	
Phaseolus vulgaris cv Wakaba	*Leguminosae*	0.99	
Coleus blumei	*Labiatae*	6	
Solanum lycopersicum	*Solanaceae*		0.88-3.33
Ampelopsis brevipedunculata	*Vitaceae*		0.6
Tropaeolum majus	*Tropaeolaceae*	0.5-5.0	
Rosa sp.	*Rosaceae*	0.5-5.1	
Magnolia soulangeana	*Magnoliaceae*	0.5-5.2	
Glycine max	*Leguminosae*	0.5-5.3	
Phaseolus limensis	*Leguminosae*	0.5-5.4	
Pisum sativum	*Leguminosae*	0.5-5.5	
Zea mays	*Gramineae*	0.5-5.6	
Erodium cicutarium	*Geraniaceae*	0.5-5.7	
Geranium viscosissimum	*Geraniaceae*	0.5-5.8	
Ricinus communis	*Euphorbiaceae*	0.5-5.9	
Oxalis corriculata	*Oxalidaceae*	0-0.5	
Medicago sativa	*Leguminosae*	0-0.5	
Triolium pratense	*Leguminosae*	0-0.5	
Aesculus hippocastanum	*Hippocastanaceae*	0-0.5	
Hordeum vulgare	*Gramineae*	0-0.5	
Geranium richardsonii	*Geraniaceae*	0-0.5	
Chrysanthemum sp.	*Compositae*	0-0.5	
Beta vulgaris	*Chenopodiaceae*	0-0.5	
Citrus sp.	*Rutaceae*		0.38(g/L)
Pyrus sp.	*Rosaceae*		0.02
Phaseolus vulgaris cv Morocco	*Leguminosae*	0	
Ampelopsis aconitifolia	*Vitaceae*		0
Arabidopsis thaliana	*Brassicaceae*	0	

1.4 Tartaric acid in grape and wine

During the grape berry development, TA is produced rapidly during the first 4 weeks after anthesis (Iland and Coombe 1988, Coombe and McCarthy 2000). The total amount of tartrate in *V. vinifera*, including the free form and the salt form, can reach 10 mg per berry, over twice that of malate at harvest (Coombe and McCarthy 2000, Sweetman, Deluc et al. 2009). TA, together with malic acid, may account for 70-90% of the acidity in all parts of the vine except the root tissues (Ruffner 1982). In contrast to malic acid, which is rapidly broken down and metabolized via the TCA cycle when ripening starts (Saito and Kasai 1968, Iland and Coombe 1988, Sweetman, Deluc et al. 2009), TA is inert to the metabolic pathways throughout the ripening process and remains stable during the subsequent wine fermentation and aging processes (Saito and Kasai 1968, Iland and Coombe 1988, Sweetman, Wong et al. 2012). TA accumulation in the grape berries creates a low pH and a zing flavor, affecting many quality aspects of wine such as colour and taste. TA in wine balances the sweetness created by sugar and alcohol, providing the "structure" of wine. It also reduces the microbial spoilage and oxidation of wine, thereby significantly improving wine's organoleptic and aging potentials. The reason why grape berries accumulate so much TA remains unclear. Biologically, it has been suggested that the accumulation of TA in grape berries is due to co-evolution with birds (van der Pijl 1982, Hardie 2000). The TA production contributes to the unpalatability of immature grape berries, which may protect them from the early predation by birds and other aggressors.

1.5 Industrial application and production

Due to its unique chemical and organoleptic properties, exogenous TA is widely used as a flavorant and antioxidant additive in the food, pharmacy and wine industry (Singh 1976, Freddi, Allara et al. 1996, Kassaian 2003). It has been suggested to be a substitute to the currently widely used citric acid (Rivas, Torrado et al. 2006, Manuel Salgado, Rodriguez et al. 2010). During the wine-making process, TA has been commonly added to raise the acidity of wine. This helps to create a more balanced and full-bodied wine during warmer years when the acid level in the grape berries is not high enough. In addition, as a chirally active compound, TA is also widely used in the

chiral synthesis area (Kassaian 2003). It is one of the cheapest and most readily available initial compounds for many chiral synthesis reactions. As a naturally occurring organic acid, TA occurs in both free and salt forms in large amount in grape berries (Synoradzki, Ruskowski et al. 2005). It was first isolated from tartar (a deposit from fermented grape juice), by the alchemist Jabir ibn Hayyan c.800 (Solieri and Giudici 2009). Methods of TA production were reviewed by Amerine etc (Amerine, Berg et al. 1972). The free form of TA was first separated by Swedish chemist and apothecary Carl Wilhelm Scheele. His work in 1769 laid the foundation of the modern process of TA production. TA has been mainly produced from potassium bi-tartrate (cream of tartar), a by-product from the wine making process (Rosenberg, Mikova et al. 1999, Willaert and De Vuyst 2006). The recovery of TA from the fermentation lees and in the form of cream of tartar usually requires the addition of calcium salts to precipitate the insoluble calcium tartrate (Amerine, Berg et al. 1972). Another method of TA production involves the chemical hydroxylation of maleic acid. However, the TA obtained by this method occurs in the racemic form, which is much less soluble than the levo form (Milsom 1987). Therefore, this application is very limited.

With the increasing demand for TA and the rapid development of biotechnology, several fermentative methods have been developed to produce TA. *Gluconobacter suboxydans* has been used to produce TA by the fermentation of panthothenic acid (Prave, Faust et al. 1987). The hydrolysis of cis-epoxysuccinic acid was also utilized as a method for the production of TA (Synoradzki, Ruskowski et al. 2005). In addition, biotechnological methods for TA formation have been described. 5-oxogluconic acid and glucose have been used as a starting compound to produce TA by *Acetobacter* and *Gluconobacter* species (Kotera, Yamada et al. 1972). However, these methods are of no practical importance for the large scale production of TA.

TA is an economically important compound. The world production of TA in 2001 was estimated to be 35,000 tons with a price of about € 6.00/kg (Lichtenthaler 2002). Based on an across-Australia average addition level of 4 g TA per liter of juice or must, the cost of the addition of TA was estimated at $AU 40 M for Australia in 2001 (DeBolt 2006). The current major TA production method is limited by the source of

wine lees. It also consumes a large amount of energy and potentially causes pollution to the environment. A more efficient and environmentally friendly method needs to be developed for its industrial production.

1.6 Tartaric acid biosynthesis

Earlier studies based on radioisotope tracing analyses in the 70s and 80s showed that the TA biosynthesis pathway lies outside the TCA cycle. Unlike the biosynthesis of malate and citrate, it does not involve the oxidative metabolism of sugar (Loewus and Stafford 1958). Ascorbic acid was first recognized as an effective precursor for the production of TA in the *Vitaceae* and *Geraniaceae* plants (Saito and Kasai 1969, Wagner and Loewus 1973, Wagner, Yang et al. 1975). In these plants, TA either originates from the C1-C4 or the C3-C6 fragment of the six-carbon chain of ascorbic acid (Saito and Kasai 1969, Wagner and Loewus 1973). The discovery of ascorbic acid as the precursor for TA biosynthesis in plants has attracted great interest to TA research. In addition to ascorbic acid, later study in some *Leguminoceae* species showed that TA could also be produced from D-glucose (Saito and Loewus 1989). It is now recognized that three distinct TA biosynthesis pathways exist, each contributing differently to the total accumulation of TA in different plants.

1.6.1 The primary TA biosynthetic pathway in grapevine

V. vinifera accumulates the highest amount of TA among higher plants (Stafford 1959). TA in grapevine is in the L form and is mainly produced from ascorbic acid via the cleavage at the C4/C5 position (Saito and Kasai 1969, Wagner, Yang et al. 1975). A schematic graph of this pathway is shown in Figure 3. Ascorbic acid is consecutively converted into 2KLG, L-idonic acid, 5KGA (Saito and Kasai 1982, Saito and Kasai 1984, Malipiero, Ruffner et al. 1987). Then 5KGA is split between C4 and C5 into a four-carbon so-called TA semialdehyde and two-carbon putative glycoaldehyde, with the TA semialdehyde being further reduced to TA (Saito 1979, Saito and Kasai 1984). The first proposal of ascorbic acid as the precursor of TA synthesis in grapevine was made by Hough and Jones in 1956, based on the observation that the C2-3 of TA originate from the C4-5 of ascorbic acid (Hough and Jones 1956). Since then, various forms of radio-labeled ascorbic acid have been tested to determine the metabolic

origin of TA. When the 6-^{14}C labeled ascorbic acid was fed into grape leaves, a large amount of the radiolabel was incorporated into different sugars including sucrose and glucose. No significant radiolabel was detected in TA (Loewus 1957, Saito and Kasai 1968). These results, in the beginning, led to the speculation that there was no direct TA synthesis pathway from ascorbic acid. However, in a later study, when the 1-^{14}C labeled ascorbic acid was administered to immature grape berries, 72% of the radiolabel was recovered as the 1-^{14}C labeled TA, providing strong support that ascorbic acid is an efficient precursor for TA production in grapevine (Saito and Kasai 1969). Additional studies using both 1-^{14}C and 6-^{14}C labeled ascorbic acid showed that TA produced in grapevine is in the levo form and originates from the split between C4 and C5 of ascorbic acid (Wagner 1974, Wagner, Yang et al. 1975). This is further confirmed by another study in which 1-^{14}C and 4-^{14}C labeled ascorbic acid were tested. Both of these two compounds were found to be equivalent precursors to carboxyl labeled L TA (Williams and Loewus 1978).

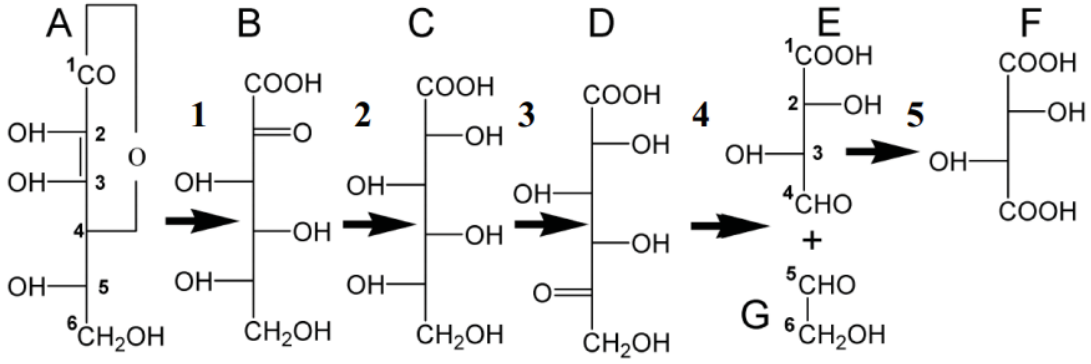
The metabolic intermediates between ascorbic acid and TA were determined in a separate study. After the 1-^{14}C labeled ascorbic acid was fed to the slices of immature grape berries, iodoacetic acid was applied to inhibit the potential enzyme activity, thereby allowing the analyses of the occurrence of the radiolabel in the putative intermediates (Saito and Kasai 1982). Three radiolabeled compounds (L-idonic acid, L-idono- γ-lactone and 2KLG) were identified (Saito and Kasai 1982). Further, when radiolabeled L-idonic acid and 2KLG were tested, another compound, which was later characterized as 5KGA, was identified (Saito and Kasai 1982, Saito and Kasai 1984). In a comprehensive time-course study on the conversion of ascorbic acid to TA, consecutive formation of 2KLG, L-idonic acid and 5KGA was observed (Malipiero, Ruffner et al. 1987). This study also revealed that the conversion of L-idonic acid to 5KGA is the rate-limiting reaction in the production of TA from ascorbic acid (Malipiero, Ruffner et al. 1987). This is consistent with a later report that the accumulation of 5KGA controlled the rate of TA synthesis in some bean plants, whereby the formation of ascorbic acid increased by almost 200-times when 5KGA was fed to cultivars which accumulate no significant TA (Saito and Loewus 1989).

These studies not only led to the identification of the putative chemical intermediates, they also provided the initial evidence that the production of TA in grapevine is an enzymatically controlled process, which promoted later studies trying to identify the genes and enzymes responsible for this pathway. After the split at C4/C5, 5KGA is speculated to be converted to a four-carbon threonate (TA semialdehyde) and a two-carbon glycoaldehyde (Williams, Saito et al. 1979, Loewus 1999, Salusjarvi, Povelainen et al. 2004). The threonate was further oxidized to yield TA while the glycoaldehdye was recycled back to the triose pathway (Wagner 1974, Williams, Saito et al. 1979).

Ascorbic acid is a ubiquitous antioxidant with multiple functions in plants. It serves as important enzyme co-factor and cellular antioxidant and plays a critical role in many metabolic processes including photosynthesis, photoprotection, cell growth, stress-resistance and the synthesis of ethylene, gibberellins, anthocyanins and hydroxyproline (Smirnoff and Wheeler 2000). Ascorbic acid is synthesized via multiple pathways. The most prominent one is via the oxidation of L-galactose with L-galactono-1,4-lactone as the intermediate (Hancock and Viola 2005). Other ascorbic acid biosynthesis pathways include the L-gulonic acid pathway and the D-galacturonic acid pathway (Hancock and Viola 2005). In contrast to the elaboration of its biosynthesis pathways, the metabolic degradation of ascorbic acid in plants has remained elusive for a long time. The biosynthesis of TA from ascorbic acid therefore represents an important pathway for the degradation of ascorbic acid in plants (Green and Fry 2005, DeBolt, Melino et al. 2007). The conversion of ascorbic to TA has been suggested to occur via dehydroascorbate as an intermediate (Green and Fry 2005). This is observed in many oxalate-accumulating plants including grapevine where ascorbic acid is oxidized to dehydroascorbate before its further hydrolysis to TA and oxalic acid (Yang and Loewus 1975). In grapevine, 1-^{14}C labeled dehydroascorbate has been found to be equally efficient as ascorbic acid for the production of radiolabeled TA. Notably, ascorbic acid in plants could be readily converted to dehydroascorbate. This process is widely reversible (Yang and Loewus 1975) and may be utilized by plants to create a reserve for ascorbic acid. At the moment, the

biochemical mechanism between ascorbic acid and 2KLG in the TA pathway remains unclear. Whether or not dehydroascorbate is involved needs to be further characterized.

Figure 3. **The primary pathway leading to TA formation; AA is cleaved between C4 and C5.** (**A**: ascorbic acid; **B**: 2-keto-L-gulonic acid; **C**: L-idonic acid; **D**: 5-keto-D-gluconic acid; **E**: L-tartaric acid semialdehyde; **G**: putative glycoaldehyde; **F**: L-tartaric acid.) (Adapted from (DeBolt, Cook et al. 2006))



1.6.2 TA biosynthetic pathway in *Geraniaceae* plants

The second pathway of TA biosynthesis is outlined in Figure 4. This pathway shares the metabolic processes from ascorbic acid to 5KGA with the first pathway described above. Instead of the cleavage at the C4/C5 position of 5KGA, a cleavage at the C2/C3 position occurs, yielding oxalic acid and threonic acid respectively, with threonic acid being further oxidized into TA. Wagner and Loewus (Wagner and Loewus 1973) showed that, in some *Pelargonium crispum* plants, the radiolabel of 1-^{14}C AA and 6-^{14}C AA could be specifically converted into oxalic acid and TA respectively. They found that C6 of L-ascorbic acid or its precursor L-galactono-1,4-lactone was converted to one of the carboxyl carbons of the TA (Wagner and Loewus 1973). This differs from Saito and Kasai's report that C1 of ascorbic acid becomes the carboxyl carbon of TA (Saito and Kasai 1969) and also the report by Loewus and Stafford that C6 of ascorbic acid could not be converted to TA in grapevine (Loewus and Stafford 1958). In addition, C1 of ascorbic acid was found to be converted to oxalic acid (Wagner and Loewus 1973) while the two-carbon

fragment resulted from the cleavage of ascorbic acid in grapevine was speculated to be glycoaldehyde and was metabolized through the triose pathway (Wagner 1974, Williams, Saito et al. 1979). For a period of time, it was believed that this pathway only exists in the family *Geraniaceae*. Recently, both oxalic acid and TA were shown to be formed from ascorbic acid in berries of *V. vinifera* (DeBolt, Hardie et al. 2004, Green and Fry 2005). These results suggest that both pathways are functional in grapevine, even though the second one may account for only a small proportion of the total ascorbate metabolized.

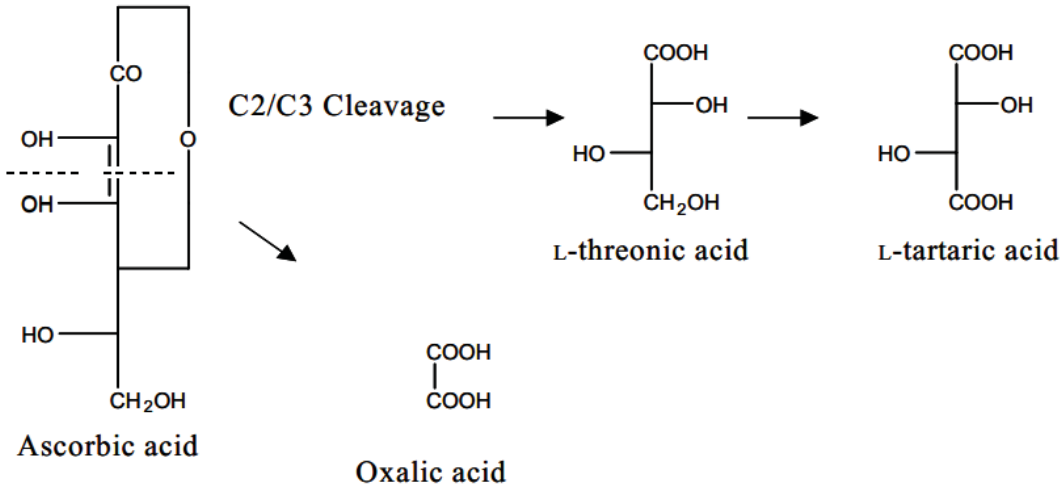


Figure 4. The secondary pathway leading to TA formation. AA is cleaved between C2 and C3 (Saito, Ohmoto et al. 1997).

1.6.3 TA biosynthetic pathway identified in *Leguminoceae* plants

For the biosynthesis of TA in higher plants, a third pathway (Figure 5) was identified in some *Leguminosae* plants, in which TA was produced from D-glucose via D-gluconic acid and 5KGA as the intermediates (Saito and Loewus 1989). This pathway was confirmed by Saito and Loewus (Saito and Loewus 1989) in *Phaseolus vulgaris*, a *Leguminosae* plant. The authors suggested that TA in this plant is synthesized from a hexose precursor (Saito and Loewus 1989). In their study, D-gluconate was found to be an equivalent precursor as 5KGA for the production of TA in bean leaves. D-glucose has long been long suggested as a precursor for TA synthesis in many plants (Loewus and Stafford 1958, Saito and Kasai 1969, Wagner,

Yang et al. 1975, Saito 1984, Saito and Loewus 1989). The use of D-glucose for the production of TA was first confirmed by Saito and Kasai in *Pelargonium* plants (Saito 1984). It was shown that the conversion of D-glucose to 5KGA proceeds via a hexonic acid which was speculated most probably to be D-gluconic acid (Saito 1984). However, this pathway only accounts for a very limited proportion of the total production of TA in *Pelargonium* plants (Saito 1984). In addition to *Pelargonium* plants, this pathway has also been shown to be functional in *Vitaceae* plants (Saito and Kasai 1969, Wagner, Yang et al. 1975, Saito 1984, Saito and Loewus 1989), contributing to less than 9% of the total TA accumulated (Saito and Loewus 1989).

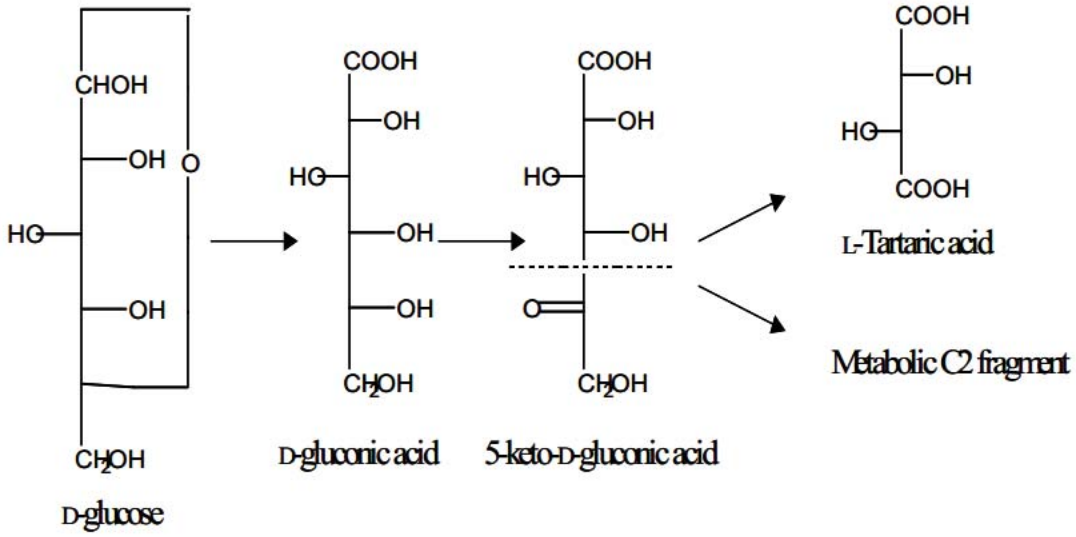


Figure 5. The third pathway leading to TA formation. D-glucose is converted to D-gluconic acid and further to 5-keto-D-gluconic acid which is then cleaved between C4 and C5 (Saito and Loewus 1989).

1.7 Enzymes responsible for tartaric acid biosynthesis

Earlier research using radiolabel tracing has contributed to the identification of the chemical intermediates in TA biosynthesis. However, the underlying biochemical mechanism is still poorly understood. The hydrolysis of ascorbic acid to tartaric acid has been suggested to be able to operate both enzymatically and non-enzymatically (Green and Fry 2005). The non-enzymatic pathway may account for the vitamin losses during cooking (Green and Fry 2005). In grapevine, the accumulation of

tartaric acid occurs rapidly during the first four weeks of grape berries development and levels out throughout the later stages (Saito and Kasai 1968, Iland and Coombe 1988, Coombe and McCarthy 2000, Sweetman, Wong et al. 2012), suggesting a highly regulated biological process. In the comprehensive kinetic analyses of the TA metabolic pathway, the conversion of L-idonate to 5KGA was determined as the rate-limiting reaction (Malipiero, Ruffner et al. 1987). The feeding of 1-^{14}C labeled 5KGA to grapevine leaves led to a more than ten-fold increase of radiolabel in the TA product (Saito and Kasai 1984). This indicates that, at least, the production of 5KGA has an enzymatic basis.

By combining the transcript profiling of expressed sequence tag (EST) data and metabolite profiling of organic acids, DeBolt etc (DeBolt, Cook et al. 2006) successfully identified an enzyme named LIDH (E.C 1.1.1.264) which was found to be responsible for the conversion of L-idonate to 5KGA. The transcription of LIDH gene is active in young berries and leaves and decreases in later developmental stages, positively correlated with the production of TA (DeBolt, Cook et al. 2006). LIDH shares a very high similarity with plant sorbitol dehydrogenase such as *A. thaliana* SDH (Gene ID: AT5G51970) and tomato SDH (Gene ID: 778312) and is considered a member of the plant SDH family (DeBolt, Cook et al. 2006, Jia, Wong et al. 2015). However, the highest activity of LIDH was observed against L-idonate with NAD(H) as the coenzyme while most of the currently determined plant SDHs prefer sorbitol as the substrate (DeBolt, Cook et al. 2006). In addition to the LIDH, no enzyme has been characterized responsible for the other steps of the TA biosynthesis pathway. It has been shown that the conversion of 5KGA to TA could proceed by chemical means under both acidic and alkaline conditions (Barch 1933, Isbell and Holt 1945) or be catalyzed by vanadate in some prokaryotic organism such as *G. oxydans,* a TA-accumulating strain (Klasen, Bringermeyer et al. 1992). This causes speculation that the last two steps of TA production pathway may be non-enzymatic. However, significant production of TA from 5KGA was also observed in the absence of vanadate in *G. oxydans* (Salusjarvi, Povelainen et al. 2004). The reaction from 5KGA to TA was proposed to proceed via an intermediate called TA semialdehyde, catalyzed

sequentially by a putative TA transketolase and a putative TA semialdehyde dehydrogenase (Salusjarvi, Povelainen et al. 2004). During the screening for the candidate genes responsible for TA production, DeBolt also identified the grapevine homologous genes putatively encoding the above two candidate enzymes (DeBolt 2006). However, none of these enzymes were shown to be able to catalyze the corresponding reaction at the significantly level (DeBolt 2006). Further studies are needed to investigate the underlying metabolic basis.

The second step of the ascorbic acid inclusive TA biosynthesis pathway involved the conversion of 2KLG to L-idonic acid. An earlier radiolabelling trace study indicated that this reaction was also enzyme catalyzed (Saito 1984). While the corresponding enzyme responsible for this reaction in grapevine remains to be characterized, an NADPH-dependent 2-keto-D-gluconate reductase (2KGR, EC 1.1.99.3) able to perform the same reaction has been reported in *E. herbicola* (Truesdell, Sims et al. 1991), *B. ketosoreductum* (Yum, Bae et al. 1998), *E. coli* (Yum, Lee et al. 1998) and various acetic bacteria (Ameyama and Adachi 1982). 2KGRs in bacteria are involved in ketogluconate metabolism and have broad substrate specificity. They catalyze the reversible reduction of 2-keto-D-gluconate (2KGA) to D-gluconate, 2, 5-diketo-D-gluconate to 5KGA and 2KLG to L-idonate, using NADP(H) as the preferred cofactor and generally have the highest activity on 2KGA (Ameyama and Adachi 1982, Truesdell, Sims et al. 1991, Yum, Bae et al. 1998, Yum, Lee et al. 1998). Enzymes of this kind belong to the D-isomer-specific 2-hydroxyacid dehydrogenase (2KDH) superfamily; the family members catalyze the reversible reduction of 2-oxoacids to the D-isomers of the respective 2-hydroxyacids. Based on these observations, Burbidge (Burbidge 2011) identified a candidate enzyme from grapevine by homology search against several ESTs datasets using the *E. coli* 2KGR sequence as a query. Preliminary in-vitro enzymatic tests of the purified recombinant protein showed that it is able to reduce 2KLG using both NADH and NADPH as the coenzymes (Burbidge 2011). However, no comprehensive enzyme characterization of this protein was performed. Initial amino acid sequence analysis suggests it has a close relationship with plant hydroxypyruvate/glyoxylate reductases (HPR). HPR

belongs to the 2KDH superfamily and displays broad substrate specificity (Janiak, Petersen et al. 2010). It specifically acts on the C2 hydroxyl group of donor with NADPH as the preferred acceptor (Janiak, Petersen et al. 2010), a reaction pattern also consistent with the conversion of 2KLG to L-idonic acid. Plant HPRs are responsible for the reduction of hydroxypyruvate to D-glycerate in the Calvin cycle and play a critical role in the photorespiration pathway (Mano, Hayashi et al. 1997, Timm, Nunes-Nesi et al. 2008, Timm, Florian et al. 2011). Multiple copies of HPR genes exist in *A. thaliana* and encode proteins targeted to different subcellular locations (Mano, Hayashi et al. 1997, Timm, Nunes-Nesi et al. 2008, Timm, Florian et al. 2011). Whether or not HPR may be involved in more than one metabolic pathway is unknown. Notably, despite its preference for hydroxypyruvate as the substrate, another plant HPR characterized from *C. blumei* has been shown to be responsible for the production of rosmarinic acid, catalyzing the reduction of 4-hydroxyphenylpyruvate to the corresponding 4-hydroxyphenyllactic acid (Kim, Janiak et al. 2004, Janiak, Petersen et al. 2010). Whether or not the 2KGR homolog, a putative HPR identified from grapevine may contribute to the production of TA, remains to be investigated.

1.8 Gene duplication, protein structural variation and plant phenotypic diversity

The identification of LIDH gene from grapevine was a breakthrough in the study of the molecular mechanism of TA biosynthesis in higher plants. Primary structural analyses showed that LIDH belongs to the plant SDH family (DeBolt, Cook et al. 2006). LIDH shares a close homology to plant SDH. For example, LIDH exhibits ~77% identity with SDHs from *S. lycopersicum* (Gene ID: 778312) and *A. thaliana* (Gene ID: AT5G51970) at the amino acid sequence level (DeBolt, Cook et al. 2006, Jia, Wong et al. 2015). Plant SDH, together with sorbitol-6-phosphate dehydrogenase (EC 1.1.1.140), are the key enzymes responsible for sorbitol metabolism in plants (Nosarzewski, Downie et al. 2012). SDH catalyzes the inter-conversion of sorbitol and D-fructose (Negm and Loescher 1979, Oura, Yamada et al. 2000, Maria, Diego et al. 2013). It has the highest reaction rate for the oxidation of sorbitol but could also utilize other polyols such as ribitol, xylitol, iditiol to a lower degree (Ng, Ye et al.

1992, Oura, Yamada et al. 2000, Maria, Diego et al. 2013). Despite its strong homology with SDH, LIDH has been shown to have principal activity against L-idonate and displays a low reaction rate with sorbitol (DeBolt, Cook et al. 2006). The unique substrate specificity of LIDH was suggested to be due to small changes in amino acid sequence encoded by paralogous genes in grapevine (DeBolt, Cook et al. 2006). It is possible that LIDH represents a functional divergence of SDH in grapevine, which may have experienced a shift in its metabolic function after the accumulation of amino acid changes. The evolutionary relationship between LIDH and SDH remains to be determined.

Functional divergence via gene duplication is the major mechanism by which genes with novel function evolve, thereby contributing to the phenotypic diversity of higher plants (Hughes 1994, Zhang 2003, Flagel and Wendel 2009). The unequal distribution of TA exemplifies the great phenotypic diversity of the angiosperm species, which acquired the most extensive diversification among the plant kingdom during evolution (Van de Peer, Fawcett et al. 2009). The predominance of flowering plants has been attributed to the wide occurrence of polyploidy or whole-genome duplication (WGDs) (De Bodt, Maere et al. 2005), which generates the genetic basis for their adaptation to the surrounding environment. Recent whole genome sequence analyses revealed that eudicots and monocots have experienced a triplication and two WGD, respectively before their subsequent speciation (Jaillon, Aury et al. 2007, Tang, Bowers et al. 2010). Comparative genomic studies showed that gene families were largely conserved across species, including those phylogenetically alienated plants (Rensing, Lang et al. 2008). On the other hand, variations in gene family size are very common among various species lineages (Flagel and Wendel 2009). Therefore, it has been suggested that it is the duplication and adaptive specialization of pre-existing genes instead of the emergence of novel genes that have contributed to the diverse phenotypes in higher plants (Flagel and Wendel 2009). Duplicated genes could arise through different mechanisms including unequal crossing over, retrotransposition, segmental duplication and chromosomal or whole genome duplication (Zhang 2003, Hurles 2004). Most duplicated genes are lost due to the accumulation of mutations

that render them non-functional (pseudogenization) (Zhang 2003). However, they can be retained under certain circumstances whereby the acquisition of beneficial mutations leads to novel function (neofunctionalization), which requires positive natural selection, or through adoption of part of the functions of the ancestral gene (sub-functionalization). This may be achieved by expression divergence or functional specialization of the protein, which usually lead to changes in the substrate specificity of the enzymes (Force, Lynch et al. 1999; Zhang 2003; Conant and Wolfe 2008; Flagel and Wendel 2009).

1.9 Protein structural variation and biological function

Protein structural analyses are very important to understand the biological function of the candidate enzymes. Proteins being able to perform the same biological functions tend to adopt a similar structure (Petsko and Ringe 2004; Worth, Gong et al. 2009). The critical amino acids directly involved in the catalytic processes are usually highly conserved among homologous proteins from different organisms (Bartlett, Porter et al. 2002; Petsko and Ringe 2004). Changes to these amino acids could lead to either the de-activation or the shift of function of the target protein (Zhang 2003). While the de-activation mutations usually result in the dysfunction of the plant and are lost during the natural selection process, those mutations generating moderate functional changes may create some advantages for plants to better adapt to the diverse and constant-changing environment and are therefore retained (Zhang 2003; Flagel and Wendel 2009). This serves as the structural basis for the gene evolution and functional divergence. Given that the chemical intermediates in the TA biosynthesis pathway have been well determined, the candidate enzymes responsible for the corresponding reactions could be identified by homology search for proteins being able to catalyze the similar reactions in other organisms. For example, LIDH, the only enzyme currently identified in the TA pathway, has been reported to share very high similarity with plant SDHs (DeBolt, Cook et al. 2006). SDH catalyzes the oxidation of the C2 hydroxyl of sugar alcohols into their corresponding ketoses, preferring polyols with a d-cis-2,4-dihydroxyl (2S,4R) configuration (Negm and Loescher 1979; Lindstad, Hermansen et al. 1992; Lindstad, Koll et al. 1998; Oura, Yamada et al. 2000; Maria,

Diego et al. 2013). L-idonic acid has the same molecular configuration with D-sorbitol from C1 to C4 and differs only at C5 (D and L chirality) and C6 (a hydroxyl group in sorbitol is replaced by a carboxyl group in L-idonic acid). The conversion of L-idonic acid to 5KGA follows the same reaction patterns catalyzed by SDHs, in good agreement with the close homology between LIDH and SDH. Protein structural analyses, such as protein crystallography, are needed to investigate the structural basis for the observed divergent enzyme kinetic profiles between the two proteins. Similarly, the candidate enzyme responsible for the other steps of the TA synthesis pathway could also be identified by homology search for proteins being able to perform the same types of reactions. Using the *E. coli* 2KGR sequence as a query, Burbidge (Burbidge 2011) identified a candidate enzyme being able to reduce 2KLG to L-idonic acid. This protein was found to be highly homologous to previously characterized plant HPRs (Burbidge 2011). The detailed enzymatic profile of this candidate protein, such as its activity with hydroxypyruvate, has not been determined. In-vivo functional characterization of this protein is also needed to determine its metabolic function. Noteworthy, another plant HPR from *C. blumei* has been shown to be involved in the rosmarinic acid synthesis pathway, despite its higher enzymatic activity with the hydroxypyruvate reduction (Kim, Janiak et al. 2004). Interestingly, *C. blumei* has been identified as TA-accumulating plant as well (Stafford 1959). Whether or not the identified grapevine 2KGR assumed to be involved in TA biosynthesis pathway represents a divergent metabolic role for plant HPR remains unclear. The 3-dimensional structure of the *C. blumei* HPR has recently been determined (Janiak, Petersen et al. 2010). The reaction mechanism of *C. blumei* HPR and its related structural basis have been discussed in detail. As such, a comprehensive structural analysis of the identified 2KGR homolog may enable us to gain more insight into its in-planta function in grapevine.

1.10 Conclusion and aims

L-tartaric acid is a naturally occurring organic acid of enormous economic value. It has a broad application in the food, wine and pharmaceutical industries (Singh 1976, Freddi, Allara et al. 1996, Kassaian 2003). TA is commonly added during the

wine-making process to improve the organoleptic property of wine and also to reduce the risks of oxidation and microbial spoilage (Jackson 2000). Unlike other common organic acids found in plants, TA only accumulates to a significant level in a limited number of families such as *Vitaceae*, *Geraniaceae* and *Leguminoceae* plants (Stafford 1959). In grapevine, TA is rapidly synthesized during the early stage of berry development and remains metabolically stable throughout the berry ripening process (Saito and Kasai 1968, Iland and Coombe 1988, Coombe and McCarthy 2000, Sweetman, Wong et al. 2012). The biological explanation of why grapevine accumulates so much TA remains unclear, with some suggesting that it may be due to the co-evolution with birds or other aggressors, whereby the presence of TA prevents young grape berries from early predation (van der Pijl 1982, Hardie 2000). No primary metabolic function has been identified for TA. Earlier radioisotope labeling studies showed that TA in grapevine is produced from ascorbic acid through the cleavage at either C4/C5 or C2/C3 (Loewus 1999). Therefore, the TA pathway may be used by some plants as an alternative degradation pathway for ascorbic acid (Loewus 1999, Green and Fry 2005, DeBolt, Melino et al. 2007).

The metabolic intermediates in the TA pathway have also been well-characterized. However, the underlying genetic basis is still poorly understood. LIDH is the only enzyme that has been characterized so far to be involved in the TA synthesis pathway (DeBolt, Cook et al. 2006). Despite its close homology with plant SDHs, LIDH displays a distinct enzyme kinetic profile from common plant SDHs, which leads to the speculation that LIDH has acquired a novel function due to the amino acid changes between proteins encoded by paralogous genes (DeBolt, Cook et al. 2006). The evolutionary origin of LIDH needs to be investigated to better understand how and why TA is synthesized in grapevine. In addition to LIDH, candidate enzymes responsible for the other steps of TA pathway have been deduced based on the types of putative reactions for each step. For example, a ketogluconate reductase, a transketolase and a semialdehyde dehydrogenase have been suggested to potentially catalyze the step 2, 4 and 5 (Figure 3) reactions in the primary TA synthesis pathway in grapevine (Salusjarvi, Povelainen et al. 2004), although none of these enzymes

have been characterized so far. The development of new sequencing technology has facilitated the elaboration of more and more plant genomes. Comparative genomic studies highlight the importance of genetic variations between homologous genes from different species, which have been shown to contribute significantly to the great phenotypic diversity among higher plants (Rensing, Lang et al. 2008, Flagel and Wendel 2009). The unequal distribution of TA in different plants makes it reasonable to speculate that the TA biosynthesis pathway may involve homologous enzymes that have multiple functions in TA-accumulating species and non-TA-accumulating species. Many functional enzymes are able to perform multiple functions due to their broad substrate specificity. For example, SDH in *A. thaliana* has been shown to be involved in both sorbitol and ribitol metabolism pathways (Nosarzewski, Downie et al. 2012). HPRs from *A. thaliana* (Timm, Nunes-Nesi et al. 2008, Timm, Florian et al. 2011) and *C. blumei* (Kim, Janiak et al. 2004) participate in the photorespiration pathway and the rosmaric acid biosynthesis pathway. A shift in substrate specificity due to the accumulation of beneficial genetic mutations may also lead to a divergence of the previous function of homologous genes (Zhang 2003). As such, enzyme kinetic characterization and protein structural analyses including the determination of the 3-dimensional structure are critical to understand the *in vivo* functions of candidate enzymes. The present study attempts to improve our understanding of TA biosynthesis at the molecular level, particularly from the evolutionary and protein structural perspectives. The specific aims of this study include:

- To determine the evolutionary origin of the grapevine LIDH and its phylogenetic relationship with the plant SDH family;
- To identify and validate the critical amino acid residues that may be responsible for the unique enzymatic activities of LIDH;
- To identify and characterize the candidate enzyme responsible for the second step of the primary TA biosynthesis pathway in grapevine, at both the enzyme kinetic level and the protein structural level;
- To identify and characterize the candidate enzyme responsible for the last

step of the TA biosynthesis pathway in grapevine.

Chapter 2 New insights into the evolutionary history of plant sorbitol dehydrogenase

Yong Jia, Darren CJ Wong, Crystal Sweetman, John B Bruning, Christopher M Ford. New Insights into the Evolutionary History of Plant Sorbitol Dehydrogenase. *BMC Plant Biology*. 2015;**15**(1):101.

Please see Manuscript attached as PDF file to the end of this document

Chapter 3 Amino acid substitutions at catalytic site of plant sorbitol dehydrogenase lead to a shift in substrate specificity

3.1 Introduction

Sorbitol dehydrogenase (SDH, EC. 1.1.1.14) belongs to the NAD(H)-dependant medium-chain dehydrogenase/reductase (MDR) superfamily and shares a distant homology with the classical alcohol dehydrogenase (ADH, EC 1.1.1.1) (Jornvall, Persson et al. 1981; Nordling, Jornvall et al. 2002; Persson, Hedlund et al. 2008). The crystal structures of human and sheep liver SDHs have been determined with and without NADH ligand respectively. Native SDHs are homotetramers (Jeffery, Cummins et al. 1981; Johansson, El-Ahmad et al. 2001; Pauly, Ekstrom et al. 2003; Yennawar, Moller et al. 2011) with each subunit containing one catalytic zinc atom that is required for catalysis (Jeffery, Chesters et al. 1984; Maret 1997; Pauly, Ekstrom et al. 2003; Yennawar, Moller et al. 2011). A monomer SDH comprises a GroES-like fold at the N-terminus and a coenzyme-binding domain conforming to a Rossmann fold at the C-terminus. SDH catalyses the reversible oxidation of D-sorbitol to D-fructose at C2. The reaction follows a Theorell-Chance compulsory order mechanism with coenzyme binding first and then the second substrate, similar as those observed for ADH (Lindstad, Hermansen et al. 1992). The C1 and C2 oxygen atom of sorbitol are positioned toward the catalytic zinc atom which is tetra-coordinated by highly conserved Cys, His, Glu and a water molecule. During the reaction, the water molecule acts as a general base and abstracts the proton from the C2 hydroxyl, thereby creating an electron flow to NAD^+. In the end, the C2 hydroxyl is oxidized to form ketone carbonyl with NAD^+ reduced to NADH (Pauly, Ekstrom et al. 2003). The dissociation of NADH from the enzyme-coenzyme-product complex has been determined as the rate-limiting factor (Lindstad, Hermansen et al. 1992). In addition to sorbitol, alternative substrates such as ribitol and xylitol could also be oxidized at a lower rate. These substrate molecules share a common d-cis-2,4-dihydroxyl (2S,4R) configuration.

NAD(H)-dependant SDH is an essential enzyme for the normal growth of higher plants. Since the first report on the characterization of SDH from the apple tissue

cultures (Negm and Loescher 1979), continuous efforts have been made to explore the biological functions of SDH. Most of these studies were dedicated to a few *Rosaceae* fruit species including apple (Yamaguchi, Kanayama et al. 1994, Yamada, Oura et al. 1998, Yamada, Mori et al. 1999, Bantog, Yamada et al. 2000, Park, Song et al. 2002, Nosarszewski, Clements et al. 2004, Nosarzewski and Archbold 2007, Wang, Xu et al. 2009, Wu, Li et al. 2010), pear (Oura, Yamada et al. 2000, Iida, Bantog et al. 2004, Ito, Hayama et al. 2005, Kim, Ahn et al. 2007), loquat (Bantog, Shiratake et al. 1999, Bantog, Yamada et al. 2000), peach (Yamada, Niwa et al. 2001, Sun, Ma et al. 2013, Matías D Hartman 2014), plum (Guo, Pan et al. 2012), in which sorbitol is utilized as the primary photosynthetic product and translocating carbohydrate (Loescher 1987). Sorbitol transported to sink tissues such as ripening fruits is oxidized to fructose by NAD-SDH (Loescher, Marlow et al. 1982, Beruter 1985, Yamaki 1986, Yamaguchi, Kanayama et al. 1996, Bantog, Shiratake et al. 1999). This SDH-mediated process facilitates sorbitol translocation among different organs and plays a critical role in sink-strength regulation (Nosarzewski and Archbold 2007). In addition to those *Rosaceae* plants, SDH has also been reported in other plants where sorbitol is not the main translocated carbon form. Recent characterization of SDH in *Arabidopsis thaliana* reveals a role in abiotic stress resistance (Nosarzewski, Downie et al. 2012, Maria, Diego et al. 2013), relating to its ability in regulating the levels of sugar alcohols which acts as important osmolytes under stress conditions (Yancey, Clark et al. 1982). In some herbaceous plants such as soybean (Kuo, Doehlert et al. 1990) and maize (Doehlert 1987, de Sousa, Paniago et al. 2008), SDH activity has also been shown to be responsible for the conversion of sucrose into sorbitol in germinating seeds using fructose as an intermediate. In addition, with NAD(H) as coenzyme, the reversible reaction catalysed by SDH also helps to maintain a balance of cell redox environment (Matías D Hartman 2014), which may be true to all plant organisms.

Members of the MDR superfamily have a long-standing role in evolutionary studies. The classical ADH family has been a choice of model enzyme to study functional evolution (Jornvall, Danielsson et al. 1993, Strommer 2011). At least 7 separate classes of ADH derived from gene duplication have been established based on

sequence homology, enzyme catalytic properties and gene expression patterns (Danielsson, Atrian et al. 1994; Duester, Farres et al. 1999; Jornvall, Nordling et al. 2003). The distribution of SDH genes in higher plants exhibits a species-specific pattern similar as those observed for ADH gene family (Strommer 2011; Jia, Wong et al. 2015). Most recently, we performed comprehensive phylogenetic and evolutionary dynamics analyses on plant SDH family and found that core Eudicot SDH could be divided into two distinct classes: Class I and Class II (Jia, Wong et al. 2015) (chapter 2 of this thesis). While Class I SDH genes are found to be conserved in all core Eudicot plants, only a few species retain Class II SDH genes (Jia, Wong et al. 2015). They differ with each other at both the amino acid sequence and gene transcription levels. Most of the currently characterized SDHs, including *Malus x domestica* MdSDH2-4 (Park, Song et al. 2002) and SDH2-9 (Nosarszewski, Clements et al. 2004), *Pyrus pyrifolia* PpySDH1-4 (Ito, Hayama et al. 2005), *Prunus salicina* SDH(PsSDH) (Guo, Pan et al. 2012), *Lycopersicon esculentum* LeSDH (Ohta, Moriguchi et al. 2005), *Prunus persica* PpeGolDHase (Matías D Hartman 2014), *A. thaliana* SDH (AtSDH) (Nosarzewski, Downie et al. 2012; Maria, Diego et al. 2013; M. Francisca Aguayo 2015), and *Fragaria ananassa* NAD-SDH (FaSDH) (Sutsawat, Yamada et al. 2008), belong to SDH Class I.

It appears that, in most cases, the currently recognized biological functions for plant SDH tend to be associated with SDH Class I instead of SDH Class II. At the moment, only a few Class II SDHs (*M. domestica* SDH1 (Yamada, Oura et al. 1998) and MdSDH1 (Park, Song et al. 2002) (SDH1 and MdSDH1 are two different proteins), *P. pyrifolia* PpySDH5 (Ito, Hayama et al. 2005), *E. japonica* NAD-SDH (EjSDH) (Bantog, Shiratake et al. 1999)) have been identified. Transcriptional analyses in apple (Park, Song et al. 2002; Nosarzewski and Archbold 2007; Wu, Li et al. 2010), pear (Ito, Hayama et al. 2005) and microarray data in grapevine, poplar and citrus (Jia, Wong et al. 2015) showed that Class I and Class II SDHs are under independent transcriptional regulation. The biological function of Class II SDH and the reason why only some plants preserve this class of SDH gene remain unclear. Earlier studies on the biosynthesis of tartaric acid in grapevine have shed some light on this novel

class of plant SDH. The previously characterized L-idonate-5-dehydrogenase (LIDH, EC 1.1.1.264) involved in tartaric acid production in grapevine was found to be a Class II SDH, suggesting that Class II SDH may have acquired a divergent function (Jia, Wong et al. 2015). LIDH catalyses the reversible oxidation of L-idonate into 5-keto-D-gluconate (5KGA), the proposed rate-limiting step in the tartaric acid biosynthesis pathway in grapevine (DeBolt, Cook et al. 2006). Preliminary analyses based on the distribution of Class II SDH genes and the occurrence of tartaric acid in corresponding plants suggest that Class II SDH genes in other species such as *Solanum tuberosum* and *Pelargonium x hortorum* may also be involved in tartaric acid production (Jia, Wong et al. 2015).

The substrate specificity of the purified form SDH from *P. pyrifolia* (Oura, Yamada et al. 2000) and *M. domestica* (Yamaguchi, Kanayama et al. 1994) and the recombinant AtSDH (Maria, Diego et al. 2013; M. Francisca Aguayo 2015), PsSDH (Guo, Pan et al. 2012), LeSDH (Ohta, Moriguchi et al. 2005) has been extensively analysed. Similar with their animal counterparts, plant SDHs display broad substrate specificity. All plant SDHs exhibit the highest relative activity with sorbitol (100%) while 5-carbon xylitol and ribitol are oxidized at lower efficiency (80% and 98% for AtSDH (Maria, Diego et al. 2013), 62% and 15% for a PsSDH (Guo, Pan et al. 2012), 29% and 60% for LeSDH (Ohta, Moriguchi et al. 2005), 77% and 14% for purified *P. pyrifolia* SDH (Oura, Yamada et al. 2000), 40% and 7% for purified *M. domestica* SDH (Yamaguchi, Kanayama et al. 1994) respectively). Of these, AtSDH, PsSDH and LeSDH belong to SDH Class I while the purified SDHs isolated from plant tissues are unknown. Of note is that grapevine LIDH, which has been recognized as a Class II SDH, displays principle activity against L-idonate and has a low reaction rate on sorbitol oxidation (DeBolt, Cook et al. 2006), suggesting divergent enzymatic properties from the common plant SDHs. Of the few Class II SDHs identified so far, a recombinant *M. domestica* SDH for which kinetic data is available also demonstrates an extremely weak affinity toward sorbitol (K_m 247mM) (Yamada, Oura et al. 1998). While our knowledge of Class II SDHs is still very limited, current observation indicates that Class II SDHs may have a distinct enzymatic profile from Class I SDHs,

supporting a divergent biological function for this class of SDH. In regards to the unique substrate specificity of LIDH, it has been speculated that it was due to small changes in amino acid sequence encoded by paralogous genes (DeBolt, Cook et al. 2006). Recently, we performed natural selection tests on plant SDH family and identified six amino acids that have been under positive natural selection in Class II SDH (Jia, Wong et al. 2015) (Chapter 2 of this thesis). Protein structural modelling analyses of SDHs from *V. vinifera* revealed three amino acid substitutions between Class I and Class II SDHs at the catalytic site, which are speculated to change the substrate specificity of SDH (Jia, Wong et al. 2015).

In this study, we aim to investigate the enzymatic properties of plant SDH and validate the amino acid residues that have been assumed to affect its substrate specificity. SDH genes from *V. vinifera*, *Solanum tuberosum*, *Theobroma cacao* were cloned and site-mutated. Recombinant protein of Class I and Class II SDHs were expressed in vivo using the *E. coli* expression system. Substrate specificities of both wild-type and mutant SDHs were analysed. In addition, homology models of Class I and Class II SDHs from *V. vinifera* were created. Ligand docking was performed to investigate their predicted kinetic performance toward different substrates. Enzymatic characterization of plant SDHs will enable us to better understand their biological functions, especially those associated with SDH Class II.

3.2 Methods

3.2.1 Chemicals

Polyol substrates, 5KGA (K^+ salt) and coenzymes (NADH, NAD^+, NADPH, $NADP^+$) were purchased from Sigma, BDH, Merck and Gold Biotechnology and were of analytical grade or higher. Glycine, Sodium hydroxide, tris-base, acetic acid were from Sigma. L-idonate (Na^+ salt) was obtained from Kazumi Saito (Kyoto University, Kyoto, Japan). Substrates and coenzyme stocks were prepared in deionized water to different concentrations and were diluted as needed for enzyme assay.

3.2.2 Sequence alignment and phylogenetic analyses

The amino acid sequences of previously characterized plant SDHs and mammalian SDHs, including *A. thaliana* AtSDH (Maria, Diego et al. 2013), *M. domestica* MdSDH1 and MdSDH3 (Park, Song et al. 2002), *V. vinifera* LIDH (DeBolt, Cook et al. 2006), *P. persica* PpeGolDHase (Matías D Hartman 2014), *P. salicina* PsSDH (Guo, Pan et al. 2012), *L. esculentum* LeSDH (Ohta, Moriguchi et al. 2005), *F. ananassa* FaSDH (Sutsawat, Yamada et al. 2008), human and sheep liver SDHs, were retrieved from Uniprot database (http://www.uniprot.org/). The AtSDH sequence was used as a query to BLASTp against the the Phytozome 10.3 database (http://phytozome.jgi.doe.gov/pz/portal.html#!search) to obtain the cDNA and amino acid sequences of homologous SDHs from *V. vinifera*, *S. tuberosum*, *T. cacao*. Amino acid sequence alignment was performed using ClustalW2 (Larkin, Blackshields et al. 2007) and was further modified to remove the upstream non-catalytic region. Bayesian phylogeny was searched on a single Markov chain for 12,750,000 generations in BEAST2.0 (Bouckaert, Heled et al. 2014), using the Yule tree prior and saving every 1000th tree. JTT+G substitution model (Jones, Taylor et al. 1992) was used as suggested by the "Find Best DNA/Protein Models" in MEGA6.0 (Tamura, Stecher et al. 2013). At convergence, a consensus tree was built by discarding the first 10% trees. Posterior Bayesian probabilities were calculated and displayed above each branch. The final phylogeny was annotated using FigTree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). Alignment of the active site residues as identified by Jia (Jia, Wong et al. 2015) was performed using the ESPript3.0 tool (http://espript.ibcp.fr/ESPript/ESPript/).

3.2.3 Molecular Cloning and Transformation

The Class I SDH gene (Gene ID: GSVIVT01010642001, cloned into the pET14b vector (Novagen, Germany)) from *V. vinifera* cv Pinot Noir was kindly provided by Emily Higginson, University of Adelaide. The pET14b-LIDH construct was generated through previous work in our lab. Synthetic oligoucleotides encoding the SDH genes from *T. cacao* (Tc03_g019270, Tc03_g019280) and *S. tuberosum* (PGSC0003DMT400081907) were synthesized using the Gene Art service at Life

Technology (ThermoFisher Scientific, USA). Genetic codes of these genes were optimized for protein expression in the *E. coli* system using the GeneOptimizer codon optimization tool at the Life Technology website. Synthesized oligonucleotides containing the BamHI and NdelI restriction sites were delivered in pMK vector (GeneArt, ThermoFisher Scientific, USA; kanamycin resistance) and were subsequently digested and ligated into pET14b vector (Novagen, Germany; ampicillin resistance). Plasmid propagation was carried out using the *E. coli* DH5α strain (Invitrogen, ThermoFisher Scientific, USA). Successful transformant colonies were selected on LB agar plates containing the desired antibiotics depending on the specific vectors being used. Plasmid extraction was performed using ISOLATE II Plasmid Mini Kit (Bioline, UK). Sequences of the constructs were verified by sequencing before each transformation step and were finally transformed into *E. coli* BL21(DE3) pLysS cells (Invitrogen, ThermoFisher Scientific, USA) for recombinant protein production.

3.2.4 Site-directed mutagenesis

Site-directed mutagenesis was carried out using the Q5® Site-directed Mutagenesis Kit (New England BioLabs). The sequence-verified plasmids (pET14 vector harbouring the wild type SDH genes) were used the templates for site mutation PCR. Mutagenic primers containing mutated nucleotides encoding the target amino acid residues were designed using the NEBaseChanger online tool at http://nebasechanger.neb.com/ and synthesized at AGRF (Adelaide, Australia). PCR reaction was setup to a total volume of 12.5 ul and was performed in a MyCycler™ Thermal Cycler (BioRad, USA). For mutagenesis at two locations (H42F-Y43L and G112S-A113T), mutations were performed in a step-by-step manner. All mutated sites were confirmed by DNA sequencing. The final mutant constructs were transformed into *E. coli* BL21(DE3) pLysS for recombinant protein production.

3.2.5 Recombinant protein purification

The production of the recombinant proteins was carried out in 5 L flasks. One liter of LB medium inoculated with *E. coli* cells harboring the desired plasmids were

incubated at 37 ℃ with shaking of 180 rpm until the OD_{600} reached about 0.6. Protein expression was initiated with the addition of IPTG to a final concentration of 0.5 mM. Cells were continuously grown overnight at 16 ℃ and harvested by centrifugation (20 min, 4000 rpm, 4 ℃). The separated cell pellet was then suspended in 50 ml suspension buffer (20 mM Tris-HCl pH8.0, 50 mM NaCl, 1 mM DTT) and passed through a Microfluidic cell disrupter. The cell lysate was centrifuged at 40, 000 x g, 4 ℃ for 25 mins. The clarified lysate was applied to a "His Trap FF Crude, 5 x 1 ml" column (GE Healthcare Life Sciences) using the NGC™ Chromatography System (Bio-Rad), and eluted with a linear imidazole gradient (10 mM - 250 mM). The fractions containing the target protein were collected and combined. The purity of the protein sample was checked by SDS–PAGE, stained by Coomassie blue. The enzyme concentration was determined using a NanoDrop™ spectrophotometer (Thermo Fisher Scientific). 10% of glycerol was added before the protein sample was flash-frozen in liquid nitrogen and stored at -80 ℃ for kinetic tests.

3.2.6 Enzyme assay

The activity of recombinant SDH was tested by monitoring the rate of changes in absorbance at 340nm for NAD^+ reduction at room temperature. The test was performed using the Multiscan™ spectrum (Thermo Fisher Scientific) in Cellstar® 96 well plates. Each reaction was carried out in 3 replicates. The optimal pH condition for the forward reaction was determined using a range of buffers (100 mM Tris-Acetate pH 5.0-5.5-6.0-6.5, 100 mM Tris-HCl pH 7.0-7.5-8.0-8.5-9.0, 100 mM NaOH-Glycine pH 9.5-10.0-10.5-11.0-11.5) with 20 mM D-sorbitol, 1 mM NAD^+. For the substrate specificity analyses, enzyme activity was examined at the determined optimal pH condition. Each reaction contains 10 ul enzyme sample, 1 mM NAD^+ and 20 mM substrate in the total volume of 200 ul. To calculate the K_m and V_{max} values, enzyme activity was examined with varying substrate concentrations in triplicates. The initial rates were fitted into the Michaelis–Menten equation using GraphPad Prism (version 6.0; GraphPad Software, Inc. CA, USA) software.

3.3 Results and discussion

3.3.1 Bioinformatic analyses

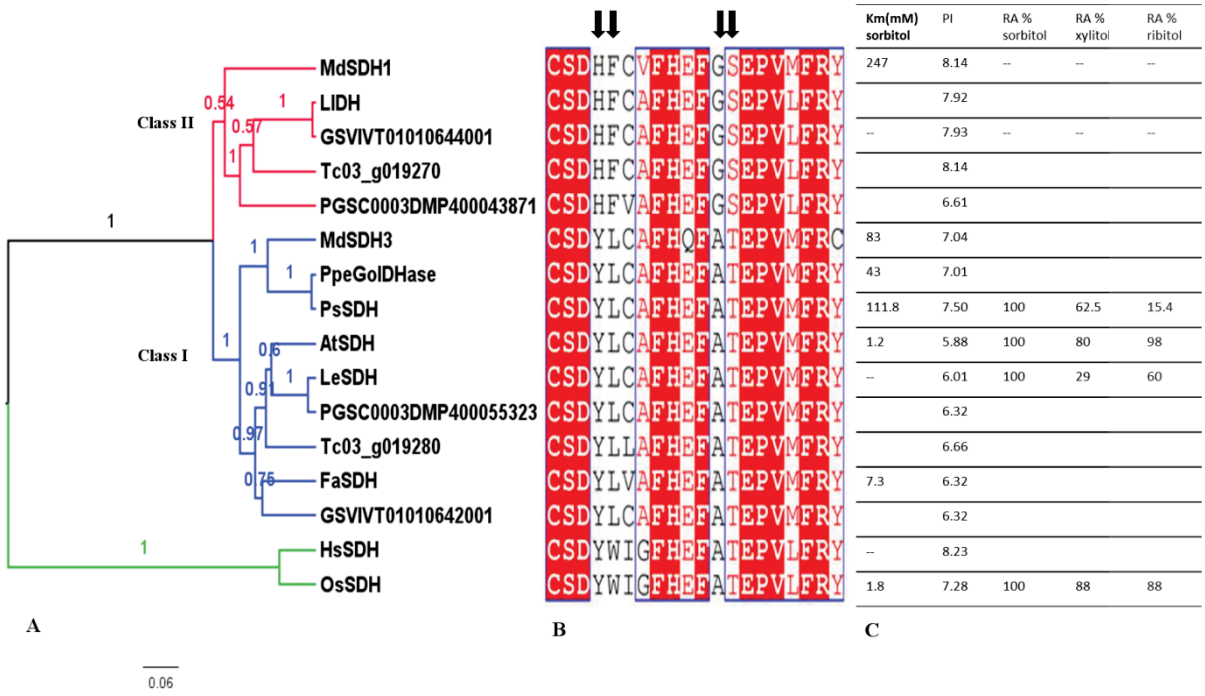
Recent evolutionary characterization of plant SDH family indicated that core eudicot SDHs could be classified into two distinct phylogeny groups (Class I and Class II) which may have divergent enzyme kinetic profiles (Jia, Wong et al. 2015). In order to validate this hypothesis, a Bayseian phylogeny of previously characterized plant SDHs and SDH homologs from *V. vinifera*, *T. cacao* and *S. tuberosum* was developed (**Figure 1A**). Three (Gene ID: GSVIVT01010642001, GSVIVT01010644001, GSVIVT01010646001/LIDH), two (Tc03_g019270, Tc03_g019280) and two (PGSC0003DMT400065063, PGSC0003DMT400081907) putative SDHs were identified from *V. vinifera*, *T. cacao* and *S. tuberosum*, respectively. These three species were selected due to their retentions of both classes of SDHs (Jia, Wong et al. 2015) and accumulation of TA (Stafford 1959). Mammal SDH orthologs from human (HsSDH) and rat (OsSDH) were included as an out-group for the phylogeny analysis. The final phylogeny was rooted on mammalian SDHs. As shown in **Figure 1A**, the target plant SDHs separated into two clades at the basal node with strong support, corresponding to SDH Class I and Class II, respectively. Within both Class I and Class II clades, Rosaceae SDHs formed separate branches from other SDHs, consistent with previous phylogenetic analyses (Jia, Wong et al. 2015). Of these SDHs, three (MdSDH3, PpeGolDHase, PsSDH) belongs to Class I, and one (MdSDH1) was classified as Class II. In addition to *Rosaceae* SDHs, three SDHs from *V. vinifera*, *T. cacao* and *S. tuberosum* and previously determined SDHs from *A. thaliana* (Maria, Diego et al. 2013), *L. esculentum* (Ohta, Moriguchi et al. 2005), *F. ananassa* (Sutsawat, Yamada et al. 2008) were grouped into Class I while the rest fall into Class II.

As shown in **Figure 1C**, the enzyme kinetic data for previously characterized SDHs were retrieved from corresponding literature and displayed along the phylogenetic tree. The Class II Rosaceae MdSDH1 has a much higher K_m value (247 mM) on D-sorbitol compared to the Class I Rosaceae MdSDH3 (83 mM), PreGolDHase (43

mM), PsSDH (111.8 mM), suggesting a significant enzymatic divergence between the two classes of SDHs. This suggestion is corroborated by the report that *V. vinifera* LIDH has the highest activity with L-idonate but a low reaction rate on D-sorbitol (DeBolt, Cook et al. 2006). Within the Class I SDH clade, other SDHs including AtSDH (1.2 mM) and FaSDH (7.3 mM) displayed much higher affinity toward D-sorbitol than rosaceae SDHs, supporting a positive correlation between phylogeny pattern and enzymatic profiles (**Figure 1C**). In addition, the protein isoelectric points (PI) were also calculated based on their amino acid composition. Results showed that Class II SDHs (except PGSC0003DMT400065063) generally have PI values at around 8.0. This is clearly distinguished from that for Class I SDHs, most of which have PI values between 6.0 and 7.0 (**Figure 1C**). DeBolt (DeBolt, Cook et al. 2006) speculated that the unique enzymatic characteristic for LIDH might be due to amino acid substitution between paralogous proteins. Based on the modelled structures of SDH homologs from *V. vinifera*, we identified three amino acid substitutions at the catalytic sites of Class I and Class II SDHs (Jia, Wong et al. 2015), providing a very positive indication that the putative shift in substrate specificity for plant SDHs might be due to amino acid changes between paralogous proteins.

These catalytic residues were aligned and displayed along the phylogenetic tree in this study. As shown in **Figure 1B**, the majority of the catalytic residues were highly conserved across plant SDHs and mammalian SDHs, except 42His, 112Gly and 113Ser (in Class II SDH) which were replaced with 42Tyr, 112Ala and 113Thr in Class I SDH. In addition, the 43Phe adjacent to 42His in Class II SDHs was also uniformly substituted with 43Leu in Class I SDHs. This residue was replaced by a Trp in mammalian SDHs (**Figure 1B**). Another amino acid site that exhibits significant difference between Class I and Class II SDHs is 268Leu in Class II SDHs except MdSDH1 which has a Met at this site. This residue was replaced by a Met in all Class I SDHs (Figure 1B). However, Leu and Met have similar property (hydrophobic side chain). This amino acid site is positioned to coordinate the coenzyme NAD(H) in the modelled *V. vinifera* SDH structures (Jia, Wong et al. 2015), and therefore may not affect the substrate specificity. To validate the putative

enzymatic divergence between Class I and Class II SDHs and also the critical amino acid residues responsible for substrate specificity, homologous genes encoding SDHs from both classes, including GSVIVT01010646001/LIDH, Tc03_g019270, PGSC0003DMT400065063 (Class II) and GSVIVT01010642001, Tc03_g019280 (Class I) were cloned in this study. Site-directed mutation of the identified critical amino acid sites was also performed.



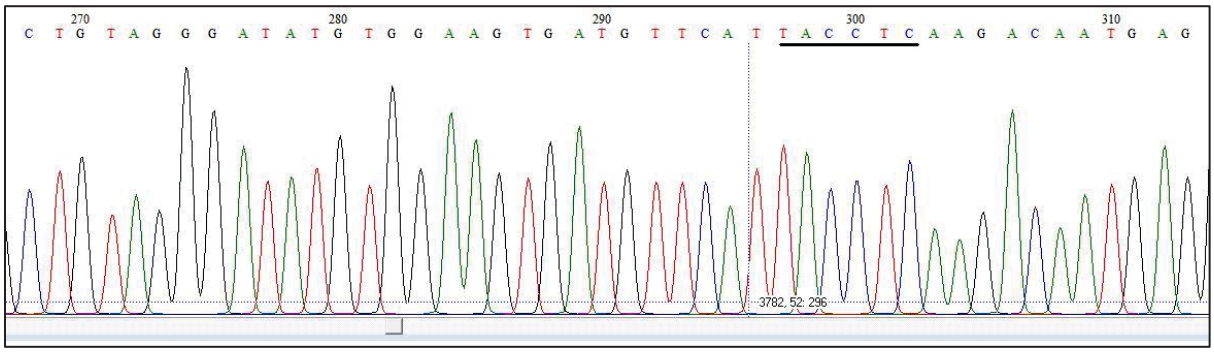
Km(mM) sorbitol	PI	RA % sorbitol	RA % xylitol	RA % ribitol
247	8.14	--	--	--
	7.92			
--	7.93	--	--	--
	8.14			
	6.61			
83	7.04			
43	7.01			
111.8	7.50	100	62.5	15.4
1.2	5.88	100	80	98
--	6.01	100	29	60
	6.32			
	6.66			
7.3	6.32			
	6.32			
--	8.23			
1.8	7.28	100	88	88

Figure 1. Phylogenetic, active residues and enzyme properties analyses of the target SDHs. **A**. A rooted Bayesian phylogeny of target SDHs. The phylogenetic tree was rooted at animal SDH. Bayesian posterior probabilities of corresponding clades are represented by values above each branch. Branch length is proportional to the number of amino acid substitutions. **B**. The alignment of the putative catalytic site residues as determined by Jia etc. Residue sites exhibit substitutions between Class I and Class II are marked with arrows. **C**. Enzyme kinetic data and protein pI value of the corresponding SDHs, either retrieved from previous studies or calculated in the present study. RA refers to relative activity.

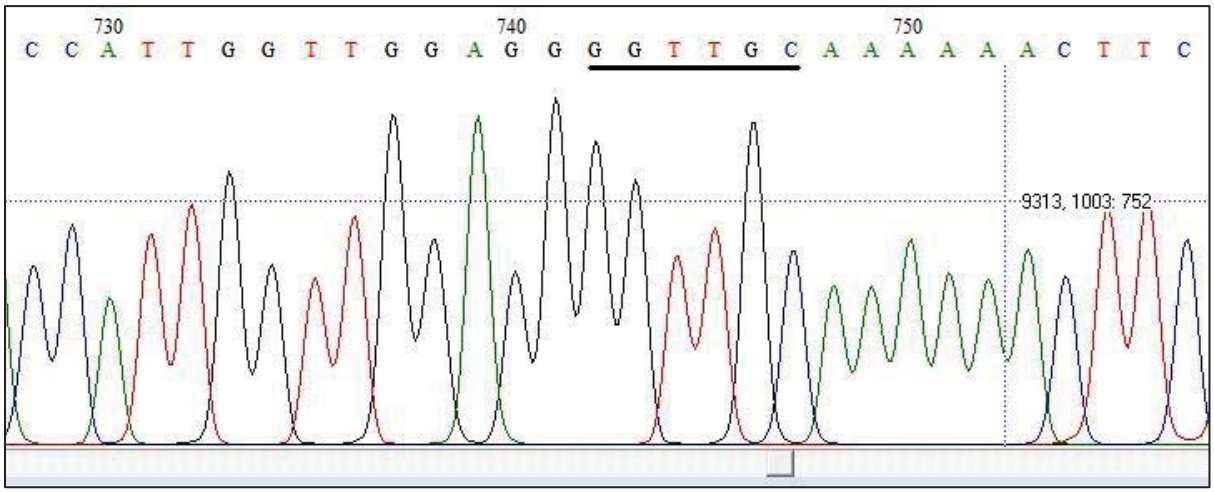
3.3.2 Molecular cloning and recombinant protein purification

The original coding DNA sequence (CDS) for Tco3_g019270, Tco3_g019280 and PGSC0003DMT400065063 were optimized for codon quality (codon usage frequency) and GC content for expression in *E. coli* system. NdeI (CATATG) and BamHI (GGATCC) restriction sites were introduced to the 5' and 3' ends of the optimized DNA sequences. The DNA fragments were successfully ligated into the pET14b vector using the above restriction sites. DNA sequencing of the extracted plasmid constructs showed that the target genes were cloned into the right frame for 6-his tagged recombinant protein production. To evaluate the effect of the identified amino acid substitutions on substrate specificity, overlapping primers (**Table 1**) were designed to mutate 42His, 43Phe, 112Gly and 113Ser (numbering in LIDH without the first 20 amino acids) of Class II SDHs including LIDH, Tco3_g019270 and PGSC0003DMT400065063 to 42Tyr, 43Leu, 112Ala and 113Thr as in their Class I counterparts. Mutation PCR was performed in a step-wise process. In the first step, 42His-43Phe of LIDH, Tco3_g019270 and PGSC0003DMT400065063were mutated to 42Tyr-43Leu, resulting in pET14b_LIDH_YL, pET14b_COC270_YL, pET14b_PtLIDH_YL constructs. The sequences of the target mutation sites were confirmed by sequencing. In the second step, the pET14b_LIDH_YL, pET14b_COC270_YL, pET14b_PtLIDH_YL constructs were used as PCR template, 112Gly-113Ser were mutated to 112Ala-113Thr, leading to the pET14b_LIDH_YLAT, pET14b_COC270_YLAT, pET14b_PtLIDH_YLAT constructs which were also sequencing confirmed (**Figure 2**). In the end, the wild type LIDH, GSVIVT01010642001, Tco3_g019270, Tco3_g019280, PGSC0003DMT400065063 constructs and their mutation forms were transformed into *E. coli* BL21(DE3) plysS cell for recombinant protein production.

VVLIDH



C T G T A G G G A T A T G T G G A A G T G A T G T T C A T ***C A C T T C*** A A G A C A A T G A G (H F - Y L)

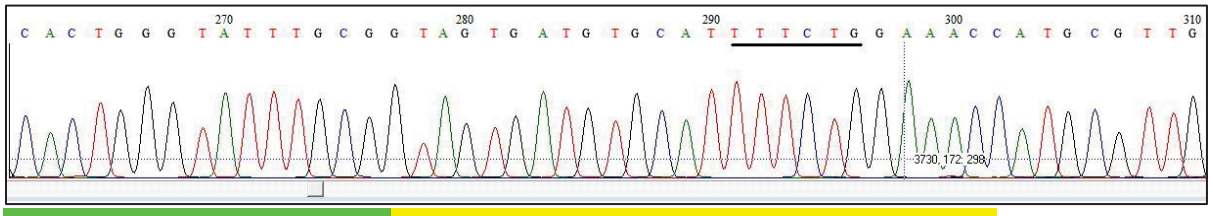


G A A G T T T T T T ***G G A T C T*** C C T C C A A C C A A T G G (G S - A T)

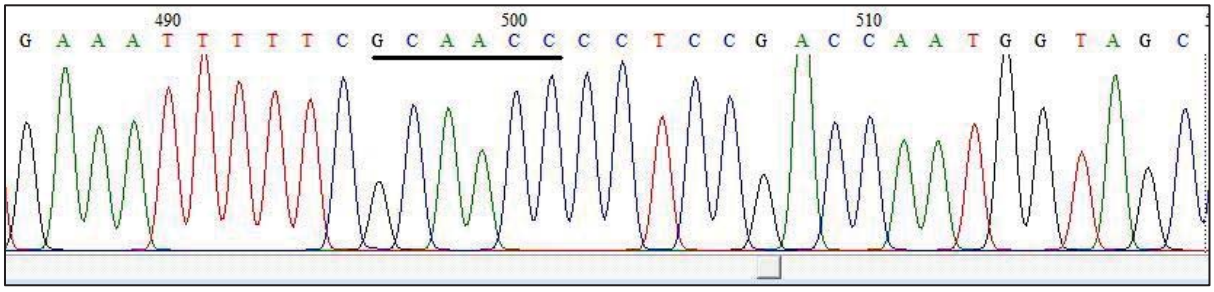
Reverse & complement:

C C A T T G G T T G G A G G ***A G A T C C*** A A A A A A C T T C

Tco3_g019270（COC270）



C A C T G G G T A T T T G C G G T A G T G A T G T G C A T ***C A T T T T*** G A A A C C A T G C G T T G (H F - Y L)



G A A A T T T T T C ***G G T A G C*** C C T C C G A C C A A T G G T A G C (G S - A T)

PGSC0003DMT400065063 (PTLIDH)

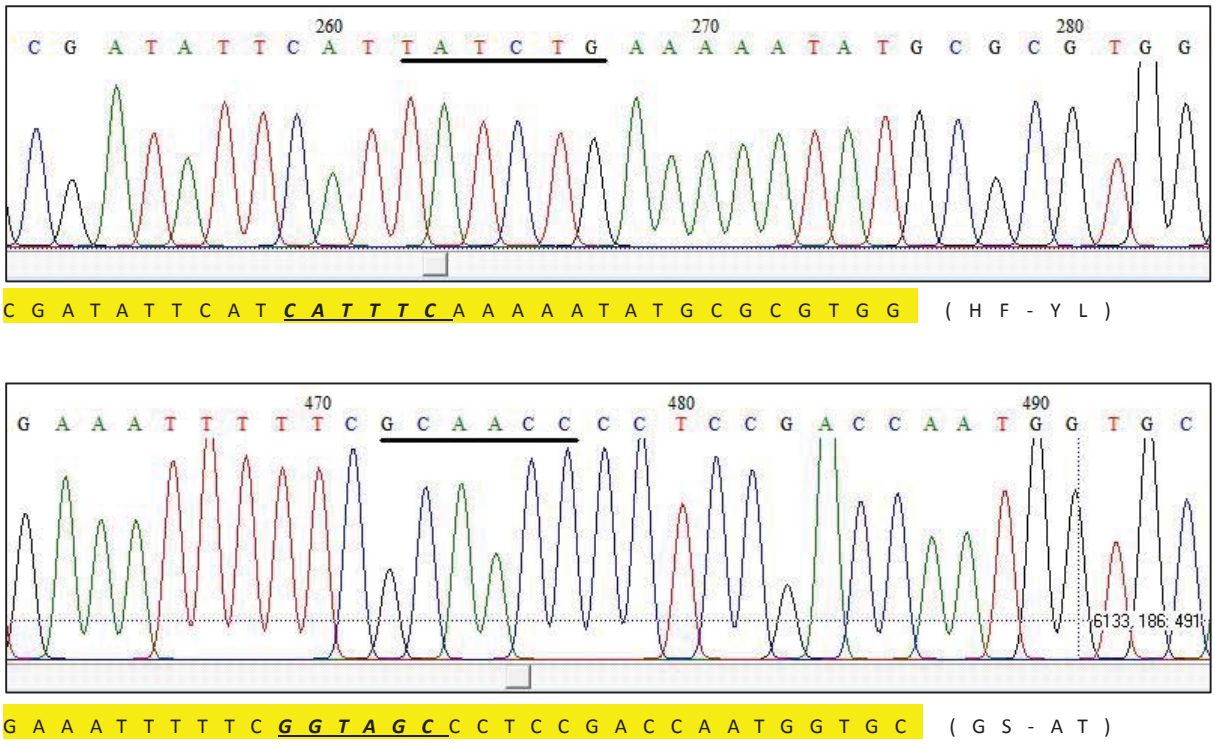


Figure 2. **Sequencing confirmation of the target mutation sites of VVLIDH, Tco3_g019270(COC270) and PGSC0003DMT400065063 (PTLIDH)**. Sequences in shade refer to the overlapping regions of the wild-type genes. Target mutation sites were underlined accordingly.

One litre of LB culture was inoculated for each construct transformant for recombinant protein production. The supernatant solution of the disrupted cells after centrifugation was loaded onto nickel columns. The 6his-tagged protein eluted out when the concentration of imidazole reached around 50 mM (data not shown). The elution fractions containing the target protein were combined and checked for enzyme purity on the SDS-PAGE gel. As shown in **Figure 3**, recombinant protein for VVLIDH and Tco3_g019280(COC280) were purified to high homogeneity. The molecular mass was estimated to be around 40 kDa for both, which is very similar to the calculated molecular mass 37.5 kDa and 39.1 kDa respectively. After the addition of 10% glycerol for long-term storage, the concentration of the enzyme samples were measured to be at 2.9 mg/ml and 3.4 mg/ml respectively. There concentrations were used for enzyme kinetic calculation in subsequent enzymatic assays. No biologically

active product could be obtained for the other candidate enzymes. Protein induction experiments were repeated using the corresponding transformed strains. Same results were obtained. The cell lysis pellet was also checked on a SDS-PAGE gel to investigate whether the target recombinant proteins are insoluble and occur in the inclusion body. No obvious target band could be detected at the predicted position of the SDS-PAGE gel. These results indicated that the production of the recombinant proteins was not initiated at the used culturing condition. The strains harbouring the target plasmid constructs need to be checked for the presence of the desired vectors. And the sequences of these vector should also be validated to make sure that the gene inserts are in the right frame with the 6HIS-tag.

Table 1. **Summary of the site-directed mutation primers.** Primers were designed to overlap the target mutation sites. The desired mutations were introduced in the forward primers, as indicated by the lower-case letters. COC270 and PtLIDH stand for genes Tco3_g019270 and PGSC0003DMT400065063 respectively. "Ta" refers to "annealing temperature" used for site-directed mutagenesis PCR.

Gene ID	Mutation	Overlaping region	Forward primer	Reverse primer	Ta (℃)
VVLIDH	HF/YL	TGATGTTCATCACTTCAAGACAATGAGCTGTAGGGATATGTGGAAG	TGATGTTCATtacctcAAGACAATGAG	CTTCCACATATCCCTACAG	60
	GS/AT	GAAGTTTTTTGGATCTCCTCCAACCAATGGCAATACAATCTATGCAG AGAAAT	GAAGTTTTTTgcaaccCCTCCAACCAATGG	ATTTCTCTGCATAGATTGTATTG	59
COC270	HF/YL	TGATGTGCATCATTTTGAAACCATGCGTTGCACTGGGTATTTGCGGT AG	TGATGTGCATtttctgGAAACCATGCGTTG	CTACCGCAAATACCCAGTG	62
	GS/AT	GAAATTTTTCGGTAGCCCTCCGACCAATGGTAGCGTTATAATCTGTG TCCGGAAAT	GAAATTTTTCgcaaccCCTCCGACCAATGGTAGC	ATTTCCGGACACAGATTATAAC	59
PtLIDH	HF/YL	CGATATTCATCATTTCAAAAATATGCGCGTGGCGTGGGTATTTGCGG CAG	CGATATTCATtatctgAAAAATATGCGCGTGG	CTGCCGCAAATACCCACG	64
	GS/AT	GAAATTTTTCGGTAGCCCTCCGACCAATGGTGCGTTATAATCTGTGC CGCAAAAT	GAAATTTTTCgcaaccCCTCCGACCAATGGTGC	ATTTTGCGGCACAGATTATAAC	59

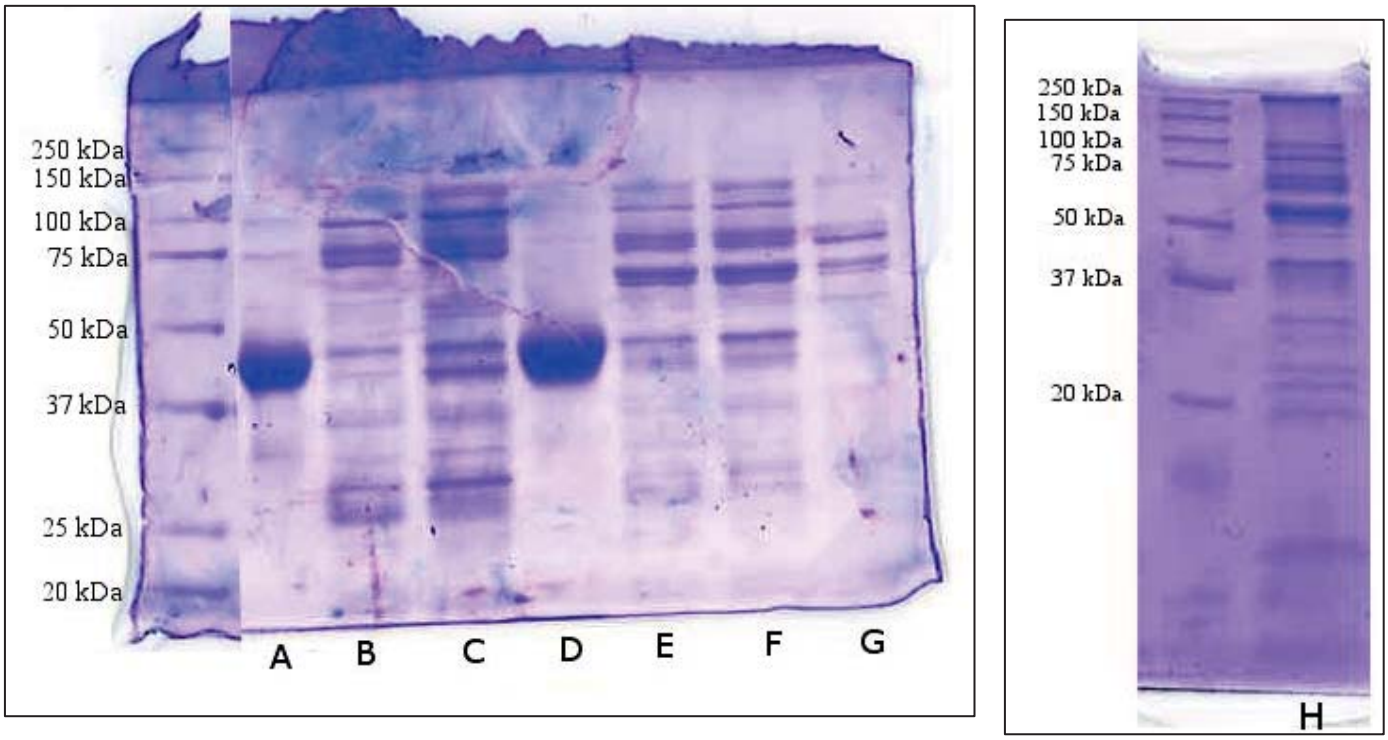


Figure 3. Purification of recombinant SDH proteins checked by SDS-PAGE. The elution fractions containing the target recombinant proteins (A-H) were loaded on SDS-PAGE. A: VVLIDH; B: VVLIDH_YLAT; C: VVSDH; D: COC280; E: COC270; F: COC270_YLAT; G: PTLIDH; H: PTLIDH_YLAT. Only A and D show significant recombinant protein products at the desired position.

3.3.3 Enzymatic substrate specificity of SDHs

The identification of *V. vinifera* LIDH represents a breakthrough in the study of TA biosynthesis pathway. Despite its significant homology with plant SDHs, LIDH has highest activity for the oxidation of L-idonate to 5KGA, the rate limiting step in primary TA biosynthesis pathway in grapevine, and only very weak activity with D-sorbitol as the substrate (DeBolt, Cook et al. 2006). In addition to D-sorbitol, plant SDHs have also been known for the capability to use other polyols such as xylitol, ribitol, iditol at different efficiency levels. Extensive information of LIDH's kinetics with these polyols would improve our understanding on the biological function of LIDH and its relationship with plant SDHs. In this study, recombinant LIDH was expressed and purified for detailed enzymatic characterization. All assays were performed using NAD^+ as coenzyme. No activity could be observed when $NADP^+$ was used (data not shown). This is consistent with previous studies which suggested that plant SDHs are NAD-dependent. In contrast to previous report by DeBolt, our results showed that LIDH still retain significant activity with D-sorbitol. Using sorbitol as the substrate, LIDH activity peaked at pH 9.0 for the forward reaction (**Figure 4A**). The relative activities of LIDH with different substrates at the optimal pH condition are shown in **Figure 4B**. Indeed, the highest activity was observed with L-idonate, which demonstrates a reaction rate over twice that for D-sorbitol. The relative activity toward xylitol and iditol stood at 73.2% and 103.6%, respectively. No significant activity could be detected for ribitol and arabitol. To determine the kinetic profile of LIDH, the initial reaction rates with L-idonate and D-sorbitol were determined at various substrate concentrations. As shown in **Figure 4C**, plotting the reaction rates against the substrate concentration revealed K_m values of 3.13 mM and 32.72 mM for L-idonate and D-sorbitol, respectively. DeBolt reported a K_m value of 2.2 mM for L-idonate at pH 8.0 (DeBolt, Cook et al. 2006), which is very similar to our results. The moderate difference may be caused by the different pH condition being used. As such, our study confirmed that *V. vinifera* LIDH has a much higher affinity toward L-idonate (K_m=3.13 mM) over D-sorbitol (K_m=32.72 mM) and could use L-idonate much more efficiently than other common substrates. Notably, our data

showed that LIDH still retains significant activity with D-sorbitol. This is consistent with its strong homology with other plant SDHs. Previous studies showed that SDH has a preference with substrates with a D-cis-2,4-dihydroxyl (2S,4R) configuration and a C1 hydroxyl group, such as D-sorbitol, xylitol and L-iditol, while it displays weak or relatively lower activity with other polyols like L-arabitol, D-mannitol (Lindstad, Koll et al. 1998, Oura, Yamada et al. 2000, Maria, Diego et al. 2013). This is consistent with our observations that LIDH has 73.2% and 103.6% activity with xylitol and L-iditiol but no activity with L-arabitiol (**Figure 4B**). LIDH in our tests displayed no activity at all with ribitol which also has a D-cis-2,4-dihydroxyl configuration but has a trans hydroxyl at C3. Additional tests on ribitol confirmed this observation (data not shown). Noteworthy, enzyme characterization of SDHs from apple (Negm and Loescher 1979) and pear (Oura, Yamada et al. 2000) also showed a relatively weak activity with ribitol, around 13% that of sorbitol. This, however, contrasts with the data obtained for *A. thaliana* SDH which has comparable activity with sorbitol and ribitol (Maria, Diego et al. 2013). Further enzyme structural analyses are needed to investigate the underlying protein structural basis for the above kinetic profiles. The K_m value of LIDH on sorbitol (32.72 mM) is relatively close to that reported for Class I rosaceae SDHs (PpeGolDHase,MdSDH3, PsSDH at 43 Mm, 83 mM and 111.8 mM, respectively), which is much higher than that for non-rosaceae Class I SDH (AtSDH, FaSDH at 1.2 mM and 7.3 mM, respectively).

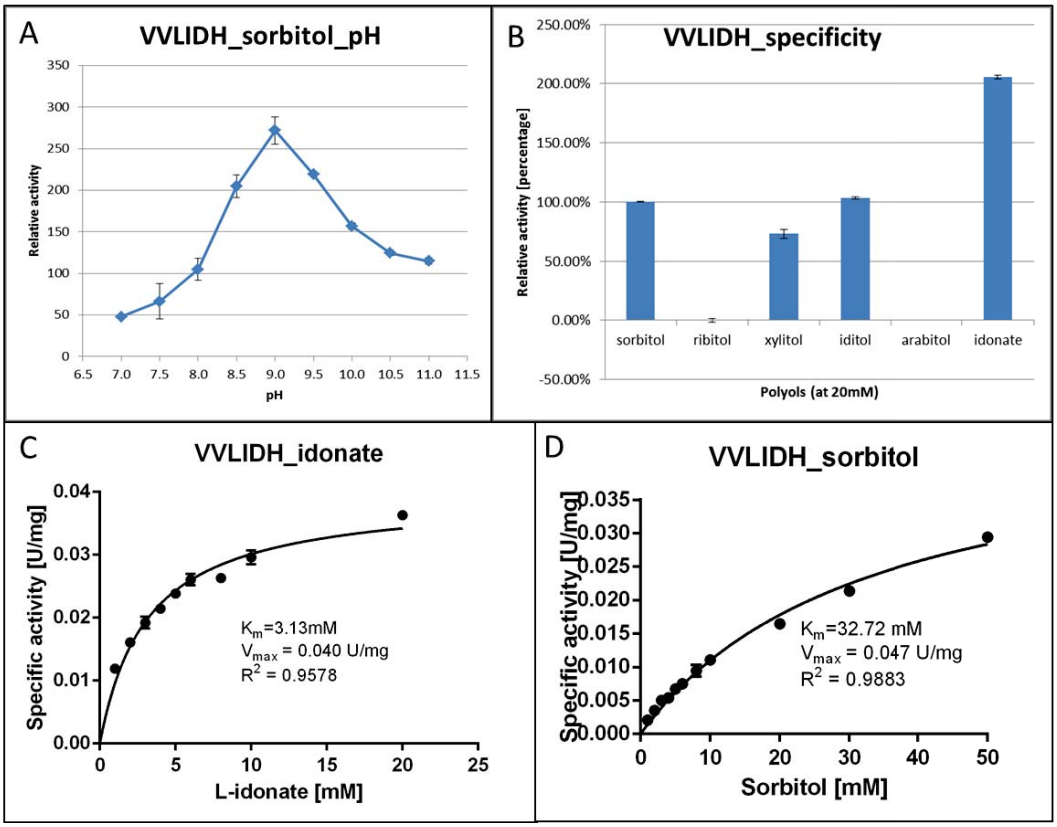


Figure 4. Enzymatic characterization of *V. vinifera* LIDH. **A**. Effects of pH on enzyme activity (data from 1 mM NAD^+ with 20 mM sorbitol); **B**. Relative enzyme substrate activity with different substrates (data from 1 mM NAD^+ with 20 mM corresponding substrates); **C**. Initial reaction rates with L-idonate; **D**. Initial reaction rates with D-sorbitol.

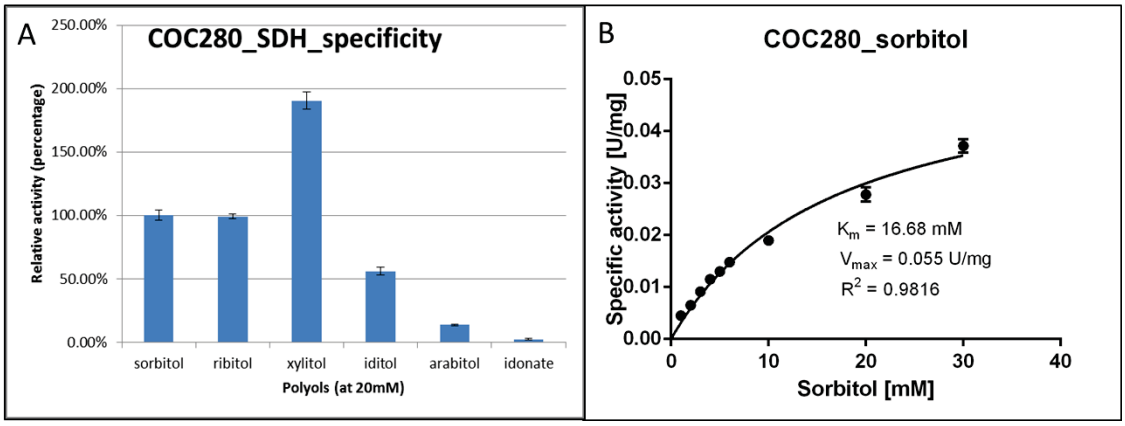


Figure 5. Enzymatic characterization of *T. cacao* Tco3_g019280 (COC280). **A**. Relative enzyme substrate activity with different substrates(data from 1 mM NAD^+ with 20 mM corresponding substrates at pH 9.0); **B**. Initial reaction rates with D-sorbitol.

The other SDH for which an enzymatic profile was characterized in this study is

Tco3_g019280 (COC280), which has been classified as a Class I SDH (**Figure 1A**). COC280 also demonstrated the maximum rate at pH 9.0 in Tris-HCl buffer. Most of the previously characterized plant SDHs have optimal pH conditions ranging from pH 9.0 to pH 9.6, such as those identified from apple (Negm and Loescher 1979, Yamaguchi, Kanayama et al. 1994), pear (Oura, Yamada et al. 2000), maize (Doehlert 1987), while the kinetics of others including apple AtSDH (Maria, Diego et al. 2013), MdSDH3 (Park, Song et al. 2002), FaSDH (Sutsawat, Yamada et al. 2008) were also measured at pH condition of 9.0 or 9.5. This is consistent with our observation with *T. cacao* Class I SDH (COC280). The only discrepant report was made with tomato LeSDH which has been shown to have maximum reaction rate at pH 10.5 (Ohta, Moriguchi et al. 2005). The enzyme activities of COC280 toward different polyols and L-idonate were determined and shown in **Figure 5A**. While COC280 also retains significant activity with sorbitol, its relative activities toward other polyols and L-idonate varied dramatically with that found with LIDH. COC280 exhibit the maximum reaction rate with xylitol (190.43%) in comparison with sorbitol. It has comparable activity with ribitol (99.22%). This is in great contrast with LIDH which was shown to be inactive on ribitol. In addition, COC280 also retains significant activity with L-iditol and relatively weak activity with D-arabitol (13.8%). This is consistent with previous reports that SDH prefers to utilize polyols with a D-cis-2,4-dihydroxyl (2S,4R) configuration. The substrate specificity profile of COC280 is most similar to that reported for AtSDH which also has comparable activity with sorbitol and ribitol. However, unlike COC280, AtSDH has around 80% and 60% activity with xylitol and L-arabitol, respectively (Maria, Diego et al. 2013). Most strikingly, COC280 in our tests showed only slight activity with L-idonate (2.47%), in great contrast with LIDH which has the highest activity with L-idonate. To our knowledge, this is the first report on a Class I SDH's activity with L-idonate. Therefore, our data for LIDH and COC280 supports a prominent divergence of L-idonate activity between Class I and Class II SDHs, lending further support that Class II SDHs may be responsible for TA biosynthesis not only in grapevine but also in other species such as *T. cacao*. The K_m value of COC280 for sorbitol was

determined to be 16.68 mM (**Figure 5B**), which is most close to that reported for FaSDH (at 7.3 mM) (Sutsawat, Yamada et al. 2008). This is also consistent with their close distance in the phylogenetic tree (**Figure 1A**).

3.4 Conclusion

In combination with enzyme kinetic data from previous reports, phylogenetic and protein property calculation analyses in the present study indicate a clear enzymatic divergence between Class I and Class II SDHs. Homologous SDH genes from *V. vinifera*, *S. tuberosum* and *T. cacao* encoding both Class I and Class II SDHs were cloned and transformed into *E. coli* for the production of recombinant protein. Site-directed mutagenesis was carried out targeting the critical amino acids at the catalytic site of SDH to evaluate their effects on enzyme activity. Enzyme kinetic tests showed that *V. vinifera* LIDH has the highest activity with L-idonate but still retains significant activity with D-sorbitol. *T. cacao* COC280 (Class I SDH) displays the highest activity with xylitol but could hardly utilize L-idonate as the substrate. Both *V. vinifera* LIDH and COC280 show a relatively low activity with ribitol. Our results provide direct support for the enzymatic divergence between Class I and Class II SDHs, supporting potentially different biological functions for each group. Attempts to obtain purified recombinant protein for other candidate enzymes were not successful in this study. The presence of each plasmid constructs in the utilized strains and their corresponding sequences need to be verified in following studies.

3.5 Abbreviations

SDH	Sorbitol dehydrogenase
MDR	Medium-chain dehydrogenase/reductase
ADH	Alcohol dehydrogenase
NAD(H)	Nicotinamide adenine dinucleotide
NADP(H)	Nicotinamide adenine phosphate dinucleotide
MdSDH	*Malus domestica* sorbitol dehydrogenase
PpySDH	*Pyrus pyrifolia* sorbitol dehydrogenase
PsSDH	*Prunus salicina* sorbitol dehydrogenase

LeSDH	*Lycopersicon esculentum* sorbitol dehydrogenase
PpeGolDHase	*Prunus persica* sorbitol dehydrogenase
AtSDH	*Arabidopsis thaliana* sorbitol dehydrogenase
FaSDH	*Fragaria ananassa* sorbitol dehydrogenase
LIDH	L-idonate-5-dehydrogenase
5KGA	5-keto-D-gluconate
cDNA	Complementary DNA
PCR	Polymerase chain reaction
IPTG	Isopropyl β-D-1-thiogalactopyranoside
DTT	Dithiothreitol

Synonyms

PTLIDH	PGSC0003DMT400065063
COC270	Tco3_g019270
COC280	Tco3_g019280

Chapter 4 Identification and structural characterization of a putative 2-keto-L-gulonic acid reductase from *Vitis vinifera*

Identification and structural characterization of a putative 2-keto-L-gulonic acid reductase from *Vitis vinifera*

Yong Jia[1], John B Bruning[2], Crista A Burbidge[1], Crystal Sweetman[3], Christopher M Ford[1]

[1]School of Agriculture, Food & Wine, University of Adelaide, Adelaide, 5005, Australia

[2]School of Biological Sciences, University of Adelaide, Adelaide, 5005, Australia

[3] School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide 5001, Australia

yong.jia@adelaide.edu.au

john.bruning@adelaide.edu.au

christopher.ford@adelaide.edu.au

4.1 Abstract

Vitis vinifera is well known for its accumulation of a large amount of tartaric acid in young leaf and berry tissues. The metabolic processes leading to the production of tartaric acid have been well characterized, which includes the putative enzymatic conversion of 2-keto-L-gulonic acid to L-idonic acid. However, the corresponding gene and enzyme have not been identified. In the present study, we identified a putative 2-keto-L-gulonic acid reductase (VV2KR, Uniprot No. A5CAL1) from *V. vinifera* based on its close homology with an *Escherichia coli* 2-ketoaldonate reductase (Uniprot ID P37666). Bioinformatic analyses showed that VV2KR belongs to D-isomer-specific 2-hydroxyacid dehydrogenase (2KDH) superfamily and has the highest similarity to *Coleus blumei* Benth hydroxypyruvate reductase (CbHPR) and *Arabidopsis thaliana* hydroxypyruvate reductase isoform 2 (AtHPR2). Heterologous recombinant VV2KR protein was expressed and purified using an *E. coli* expression system. Enzymatic tests demonstrate that VV2KR is able to catalyze the reduction of 2-keto-L-gulonic acid (2KLG) to L-idonic acid (IA) with a preference for NADP(H) as the coenzyme and may be responsible for the biosynthesis of tartaric acid in grapevine. We went further to determine the crystal structure of VV2KR to a resolution of 2.1 Å. Structural analyses show that VV2KR adopts a typical 2KDH protein fold, consisting of a larger coenzyme binding domain and a smaller substrate binding domain. Detailed structural characteristics of VV2KR are described and compared to other known structures of this family. The potential biological function of VV2KR in tartaric acid synthesis in grapevine is discussed.

4.2 Introduction

L-tartaric acid (TA) is a four-carbon, strong organic acid commonly found in plants (Stafford 1959, Mattick, Plane et al. 1980). Despite its wide distribution, only a few species such as the cultivated grapevine *Vitis vinifera* accumulate TA as the dominant acid (Stafford 1959). In *V. vinifera*, the amount of tartrate can reach 10 mg per ripe berry (Iland and Coombe 1988, Sweetman, Deluc et al. 2009), over twice that of malate. During grape berry development, TA is produced rapidly during the first 4 weeks after anthesis and remains metabolically stable throughout the ripening process

(Iland and Coombe 1988, Coombe and McCarthy 2000). It confers a low pH and a zing flavor to berries and juice, affecting many quality aspects of wine such as colour, taste, microbial stability and aging potential. As such, exogenous TA is widely used as a flavorant and antioxidant additive in the food and wine industry (Manuel Salgado, Rodriguez et al. 2010).

The biosynthesis of TA in grapevine has attracted great research interest since L-ascorbic acid (AA, Vitamin C) was identified as the metabolic precursor (Loewus and Stafford 1958, Saito and Kasai 1969). The intermediates involved in TA biosynthetic pathway have been well characterized. Earlier radioisotope tracer studies showed that AA fed to grapevine is consecutively converted to 2-keto-L-gulonic acid (2KLG), L-idonic acid (IA) and 5-keto-D-gluconic acid (5KGA) (Saito and Kasai 1984, Malipiero, Ruffner et al. 1987). The six-carbon 5KGA is then cleaved between carbon atoms 4 and 5, yielding TA and a two-carbon fragment (Saito 1979, Saito and Kasai 1984, Malipiero, Ruffner et al. 1987). The two-carbon fragment, possibly glycoaldehyde, is recycled into triose and hexose phosphate metabolism (Wagner 1974, Saito 1979, Loewus 1999). Time course study indicated that the oxidation of L-idonic acid to 5KGA is likely the rate-limiting step during TA production in grapevine (Malipiero, Ruffner et al. 1987). In addition, 2KLG, IA and 5KGA labeled with ^{14}C on position 1 were shown to be as efficient as AA for the production of C1-labeled TA (Saito and Kasai 1984). In addition to the cleavage between positions C4/C5, concurrent studies in Pelargonium plants revealed that TA could also originate from the C3-C6 fragment of AA. In this pathway, 5KGA breaks between positions C2/C3, leading to the production of oxalic acid and L-threonate (Williams and Loewus 1978, Saito 1984, Saito, Ohmoto et al. 1997). The four-carbon L-threonate is further converted to TA (Saito 1992). Recently, it was verified that both pathways are functional in grapevine, and the first pathway accounts for the major fraction of TA produced (DeBolt, Hardie et al. 2004).

In contrast to the well-established chemical intermediates in TA biosynthesis, the genetic basis underlying this pathway is still largely unknown. Notably, *Vitaceae* plants accumulate TA to be the dominant organic acid in young berries over a

relatively short period, suggesting a highly active and biologically regulated metabolic process. By combining transcription profiling of EST data and metabolic profiling of TA, DeBolt etc successfully identified a L-idonate-5-dehydrogenase (LIDH, EC 1.1.1.264) which catalyzes the rate-limiting oxidation of L-idonic acid to 5KGA (DeBolt, Cook et al. 2006). This is the only enzyme characterized so far to be involved in TA synthesis in the grapevine. Most recently, we employed a computational approach to demonstrate that LIDH belongs to the sorbitol dehydrogenase (SDH, EC 1.1.1.14) family and represents a functional divergence of SDH after gene duplication (Jia, Wong et al. 2015). The enzymatic mechanism of the cleavage of 5KGA into TA remains unclear. It has been shown that this reaction could proceed by chemical means under both acidic and alkaline conditions (Barch 1933, Isbell and Holt 1945) or be catalyzed by vanadate in *Gluconobacter oxidans* (Klasen, Bringermeyer et al. 1992), causing speculation that this penultimate step of TA synthesis may be non-enzymatic. In a more recent study, a significant production of TA from 5KGA was observed in the absence of vanadate in *G. oxidans* (Salusjarvi, Povelainen et al. 2004). It was speculated that an enzymatic mechanism involving a transketolase and a semialdehyde dehydrogenase may be responsible for TA formation (Salusjarvi, Povelainen et al. 2004). However, none of these enzymes have been identified. Based on kinetic analyses using isotopically labeled intermediates, the reduction of 2KLG to IA in TA pathway was also assumed to be an enzymatic reaction (Saito and Kasai 1984). The incorporation of ^{14}C into 2KLG from IA suggests that the reaction is reversible (Saito and Kasai 1984). While the corresponding enzyme responsible for this reaction in grapevine has not been identified, NADPH-dependent 2-keto-D-gluconate reductase (2KGR, EC 1.1.99.3) able to perform the same reaction has been reported in *Erwinia herbicola* (Truesdell, Sims et al. 1991), *Brevibacterium ketosoreductum* (Yum, Bae et al. 1998) and *Escherichia coli* (Yum, Lee et al. 1998) and various acetic bacteria (Ameyama and Adachi 1982). 2KGR enzymes in bacteria are involved in ketogluconate metabolism and have broad substrate specificities. They catalyze the reversible reduction of 2-keto-D-gluconate (2KGA) to D-gluconate, 2, 5-diketo-D-gluconate to 5KGA and

2KLG to IA using NADP(H) as the preferred cofactor and generally have the highest activity on 2KGA (Ameyama and Adachi 1982, Truesdell, Sims et al. 1991, Yum, Bae et al. 1998, Yum, Lee et al. 1998). Enzymes of this kind belong to the D-isomer-specific 2-hydroxyacid dehydrogenase (2KDH) superfamily which catalyzes the reversible reduction of 2-oxoacids to the D-isomers of the respective 2-hydroxyacids. It's very likely that the reduction of 2KLG to IA in TA pathway is catalyzed by a homologous 2KDH present in grapevine.

Many enzymes present in higher plants belong to the 2KDH family. Well-known ones include hydroxypyruvate/glyoxylate reductase (HPR) (Tolbert, Yamazaki et al. 1970, Kim, Janiak et al. 2004, Timm, Nunes-Nesi et al. 2008), D-lactate dehydrogenase (Antonyuk, Strange et al. 2009) and D-glycerate dehydrogenase (Goldberg, Yoshida et al. 1994). Of these, the reactions performed by HPR proteins come closest to the conversion of 2KLG to IA proposed to occur in the TA biosynthetic pathway. The HPR characterized from *Coleus blumei* has been shown to be able to act on a range of substrates including glyoxylate, pyruvate, hydroxypyruvate, phenylpyruvate, 4-hydroxyphenylpyruvate, 2-oxoisocaproate etc (Janiak, Petersen et al. 2010). Accepted substrates for this enzyme share a common glyoxylate skeleton and are specifically reduced at the C2 to R-hydroxy isomers (Janiak, Petersen et al. 2010). The assumed reduction of 2KLG to IA falls into this pattern of reaction as well. The biological function of plant HPR has only been characterized in a few plants. In *Arabidopsis thaliana*, three isoforms of HPR (AtHPR1-3) have been identified and were shown to be involved in the photorespiration pathway (Mano, Hayashi et al. 1997, Timm, Nunes-Nesi et al. 2008, Timm, Florian et al. 2011). AtHPR1 is targeted to the peroxisome and contributes a major role to the photorespiration core cycle, whereas AtHPR2 and AtHPR3 have recently been shown to be targeted to the cytosol and chloroplast, respectively and represent a bypass of the photorespiration core cycle. This biological function of HPR in photorespiration has also been recognized in other plants such as barley (Murray, Blackwell et al. 1989) and spinach (Tolbert, Yamazaki et al. 1970) and is considered common to all plants. However, the presence of multiple copies of this gene may suggest a multifunction potential for HPR in plants.

Notably, hydroxypyruvate reductase in *C. blumei* has been shown to be responsible for the biosynthesis of rosmarinic acid (Kim, Janiak et al. 2004), a natural product in *Lamiaceae* and *Boraginaceae* plants with important antimicrobial and antioxidant functions (Parnham and Kesselring 1985, Szabo, Thelen et al. 1999). No information is available in regard to its relation to photorespiration in this plant. In addition, orthologous HPR genes in many other plant species such as *Solanum lycopersicum* (K4BML6), *Eucalyptus gradis* (A0A059BU28), *Citrus clementine* (V4SMV5) and *Theobroma cacao* (A0A061GB54) have also been reported, however, the corresponding in-planta functions of these putative HPR have not been characterized. A quick homology search showed that three copies of HPR genes are predicted to be present in the *V. vinifera* genome, similar to *A. thaliana*. Whether a corresponding HPR in *V. vinifera* is involved in the TA biosynthesis pathway remains unknown.

AA is a ubiquitous antioxidant in plants and plays pivotal roles in many aspects of plant growth and development (Smirnoff and Wheeler 2000). TA biosynthesis represents an important degradative pathway for AA. Identification and characterization of candidate genes responsible for TA biosynthesis would enable us to better understand how TA production is regulated in grapevine and its biological impact on AA metabolism. In this study, we report the identification of a putative 2-keto-L-gulonate reductase (VV2KR) from grapevine which is able to catalyze the reduction of 2KLG to IA using NADPH as the preferred cofactor. In addition, we determined the ligand-free crystal structure of VV2KGR to 2.1 Å. Comparative structural analyses showed that VV2KR belongs to the 2KDH superfamily and shares the highest structural similarity with plant HPR. The potential in-planta function of VV2KR will be discussed.

4.3 Results and Discussion

4.3.1 Molecular characteristics of the putative 2-keto-L-gulonate reductase

tBLASTn revealed a tentative consensus sequence (TC59682) encoding a putative enzyme (313 aa, **Figure 1**) highly homologous (~39% identity) with the *E. coli* 2-ketoaldonate reductase sequence (Uniprot ID P37666). TC59682 was selected also based on its presence in EST libraries prepared from young grape berry where TA was

synthesized rapidly. The encoded putative 2-ketoaldonate reductase homolog is named as 2-keto-L-gulonate reductase (VV2KR) in the present study. The VV2KR gene was successfully amplified by PCR using *V. vinifera* cv Shiraz pre-verasion berry cDNA as the template. The 6-His-tagged recombinant VV2KR protein was purified to high homogeneity with an estimated molecular mass at ~35kDa (**Figure 2**). Sequence analyses showed that VV2KR belongs to the 2KDH family, containing typical 2KDH catalytic domain and NAD(P)(H)-binding domain signatures (IPR006139 and IPR006140). The encoded amino acid sequence has the highest similarity with *C. blumei* HPR (CbHPR, Uniprot ID Q65CJ7; 78% identity) and *A. thaliana* HPR isoform2 (AtHPR2, Uniprot ID Q9CA90; 75% identity), while sharing 48% and 34% identity with AtHPR3 and AtHPR1, respectively. In addition, VV2KR is also highly homologous with *Homo sapiens* HPR (GRHPR, ~38% identity) (Booth, Conners et al. 2006), *Ketogulonicigenium vulgare* 2-hydroxyacid dehydrogenase (Kv2DH, ~41% identity) (Han, Xiong et al. 2014) and *Pyrococcus horikoshii* glyoxylate reductase (Ph2GR, ~45% identity) (Yoshikawa, Arai et al. 2007). This is consistent with our prediction that the reduction of 2KLG to IA is catalyzed by a putative enzyme belonging to the 2KDH family. Notably, *K. vulgare* Kv2DH has been shown to be able to convert 2KLG to IA (Han, Xiong et al. 2014).

```
     ACCAGCGAAGTCCCCAAAGATATCGGACGGCTCATCCGCCCCACCGTCACCGCCGCTGCA
   1                                                       M   E
     CTCCCACCTCCGTGCGCCGCCGAGTGTATAGTGCGTCTTTCACAAAGCGAAACCATGGAA
  22 S   I   G   V   L   L   T   C   P   M   N   P   Y   L   E   Q   E   L   D   K
     AGCATCGGGGTACTGTTGACTTGCCCAATGAACCCATACCTGGAACAGGAACTGGACAAG
  42 R   F   K   L   F   R   F   W   D   F   P   S   A   N   D   L   F   R   E   H
     CGCTTCAAGCTCTTCCGCTTCTGGGACTTTCCAAGCGCCAACGATCTTTTCAGGGAGCAT
  62 S   N   S   I   R   A   V   V   G   N   S   F   I   G   A   D   A   Q   M   I
     TCAAATTCGATCCGAGCTGTGGTTGGAAACTCCTTCATCGGCGCCGACGCCCAGATGATC
  82 E   A   L   P   K   M   E   I   V   S   S   F   S   V   G   L   D   K   I   D
     GAGGCGTTGCCCAAGATGGAGATTGTGTCGAGTTTCAGCGTTGGGTTGGACAAGATCGAT
 102 L   V   R   C   K   E   K   G   I   R   V   T   N   T   P   D   V   L   T   E
     TTGGTGAGGTGCAAGGAGAAGGGAATTAGGGTTACGAACACTCCGGATGTGCTGACGGAG
 122 D   V   A   D   L   A   L   A   L   I   L   A   T   L   R   R   I   C   E   S
     GACGTGGCGGACTTGGCACTTGCTTTGATTTTGGCGACTCTGAGACGTATTTGTGAAAGT
 142 D   R   Y   V   R   S   G   S   W   K   K   G   D   F   K   L   T   T   K   F
     GATCGTTATGTGAGGAGTGGGTCGTGGAAGAAAGGGGATTTCAAGTTGACTACCAAGTTC
 162 T   G   K   S   V   G   I   I   G   L   G   R   I   G   S   A   I   A   K   R
     ACTGGAAAATCAGTTGGCATTATAGGGTTGGGTAGGATTGGCTCAGCAATTGCCAAGAGA
 182 A   E   G   F   S   C   P   I   S   Y   H   S   R   T   E   K   P   G   T   N
     GCCGAGGGATTTAGCTGTCCAATTAGTTACCATTCCAGAACAGAGAAACCAGGGACAAAC
 202 Y   K   Y   Y   P   S   V   V   E   L   A   S   N   C   Q   I   L   V   V   A
     TACAAGTACTATCCTAGTGTCGTTGAATTGGCCTCCAACTGTCAAATCCTGGTTGTTGCT
 222 C   A   L   T   P   E   T   R   H   I   I   N   R   E   V   I   N   A   L   G
     TGCGCGTTAACACCAGAAACCCGCCACATCATCAACCGTGAAGTCATCAATGCACTGGGT
 242 P   K   G   V   V   I   N   I   G   R   G   L   H   V   D   E   P   E   L   V
     CCAAAGGGTGTGGTCATCAACATCGGAAGGGGATTACATGTGGATGAACCTGAGCTTGTA
 262 S   A   L   V   E   G   R   L   G   G   A   G   L   D   V   F   E   N   E   P
     TCCGCACTGGTTGAAGGCCGGTTGGGAGGTGCTGGACTTGATGTGTTTGAAAATGAGCCT
 282 N   V   P   E   E   L   L   A   M   D   N   V   V   L   L   P   H   V   G   S
     AATGTACCTGAAGAGCTGTTAGCAATGGACAATGTAGTCCTTTTGCCTCATGTTGGAAGC
 302 G   T   V   E   T   R   K   D   M   A   D   L   V   L   G   N   L   E   A   H
     GGAACGGTGGAAACCCGGAAAGACATGGCTGACCTGGTACTTGGAAACTTAGAGGCTCAC
 313 F   L   N   K   P   L   L   T   P   V   V   *
     TTTCTGAACAAACCACTGTTAACTCCAGTGGTTTAA
```

Figure 1. Tentative consensus sequence TC59682 encoding a putative 2-keto-L-gulonate reductase.

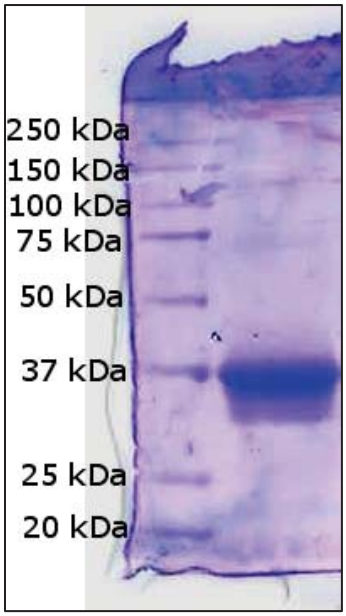


Figure 2. Enzyme purification of the recombinant VV2KR.

4.3.2 Enzymatic activity of VV2KR

As shown in **Figure 2**, the 6-His-tagged VV2KR was purified to high homogeneity, with an estimated molecular mass of around 33 kDa, consistent with the computational prediction. The concentration of the enzyme sample used for enzymatic tests was measured at around 2.6 mg/ml. VV2KR exhibits the highest activity at pH 7.5 in 100 mM HEPES buffer at 37 ℃. The optimal pH for VV2KR is similar to that reported for *C. blumei* CbHPR (pH 7.0) (Kim, Janiak et al. 2004) and human GRHPR (pH 7.5) (Mdluli, Booth et al. 2005) but differs from the peroxisomal spinach HPR1 (pH 5.1 and pH 6.2 with NADPH and NADH, respectively) (Tolbert, Yamazaki et al. 1970) and the cytosol spinach HPR2 (pH 5.5-6.5) (Kleczkowski and Randall 1988). The specific activities of VV2KR on glyoxylate, pyruvate, hydroxypyruvate and 2KLG were determined with NADH and NADPH at the optimal condition. Both NADH and NADPH could be used by VV2KR as the cofactor for the reduction of glyoxylate, pyruvate and 2KLG. However, no significant activity could be detected with hydroxypyruvate ranging from 0.05 mM to 25 mM using NADH. The substrate specificity of VV2KR was shown in **Table 1** (see detailed data in Figure S1). VV2KR shows a preference for NADPH over NADH with the reduction of glyoxylate, pyruvate and hydroxypyruvate. Comparable substrate specificity was observed with 2KGA using NADH and NADPH as cofactors. VV2KR exhibits the highest substrate specificity with glyoxylate which is about 5, 10 and 100 times that of hydroxypyruvate, 2KGA and pyruvate, respectively. As shown in **Table 1**, VV2KR has lower K_m values for NADPH than NADH for all substrates tested in this study, which is similar with that reported for human GRHPR (Mdluli, Booth et al. 2005). The variations were at around 6, 3 and 2 times for glyoxylate, pyruvate and 2KGA, respectively. The K_m value of VV2KR on hydroxypyruvate with NADPH as cofactor was 0.096 mM, which is very close to that reported for the native barley NADPH-preferring HPR (0.08 mM) (Kleczkowski, Edwards et al. 1990), the purified maize HPR1 (0.11 mM) and HPR2 (0.09 mM) (Kleczkowski and Edwards 1989), and about half of that for the purified spinach HPR1 (0.24 mM) (Kleczkowski and Edwards 1989) and higher than human GRHPR (0.058 mM) (Mdluli, Booth et al.

2005). Notably, VV2KR differs from these enzymes for its lack of NADH-dependent hydroxypyruvate activity. The K_m value (0.044 mM) of VV2KR on glyoxylate with NADPH is closest to that reported for the native barley NADPH-preferring HPR (0.07 mM) (Kleczkowski, Edwards et al. 1990) but significantly lower than that for spinach HPR1 (0.15 mM) (Kleczkowski and Edwards 1989). The human GRHPR showed a K_m value of 0.24 mM for glyoxylate with NADPH, which is much higher than its plant counterparts (Mdluli, Booth et al. 2005).

In addition to glyoxylate and hydroxypyruvate, VV2KR in this study is able to catalyze the reduction of 2KGA efficiently using both NADH and NADPH as cofactor, supporting its potential role in the TA biosynthesis pathway in the grapevine. The K_m value for 2KGA with NADPH is measured to be 0.70 mM which is around half of that with NADH(1.561 mM). The substrate specificity of VV2KR for 2KGA with NADH is slightly higher than that with NADPH. Using NADPH as the cofactor, the V_{max} for 2KGA (7.544 U/mg) is higher than that for hydroxypyruvate (6.295 U/mg) while the substrate specificity for 2KGA is around 1/6 that for hydroyxpyruvate.

Table 1. Enzymatic activity of VV2KR with different substrates.

Substrate	**Cofactor**	**K_m (mM)**	**V_{max} (U/mg)**	**V_{max}/K_m (U/mg*mM)**
Glyoxylate	NADH	0.248	69.380	279.758
	NADPH	0.044	13.820	314.091
Pyruvate	NADH	3.453	0.442	0.128
	NADPH	1.371	0.548	0.400
Hydroxypyruvate	NADH	--	--	--
	NADPH	0.096	6.295	65.573
2KGA	NADH	1.561	17.190	11.012
	NADPH	0.700	7.544	10.777

4.3.3 Crystallization

Crystals for data collection were produced with the well solution containing 1260 mM ammonium sulfate, 100 mM sodium acetate/acetic acid pH 4.5, 200 mM sodium chloride using the sitting drop (1µl+1µl) method. All crystals tend to grow into thin sheet shape (**Figure S2**). Preliminary diffraction data processing revealed an apparent

lattice group of P222; however, models generated from this solution could not be well refined. The structure was solved by molecular replacement method using diffraction data to a resolution of 2.1 Å. Data collection and refinement statistics are summarized in **Table 2**. Twinning analysis using Xtriage (Adams, Afonine et al. 2010) showed that the apo VV2KR crystal was pseudo-merohedrally twinned. As such, a twinning refinement (twin operator: h, -k, -l) was applied during refinement, which allowed the structure to be refined to completion. The ligand-free VV2KR crystal has the symmetry primitive monoclinic space group $P2_1$, with unit-cell parameters a=73.05, b=85.72, c=112.90. Enzyme cofactor (NADPH, at 2 mM) and substrate (2KLG, at 20-50 mM) were added both separately and together to the crystallization solution, in an attempt to obtain co-crystals. In addition, the same cofactor and substrate concentrations were also used for crystal soaking. Despite exhaustive trials, no sufficient electron density could be observed for NADPH and 2KLG in the electron density map. This might be due to the weak binding affinity of VV2KR for these ligands.

Table 2. Summary of X-ray diffraction data and refined model statistics for grapevine VV2KR crystal structure. (For the resolution range, the number in the parenthesis represents the highest shell**)**

Data collection and processing	**Apo-VV2KR**
Wavelength (Å)	0.9537
Space group	$P2_1$
Unit-cell parameters (Å)	a=73.05, b=85.72, c=112.90
	α=90.00, β=89.91, γ=90.00
No. of observations	284489
Resolution range (Å)	47.15-2.10 (2.14-2.10)
Rmerge (%)	7.6 (5.9 / 11.7)
Rpim (%)	7.2 (5.8 / 11.1)
Completeness (%)	99.3 (76.8 / 100.0)
Average I/sd(I)	9.2 (10.8 / 6.4)
Multiplicity	3.5 (2.7 / 3.6)
Refinement	
Rwork (%)	19.34
Rfree (%)	25.34
Refined residues	1248 (312*4)
Water molecules	1052
Mean B factors water ($Å^2$)	20.96
Mean B factors chain A/B/C/D ($Å^2$)	22.37 / 23.52 / 22.93 / 21.82
R.m.s.d bond angles (Å)	0.884
R.m.s.d bond distances (Å)	0.004
Protein geometry	
Ramachandran outliers (%)	0.32
Ramachandran favored (%)	94.16

4.3.4 Overall structure

The determined VV2KR structure has four subunits (chain A-D) in the asymmetric unit (**Figure S3**). The four subunits share high structural similarity with an average main-chain atom root mean square deviation (RMSD) of 0.365 Å and are considered identical with each other. The topology of VV2KR structure exhibits typical characteristics of the 2KDH family, with each monomer comprising a larger coenzyme-binding domain (CBD) and a smaller substrate-binding domain (SBD) (**Figure 3**). Based on structural superimposition with known 2KDH structures, the catalytic site is identified to be located at the cleft between the two domains. The two

domains are joined by two loop hinges (**Figure 3**), which is a common observation in other 2KDH protein structures. Previous studies on the human GRHPR structure indicated that the flexibility of these two loop hinges could affect the enzyme catalysis process (Booth, Conners et al. 2006). The four subunits of VV2KR demonstrate no significant angular changes between domains (~2º between chain A and D, calculated by the DynDom program (Hayward and Berendsen 1998)), suggesting a stable and uniform catalytic site for VV2KR subunits. In addition, domain motion analyses also showed no significant difference between VV2KR and other close homologous structures such as *C. blumei* hydroxypyruvate reductase (CbHPR), human hydrxypyruvate/glyoxylate reductase (GRHPR), *P. horikoshii* glyoxylate reductase (Ph2GR), except *K. vulgare* 2-hydroxyacid dehydrogenase (Kv2DH) which demonstrates a larger domain angle (Han, Xiong et al. 2014). The VV2KR monomer has the highest structural similarity (all atom RMSD 0.452 Å) with *C. blumei* CbHPR (PDB: 3BA1) and demonstrates a relatively higher level of deviation with its animal and prokaryotic counterparts human GRHPR (PDB: **2GCG**, RMSD 1.203 Å), *K. vulgare* Kv2DH (PDB: **4LSW**, RMSD 1.575 Å) and *P. horikoshii* Ph2GR (PDB**: 2DBZ**, RMSD 1.072 Å). In addition to HPR proteins, VV2KR structure also shares high homology with other 2KDH proteins including D-lactate dehydrogenase (**3WX0**), D-glycerate dehydrogenase (**1GDH**), phosphateglycerate dehydrogenase (**2G76**, **1YBA**) and transcription co-repression CtBP dehydrogenase (**1HKU**, **1MX3**), with C-alpha RMSD values ranging from 1.165 Å to 4.400 Å (**Table 3**).

During the cloning process, a peptide of 55 aa including the 6×His tag was introduced to the N-terminus of the VV2KR. These additional amino acids did not seem to affect the global folding of the functional domains. No sufficient electron density was observed for the introduced peptide, which suggests a high mobility for this region, which is a frequent observation for other protein crystals containing this domain. As shown in the schematic graph for VV2KR (**Figure 3**), residues 2-94 and 285-311 from the N-terminal and the C-terminal regions respectively form the smaller SDB domain, while the larger CBD domain consists of residues 100-276. Both the CBD

and SBD domains are marked by a typical α/β/α pattern (**Figure 3**), similar to those observed for other 2KDH members. In particular, the CBD is composed of 7-stranded parallel beta-sheet at the core, flanked by 5 and 4 alpha-helices on each side, forming a typical NAD(P)-binding Rossmann fold. This feature is strictly conserved in the CbHPR structure (Janiak, Petersen et al. 2010). Booth etc reported only 6 strands at the core of the CBD of human GRHPR structure (Booth, Conners et al. 2006). The additional β-strand in VV2KR and CbHPR, corresponding to β-strand 8 (**Figure 4**), consists of only 3 residues (184-186) and was not defined as a strand in some 2KDH structures due to its short length. An extended loop (residues 131-146 in VV2KR), inserted between α-helix M and β-strand 6, is also commonly observed in 2KDH proteins including HPR and D-glycerate dehydrogenase. This loop protrudes into the catalytic site of its dimer unit and has been suggested to contribute to the dimer formation (Booth, Conners et al. 2006). Compared to the apo form of *C. blumei* CbHPR, VV2KR lacks an additional short α-helix (residues 207-209 in the apo CbHPR structure) which is located adjacent to the substrate binding pocket. Interestingly, this helix is also not observed in the *C. blumei* NADPH-CbHPR structure (PDB: 3BAZ, residues 207-209) (**Figure 4**). A corresponding short α-helix was found in *K. vulgare* Kv2DH (PDB: 4LSW) and *P. horikoshii* Ph2GR (PDB: 2DBZ) but not in human GRHPR (PDB: 2GCG) and *P. thermophile* 2GR (Pt2GR; PDB: 3WNV), suggesting a high flexibility in helix formation for this region (**Figure 4**). Due to its close position to the catalytic site, the structural variation at this region may have significantly impacted the substrate binding. The SBD domain of VV2KR exhibits a 2-α/5-β/4-α motif, forming a flavodoxin-liking fold (**Figure 3**). While this characteristic is strictly conserved in the *C. blumei* CbHPR and *K. vulgare* Kv2DH structures, the short α-helix L (residues 29-31 in VV2KR) is not observed in human GRHPR (PDB: 2GCG), *P. horikoshii* Ph2GR (PDB: 2DBZ) and *P. thermophile* Pt2GR (PDB: 3WNV) and may represent a unique structural characteristic for plant HPR proteins. Notably, *P. thermophile* Pt2GR contains a distinct α-helix between β-strand 3 and α-helix C (**Figure 4**). This α-helix is not observed in any other glyoxylate reductase structure. The CBD and SBD of VV2KR are joined by two loops

(residues 95-99 and 277-284) located at β-strand 5/ α-helix E and β-strand 12/ α-helix K, respectively (**Figure 3**). The catalytic site, located at the cleft, is flanked by loops from both domains.

Table 3. The pair-wise all atom RMSD value (Å) of the respective monomers and other 2KDH structures.

		VV2KR chain A
VV2KR subunit	B	0.344
	C	0.273
	D	0.431
Hydroxypyruvate/glyoxylate reductase	3BA1	0.452
	3BAZ	0.500
	4LSW	1.575
	2GCG	1.203
	2DBZ	1.072
D-lactate dehydrogenase	3WX0	4.400
	3WWZ	1.883
	3KB6	1.380
	4CUJ	4.275
D-glycerate dehydrogenase	1GDH	3.347
Phosphateglycerate dehydrogenase	2G76	1.414
	1YBA	2.041
Transcription co-repression Ctbp dehydrogenase	1HKU	1.239
	1MX3	1.235
	4U6S	1.165

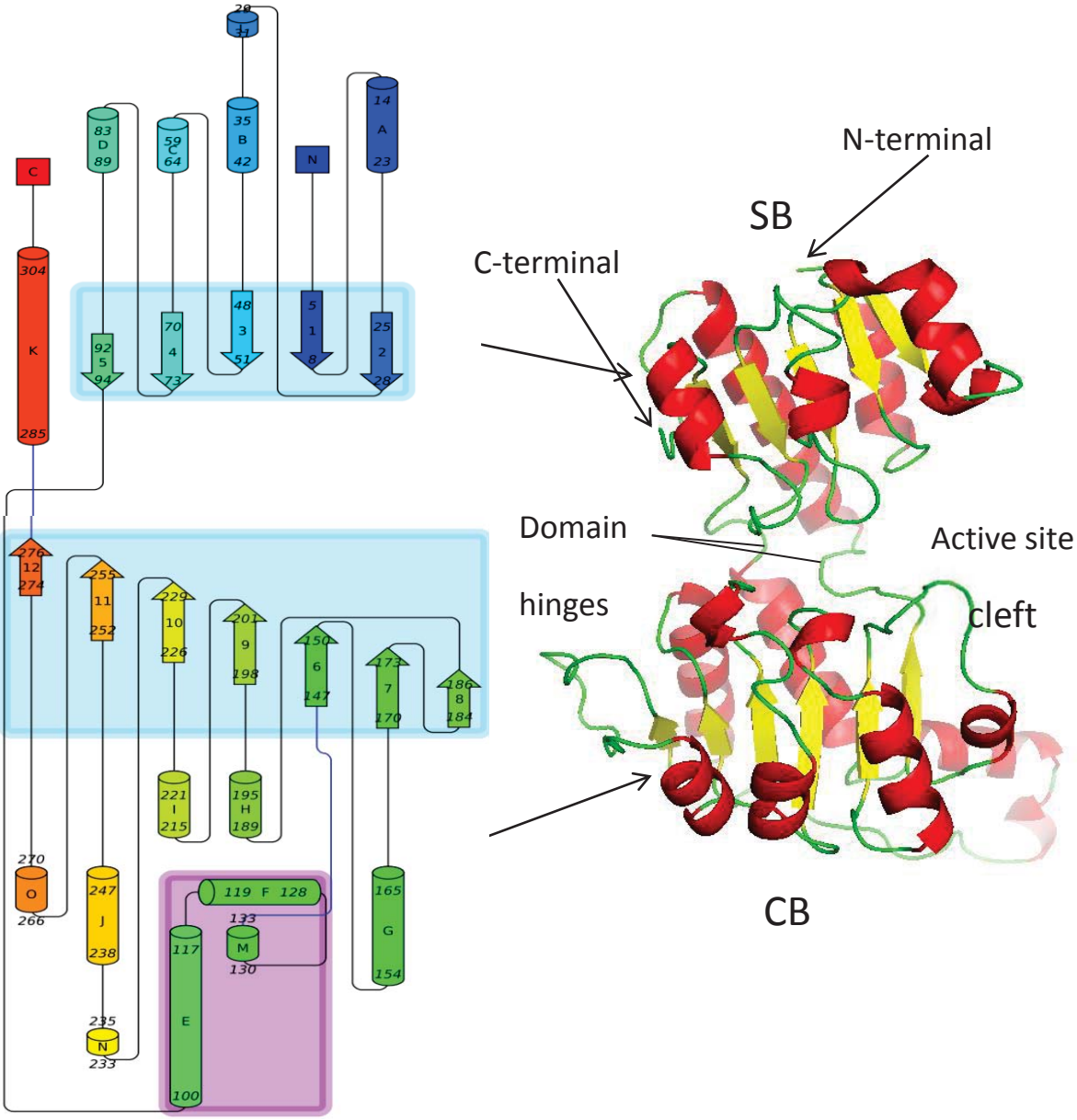


Figure 3. A schematic graph of VV2KR generated by the Pro-Origami program and the global folding of VV2KR monomer. β-strand and α-hellix clusters were shaded in light blue and purple colour, respectively. Residue numbers of each secondary structural element were labelled sequentially. CBD and SBD stand for coenzyme binding domain and substrate binding domain, respectively.

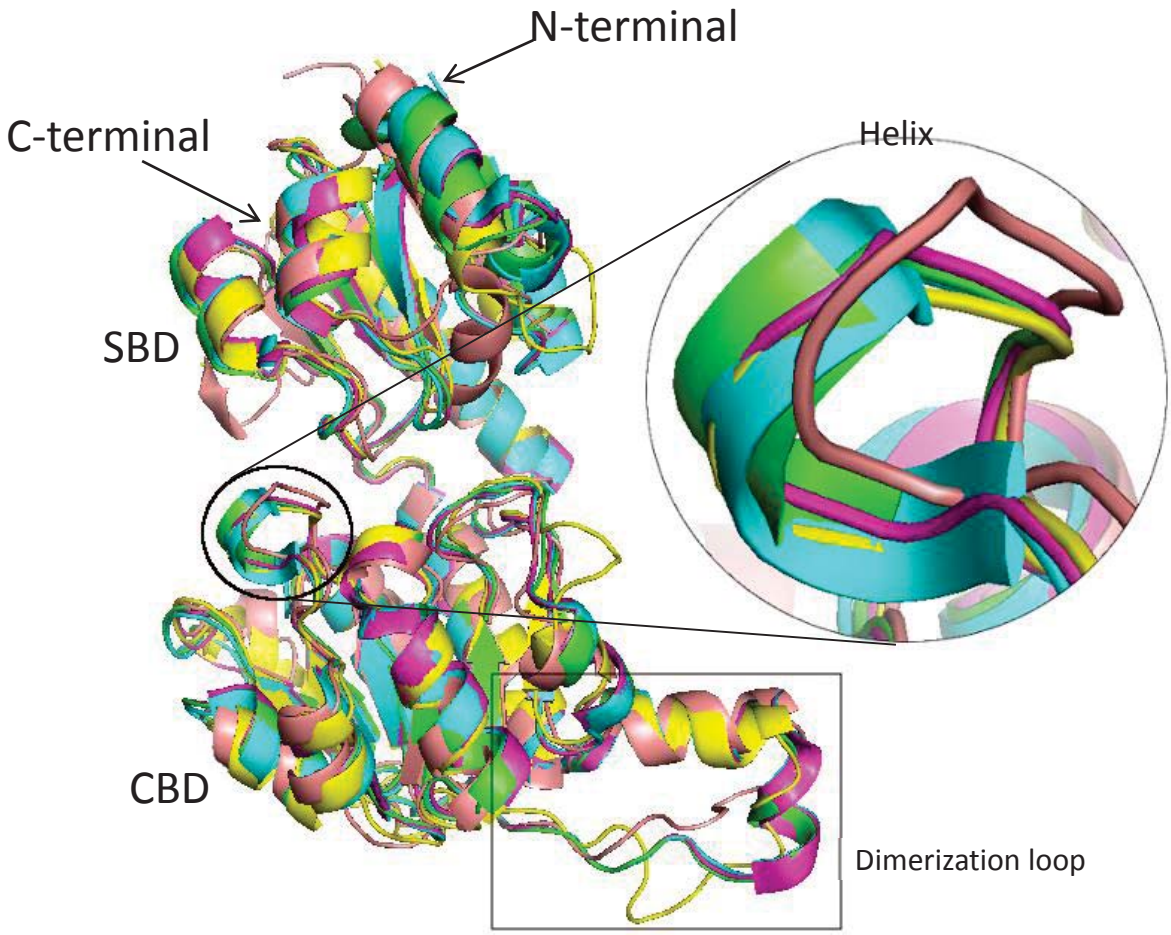


Figure 4. Structural superimposition of VV2KR with other 2KDH proteins. VV2KR (green), apo-CbHPR (cyans), binary-CbHPR (magenta), GRHPR (yellow), Pt2GR (salmon).

4.3.5 Dimer interface

Our crystal structure indicates that VV2KR may be a biologically active dimer. Of the four subunits in the asymmetric unit, chain A dimerizes with chain D with close non-covalent interactions, while Chain B and Chain C form dimers with their respective crystallographic symmetry mates (**Figure S3**). This observation is consistent with the prediction based on VV2KR's high structural similarity with *C. blumei* CbHPR, which has been shown to function as a dimer on the basis of both crystal structure (Janiak, Petersen et al. 2010) and molecular mass measurement (Kim, Janiak et al. 2004). In addition to HPR, many other members of the 2KDH family including D-glycerate dehydrogenase (Goldberg, Yoshida et al. 1994),

phosphoglycerate dehydrogenase (Thompson, Bell et al. 2005), D-lactate dehydrogenase (Antonyuk, Strange et al. 2009), transcription corepressor CtBP dehydrogenase (Nardini, Spano et al. 2003), NAD(H)-dependent formate dehydrogenase (3N7U), D-hydroxyisocaproate dehydrogenase (Dengler, Niefind et al. 1997), D-mandelate dehydrogenase (2W2K) from both eukaryotic and prokaryotic organisms could also be considered as biological dimers. The exceptions are the phosphoglycerate dehydrogenases from *E. coli* and *Mycobacterium tuberculosis*, which form tetramers with distinct topologies (Schuller, Grant et al. 1995, Dey, Grant et al. 2005).

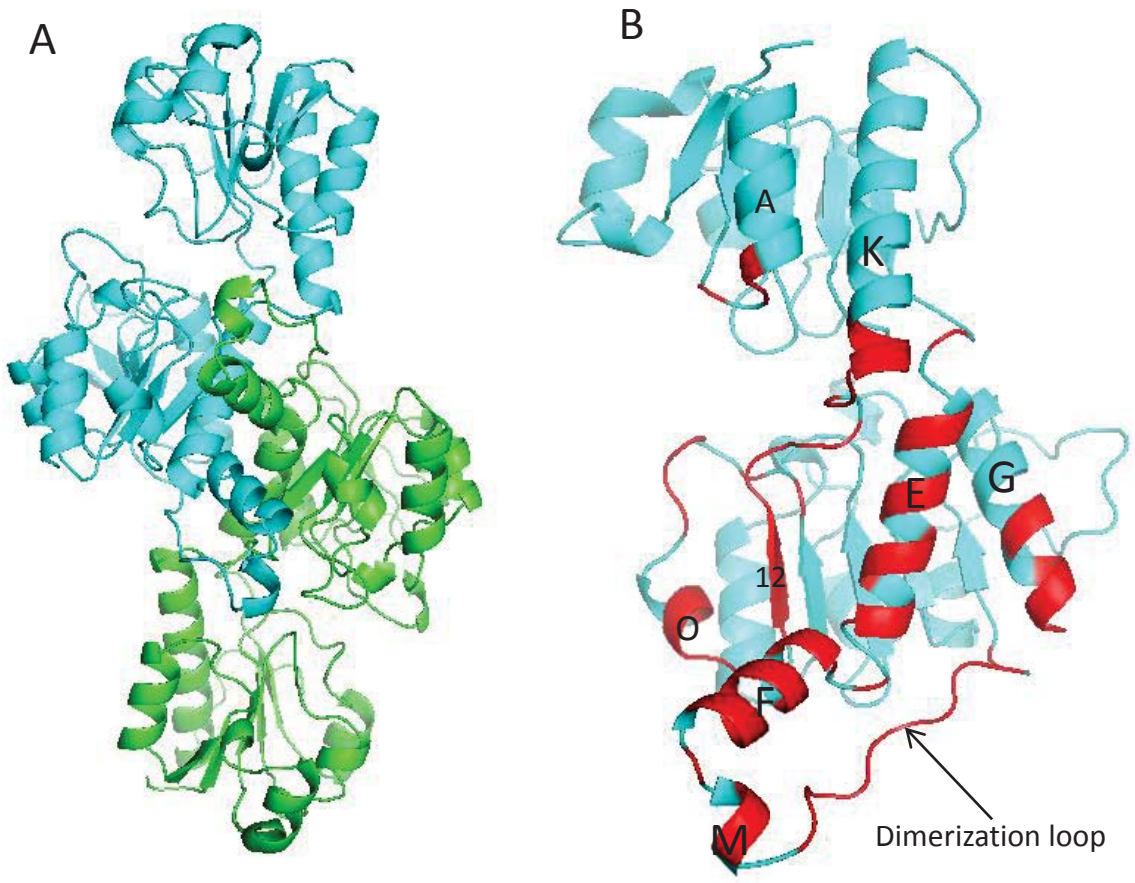


Figure 5. Displays the dimer interface of VV2KR. A. The overall structure of VV2KR dimer. **B**. The residues involved in the dimer interaction (annotated in red).

Molecular contacts between subunits of the VV2KR dimer resemble those observed in other 2KDH members and mainly involve residues from the CBD domain (**Figure**

5A). The dimer interface mainly involves 6 helices (E, F, G, K, M, O), one β-strand 12 (residues 274-276, sandwiched between helix E and O) and several loops including the so called dimerization loop (**Figure 5B**). In particular, helix G stacks against its oligomeric counterpart in an anti-parallel manner within hydrogen binding distance. Helix E cross-packs with its dimer counterpart (a head-to-tail angle of approximately 120°), which is located at the center of the contact interface. Helix F cross-stacks against β-strand 12 vertically and is also within close distance to Helix O from the other molecule. The supplementary helix M to helix F interacts with the hinge loop (residues 252-255) between helix O and β-strand 11 from the other monomer. The dimerization loop (residues 131-146) wraps around the neighbor subunit at the domain-joining hinge and is believed to contribute significantly to the dimer formation. This loop is also positioned close to the substrate-binding site of the other subunit and may affect substrate binding specificity in the native protein. In addition, several residues from the helix A and K also contribute to the dimerization (**Figure 5B**). The oligomeric interface of VV2KR was analyzed using the PDBePISA program (**Figure S4**). The buried dimer interface area was calculated to be 2439 $Å^2$ and 2411 $Å^2$ for each monomer, corresponding to 16.6% and 16.4% of the total surface area (14699 $Å^2$ and 14675.5 $Å^2$, respectively) of the respective subunit. 66 and 65 residues from chain A and D, respectively are located at the interface. The majority of these residues belong to the CBD domain, with only a few from the SBD domain (residues 10-13, 51, 282-287). In addition, 29 hydrogen bonds and 8 salt bridges (**Figure S4**) are formed at the interface, constituting the main force for the dimer interaction. This is comparable to the 28 hydrogen bonds and 6 salt bridges observed for the *C. blumei* CbHPR dimer. Superimposition of the dimer forms of *C. blumei* CbHPR and VV2KR revealed a whole-structure main-chain atom RMSD of 0.550 Å, suggesting a strong similarity in dimer arrangement for the two proteins. A total of 21 residues are involved in hydrogen binding interaction contributing to the dimer formation, majority of which are conserved in *C. blumei* except residues 102, 143, 263 and 269. No disulfide bonds or covalent bonds were found between the two dimer subunits.

4.3.6 Co-enzyme binding

Since no significant electron density could be observed for the cofactor after either apo-crystallization or co-crystallization, the preferred coenzyme $NADP^+$ was docked into the apo form VV2KR structure using the ICM-Pro software (Molsoft LLC, La Jolla, CA, USA). A number of putative conformations of comparable free energy value were generated. The conformation with the lowest free-energy is analyzed (**Figure 6**). Angular variation between the two major domains of 2KDH members upon coenzyme binding has been reported in several structures. B-factor calculation for the human GRHPR structure also indicated high mobility between the two domains (Booth, Conners et al. 2006). However, recent study on the *C. blumei* CbHPR structure, which shares the highest similarity with VV2KR, showed no significant change in domain angle upon NADP(H) binding, suggesting that VV2KR might follow a similar pattern (Janiak, Petersen et al. 2010). This indicates that the apo VV2KR structure determined in this study is suitable for coenzyme docking analyses.

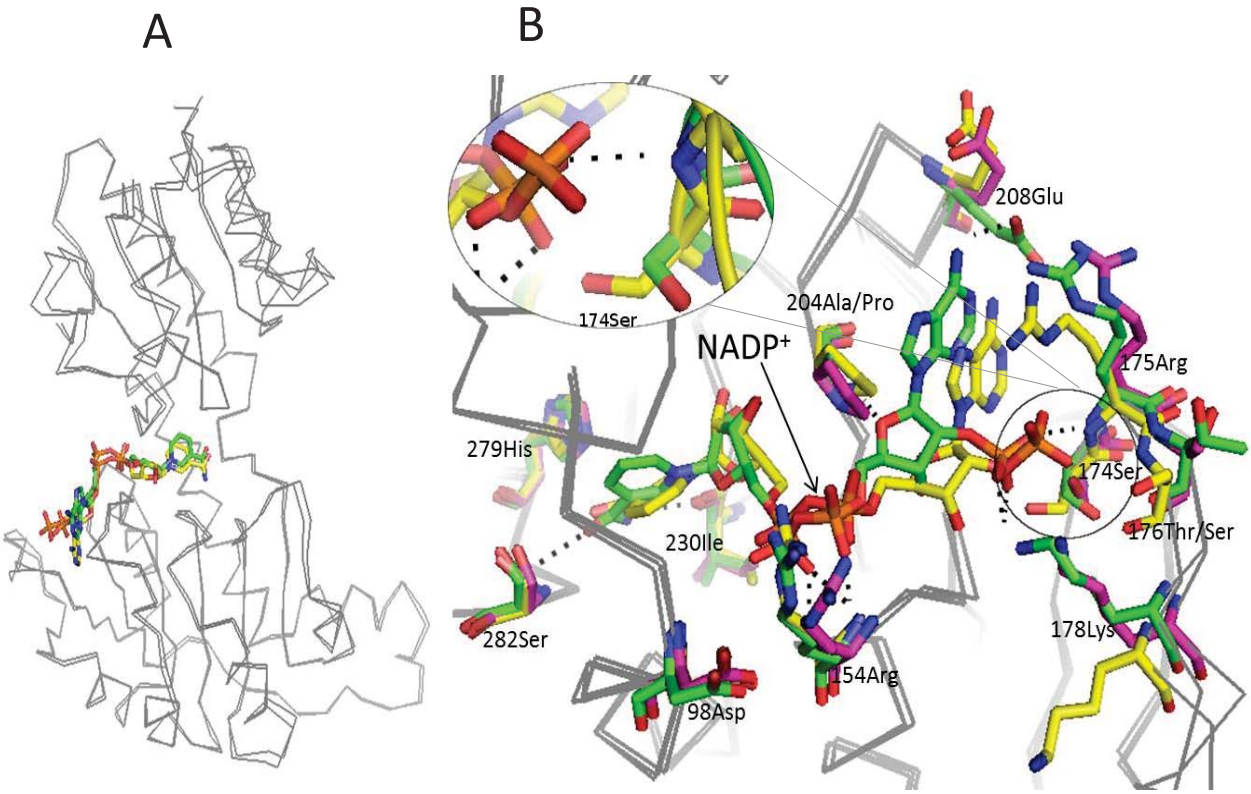


Figure 6. Co-enzyme docking analyses of VV2KR and superimposition with CbHPR. A. Overall position of coenzyme $NADP^+$ in the VV2KR and *C. blumei*

CbHPR structures (Ribbon representation). B. Structural superimposition of coenzyme binding residues of VV2KR (green), 3BAZ (yellow) and 3BA1 (magenta). Residues with interacting potentials with coenzyme are represented in sticks.

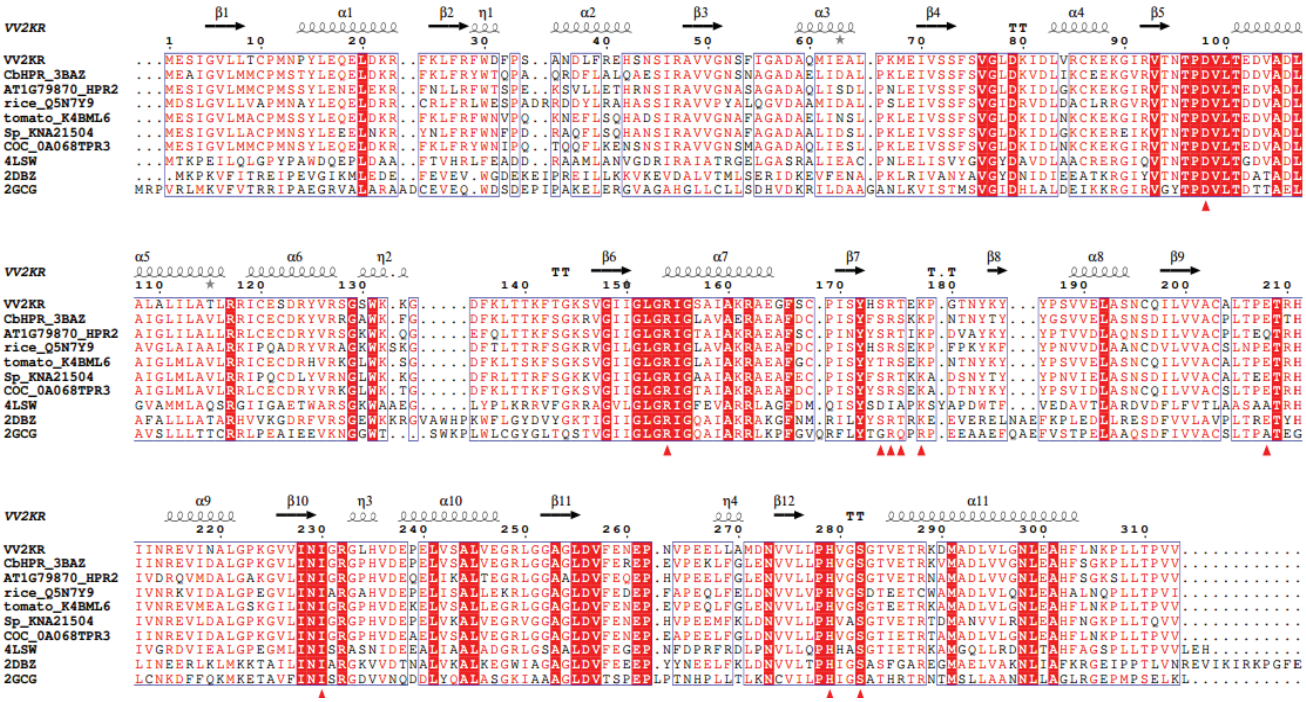


Figure 7. Sequence alignment of hydroxypyruvate/glyoxylate reductase homologs from different organisms. Residues identified as NADPH-binding in the present study were marked with red-triangle.

In the modeled VV2KR-NADPH complex, NADPH is located close to the catalytic site and occupies a comparable position to the NADPH in the binary structure of *C. blumei* CbHPR (**Figure 6**). This is consistent with the high structural similarity between VV2KR and CbHPR. The most significance difference is observed with the amide ring, which exhibits a shift of around 3.5 Å. In addition, the niacin ring at the other end of the NADPH molecule displayed a moderate rotation of around 30º along the axis of the niacin ring panel. A total of 10 hydrogen bonds are detected between the modeled NADPH and VV2KR. These hydrogen bonds involve the interactions with the main-chains and side-chains of residues (208E, 175R, 178K, 154R, 230I, 282S). The number of these predicted interactions is significantly less compared to the 16 hydrogen bonds reported for NADPH in the *C. blumei* CbHPR structure. In the CbHPR co-crystal, additional residues (98D, 176S, 174S, 279H) also form hydrogen

bonds with NADPH (Janiak, Petersen et al. 2010). Structural superimposition shows that all of the 10 coenzyme binding residues in *C. blumei* CbHPR mentioned above are strictly conserved in VV2KR, except 176S, which is replaced by a structurally similar residue 176T. While most of these NADPH binding residues are highly conserved among plant HPR proteins (*C. blumei*, *A. thaliana*, rice, tomato, spinach and cocao), significant variations were also observed with other 2KDH member proteins (**Figure 7**), which may reflect the differences in NADPH binding for different proteins. In addition to hydrogen bonding interactions with catalytic residues, NADPH in the *C. blumei* CbHPR structure is also coordinated by hydrogen bonds to a significant number of solvent water molecules. Given the strong homology between VV2KR and *C. blumei* CbHPR, this is also likely to be true for the native NADPH-binding VV2KR structure. While the spatial positions of the main chain of the NADPH-coordinating residues in VV2KR are highly conserved in the *C. blumei* CbHPR structure, significant side chain movements are observed for residues 208E, 175R, 178K, 176S, 154R (**Figure 6**). Specifically, the side chains of 208E and 178K shift dramatically away from the amide ring of NADPH in the co-crystal of *C. blumei* CbHPR, while the 175R side chain is attracted to NADPH to form a hydrogen bond (**Figure 6**). In addition, the 176S (equivalent to 176T in VV2KR) side chain flips around 90º and also establishes a hydrogen binding interaction with NADPH. Notably, the apo-form *C. blumei* CbHPR structure resembles VV2KR in all of these side chain changes. Thus our determined VV2KR structure provides solid support that significant side chain movement at the coenzyme binding site occurs during the NADPH binding process. It also explains the above mentioned movement of NADPH in comparison with *C. blumei* CbHPR and may be the underlying cause for the reduced number of coenzyme hydrogen bonding interactions detected in our modeled NADPH-VV2KR complex.

Most proteins from the 2KDH family employ NADH as the coenzyme and have no activity with NADPH. This includes the bacterial D-glycerate dehydrogenase (Goldberg, Yoshida et al. 1994), D-lactate dehydrogenase (Antonyuk, Strange et al. 2009), formate dehydrogenase (Baack, Markwell et al. 2003), phosphoglycerate

dehydrogenase (Grant 1989) and D-mandelate dehydrogenase (Wada, Iwai et al. 2008). The enzyme assay in our study showed that VV2KR can utilize both NADH and NADPH as cofactor and has a preference for NADPH. Using 2KLG as the substrate, the V_{max} of VV2KR with NADPH as the coenzyme is almost twice that with NADH. Similar results have been reported for *C. blumei* CbHPR (Kim, Janiak et al. 2004), human GRHPR (Mdluli, Booth et al. 2005) and *K. vulgare* Kv2DH (Han, Xiong et al. 2014), which also demonstrate substantially higher affinity for NADPH than NADH. Bernard et al identified the critical amino acid responsible for the absence of NADPH activity to be the Asp in a 2KDH protein from *Lactobacillus delbrueckii* subsp. bulgaricus (corresponding to 174Ser in VV2KR; **Figure 6**) (Bernard, Johnsen et al. 1994). The negatively charged carboxyl group of the Asp side chain, positioned at the coenzyme binding site, was assumed to prevent the binding of NADPH due to its similarly negatively-charged phosphate group (Bernard, Johnsen et al. 1994). The Asp is commonly conserved in the above mentioned NAD(H)-dependent 2KDH proteins. In our modeled VV2KR-NADP complex, a Ser is found in place of Asp, which is positioned close to the phosphate group of NADPH with hydrogen binding potential (**Figure 6**), thereby facilitating the preferential binding of NADPH over NADH. Structural superimposition of VV2KR with *C. blumei* CbHPR, human GRHPR and *K. vulgare* Kv2DH revealed a highly conserved Ser at the same position of the catalytic site (**Figure 6**). NADP(H) co-crystals of *C. blumei* CbHPR and human GRHPR showed that Ser forms hydrogen bonds to the phosphate group of the bound NADP(H) molecule, which may contribute to a higher NADP(H) binding affinity for these proteins. Notably, significant variation of coenzyme binding affinity also occurs among NADP(H)-dependent 2KDH proteins. While *C. blumei* CbHPR and human GRHPR both exhibit over 20-fold higher affnity for NADPH than NADH, *C. blumei* CbHPR binds NADPH (K_m=0.02 mM) (Kim, Janiak et al. 2004) substantially tighter than human GRHPR (K_m=0.11mM) (Mdluli, Booth et al. 2005). In addition, the K_m value of *K. vulgare* Kv2DH toward NADH is just over two times that for NADPH (Han, Xiong et al. 2014). Interestingly, counterpart glyoxylate reductases from the thermostable archaeal bacterial

Thermococcus litoralis (Ohshima, Nunoura-Kominato et al. 2001) and *P. horikoshii* (Yoshikawa, Arai et al. 2007) have been shown to utilize NADH more efficiently than NADPH, while also demonstrating significant levels of NADPH-dependent activities. Sequence alignment and structural inspection showed that all of these proteins retain Ser at the equivalent site to the Asp in the NAD(H)-dependent D-glycerate dehydrogenase. This indicates that, while the substitution of Asp with Ser at this site favors the binding of NADP(H), the respective binding affinity for NAD(H) and NADP(H) may also be affected by other factors as well. As such, the enzyme assay and structural characterization of VV2KR in the present study will facilitate future investigation into the structural basis for the observed coenzyme kinetic variation among different 2KDH proteins.

4.3.7 The catalytic site of VV2KR and substrate binding

The catalytic site of VV2KR is located at the cleft between the two major structural domains, close to the predicted coenzyme binding site (**Figure 3**). Based on previous study on other 2KDH structures, three electrostatically charged amino acids (232Arg, 261Glu and 279His) have been identified to form the substrate binding site of VV2KR (**Figure 8**), corresponding to 232Arg-261Glu-279His in *C. blumei* CbHPR, 269Arg-274Glu-293His in human GRHPR, 234Arg-263Glu-281His in *K. vulgare* Kv2DH. These 3 active site residues are also commonly conserved in most other 2KDH members including D-glycerate dehydrogenase, D-lactate dehydrogenase, phosphateglycerate dehydrogenase, Phosphite dehydrogenase, transcription co-repression Ctbp dehydrogenase, D-mandelate dehydrogenase and D-2-hydroxyisocaproate dehydrogenase (**Figure S3 sequence alignment**). Structural superimposition of the spatial arrangement of these two residues with *C. blumei* CbHPR, human GRHPR, *K. vulgare* Kv2DH is shown in **Figure 8**. Of the 3 active site residues in VV2KR, the OE1 and OE2 atom of Glu261 are hydrogen bonded (3.7 Å and 3.0 Å) with the ND1 atom of 279His. This hydrogen bonding interaction has been suggested to establish a charge relay system (Lamzin, Dauter et al. 1994, Booth, Conners et al. 2006, Janiak, Petersen et al. 2010). It helps to protonate the imidazole ring of 279His, which acts as the acid/base catalyst during the catalytic process and

attracts electron flow from the C2 carbonyl group of the substrate. 2KDH proteins have been known for their broad substrate specificity and are able to catalyze the reduction of a series of 2-oxyacids to the D-isomers of the respective 2-hydroxyacids. The side chain of 232Arg in VV2KR overlaps well with that of the corresponding catalytic Arg in *C. blumei* CbHPR, human GRHPR and other HPR proteins (**Figure 8**). Potential hydrogen bonding interactions between the side chain of 232Arg and the carboxyl oxygen atom and the carbonyl oxygen atom of the substrates help to orient the 2-keto group for electron attack, leading to the production of 2-hydroxyl acid.

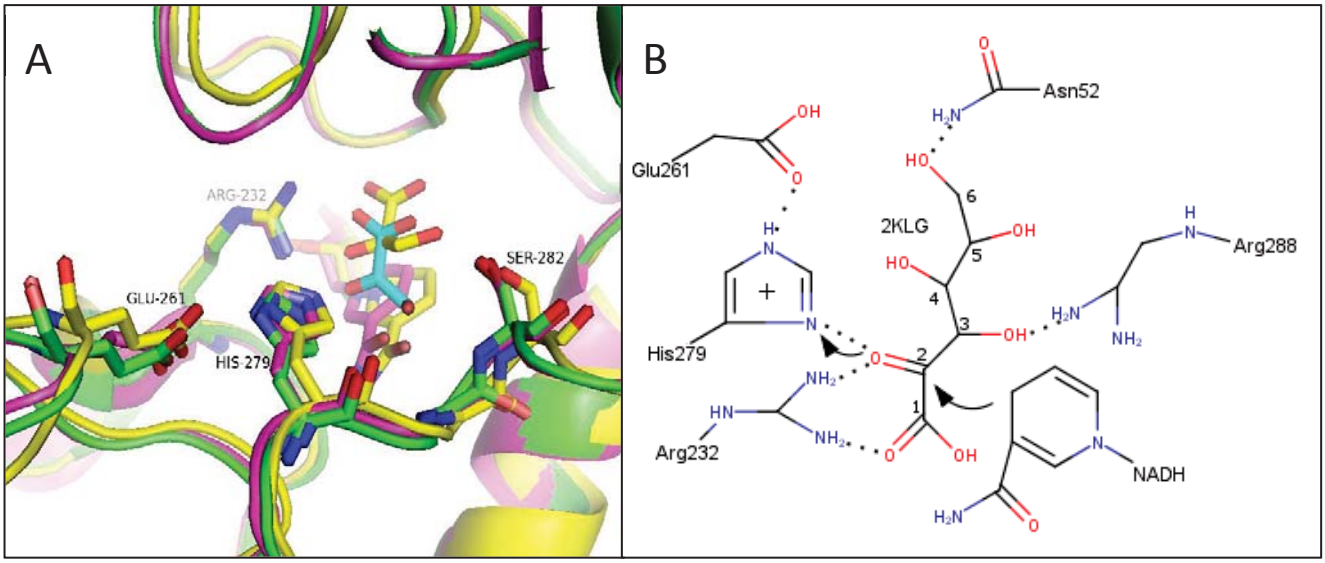


Figure 8. Active sites of VV2KR and the proposed reaction mechanism of VV2KR on the oxidation of 2KLG. A. Structural superimposition of the active site residues of VV2KR with other 2KDH structures; VV2KR (green, docked with hydroxypyruvate (cyan)), *C. blumei* CbHPR (magenta, bound with NADPH), human GRHPR (yellow, bound with NADPH + hydroxypyruvate). **B**. The proposed reaction mechanism of VV2KR on the oxidation of 2KLG (based on the substrate docking results in the present study). Dotted lines indicate hydrogen bonding interactions. Arrows display the electron transferring potential. (Residues numbering according to VV2KR)

Although the crystal structures of 2KDH proteins have been available for a long time, most of the early-determined structures correspond to apo or holo forms. These include the crystal structures for D-glycerate dehydrogenase (Goldberg, Yoshida et al. 1994), D-formate dehydrogenase (Lamzin, Dauter et al. 1994). Other determined

ternary structures contain substrate analogues but not native substrates, such as the D-2-hydroxyisocaproate dehydrogenase complexed with NAD^+ and 2-oxo-4-methylpentanoic acid (Dengler, Niefind et al. 1997) and the human transcription corepressor CtBP bound with NAD and acetic acid (Kumar, Carlson et al. 2002). Initial speculations on the substrate binding mode for hydroxypyruvate reductase or D-glycerate dehydrogenase were mainly deduced from kinetic data obtained using mutant enzymes and the binary structure of the *Lactobacillus helveticus* D-lactate dehydrogenase (PDB: 2DLD; bound with NAD^+ and oxalate). In this structure, the side chain of the 171Arg forms two hydrogen bonds with both of the two oxygen atoms of the oxalate carboxyl group. An electrostatic interaction is also believed to exist between the positively charged Arg and the electro-negative carboxyl group. The oxalate molecule is further stabilized by the hydrogen bond between the oxygen atom of its keto group and the ND2 atom from the His imidazole ring. As shown in **Figure 9A**, the nicotine amide ring of the coenzyme is positioned under the panel formed by the substrate and its coordinating residues, which makes the hydride transfer from NADH to the C2 keto group of the substrate possible. This binding mode (defined as mode A in present study) is consistent with the acetic acid-bound human CtBP structure (PBD: 1MX3; NAD^+ and acetic acid bound; 25% identity with VV2KR), in which both oxygen atoms from the carboxyl group were positioned to form hydrogen bonds with the side chain of 266Arg (**Figure 9A**). In addition, the O1 atom of the acetic acid molecule was further fixed by hydrogen binding with the side chain of another 97Arg and the backbone nitrogen atoms of 100Ser and 101Gly (**Figure 9**). However, the occurrence of binding mode A might be due to the small size of the bound substrate, which allows flexible binding adaptability for the acetic acid molecule to its surrounding electrostatic environment. Despite this, further support was observed for binding mode A in the human D-3-phosphateglycerate dehdydrogenase structure (PDB: 2G76; NADH and malate bound; 32% identity with VV2KR), in which the substrate malate overlaps well with the oxalate molecule in *L. helveticus* D-lactate dehydrogenase.

It was only until the revelation of the D-glyoxyate-bound human GRHPR structure

that direct evidence for the precise substrate binding mode for HPR proteins became available. This structure supports an alternative binding mode B distinct from the above mentioned mode A. In this second binding mode B, although the highly conserved catalytic 245Arg still coordinates the substrate (D-glyoxylate) with two hydrogen bonds, only one oxygen atom of the carboxyl group is involved. The other hydrogen binding oxygen atom belongs to the C2 carbonyl group (**Figure 9B**). For the structures of both GRHPRs, the C2 oxygen atom of the substrate also forms hydrogen bond with the ND2 of the imidazole ring of the catalytic His, facilitating the keto reduction at the C2 position. Binding mode B is corroborated by the tertiary structure of a *Lactobacillus pentosus* D-lactate dehydrogenase (1GDH) (Stoll, Kimber et al. 1996) and a *Lactobacillus bulgaricus* D-lactate dehydrogenase (1J49) (Razeto, Kochhar et al. 2002). The spatial arrangement of the predicted pyruvate for 1GDH and 1J49 resembles that observed for the D-glyoxylate in human GRHPR structure in relation to the three highly conserved catalytic residues, with two oxygen atoms from the carboxyl group and C2 carbonyl group, respectively oriented to the side chain of Arg by hydrogen binding interactions (Stoll, Kimber et al. 1996, Razeto, Kochhar et al. 2002). The C2 keto oxygen atom of pyruvate is also coordinated by a hydrogen bond with the ND2 atom of the His in *L. bulgaricus* D-lactate dehydrogenase (Razeto, Kochhar et al. 2002). In both human GRHPR and the modeled *L. bulgaricus* D-lactate dehydrogenase tertiary structures (Razeto, Kochhar et al. 2002, Booth, Conners et al. 2006), the substrates are located at the si face (prochirality) side of the nicotinamide ring, leading to the production of R-isomer of product instead of S-isomer (as in the *L. helveticus* D-lactate dehydrogenase, 2DLD). Later studies on the *Aquifex aeolicus* D-lactate dehydrogenase (PDB: 3KB6; NAD and D-lactic acid bound; 29% identity with VV2KR) also lend direct support to binding mode B.

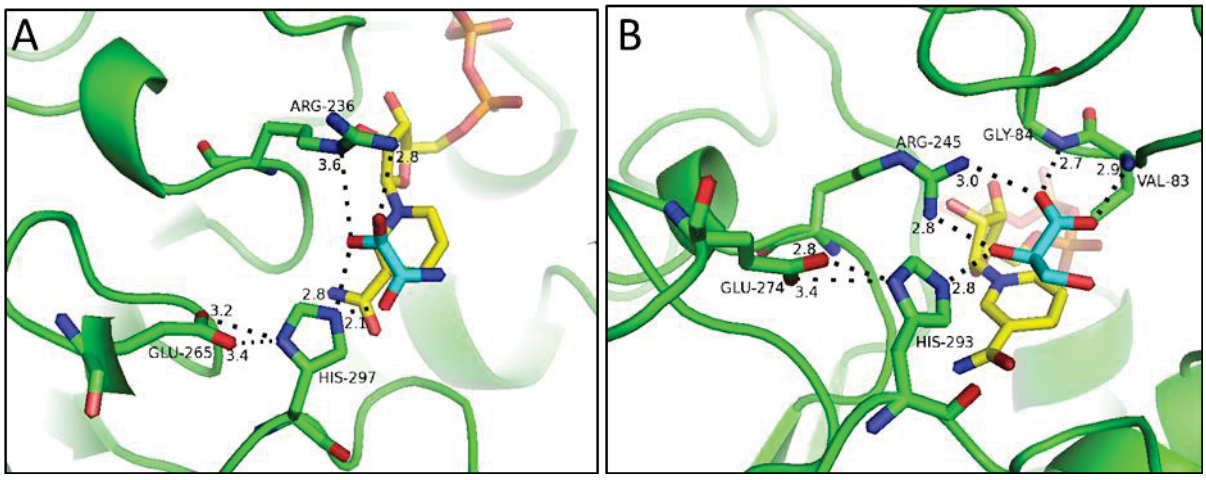


Figure 9. The substrate binding mode A and mode B. A: *L. helveticus* D-lactate dehydrogenase (green), PDB: 2DLD, bound with NADH (yellow) and oxalate (cyan) in mode A where both oxygen atoms of the C1 carboxyl group form hydrogen bonds with ARG-236. **B**: GRHPR, PDB: 2GCG, bound with NADP(H) and D-glycerate (cyan) in mode B where only one oxygen atom from the C1 carboxyl group together with the C2 oxygen atom form hydrogen bonds with ARG-245. (residues numbered according to respective protein)

In addition to the above mentioned distinct substrate binding modes for different 2KDH proteins, simultaneous occurrence of binding mode A and B was recently reported in the human CtBP structure (PDB: 4U6S; NAD^+ and phenylpyruvate bound; 25% identity with VV2KR) (Hilbert, Morris et al. 2015). Phenylpyruvate is known as a commonly accepted native substrate for 2KDH proteins. In the tertiary structure of this human CtBP, two distinct conformations of phenylpyruvate were present in approximately equal proportions (Hilbert, Morris et al. 2015). In the first conformation, the two carboxylate oxygen atoms are oriented to the side chain of the catalytic 266Arg within hydrogen binding distance, similar to that of human CtBP structure (Kumar, Carlson et al. 2002) and *L. helveticus* D-lactate dehydrogenase (2DLD), typical characteristics for binding mode A. However, the second conformation follows the pattern of binding mode B as mentioned above: the phenylpyruvate carbonyl is coordinated by two hydrogen bonds with the side chains of 266Arg and 315His, respectively, while only one oxygen atom from the C1 carboxyl group has hydrogen binding potential with 266Arg. This resembles the spatial arrangement of D-glyoxylate in human GRHPR, pyruvate in *L. bulgaricus*

D-lactate dehydrogenase and lactate in *A. aeolicus* D-lactate dehydrogenase. The presence of two conformations at the same time provides direct evidence that both substrate binding modes are possible for 2KDH proteins in the native structure. Depending on the specific active site environment and also the substrate molecule property, different proteins may adopt specific substrate binding modes to a variable degree.

Our VV2KR structure has the highest structural similarity with CbHPR and represents the second plant HPR ortholog characterized to date. Unlike CbHPR which is involved in the rosmarinic acid synthesis pathway, VV2KR is suspected to participate in the TA biosynthesis process in grapevine. However, both proteins were solved in either the apo form or the binary form (complexed with NADP(H)) and no direct evidence of substrate binding mode could be derived. Despite this, computational docking of a range of accepted substrates to CbHPR supports the possibility of two binding modes, demonstrating comparable binding energy for several validated native substrates. Noteworthy, in contrast to the case with *L. helveticus* L-lactate dehydrogenase, which adopts the first substrate binding mode and sees the re face of the bound oxalate facing to the nicotinamide ring of the coenzyme (leading to the production of S-isoform product), the structural information from the phenylpyruvate-bound human CtBP (two substrate binding modes present simultaneously) suggests that only the second substrate binding mode would facilitate the C2 keto reduction, leading to the production of R-isoform product. In the first binding mode conformation, hydride transfer to the C2 carbonyl from the nicotinamide ring of coenzyme becomes impossible. Similar conclusion has been made for the human GRHPR (Booth, Conners et al. 2006), CbHPR structures (Janiak, Petersen et al. 2010), *L. bulgaricus* D-lactate dehydrogenase (Razeto, Kochhar et al. 2002), *A. aeolicus* D-lactate dehydrogenase (Antonyuk, Strange et al. 2009). This may serve as the distinct substrate binding feature between 2KDH proteins and the L-specific oxidoreductases (represented by L-lactate dehydrogenase). In addition to *C. blumei* CbHPR, VV2KR also demonstrates relatively strong homology with huaman GRHPR, *P. horikoshii* Ph2GR and *K. vulgare* Kv2DH. Structural superimpositions of

VV2KR with these proteins suggest highly conserved active site residues, indicating that a similar substrate binding mechanism could be postulated for VV2KR.

4.3.8 Substrate docking and substrate specificity

To investigate the substrate binding mode and also to deduce the substrate specificity for VV2KR, the most commonly accepted substrates for HPR proteins (D-glyoxylate, pyruvate, 3-hydroxypyruvate), the proposed VV2KR substrate in the TA synthesis pathway in grapevine (2KLG) and its related five carbon homolog (L-ribulosonic acid) were docked into our VV2KR structure. L-ribulosonic acid was included in this study to investigate the effect of carbon chain length of 2-oxyacids on substrate binding affinity and specificity. Since significant position movement for the docked NADP was predicted due to the potential side chain movement of the coenzyme binding residue upon NADP(H) binding, the apo form VV2KR instead of the modeled binary NADP-VV2KR structure was used to avoid any unbeneficial hindrance effect by the coenzyme. Unlike the coenzyme binding residues, no significant side chain movement is predicted for the substrate binding residues based on structural comparison of VV2KR with the apo and binary structures of *C. blumei* CbHPR and also with the tertiary structure of human GRHPR. This suggests that substrate docking to the apo VV2KR structure should be feasible. The substrate binding site residues of VV2KR were determined by structural superimposition with the human GRHPR in complex with D-glycerate. The VV2KR residues surrounding the bound D-glycerate in human GRHPR were defined as potentially ligand coordinating.

Table 4. Summary of the docking results.

Substrate	Generated conformations	Binding mode	Lowest binding energy (ranking)	Binding mode frequency	Predicted product Stereochemistry
D-glyoxylate	11	B	-17.96 (1/11)	3	R
		A	-17.94 (2/11)	5	No-protonation(NP)
Pyruvate	9	B	-22.38 (1/9)	4	R
		A	-22.19 (2/9)	5	NP
3-Hydroxypyruvate	20	B	-27.01 (1/20)	12	R
		A	-26.56 (4/20)	7	NP
L-Ribulosonic acid	47	B	-35.60 (16/47)	4	R
(C5H8O6)		A	-33.29 (34/47)	5	NP
2KLG	54	B	-39.77 (13/54)	11	R
		A	-37.52 (20/54)	3	NP

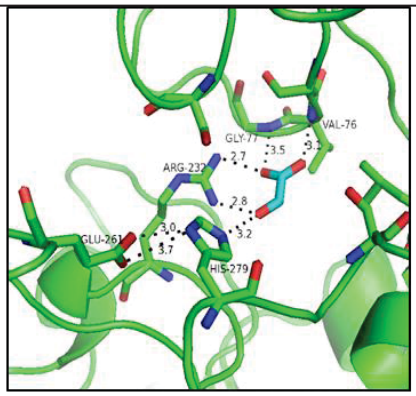
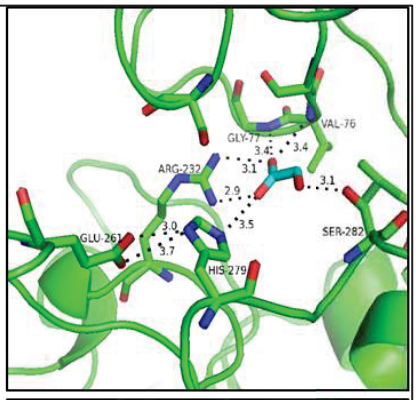
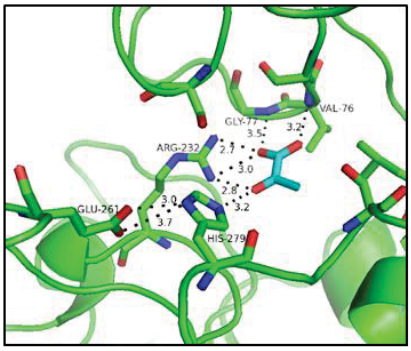
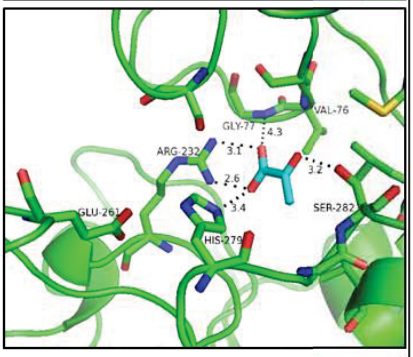
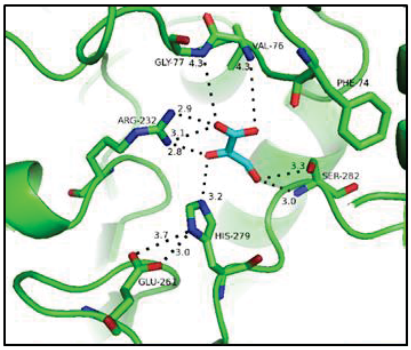
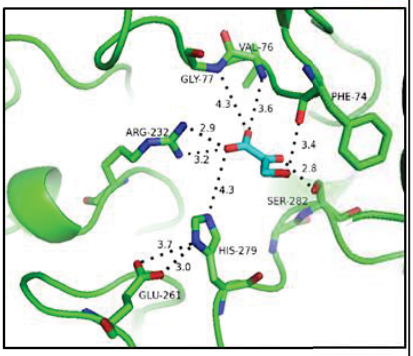
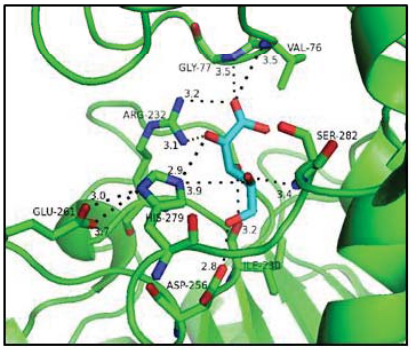
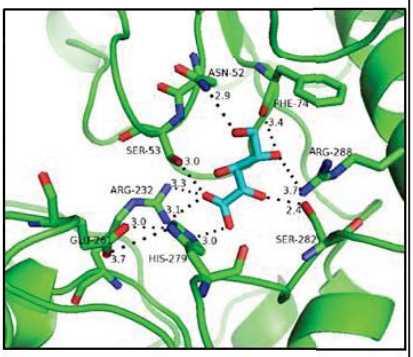
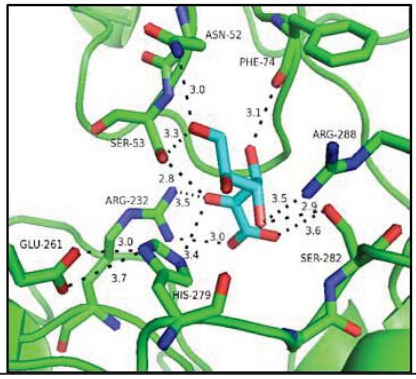
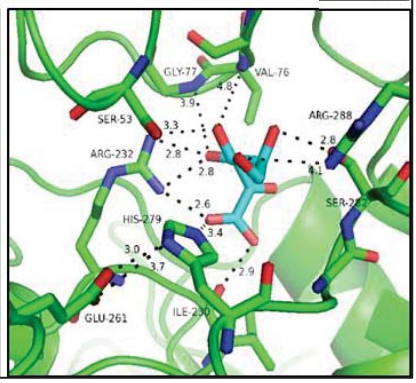
Glyoxylate		
Pyruvate		
Hydroxypyruvate		
L-Ribulosonic acid		
2KLG		

Figure 10. The two possible binding modes for each substrate. The left panel refers to binding mode B, the right panel binding mode A.

As shown in **Table 4**, varying numbers of conformations for the analyzed substrates were obtained. Generally, a higher degree of flexibility in substrate binding was observed as the number of the carbon atoms in the ligand increases. This also coincides with a decreasing substrate binding energy, which may be attributed to the binding interaction with the additional hydroxyl groups. A similar pattern was obtained for the docking analysis for *C. blumei* CbHPR (Janiak, Petersen et al. 2010). The top solution recognized as either adopting binding mode A or B for each substrate is illustrated in **Figure 10**. The corresponding binding energies for these top ranking solutions were also calculated. As shown in **Table 4**, all of the five substrates studied, including 2KLG, utilize binding mode B as the favored binding position (with the lowest binding energy), suggesting binding mode A would be the real binding position in the native structure of VV2KR for most accepted substrates. This agrees well with the substrate binding mode prediction for *C. blumei* CbHPR and other 2KDH proteins including human GRHPR. Notably, despite a preference for binding mode A, no significant binding energy difference between binding mode A and B was observed for the relatively small molecules (D-glyoxylate, pyruvate and hydroxypyruvate). Similar results have been obtained for the docking analysis with *C. blumei* CbHPR which also showed a slight preference for binding mode B, while no significant binding energy difference with binding mode A (Janiak, Petersen et al. 2010). In contrast to the relatively small substrates, docking results with L-rosmarinic acid and 2KLG revealed significantly lower binding energies for mode B than mode A (**Table 4**). More hydrogen binding interactions were predicted for these two molecules. Notably, unlike the relatively small substrate, docking L-rosmarinic acid and 2KLG to the binding pocket produced a number of conformations that could not be classified as either mode A or mode B. The top recognized binding mode B conformations only ranked 16/47 and 13/54 for L-rosmarinic acid and 2KLG, respectively. This may lead to a decreased catalytic efficiency for VV2KR on these two molecules. For binding mode A predicted for all five substrates, the C2 ketone oxygen atom was positioned

away from the nicotinamide ring of the 279His side chain, which acts as the catalytic base (**Figure 10**). This hinders the reduction at the C2 position and leads to no product. Notably, the top binding mode A solution for 2KLG has only one oxygen atom from the C1 carboxyl group positioned to 232Arg while normal binding mode A predicts both oxygen atoms of the carboxyl group binding to 232Arg side chain (**Figure 10**). None of the 54 conformations generated for 2KLG follow the normal binding mode A interactions. This might be due to the potential hindrance caused by its longer chain. In addition to the above mentioned active triad 232Arg-261Glu-279His, additional active site residues including 52Asn, 53Ser, 74Phe, 282Ser and 288Arg may also be involved in 2KLG binding and affect the catalytic process (**Figure 10**). In regards to substrate specificity, Verena etc demonstrated that *C. blumei* CbHPR has a preference for hydroxypyruvate over 4-hydroxyphenylpyruvate as the substrate despite that only the latter was involved in the rosmarinic acid synthesis pathway (Janiak, Petersen et al. 2010). Docking analyses with CbHPR revealed a lower binding energy for 4-hydroxyphenylpyruvate than hydroxypyruvate (Janiak, Petersen et al. 2010). A similar observation was made for VV2KR which showed much lower binding energy for 2KLG than hydroxypyruvate (**Table 4**). As mentioned above, the calculated binding energy appears to depend on the number of hydroxyl groups present in the substrate. This is supported by our docking analyses results which showed that the docked 2KLG exhibits more hydrogen bond interactions that other studied molecules (**Figure 10**). Although a lower binding energy would indicate a higher binding affinity for the substrate, it could also lead to a lower turnover rate for the reduction products. Therefore, it's reasonable to speculate that, while VV2KR may also have a higher binding affinity toward 2KLG, it would still use hydroxypyruvate more efficiently as a substrate than 2KLG. Further enzymatic characterization is needed to validate our hypothesis.

Kinetic data of human GRHPR has suggested a much higher affinity for hydroxypyruvate over pyruvate or glyoxylate (Mdluli, Booth et al. 2005). Booth etc reported that the preference for hydroxypyruvate of human GRHPR might be due to potential hydrogen bond interactions between the hydroxymethyl group of

hydroxypyruvate and 296Ser, 141Trp (from the dimer molecule), while the lower binding affinity for pyruvate and glyoxylate are more affected by their steric clash with 59Leu (Booth, Conners et al. 2006). Structural comparison of VV2KR with human GRHPR showed that only 296Ser was conserved in VV2KR while 59Leu and 141Trp were replaced by Ser and Gly, respectively. Further structural inspection showed that 141Trp in the dimerization loop of human GRHPR is positioned close to the substrate binding pocket. Based on this observation, Booth et al suspect that the substrate specificity of GRHPR might be dependent on the dimerization processes of the enzyme (Booth, Conners et al. 2006). This, however, does not seem to be the case with VV2KR and *C. blumei* CbHPR, for which the corresponding dimerization loops are much shorter and swung much further away from the substrate binding sites (**Figure S5**). Sequence alignment showed that GRHPR has an insertion of additional residues at this position (**Figure 7**). Interestingly, the archeal *P. horikoshii* Ph2GR resembles human GRHPR and also demonstrates a longer dimerization loop that is positioned close to the active site (**Figure S5**). Further structural comparison shows that *K. vulgare* Kv2DH and *P. thermophile* PtGR follow the same pattern with VV2KR and *C. blumei* CbHPR, with *P. thermophile* PtGR having the dimerization loop positioned even further away from the active site (**Figure S5**). Based on this observation, it seems that the mammalian HPRs have a closer relationship with its archeal counterparts while the plant HPRs, represented by VV2KR and *C. blumei* CbHPR, are more similar to its fungal and bacterial homologs. The variation of the spatial arrangement of the dimerization loop represents a notable difference among HPR from different organisms and may contribute to their distinct enzymatic profiles. In vivo enzymatic tests and site mutation studies are needed to validate these hypotheses.

4.3.9 The potential role of VV2KR in TA biosynthesis in the grapevine

The biological functions of plant HPRs are still largely unknown. Current knowledge on this protein family indicates that it may have multiple roles among different species. Three isoforms (AtHPR1-3) of HPR genes have been identified in *A. thaliana*, responsible for the conversion of hydroxypyruvate to D-glycerate. Each isoform is

targeted to the peroxisome (AtHPR1) (Mano, Hayashi et al. 1997), chloroplast (AtHRP3) (Timm, Florian et al. 2011) and cytosome (AtHRP2) (Timm, Nunes-Nesi et al. 2008), respectively, performing a complementary function in the photorespiratory cycle of plant. This biological role for HRP in plants has been corroborated with evidence from similar studies in barley (Murray, Blackwell et al. 1989), and spinach (Tolbert, Yamazaki et al. 1970).

The identified VV2KR in the present study has the highest similarity with AtHPR2 (74.8% sequence identity) and also shares 48% and 34% identity with AtHPR3 and AtHPR1, respectively. This suggests that VV2KR may also be targeted to the cytosol in the grapevine as AtHPR2 in *A. thaliana*. Given the strong homology between VV2KR and AtHPR2, it's very likely that VV2KR may also be involved in the photorespiratory cycle, a common metabolism process in higher plants. However, its potential role in other biological processes, such as TA biosynthesis in grapevine, could not be excluded. Notably, *C. blumei* CbHRP, sharing 78% identity with VV2KR, has been shown to be responsible for the rosmarinic acid synthesis in this plant, converting 4-hydroxyphenylpyruvate to 4-hydroxyphenylllactate, despite its preference for hydroxypyruvate as the substrate (Kim, Janiak et al. 2004, Janiak, Petersen et al. 2010). A notable characteristic for proteins of this class is that they generally demonstrate broad substrate specificity. They can catalyze the reduction of a range of 2-oxoacids to their 2-hydroxy forms, including D-glyoxylate, pyruvate, hydroxypyruvate and in this study 2KLG etc. Enzymes with broad substrate specificity playing multiple roles are frequent observations in various living organisms. The exact function of these proteins depends on transcriptional time, subcellular location and enzymatic properties. Noteworthy, of the 3 HPR isoforms in *A. thaliana*, AtHPR2, which shares the highest similarity with VV2KR, represents a cytosolic bypass to the photorespiratory core cycle and may only play a complementary role to the peroxisomal enzyme *A. thaliana* AtHRP1 (Timm, Nunes-Nesi et al. 2008). Therefore, it's possible that it may be associated with additional biological function. VV2KR corresponds to the AtHPR2 homolog in grapevine and, based on EST data, is mainly transcribed in the early stage of leaf

development, coinciding with the accumulation of TA in young leaves.

Substrate docking of 2KLG to our VV2KR structure revealed a favorable binding pattern for the proposed catalytic mechanism for proteins of this class, leading to the production of IA. The bacterial homologous enzymes being able to perform the same reaction have been identified from *E. coli* and *K. vulgare*. Notably, both of these enzymes belong to the 2KDH family. The crystal structure of *K. vulgare* Kv2DH (41% sequence identity) has also been determined and revealed highly conserved substrate binding site with VV2KR. Based on these observations, it's reasonable to speculate that VV2KR could be the most likely candidate enzyme responsible for the conversion of 2KLG to IA in grapevine. Gene transcriptional analyses (*in vivo* gene overexpression or knockout) are currently underway in our laboratory to validate the biological function of VV2KR in grapevine.

4.4 Conclusion

The unique accumulation of tartaric acid to be the dominant organic acid in grape berry and its intriguing relationship with AA metabolism have made it a prompt issue to understand the enzymatic mechanism of TA synthesis in grapevine. We identified a putative 2-keto-L-gulonic acid reductase (VV2KR) from grapevine which is able to reduce 2KLG to IA using NADP(H) as the preference coenzyme. We have determined the crystal structure of apo form VV2KR to a resolution of 2.1 Å. Structural analyses showed that 2KGR belongs to 2KDH superfamily. It has the highest structural similarity to *C. blumei* CbHPR. VV2KR represents the second plant structure being determined for proteins of this class. Substrate docking of 2KLG to VV2KR structure indicated a favorable binding mode for its reduction to IA. Therefore, VV2KR is very likely the candidate enzyme responsible for the conversion of 2KLG to IA in TA synthesis pathway in grapevine. Further *in vivo* transcriptional analyses, such as gene overexpression or knockout, are needed to validate this hypothesis.

4.5 Materials and Methods

4.5.1 Identification of a putative 2-keto-L-gulonate reductase gene from grapevine

The amino acid sequence of the *E. coli* 2-ketoalodnate reductase (Uniprot ID P37666),

which has been confirmed for its ability to catalyze the reversible reduction of 2KLG to IA, was used as a query to tBLASTn against the Gene Index Project: Gene Indices: Grape database. Homologous tentative consensus sequences (TC, aligned EST's forming a putative gene) with E-value of zero were selected and further refined to those ESTs occurring in the early stage of berry development. The encoded amino acid sequences of the selected TCs were analyzed for the presence of D-isomer 2-hydroxyacid reductase and NAD(P)H-binding domain signatures using Interpro online tool (http://www.ebi.ac.uk/interpro/) .

4.5.2 Molecular cloning

Gene specific primers containing NdeI (Forward CTACATATGATGGCGATGATGAAGCGAGTTGCTGAG) and BamHI (Reverse CCAGGATCCTTACTGGTACTGGCTTGCTAGTTGGCC) restriction sites were designed based on the obtained TC sequence to clone the putative 2-keto-L-gulonic acid reductase gene. PCR was performed using a pre-verasion *V. vinifera* cv Shiraz berry cDNA library (obtained from previous work in our lab) as the template. Digested PCR product was ligated into the pDRIVE (Qiagen, Australia) vector for sequencing confirmation. The putative grapevine 2-keto-L-gulonic acid reductase gene was finally cloned into the pET14b (Novagen, Germany) vector and transformed into BL21(DE3) pLysS-T1^{R} (Sigma-Aldrich, USA) for recombinant protein production.

4.5.3 Recombinant protein expression and purification

E. coli BL21(DE3) pLysS-T1^{R} harboring the pET14b-VV2KR construct was grown in LB culture (10 liters) containing 50 ng/ml ampicillin with shaking at 37 ℃ until OD600 reaches 0.6. Protein expression was initiated with the addition of 0.5 mM IPTG and maintained at 16 ℃ overnight. Cells were harvested by centrifugation and disrupted mechanically using a French Press disruptor. The lysis solution was centrifuged at 4 ℃， 40,000 × g for 30 mins. The supernatant was loaded onto a HisTrap FF Crude 5 × 1 ml column (GE Healthy Care Life Sciences) using a NGC^{TM} Chromatography System (Bio-Rad, Australia). The elution fractions containing the target protein were combined and loaded onto a Hitrap Q HP, 5 × 1 ml column (GE

Healthy Care Life Sciences) for further purification to remove contaminations. The fractions containing target protein were then combined and dialyzed overnight at 4 ℃ using Cellu Sep® T4 dialysis membrane. Part of the resulted enzyme solution was flash-frozen in liquid nitrogen with 10% glycerol and stored at -80 ℃ for enzyme assays, while the rest was concentrated using an Amicon Ultra-15 Centrifugal Filter (Millipore, Australia) to ~10 mg/ml for crystallization. The washing and elution buffers used for the His-tagged purification were 20 mM Tris (pH 8.0), 500 mM NaCl, 1mM 2-Mercaptoethanol (BME), 10 mM imidazole and 20 mM Tris (pH 8.0), 500 mM NaCl, 1mM 2-Mercaptoethanol (BME), 250 mM imidazole, respectively. The washing and elution buffers used for the Q anion exchange column were 20 mM Tris (pH 8.0), 50 mM NaCl, 1mM DTT and 20 mM Tris (pH 8.0), 750 mM NaCl, 1 mM DTT, respectively. The dialysis buffer contains 20 mM Tris (pH 8.0), 50 mM NaCl and 1 mM DTT. The elution fractions containing the target protein were checked by 10% SDS-PAGE.

4.5.4 Enzyme kinetic assays

The activity of recombinant VV2KR was monitored at 340 nm, 37 ℃ in 96-well flat-bottom UV plate (Costar, USA; with pathlength correction) using a FLUOstar Omega spectrophotometer (BMG Labtech, Germany). Each reaction contains 187 ul buffer solutions, 2 ul coenzyme solutions, 1 ul enzymes sample and 10 ul substrate solution. Stock solutions of coenzyme and substrate were prepared using the corresponding reaction buffer solution and were diluted to varied degrees to achieve the desired concentration in the final reaction. Enzyme samples were also diluted accordingly to facilitate the measurement of the initial reaction velocity. All assays were carried out in triplication except the test of 2KGA with NADPH, for which duplication was performed. A master reagent mix including purified enzyme, buffer and coenzyme was dispensed into the 96-well plate. The optimal pH condition was determined using a range of buffers (100 mM HEPES pH 6.5-7.0-7.5-8.0, 100 mM Tris-HCl pH 6.0-6.5-7.0-7.5-8.0-8.5-9.0, 100 mM NaOH-Glycine pH 8.0-8.5-9.0-9.5-10.0). Kinetic tests were performed at the determined optimal pH condition. Negative controls including no substrate were also tested. The rate of

absorbance changes was processed using MARS Data Analysis Software (version 3.10 R6) and converted to the quantity of specific enzyme activity using a extinction coefficient of 6.22 $mM^{-1}*cm^{-1}$ for NADH and NADPH. The obtained initial reaction rates at different substrate concentrations were fitted into the Michaelis–Menten equation using GraphPad Prism software (version 6.0; GraphPad Software, Inc. CA, USA) to calculate the K_m and V_{max}.

4.5.5 Protein crystallization

Several commercially available sparse matrix screens were tested to obtain the crystallization conditions. Crystals were grown in sitting drops by vapor diffusion in 96-well INTELLI-PLATE (Art Robbins Instruments) at 16 ℃. Each well consisted of 1-ul/1-ul ratio of well solution to protein solution and the well solution reservoir (75 ul). The screening plates were setup using a Phoenix robot system (Art Robbins Instruments). In addition to the ligand-free sample, an enzyme solution containing 2 mM NADPH was also tested to obtain NADPH-VV2KR co-crystals. Crystals were obtained in a number of solutions in the Wizzard Classic 1 and 2 block (Rigaku, USA) and polyethylene glycol (PEG)/ion screen (Hampton Research). All crystals grew into full size (sheet shape) within 2 days. Crystals for data collection were produced with the well solution containing 1260 mM ammonium sulfate, 100 mM sodium acetate/acetic acid pH 4.5, 200 mM sodium chloride.

4.5.6 Data collection, structure determination and refinement

Crystals were transferred to Methyl-2,4-pentanediol (MPD) for cryoprotection and were flash frozen to 100 K by submersion into liquid nitrogen. Data collection was performed at 100 K on the MX1 beamline at the Australian Synchrotron, Victoria, Australia. The peak wavelength was at 0.9537 Å. Complete datasets were obtained with a rotation of $\Delta\varphi$ = 0.5 °. Data were indexed, integrated and scaled using the iMosflm (Battye, Kontogiannis et al. 2011) and Scala (Winn, Ballard et al. 2011) programs within the CCP4 package. The structure was solved to a resolution of 2.1 Å, in space group $P2_1$. The phase problem was solved by molecular replacement method using the program Phaser MR (Mccoy, Grosse-Kunstleve et al. 2007). A homology model of VV2KR was created using ICM Pro (Molsoft LLC, La Jolla, CA, USA)

with the protein structure of CbHPR (PDB: 3BAZ, sequence identity at ~79%) as the template. This model was used as a search model in Phaser MR, which yielded in a single strong solution. Data quality analysis using Phenix.xtriage (Adams, Afonine et al. 2010) revealed the presence of twinning in the crystal structure, with the twin law h, -k, -l and a twinning fraction of 0.35. The generated model was then refined using Phenix.refine. The resulted model was then subjected to multiple rounds of rebuilding using Coot (Emsley, Lohkamp et al. 2010) followed by refinement in Phenix.refine until R factors reached convergence. The final model was checked for its stereochemistry using MolProbity (Chen, Arendall et al. 2010). The solved structure results along with data processing and refinement statistics are presented in **Table 2**.

4.5.7 Substrate docking

Substrate docking was performed in ICM Pro (Molsoft LLC, La Jolla, CA, USA) using the interactive docking module. Ligand molecules were downloaded either from the RCSB Protein Data Bank (Berman, Westbrook et al. 2000) or from the ChemSpider (http://www.chemspider.com/) website and were loaded into ICM Pro for electrostatic charge assignment (formal charges), pH set up (pH=7.0) and energy minimization. The VV2KR structure was superimposed with the tertiary human GRHPR structure (PDB: 2GCG) for the determination of the substrate binding pocket (VV2KR residues surrounding the GRHPR D-glycerate ligand were selected). Potential substrate conformations were generated and ranked according to their free binding energy. Each conformation was compared with the putative binding mode A and B and analyzed for the hydrogen binding interaction potential with the active site residues.

4.6 Supplementary Figures

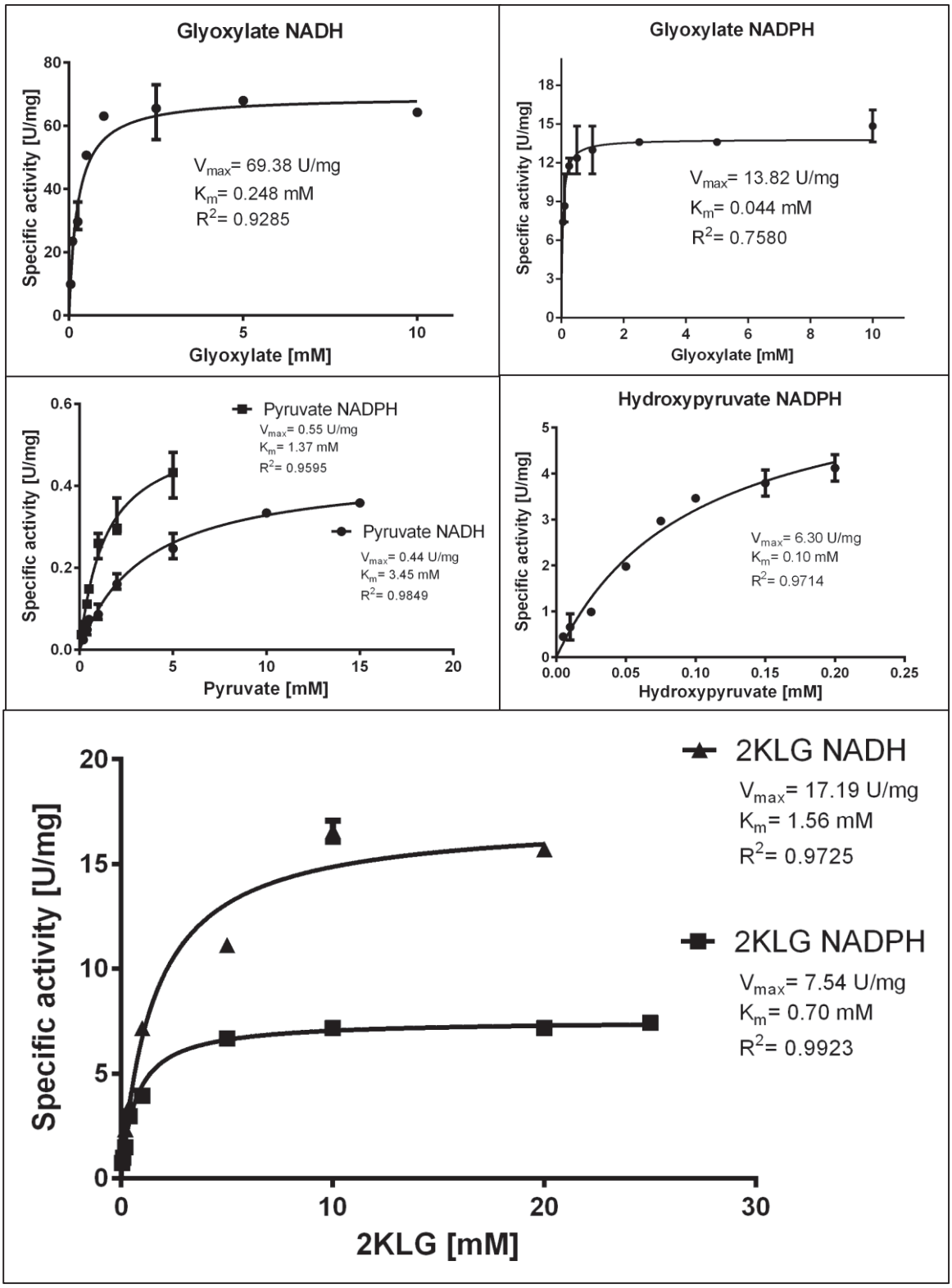


Figure S1. Enzymatic characterization of VV2KR with different substrates.

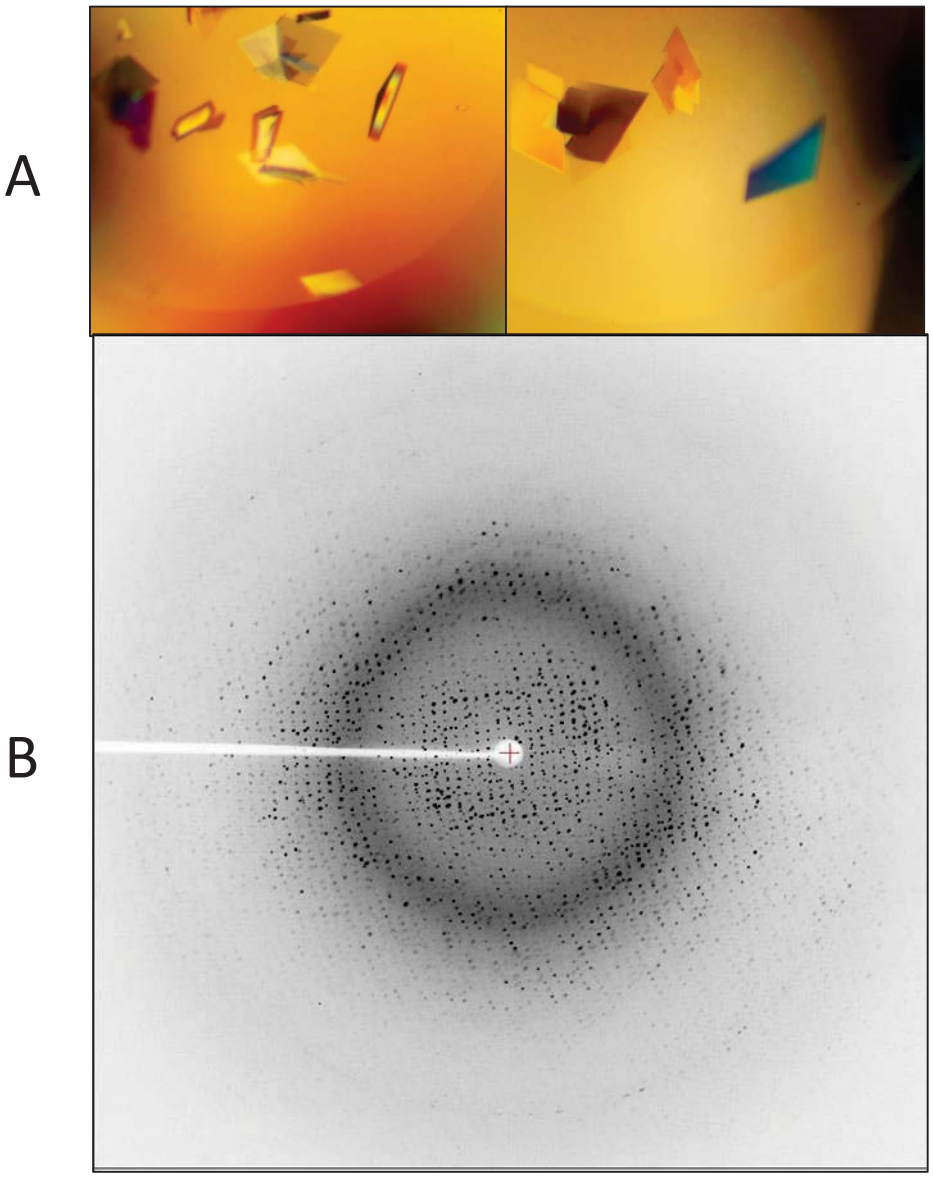


Figure S2. Pictures displaying VV2KR crystal morphology and diffraction pattern. A. VV2KR crystals growing into thin "sheet" forms. **B**. VV2KR diffraction pattern collected with synchrotron light source (wavelength 0.9537 Å, φ = 0°)。

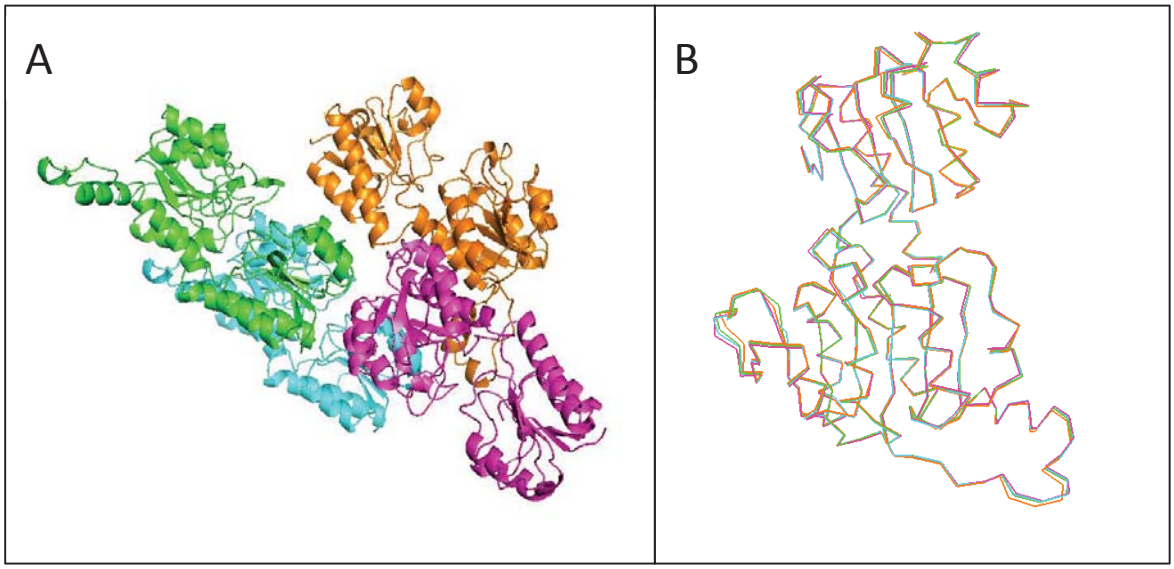


Figure S3. The spatial arrangement (A) and the structural superimposition (B) of the 4 subunits of VV2KR in the asymmetric unit. Chain A (cyans), B (green), C (orange), D (pink).

Interface Summary XML

	Structure 1		Structure 2	
Selection range	A		D	
class	Protein		Protein	
symmetry operation	x,y,z		,,	
symmetry ID	1_555		0_555	
Number of atoms				
interface	251	10.7%	247	10.5%
surface	1330	56.8%	1329	56.4%
total	2341	100.0%	2356	100.0%
Number of residues				
interface	66	21.2%	65	20.8%
surface	279	89.4%	279	89.4%
total	312	100.0%	312	100.0%
Solvent-accessible area, A				
interface	2439.0	16.6%	2411.1	16.4%
total	14698.9	100.0%	14675.5	100.0%
Solvation energy, kcal/mol				
isolated structure	-298.2	100.0%	-298.1	100.0%
gain on complex formation	-11.4	3.8%	-13.3	4.5%
average gain	-6.6	2.2%	-6.3	2.1%
P-value	0.174		0.076	

Hydrogen bonds XML

##	Structure 1	Dist. [A]	Structure 2
1	A:LYS 141[HZ1]	2.13	D:GLU 102[OE1]
2	A:ARG 117[HH12]	2.04	D:ASP 103[O]
3	A:THR 143[H]	2.26	D:ASP 106[OD2]
4	A:LEU 276[H]	2.27	D:ASP 123[OD2]
5	A:VAL 285[H]	2.17	D:THR 139[O]
6	A:ARG 162[HH21]	2.12	D:THR 143[OG1]
7	A:LYS 132[HZ3]	1.77	D:ASN 263[OD1]
8	A:ARG 127[HH11]	2.44	D:LEU 269[O]
9	A:ARG 127[HH12]	2.06	D:MET 271[O]
10	A:ARG 127[HH22]	2.25	D:VAL 274[O]
11	A:ARG 117[HH21]	2.27	D:VAL 280[O]
12	A:ARG 117[HH22]	2.24	D:GLY 281[O]
13	A:THR 139[OG1]	3.30	D:GLY 283[O]
14	A:LYS 141[H]	2.23	D:GLU 286[OE1]
15	A:LYS 145[HZ1]	2.18	D:GLU 286[OE2]
16	A:GLU 102[OE1]	1.68	D:LYS 141[HZ1]
17	A:ASP 103[O]	2.23	D:ARG 117[HH12]
18	A:ASP 106[OD2]	2.32	D:THR 143[H]
19	A:ASP 123[OD2]	2.30	D:LEU 276[H]
20	A:THR 139[O]	1.98	D:VAL 285[H]
21	A:THR 143[OG1]	2.00	D:ARG 162[HH21]
22	A:LEU 269[O]	2.21	D:ARG 127[HH11]
23	A:MET 271[O]	2.07	D:ARG 127[HH12]
24	A:VAL 274[O]	2.26	D:ARG 127[HH22]
25	A:VAL 280[O]	2.46	D:ARG 117[HH21]
26	A:VAL 280[O]	2.34	D:ARG 117[HH22]
27	A:GLY 283[O]	2.07	D:THR 139[H]
28	A:GLU 286[OE1]	2.40	D:LYS 141[H]
29	A:GLU 286[OE2]	2.39	D:LYS 145[HZ1]

Salt bridges XML

##	Structure 1	Dist. [A]	Structure 2
1	A:LYS 141[NZ]	2.96	D:GLU 102[OE1]
2	A:ARG 117[NH1]	3.86	D:ASP 103[OD1]
3	A:ARG 124[NH1]	3.42	D:ASP 123[OD1]
4	A:LYS 145[NZ]	2.98	D:GLU 286[OE2]
5	A:GLU 102[OE1]	2.52	D:LYS 141[NZ]
6	A:ASP 103[OD1]	3.57	D:ARG 117[NH1]
7	A:ASP 123[OD1]	3.51	D:ARG 124[NH1]
8	A:GLU 286[OE2]	3.21	D:LYS 145[NZ]

No disulfide bonds found
No covalent bonds found

Figure S4. Dimer interface analyses (subunit A and D) results generated with PDBePISA (http://www.ebi.ac.uk/pdbe/pisa/).

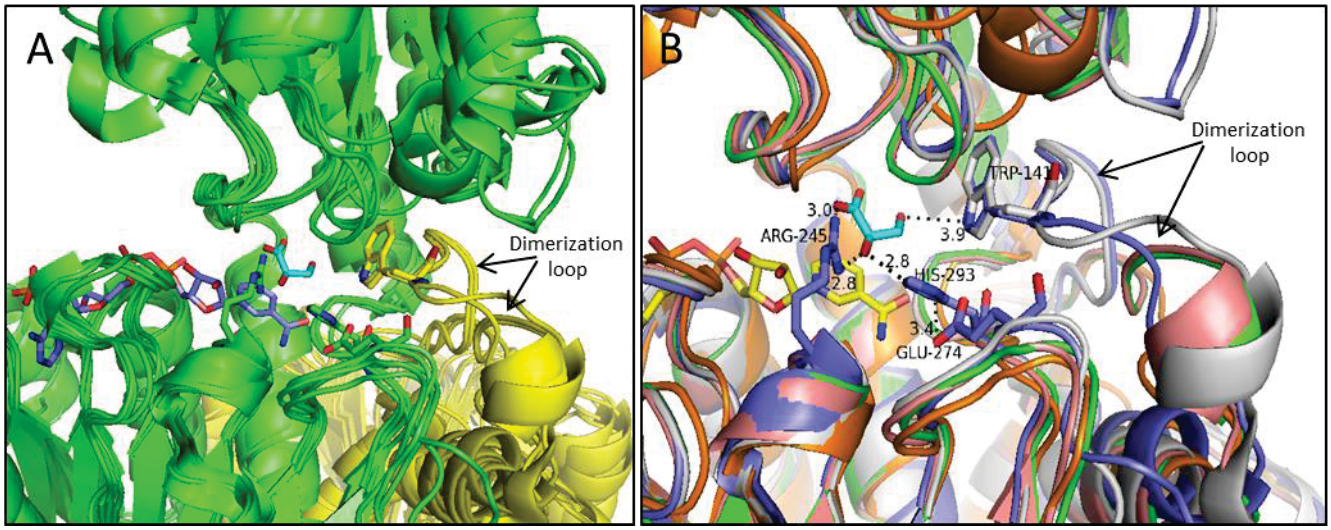


Figure S5. Structural variation of dimerization loop and its potential effect on the substrate binding. A. Structural superimposition of VV2KR, *C. blumei* CbHPR, human GRHPR, *P. horikoshii* Ph2GR and *P. thermophile* PtGR dimer structures (subunit 1 as green; subunit 2 as yellow; NADPH and D-glycerate from GRHPR as purple and cyans respectively;). B. Displays structural variation of the dimerization loops. Active site residues of GRHPR were shown in sticks and labeled correspondingly (numbering in GRHPR). VV2KR (green), CbHPR (salmon, 3BAZ), GRHPR (purple, 2GCG), Ph2GR (2DBZ, grey), PtGR (3WNV, orange).

4.7 Contributions

Author contributions: YJ designed the research. CAB performed gene cloning and transformation. YJ and JBB purified the protein for enzymatic tests and crystallization. JBB collected the diffraction data. YJ and JBB solved the structure together. YJ, CS and CAB performed the enzymatic tests. YJ performed kinetic data and protein structure analyses. YJ drafted the manuscript. JBB and CMF contributed to the editing of the manuscript. CMF and JBB supervised the project.

4.8 Acknowledgement

We thank Dr Vanessa Melino for providing the pre-veraison berry cDNA and Mr Ian Lau for help with enzyme purification and setting up the crystallization plates. We acknowledge the Australian Synchrotron staff for help with data collection.

4.9 Abbreviations

2KGA	2-keto-D-gluconate
2KLG	2-keto-L-gulonic acid
5KGA	5-keto-D-gluconic acid
AtHPR1-3	*Arabidopsis thaliana* hydroxypyruvate/glyoxylate reductase isoform1-3
CBD	Coenzyme-binding domain
CtBP	C-terminal binding protein
2KDH	D-isomer-specific 2-hydroxyacid dehydrogenase
EST	Expressed sequence tag
Kv2DH	*Ketogulonicigenium vulgare* 2-hydroxyacid dehydrogenase
AA	L-ascorbic acid
IA	L-idonic acid
TA	L-tartaric acid
LIDH	L-idonate-5-dehydrogenase
HPR	Hydroxypyruvate/glyoxylate reductase
GRHPR	*Homo sapiens* HPR
Ph2GR	*Pyrococcus horikoshii* glyoxylate reductase
Pt2GR	*Paecilomyes thermophile* 2GR
RMSD	Root mean square deviation
SDH	Sorbitol dehydrogenase
SBD	Substrate-binding domain
VV2KR	*Vitis vinifera* 2-keto-L-gulonic acid reductase

4.10 References

Adams, P. D., P. V. Afonine, et al. (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallographica Section D-Biological Crystallography **66**: 213-221.

Ameyama, M. and O. Adachi (1982). "2-Keto-D-Gluconate reductase from acetic-acid bacteria." Methods in Enzymology **89**: 203-210.

Antonyuk, S. V., R. W. Strange, et al. (2009). "Structure of D-lactate dehydrogenase from Aquifex aeolicus complexed with NAD(+) and lactic acid (or pyruvate)." Acta Crystallographica Section F-Structural Biology and Crystallization Communications **65**: 1209-1213.

Baack, R. D., J. Markwell, et al. (2003). "Kinetic behavior of the Arabidopsis thaliana leaf formate dehydrogenase is thermally sensitive." Journal of Plant Physiology **160**(5): 445-450.

Barch, W. E. (1933). "The Oxidation of 5-Keto-gluconic acid with nitric acid in the presence of vanadium." Journal of The American Chemistry Society **55**: 3653-3658.

Battye, T. G. G., L. Kontogiannis, et al. (2011). "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM." Acta Crystallographica Section D-Biological Crystallography **67**: 271-281.

Berman, H. M., J. Westbrook, et al. (2000). "The protein data bank." Nucleic Acids Research **28**(1): 235-242.

Bernard, N., K. Johnsen, et al. (1994). "NAD^+-dependent d-2-hydroxyisocaproate dehydrogenase of Lactobacillus-Delbrueckii subsp bulgaricus - gene cloning and enzyme characterization." European Journal of Biochemistry **224**(2): 439-446.

Booth, M. P. S., R. Conners, et al. (2006). "Structural basis of substrate specificity in human glyoxylate reductase/hydroxypyruvate reductase." Journal of Molecular Biology **360**(1): 178-189.

Chen, V. B., W. B. Arendall, et al. (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallographica Section D-Biological Crystallography **66**: 12-21.

Coombe, B. G. and M. G. McCarthy (2000). "Dynamics of grape berry growth and physiology of ripening." Australian Journal of Grape and Wine Research **6**(2): 131-135.

DeBolt, S., D. R. Cook, et al. (2006). "L-Tartaric acid synthesis from vitamin C in higher plants." Proceedings of the National Academy of Sciences of the United States of America **103**(14): 5608-5613.

DeBolt, S., J. Hardie, et al. (2004). "Composition and synthesis of raphide crystals and druse crystals in berries of Vitis vinifera L. cv. Cabernet Sauvignon: Ascorbic acid as precursor for both oxalic and tartaric acids as revealed by radiolabelling studies." Australian Journal of Grape and Wine Research **10**(2): 134-142.

Dengler, U., K. Niefind, et al. (1997). "Crystal structure of a ternary complex of

D-2-hydroxyisocaproate dehydrogenase from Lactobacillus casei, NAD(+) and 2-oxoisocaproate at 1.9 angstrom resolution." Journal of Molecular Biology **267**(3): 640-660.

Dey, S., G. A. Grant, et al. (2005). "Crystal structure of Mycobacterium tuberculosis D-3-phosphoglycerate dehydrogenase." Journal of Biological Chemistry **280**(15): 14892-14899.

Emsley, P., B. Lohkamp, et al. (2010). "Features and development of Coot." Acta Crystallographica Section D-Biological Crystallography **66**: 486-501.

George Wagner, F. A. L. (1974). "Ascorbic acid metabolism in Vitaceae conversion to tartaric acid and hexoses." Plant Physiology **54**(5): 784-787.

Goldberg, J. D., T. Yoshida, et al. (1994). "Crystal structure of a NAD-dependent D-glycerate dehydrogenase at 2.4-angstrom resolution." Journal of Molecular Biology **236**(4): 1123-1140.

Grant, G. A. (1989). "A new family of 2-hydroxyacid dehydrogenases." Biochemical and Biophysical Research Communications **165**(3): 1371-1374.

Han, X. D., X. H. Xiong, et al. (2014). "Crystallization and structural analysis of 2-hydroxyacid dehydrogenase from Ketogulonicigenium vulgare." Biotechnology Letters **36**(2): 295-300.

Hayward, S. and H. J. C. Berendsen (1998). "Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme." Proteins-Structure Function and Genetics **30**(2): 144-154.

Hilbert, B. J., B. L. Morris, et al. (2015). "Structure-guided design of a high affinity inhibitor to human CtBP." ACS Chemical Biology **10**(4): 1118-1127.

Iland, P. G. and B. G. Coombe (1988). "Malate, tartrate, potassium, and sodium in flesh and skin of shiraz grapes during ripening- concentration and compartmentation." American Journal of Enology and Viticulture **39**(1): 71-76.

Isbell, H. S. and N. B. Holt (1945). "Oxidation of galacturonic acid and of 5-Keto-Gluconic acid in alkaline solution." Journal of Research of the National Bureau of Standards **35**(5): 433-438.

Janiak, V., M. Petersen, et al. (2010). "Structure and substrate docking of a hydroxy(phenyl)pyruvate reductase from the higher plant Coleus blumei Benth." Acta Crystallographica Section D-Biological Crystallography **66**: 593-603.

Jia, Y., D. C. J. Wong, et al. (2015). "New insights into the evolutionary history of plant sorbitol dehydrogenase." BMC Plant Biology **15**.

Kazumi Saito, F. A. L. (1979). "The metabolism of L-[6-14C]ascorbic acid in detached grape leaves." Plant Cell Physiology **20**: 1481-1488.

Kim, K. H., V. Janiak, et al. (2004). "Purification, cloning and functional expression of hydroxyphenylpyruvate reductase involved in rosmarinic acid biosynthesis in cell cultures of Coleus blumei." Plant Molecular Biology **54**(3): 311-323.

Klasen, R., S. Bringermeyer, et al. (1992). "Incapability of gluconobacter-oxydans to produce tartaric acid." Biotechnology and Bioengineering **40**(1): 183-186.

Kleczkowski, L. A. and G. E. Edwards (1989). "Identification of hydroxypyruvate and

glyoxylate reductases in maize leaves." Plant Physiology **91**(1): 278-286.
Kleczkowski, L. A., G. E. Edwards, et al. (1990). "Enzymology of the reduction of hydroxypyruvate and glyoxylate in a mutant of barley lacking peroxisomal hydroxypyruvate reductase." Plant Physiology **94**(2): 819-825.
Kleczkowski, L. A. and D. D. Randall (1988). "Purification and characterization of a novel NADPH(NADH)-dependent hydroxypyruvate reductase from spinach leaves - Comparison of immunological properties of leaf hydroxypyruvate reductases." Biochemical Journal **250**(1): 145-152.
Kumar, V., J. E. Carlson, et al. (2002). "Transcription corepressor CtBP is an NAD(+)-regulated dehydrogenase." Molecular Cell **10**(4): 857-869.
Lamzin, V. S., Z. Dauter, et al. (1994). "High resolution structures of holo and apo formate dehydrogenase." Journal of Molecular Biology **236**(3): 759-785.
Loewus, F. A. (1999). "Biosynthesis and metabolism of ascorbic acid in plants and of analogs of ascorbic acid in fungi." Phytochemistry **52**(2): 193-210.
Loewus, F. A. and H. A. Stafford (1958). "Observations on the incorporation of C-14 into tartaric acid and the labeling pattern of D-glucose from an excised grape leaf administered L-ascorbic acid-6-C-14." Plant Physiology **33**(2): 155-156.
Malipiero, U., H. P. Ruffner, et al. (1987). "Ascorbic to tartaric acid conversion in grapevines." Journal of Plant Physiology **129**(1–2): 33-40.
Mano, S., M. Hayashi, et al. (1997). "Hydroxypyruvate reductase with a carboxy-terminal targeting signal to microbodies is expressed in Arabidopsis." Plant and Cell Physiology **38**(4): 449-455.
Manuel Salgado, J., N. Rodriguez, et al. (2010). "Improving downstream processes to recover tartaric acid, tartrate and nutrients from vinasses and formulation of inexpensive fermentative broths for xylitol production." Journal of the Science of Food and Agriculture **90**(13): 2168-2177.
Mattick, L. R., R. A. Plane, et al. (1980). "Lowering wine acidity with carbonates." American Journal of Enology and Viticulture **31**(4): 350-355.
Mccoy, A. J., R. W. Grosse-Kunstleve, et al. (2007). "Phaser crystallographic software." Journal of Applied Crystallography **40**: 658-674.
Mdluli, K., M. P. S. Booth, et al. (2005). "A preliminary account of the properties of recombinant human glyoxylate reductase (GRHPR), LDHA and LDHB with glyoxylate, and their potential roles in its metabolism." Biochimica Et Biophysica Acta-Proteins and Proteomics **1753**(2): 209-216.
Murray, A. J. S., R. D. Blackwell, et al. (1989). "Metabolism of hydroxypyruvate in a mutant of barley lacking NADH-dependent hydroxypyruvate reductase, an important photorespiratory enzyme-activity." Plant Physiology **91**(1): 395-400.
Nardini, M., S. Spano, et al. (2003). "CtBP/BARS: a dual-function protein involved in transcription co-repression, and Golgi membrane fission." EMBO Journal **22**(12): 3122-3130.
Ohshima, T., N. Nunoura-Kominato, et al. (2001). "A novel hyperthermophilic archaeal glyoxylate reductase from Thermococcus litoralis - Characterization, gene cloning, nucleotide sequence and expression in Escherichia coli."

European Journal of Biochemistry **268**(17): 4740-4747.

Parnham, M. J. and K. Kesselring (1985). "Rosmarinic acid." Drugs of the Future **10**(9): 756-757.

Razeto, A., S. Kochhar, et al. (2002). "Domain closure, substrate specificity and catalysis of D-Lactate dehydrogenase from Lactobacillus bulgaricus." Journal of Molecular Biology **318**(1): 109-119.

Saito, K. (1984). "Synthesis of L-(+)-tartaric acid from 5-Keto-D-gluconic acid in Pelargonium." Plant and Cell Physiology **25**(7): 1223-1232.

Saito, K. (1992). "Metabolism of L-threotetruronic acid by Pelargonium crispum." Phytochemistry **31**(4): 1219-1222.

Saito, K. and Z. Kasai (1969). "Tartaric acid synthesis from L-ascorbic acid-C-14 in grape berries." Phytochemistry **8**(11): 2177-2182.

Saito, K. and Z. Kasai (1984). "Synthesis of L-(+)-tartaric acid from L-ascorbic acid via 5-keto-D-gluconic acid in grapes." Plant Physiology **76**(1): 170-174.

Saito, K., J. Ohmoto, et al. (1997). "Incorporation of O-18 into oxalic, L-threonic and L-tartaric acids during cleavage of L-ascorbic and 5-keto-D-gluconic acids in plants." Phytochemistry **44**(5): 805-809.

Salusjarvi, T., M. Povelainen, et al. (2004). "Cloning of a gluconate/polyol dehydrogenase gene from Gluconobacter suboxydans IFO 12528, characterisation of the enzyme and its use for the production of 5-ketogluconate in a recombinant Escherichia coli strain." Applied Microbiology and Biotechnology **65**(3): 306-314.

Schuller, D. J., G. A. Grant, et al. (1995). "The Allosteric Ligand Site in the V-Max-Type Cooperative Enzyme Phosphoglycerate Dehydrogenase." Nature Structural Biology **2**(1): 69-76.

Smirnoff, N. and G. L. Wheeler (2000). "Ascorbic acid in plants: Biosynthesis and function." Critical Reviews in Biochemistry and Molecular Biology **35**(4): 291-314.

Stafford, H. A. (1959). "Distribution of tartaric acid in the leaves of certain angiosperms." American Journal of Botany **46**(5): 347-352.

Stoll, V. S., M. S. Kimber, et al. (1996). "Insights into substrate binding by D-2-ketoacid dehydrogenases from the structure of Lactobacillus pentosus D-lactate dehydrogenase." Structure **4**(4): 437-447.

Sweetman, C., L. G. Deluc, et al. (2009). "Regulation of malate metabolism in grape berry and other developing fruits." Phytochemistry **70**(11-12): 1329-1344.

Szabo, E., A. Thelen, et al. (1999). "Fungal elicitor preparations and methyl jasmonate enhance rosmarinic acid accumulation in suspension cultures of Coleus blumei." Plant Cell Reports **18**(6): 485-489.

Thompson, J. R., J. K. Bell, et al. (2005). "V-max regulation through domain and subunit changes. The active form of phosphoglycerate dehydrogenase." Biochemistry **44**(15): 5763-5773.

Timm, S., A. Florian, et al. (2011). "The hydroxypyruvate-reducing system in Arabidopsis: Multiple enzymes for the same end." Plant Physiology **155**(2): 694-705.

Timm, S., A. Nunes-Nesi, et al. (2008). "A cytosolic pathway for the conversion of hydroxypyruvate to glycerate during photorespiration in Arabidopsis." Plant Cell **20**(10): 2848-2859.

Tolbert, N. E., R. K. Yamazaki, et al. (1970). "Localization and properties of hydroxypyruvate and glyoxylate reductases in spinach leaf particles." Journal of Biological Chemistry **245**(19): 5129-&.

Truesdell, S. J., J. C. Sims, et al. (1991). "Pathways for metabolism of ketoaldonic acids in an Erwinia sp." Journal of Bacteriology **173**(21): 6651-6656.

Wada, Y., S. Iwai, et al. (2008). "A new family of D-2-hydroxyacid dehydrogenases that comprises D-mandelate dehydrogenases and 2-ketopantoate reductases." Bioscience Biotechnology and Biochemistry **72**(4): 1087-1094.

Williams, M. and F. A. Loewus (1978). "Biosynthesis of L-(+)-tartaric acid from L-ascorbic-4-C-14 acid in grape and geranium." Plant Physiology **61**(4): 672-674.

Winn, M. D., C. C. Ballard, et al. (2011). "Overview of the CCP4 suite and current developments." Acta Crystallographica Section D-Biological Crystallography **67**: 235-242.

Yoshikawa, S., R. Arai, et al. (2007). "Structure of archaeal glyoxylate reductase from Pyrococcus horikoshii OT3 complexed with nicotinamide adenine dinucleotide phosphate." Acta Crystallographica Section D-Biological Crystallography **63**: 357-365.

Yum, D. Y., S. S. Bae, et al. (1998). "Purification and characterization of the 2-ketoaldonate reductase from Brevibacterium ketosoreductum ATCC21914." Bioscience Biotechnology and Biochemistry **62**(1): 154-156.

Yum, D. Y., B. Y. Lee, et al. (1998). "The yiaE gene, located at 80.1 minutes on the Escherichia coli chromosome, encodes a 2-ketoaldonate reductase." Journal of Bacteriology **180**(22): 5984-5988.

Chapter 5 Summary, Future Directions and Conclusions

5.1 Summary of Research Outcomes

5.1.1 Evolutionary origin of the grapevine LIDH

V. vinifera LIDH, previously identified to catalyze the rate-limiting step of the primary tartaric acid biosynthesis pathway in grapevine, displays a close homology with SDHs from other plants (DeBolt, Cook et al. 2006). To determine the evolutionary origin of LIDH, SDH homologs were searched across a broad range of angiosperm species (**Chapter 2**). A comprehensive phylogenetic analysis was carried out on the SDH family, with those remote SDH homologs from yeasts and animals as the out-group (**Chapter 2. Figure 2**). Natural selection tests, gene synteny and transcriptional analyses on selected plant SDHs were also performed (**Chapter 2. Table 3, Figure 4, Figure 6**). The results showed that core eudicot SDHs have evolved into two independent lineages, SDH Class I and Class II. Positive natural selection was detected on the SDH Class II only and was shown to have contributed to the emergence of SDH Class II after the tandem duplication of SDH genes in the common ancestor of core eudicot plants. Most of the previously characterized SDHs belong to the SDH Class I, which was found to be universally conserved in core eudicot species and appears to be essential for the normal growth of plants. LIDH was identified as a Class II SDH. The distribution of Class II SDHs among higher plants, in general, showed a positive correlation with the production of TA. They were found to be present in those prominent TA-accumulating species including *V. vinifera*, *S. lycopersicum* and *P. hortorum* (a *Geraniaceae* plant) (**Chapter 2. Figure 1**). *Leguminosae* plants (TA-accumulating) investigated in this study were found to retain no Class II SDH, which is also in good agreement with the report that TA in *Leguminosae* plants is produced via the D-glucose pathway which doesn't involve the conversion of L-idonic acid to 5KGA (Saito and Loewus 1989). In addition, the distinct enzymatic profile of LIDH was tentatively explained by protein modeling analyses of SDH homologs from *V. vinifera* (**Chapter 2. Figure 5**). Three amino acid residue substitutions at the catalytic site between Class I and Class II SDHs were identified and may have changed the substrate specificity of Class II SDHs, thereby facilitating their putative function in

TA biosynthesis. Previous studies in apple and pear also showed that Class I and Class II SDH genes are under independent transcriptional regulation (Park, Song et al. 2002, Ito, Hayama et al. 2005). In the present study, transcriptional analyses of SDH genes in *V. vinifera* and *C. sinensis* based on microarray and RNA sequencing data also demonstrated a clear divergence between Class I and Class II (**Chapter 2. Figure 6**), supporting a functional divergence between SDH Class I and Class II at the expression level.

Taken together, it is concluded that LIDH is a Class II SDH and represents a functional divergence of SDH in grapevine as a result of positive natural selection after the tandem duplication of SDH genes in the common ancestor of core eudicot plants. This study significantly improves our understanding on the biosynthesis of TA from the evolutionary perspective. Previous studies on TA research have been concentrated on *V. vinifera*. The conservation of Class II SDH genes in other prominent TA-accumulating species would promote future studies to investigate the TA production in other species such as *S. lycopersicum* and *P. hortorum* which may be better model plants than *V. vinifera* regarding TA research.

5.1.2 Substrate specificity of SDH and its effect on TA biosynthesis in higher plants

In the first part of this study, we found that LIDH belongs to core eudicot SDH Class II. Protein modeling analyses of SDH homologs from the grapevine revealed three amino acid substitutions between Class I and Class II SDHs at the catalytic site, caused by positive natural selection. These amino acid changes may have altered the substrate specificity of Class II SDHs, generating a novel function for Class SDHs, such as their involvement in TA biosynthesis. To validate our hypothesis, homologous SDH genes from *V. vinifera*, *T. cacao* and *S. tuberosum* (all TA-accumulating species) were cloned and transformed into *E. coli* for recombinant protein production. The enzyme activities of these SDHs were tested with L-idonic acid and other common polyol substrates. Site-directed mutagenesis of the Class II SDH genes was also performed to assess the effect of the target amino acid residues on the enzyme substrate specificity (**Chapter 3**). Our results confirmed that *V. vinifera* LIDH has the highest activity against L-idonic

acid (**Chapter 3**). The substrate specificity of LIDH with L-idonic acid is over ten times higher than that with sorbitol, supporting its biological role in TA biosynthesis. Previous study on LIDH provided detailed information on its activity with the common polyol substrates. In this study, we found that LIDH still retains significant activity with sorbitol, xylitol and iditiol but has very weak activity with ribitol. The determined K_m value of LIDH with sorbitol (32.72 mM) is comparable to that reported for pear and apple Class I SDHs but significantly lower than *A. thaliana* and *F. ananassa* SDHs. The production of recombinant *V. vinifera* Class I SDH was not successful. Based on the phylogenetic analysis (**Chapter 3**), it is suspected that the Class I SDH of *V. vinifera* may have similar kinetic profile with *F. ananassa* FaSDH and *T. cacao* COC280. In the present study, *T. cacao* COC280 was determined to have a K_m value of 16.68 mM with sorbitol, most close to that reported for FaSDH (7.3 mM) (Sutsawat, Yamada et al. 2008). In dramatic contrast to LIDH, Class I *T. cacao* COC280 showed almost no activity with L-idonic acid (**Chapter 3**). To our knowledge, this is the first report of a Class I SDH's activity with L-idonic acid. Our results strongly indicate that a clear enzymatic divergence exists between Class I and Class II SDHs, potentially supporting a divergent biological function for plant Class II SDH genes. Unfortunately, despite repeated tests, the purification of the recombinant protein for *T. cacao* COC270, *S. tuberosum* LIDH and other corresponding mutant proteins was not successful in this study. This may be due to the occurrence of the recombinant proteins in the inclusion body or a lack of protein expression at the growth conditions used in these experiments. Further experimental trouble shooting is needed to obtain the recombinant proteins.

5.1.3 Enzymatic and protein structural characterization of VV2KR in grapevine

The second step of the TA biosynthesis pathway involves the conversion of 2KLG to L-idonic acid. Previous studies based on radioisotope labeling analyses indicated that this reaction also has an enzymatic basis (Saito and Kasai 1984). However, the corresponding enzyme has not been characterized. NADPH-dependent 2-keto-D-gluconate reductase (2KGR, EC 1.1.99.3) being able to perform the same reaction has been reported in *E. herbicola* (Truesdell, Sims et al. 1991), *B. ketosoreductum* (Yum, Bae et al. 1998), *E. coli* (Yum, Lee et al. 1998) and various

acetic bacteria (Ameyama and Adachi 1982). This leads to the speculation that the second step of TA pathway may be catalyzed by a 2-keto-D-gluconate reductase homolog in grapevine. Homolog search against EST data using the *E. coli* 2-ketoaldonate reductase as a query revealed a grapevine homolog (now named VV2KR) with 39% similarity at the amino acid level (**Chapter 4**). VV2KR gene was selected also based on its transcription in the early stage of berry development, coinciding with the production of TA. Sequence analyses showed that VV2KR belongs to the 2KDH family and shares the highest similarity with *C. blumei* HPR (CbHPR, Uniprot ID Q65CJ7; 78% identity) and *A. thaliana* HPR2 (AtHPR2, Uniprot ID Q9CA90; 75% identity). As determined in the present study, VV2KR prefers to use NADPH over NADH as the coenzyme (**Chapter 4**). Among the four substrates tested, VV2KR demonstrated the highest reaction rate with D-glyoxylate and a weak activity with pyruvate. This is consistent with its close homology with *C. blumei* HPR, *A. thaliana* HPR2 and human GRHPR, which have been shown to have similar enzymatic profiles. In addition, VV2KR was found to be able to reduce 2KLG to L-idonic acid effectively, providing direct support for its potential role in the TA biosynthesis pathway.

The 3-dimensional structure of VV2KR was determined to 2.1 Å resolution (**Chapter 4**). The spatial arrangement of the four subunits in the asymmetric unit indicates that VV2KR is a biologically active dimer. The monomer structure of VV2KR demonstrated the typical characteristics of 2KDH protein with a larger coenzyme binding domain and a smaller substrate binding domain. It shares a close homology with the previously determined *C. blumei* HPR structures (Janiak, Petersen et al. 2010), and also with other structures of the HPR such as human GRHPR (Booth, Conners et al. 2006). The preference of VV2KR for NADPH as coenzyme was analyzed at the structural level by docking $NADP^+$ to the apo-VV2KR structure, with 174Ser being the critical amino acid (**Chapter 4**). Previous studies indicated two substrate binding modes (mode A and B) exist for 2KDH proteins, potentially leading to the production of distinct isomers of the corresponding products (Goldberg, Yoshida et al. 1994, Stoll, Kimber et al. 1996, Kumar, Carlson et al. 2002, Razeto, Kochhar et al. 2002, Booth,

Conners et al. 2006, Janiak, Petersen et al. 2010). We docked five putative substrates to our determined VV2KR structure. Results (**Chapter 4**) showed that, while both substrate binding modes are possible for VV2KR, mode B is the preferred with all substrates. Under this binding mode, 2KLG is reduced at the C2 position, leading to the production of L-idonic acid. Our substrate docking analysis also showed that the observed lower reaction rate of VV2KR with 2KLG than that with D-glyoxylate and hydroxypyruvate may be due to the higher number of hydroxyl groups of 2KLG. These hydroxyl groups generate increased binding interaction with the surrounding amino acid residues, thereby potentially reducing the turn-over rate of the corresponding reaction.

In summary, the present study has provided a comprehensive enzymatic characterization of VV2KR, supporting its potential involvement in the biosynthesis of TA in grapevine. In-depth structural analyses of VV2KR were performed by protein crystallization and computational methods. The structural characteristics of VV2KR were discussed in detail with regards its putative catalytic mechanism and the potential biological function.

5.2 Discussion and Future Research

5.2.1 The biological function of the Class II SDHs and TA biosynthesis in higher plants

Before the elaboration of the phylogenetic relationship between LIDH and SDH in the present study, the in-planta functions of SDH were mainly attributed to its involvement in sorbitol metabolism pathway (Negm and Loescher 1979, Yamada, Oura et al. 1998, Oura, Yamada et al. 2000, Park, Song et al. 2002, Nosarszewski, Clements et al. 2004, Nosarzewski and Archbold 2007, Wang, Xu et al. 2009, Nosarzewski, Downie et al. 2012, Maria, Diego et al. 2013). SDH activity regulates the level of polyols such as sorbitol and ribitol (Nosarzewski, Downie et al. 2012, Maria, Diego et al. 2013), which serve as important osmolytes under abiotic stress conditions (Yancey, Clark et al. 1982). Therefore, SDH plays a critical role in plants' resistance to abiotic stresses such as drought and salinity. In some herbaceous plants, SDH has also been shown to be involved in sugar metabolism during the seed germination process (Doehlert 1987,

Kuo, Doehlert et al. 1990, de Sousa, Paniago et al. 2008). In addition, extensive studies on plant SDH have been dedicated to some *Rosaceae* fruit species (Yamada, Oura et al. 1998, Bantog, Yamada et al. 2000, Oura, Yamada et al. 2000, Yamada, Niwa et al. 2001, Park, Song et al. 2002, Nosarszewski, Clements et al. 2004, Ito, Hayama et al. 2005, Nosarzewski and Archbold 2007, Wang, Xu et al. 2009, Matías D Hartman 2014), in which sorbitol is the major photosynthate and phloem transported carbohydrate form (Loescher 1987). SDH in these plants affect the oxidation and translocation of sorbitol to sink tissues such as developing fruits and young leaves, thereby playing an important role in sink-strength regulation (Beruter 1985, Kuo, Doehlert et al. 1990, Yamaguchi, Kanayama et al. 1996, Bantog, Shiratake et al. 1999, Bantog, Yamada et al. 2000, Nosarszewski, Clements et al. 2004, Nosarzewski and Archbold 2007, Wang, Xu et al. 2009). The metabolic function of SDH in apple and pear is very complex. This is largely due to the extensive duplication of SDH in these species and the evolution of diverse gene expression patterns in different tissues at different developmental stages. Despite the earlier reports of the presence of a distinct copy of SDH gene in both apple and pear (Park, Song et al. 2002, Nosarszewski, Clements et al. 2004, Ito, Hayama et al. 2005), no special attention has been given to these SDH regarding to their specific biological function in the corresponding plants. The present study recognized that these distinct *Rosaceae* SDHs, together with the grapevine LIDH, belong to a separate group of SDH (classified as SDH Class II) that has evolved independently from the normal SDH genes (SDH Class I). Unlike the universal conservation of Class I SDH, Class II SDHs were found to be absent in many species including the *Brassicaceae* plants, *Leguminosae* plants and some *Asterid* plants (**Chapter 2. Figure 1&2**). The discovery of LIDH as a Class II SDH suggests that other Class II SDHs may also be responsible for the biosynthesis of TA. This is supported by the distribution of the Class II genes among higher plants, which shows a positive correlation between the presence of Class II SDHs and the accumulation of TA. The enzymatic characterization of the *T. cacao* COC280 (SDH Class I) and the grapevine LIDH also provided a positive indication for a divergent biological function of Class II SDHs from Class I SDHs (**Chapter 3**). Our knowledge on the distribution of TA among higher plants is very limited. Information

on the occurrence of TA in many species such as apple, pear, peach, citrus and *T. cacao* is still very fragmentary or absent. All of these species were shown to retain Class II SDH genes in the present study (Chapter 2. Figure 1). A comprehensive screening for TA among higher plants is needed to investigate the potential link of Class II SDHs with the TA pathway in species other than the grapevine.

In addition, studies in apple (Park, Song et al. 2002) and pear (Ito, Hayama et al. 2005) have shown that the Class II SDH genes are under independent transcriptional regulation from other SDH genes, shedding some light on the putative metabolic role of the Class II SDH genes in these species. The function of a gene could be affected at both the gene transcriptional and protein structural levels, future studies should be dedicated to characterize the transcriptional and enzyme kinetic profiles of the Class II SDH genes in species which have also been shown to accumulate a substantial amount of TA. One of those species deserves great attention is *S. tuberosum* which represents a significant expansion of the TA-accumulating species. In contrast to most plants which retain the Class I SDH genes in a commonly conserved collinear block, *S. tuberosum* only has a Class II SDH gene in this genetic region (**Chapter 2. Figure 4**). The Class I SDH gene was found to be a dispersed duplication in this plant (**Chapter 2. Table 2**). *S. tuberosum* uses sucrose as the major translocated carbohydrate (Mares and Marschner 1980) form and displays a very different carbon metabolism profile from the *Rosaceae* species that use sorbitol as the main translocated carbohydrate (Loescher 1987). Future functional characterization of SDH genes in *S. tuberosum* would not only help us to validate whether the Class II SDH gene in *S. tuberosum* is responsible for TA production or not, it could also possibly improve our understanding of the regulation of carbon metabolism in this species.

While a novel metabolic role for Class II SDH genes in TA pathway is possible, the current study also provided evidences that the Class II SDH genes may still retain some overlapping function with the Class I SDH genes in the sorbitol metabolism pathway. For example, the enzymatic characterization of LIDH showed that it still retains significant activity with the oxidation of sorbitol, despite its much higher substrate specificity with L-idonic acid (**Chapter 3**). In addition, although the

transcription of the LIDH gene has been found to be mainly restricted to the early stage of grape berry development (DeBolt, Cook et al. 2006, Wen, Li et al. 2010, Sweetman, Wong et al. 2012), recent studies based on immunoblotting analyses showed that active LIDH enzyme is still present in mature berries (Wen, Li et al. 2010). In addition to being highly expressed in source leaves as well as young leaves, the Class II SDH gene in apple was also found to be highly expressed in the maturing fruits (Park, Song et al. 2002), suggesting it has multiple metabolic roles to play and may also be responsible for the conversion of sorbitol into D-fructose in the mature apple fruit. Earlier study has shown that TA was absent or only in trace amounts in apple (Fuleki, Pelayo et al. 1995). These observations indicate that the Class II SDH gene in apple may not be responsible for TA biosynthesis. A very similar observation was made with the Class II SDH gene from pear (Ito, Hayama et al. 2005). In a recent study (Dai, Shi et al. 2015), transcriptional analyses based on RNA-seq data showed that the Class II SDH gene in pear (named as PpySDH8 in the study) was also highly expressed in both young and ripening stages of pear fruit development, although the authors did not recognize the distinct molecular characteristics of this gene and its potential correlation with TA synthesis. In contrast to that report in apple, a moderate amount of TA has been detected in pear (Hudina and Stampar 2000, Sha, Li et al. 2011), prompting speculation that the Class II SDH gene in pear may be involved in the TA biosynthesis pathway. In regard to the putative metabolic role for the Class II SDH genes in the sorbitol metabolism pathway, it should also be noted that previous enzymatic characterization of the apple Class II SDH (Yamada, Oura et al. 1998) showed that it displayed a much weaker affinity for sorbitol (K_m 247mM) as the substrate, which suggests that the apple Class II SDH may not be as effective as the Class I SDHs with the sorbitol metabolism. Similar observation was made with LIDH in this study which showed much weaker affinity for sorbitol compared to the previously characterized non-rosaceae SDHs. Further gene expression pattern and enzymatic activity analyses are needed to characterized the *in vivo* biological function of the Class II SDH genes in these species.

5.2.2 *In vivo* characterization of the putative function of VV2KR in TA biosynthesis in the grapevine

By measuring the in-vitro activity of the purified recombinant protein, we showed that a ketogluconate reductase homolog (VV2KR) from grapevine is able to effectively catalyze the interconversion of 2KLG and L-idonic acid (**Chapter 4**), the second step of the biosynthesis pathway of TA. This provided biochemical support that VV2KR may be involved in the TA production pathway. However, the primary structure analyses of VV2KR showed that it has the highest similarity with *A. thaliana* HPR2 (Timm, Nunes-Nesi et al. 2008) (**Chapter 4**). Enzymatic tests also demonstrated that VV2KR has much higher substrate specificity with the reduction of D-glyoxylate and hydroxypyruvate, the typical enzymatic characteristics for the HPR family (Kim, Janiak et al. 2004, Booth, Conners et al. 2006). These observations implied that VV2KR corresponds to the *A. thaliana* HPR2 ortholog in the grapevine. Three isoforms of HPR (HPR1-3) (Mano, Hayashi et al. 1997, Timm, Nunes-Nesi et al. 2008, Timm, Florian et al. 2011) have been identified in *A. thaliana*, with HPR1 (Mano, Hayashi et al. 1997) being targeted to the peroxisome and contributing a major role to the photorespiration cycle, whereas HPR2 (Timm, Nunes-Nesi et al. 2008) and HPR3 (Timm, Florian et al. 2011) were found to be targeted to the cytosol and chloroplast, respectively, serving as the bypass of the peroxisomal photorespiration cycle. The photorespiration cycle is one of the major pathways of plant primary metabolism. The combined deletion of HPR1-3 in *A. thaliana* led to a dramatic reduction of photosynthetic performance and severe growth retardation (Timm, Florian et al. 2011). Therefore, it's reasonable to speculate that VV2KR may also be responsible for the photochemical reaction in the grapevine. Organic acids in plants are commonly synthesized in the cytosol and then transported to the vacuole. Previous characterization of LIDH showed that it is a cytosolic enzyme, indicating that TA is mainly produced in the cytosol (DeBolt, Cook et al. 2006). Of the three HPR isoforms identified in *A. thaliana*, only HPR2 was shown to be targeted to the cytosol while HPR1 and HPR3 were directed to the peroxisome and chloroplast, respectively (Mano, Hayashi et al. 1997, Timm, Nunes-Nesi et al. 2008, Timm, Florian et al. 2011). The close homology between VV2KR and HPR2 suggests that VV2KR is very likely a cytosolic protein. This is in good agreement with its putative involvement in the TA

production in the cytosol. In regard to the potential multiple biological functions for VV2KR, support could be drawn from its homologous counterpart in *C. blumei*, CbHPR (Kim, Janiak et al. 2004). Rosmarinic acid is a natural production occurring in *Lamiaceae* and *Boraginaceae* plants including *C. blumei*, that have been shown to have important antimicrobial and antioxidant functions (Parnham and Kesselring 1985, Szabo, Thelen et al. 1999). Despite its preference for hydroxypyruvate as the substrate, *C. blumei* CbHPR has been shown to be responsible for the biosynthesis of rosmarinic acid, catalyzing the reduction of 4-hydroxyphenylpyruvate to the corresponding 4-hydroxyphenyllactic acid, a critical step in the rosmarinic acid biosynthesis pathway (Kim, Janiak et al. 2004). The crystal structures of CbHPR have been determined in both the apo-form and the coenzyme-binding form (Janiak, Petersen et al. 2010). In the present study, we solved the ligand-free structure of VV2KR with the hope to gain more insights into the catalytic mechanism and biological function of this protein. Our results showed that the VV2KR structure adopts the typical folding pattern of the 2KDH superfamily and is highly homologous with the CbHPR (Janiak, Petersen et al. 2010) and the human GRHPR (Booth, Conners et al. 2006) structures, providing clear evidence that VV2KR belongs to the HPR family. Our structure represents the second plant homologous structure that has been determined in this family. The catalytic mechanism of VV2KR on the reduction of 2KLG to L-idonic was analyzed in detail by molecular docking method (**Chapter 4**). Our results will facilitate future study on the biological function of VV2KR, particularly its potential involvement in the TA biosynthesis pathway.

The absence of the TA biosynthesis pathway in model plants such as *A. thaliana* (Stafford 1959) makes the characterization of its molecular mechanism extremely challenging. Current efforts trying to understand how TA is produced in grapevine, a perennial and woody species, has been hindered by its long flowering time and the difficulty to obtain the transgenic lines of this plant. These difficulties have been greatly reduced by the recent development of a grapevine model system (Chaib, Torregrosa et al. 2010). This system is based on the discovery of a *Pinot Meunier* mutant, called the microvine, which displays dwarf stature and rapid flowering

phenotypes (Boss and Thomas 2002). This makes rapid genetic studies possible in grapevine. Highly transformable lines have been produced to facilitate genetic transformation studies based on the *Agrobacterium tumefaciens* mediating method (Chaib, Torregrosa et al. 2010). Future studies to validate the *in vivo* metabolic function of VV2KR could take advantage of this system to generate VV2KR knockout or over-expression transgenic lines and investigate their impacts on TA accumulation. In addition, the evolutionary study of the plant SDH family in the present study indicated that a potentially similar TA biosynthesis pathway may exist in *S. tuberosum* as in *V. vinifera* (**Chapter 2**). *S. tuberosum*, which is an herbaceous perennial and could be propagated rapidly, potentially could be a better alternative system for TA research. The identification and characterization of the *S. tuberosum* orthologs of VV2KR, LIDH and other putative enzymes responsible for TA production would greatly accelerate our understanding of the molecular mechanism of TA production in higher plants. Attempts have been made to characterize the enzymatic profile of the Class II SDH in *S. tuberosum* in the present study (**Chapter 3**). However, no recombinant protein has been obtained. Further optimization of the E. coli expression system is needed to obtain the active recombinant protein.

5.2.3 Identification of the other candidate enzymes involved TA biosynthesis pathway

With the emphasis of the second and third steps of the TA-synthesis pathway, the enzymatic mechanisms for the other steps remain unclear. The conversion of 5KGA to TA has been speculated to be catalyzed consecutively by a putative transketolase and a semialdehyde dehydrogenase, via a putative TA semialdehyde as the intermediate (Salusjarvi, Povelainen et al. 2004). Putative grapevine homologs for both of these enzymes have been identified by homology search and protein domain analyses in previous studies in our lab (DeBolt 2006, Burbidge 2011). No significant enzyme activity could be detected for the corresponding biochemical reactions that are associated with these proteins. New strategy should be developed for the more direct search for the corresponding genes and enzymes related to the TA pathway. With the advancement of sequencing technology, more and more genomic information is

becoming available for plant biological research. Population genetic association techniques such as quantitative trait locus (QTL) mapping have been a powerful tool for the location and identification of candidate genes responsible for various crop phenotypes (Miles; and Wayne 2008). Grapevine is among the first few crops for which complete and well assembled genome sequence data has been available (Jaillon, Aury et al. 2007, Velasco, Zharkikh et al. 2007). This will facilitate future genetic mapping studies in this plant. Notably, significant variation of TA accumulation has been observed among different varieties of grapevine (Stafford 1959, DeBolt 2006). Future studies should consider QTL mapping analyses as an alternative choice to identify and validate other enzymes related to TA production in grapevine.

The results from the evolutionary analyses of the plant SDH family (**Chapter 2**) and the enzymatic and structural characterization of VV2KR in the present study (**Chapter 4**) highlight the biological significance of genetic variation between paralogous genes and also reveal the prevalence of multiple metabolic roles associated with a single gene due to the broad substrate specificity of the encoded enzyme. Current comparative genomic studies showed that most gene families are commonly conserved across different plants (Conant and Wolfe 2008). The great phenotypic diversification that has evolved among the angiosperm plants is mainly attributed to the functional divergence of the commonly conserved paralogous genes in different plants, either at the gene structural level or at the transcriptional level (Zhang 2003, Flagel and Wendel 2009). Due to the restriction of the TA pathway to a limited species, special attention should be given to the possibility that the enzymes responsible for the production of TA have different metabolic roles in different plants. In vivo characterization of the candidate genes, together with the in vitro biochemical evidence is required to prove their roles for TA biosynthesis.

5.3 Conclusion

The results of the present study have significantly improved our understanding on the molecular mechanism of the biosynthesis of TA in grapevine. The evolutionary origin of LIDH, the critical enzyme catalyzing the rate-limiting step of the TA production in grapevine, has been clearly characterized (**Chapter 2**). Core eudicot SDH genes have

evolved independently into two distinct lineages (Class I and Class II) as a result of positive natural selection after tandem duplication during the common ancestor of core eudicot plants. LIDH is identified as a Class II SDH and represents the functional divergence of SDH in grapevine. Three amino acid substitutions at the catalytic site were identified and may be responsible for the unique enzyme activity of LIDH which has contributed to its involvement in the TA pathway. Most importantly, the comprehensive evolutionary analyses of the plant SDH family imply the existence of a similar TA pathway in previously unrecognized species and provide a guide for future research on TA biosynthesis in other plants. Furthermore, SDH genes of different classes from *V. vinifera*, *T. cacao* and *S. tuberosum* have been cloned and transformed into *E. coli* expression system (**Chapter 3**). Site-mutation of the identified target amino acid residues was also performed with the hope to assess their impact on the enzyme substrate specificity. Initial enzymatic results of *T. cacao* COC280 and *V. vinifera* LIDH support a functional divergence between Class I and Class II SDHs. Further experiments are needed to optimize the in-vitro recombinant protein expression and purification system to characterize the enzyme kinetic profile of the other candidate proteins. Additionally, we reported a very detailed enzymatic characterization and the structural determination of a *V. vinifera* 2-ketogluconate reductase, VV2KR (**Chapter 4**). VV2KR belongs to the 2KDH superfamily and is the *A. thaliana* HPR2 counterpart in grapevine. Our determined structure of VV2KR represents the second structure being solved for the plant HPR family and will facilitate future functional characterization of proteins within this family. Despite its much higher substrate specificity for D-glyoxylate and hydroxypyruvate, VV2KR was shown to be able to catalyze the reduction of 2KLG to L-idonic acid effectively, the second step of the primary biosynthesis pathway of TA in grapevine. The catalytic mechanism of the corresponding biochemical reaction for VV2KR was investigated by molecular docking and also structural comparison to other structures of this family. While the enzymatic and protein structural analyses, including its potential subcellular location to the cytosol, provide positive indication, further *in vivo* functional characterization is needed to confirm its metabolic role for the biosynthesis of TA in grapevine.

Abbreviations

(Chapter 1 & 5)

TA	L-tartaric acid
2KLG	2-keto-L-gulonic acid
IA	L-idonic acid
5KGA	5-keto-D-gluconic acid
LIDH	L-idonate-5-dehydrogenase
VV2KR	2-keto-L-gulonate reductase
SDH	Sorbitol dehydrogenase
HPR	Hydroxypyruvate/glyoxylate reductase
CbHPR	*Coleus blumei* hydroxypyruvate/glyoxylate reductase
AtHPR	*Arabidopsis thaliana* hydroxypyruvate/glyoxylate reductase
2KDH	D-isomer 2-hydroxyacid dehydrogenase
EST	Expressed sequence tag
2KGR	2-keto-D-gluconate reductase
2KGA	2-keto-D-gluconate
WGD	Whole genome duplication
FaSDH	*Fragaria ananassa* sorbitol dehydrogenase
GRHPR	Human glyoxylate reductase/hydroxypyruvate reductase
NAD(H)	Nicotinamide adenine dinucleotide
NADP(H)	Nicotinamide adenine dinucleotide phosphate
QTL	Quantitative trait locus

Appendices

Appendix 1

(Chapter 2)

Yong Jia., Darren CJ Wong, Crystal Sweetman, John B Bruning, Christopher M Ford. (2015). "New insights into the evolutionary history of plant sorbitol dehydrogenase." BMC Plant Biology 15(1):101.

Please see the manuscript attached as a PDF file to the end of this document.

Appendix 2

2.1 Supplementary file S1: CDS and amino acid sequences of target SDHs

CDS of target SDH genes

Site-directed mutation forward primer, reverse primer, ***mutation sites***

Vitis vinifera

>GSVIVT01010642001

ATGGGAAAAGGAGGGATGTCTCAGGGCGGAGATGGAAGAGGTGAAGGAGAAGAGAACATGGCAGCATGGCTTCTTGGGGTTAACAACCTC
AAGATCCAACCTTTCATCCTCCCTCCTCTTGGTCCTCATGATGTTAGAGTTAGGATGAAGGCTGTTGGTATATGTGGAAGTGATGTTCACTACCT
CAAGAAATTGAGATGTGCAGATTTTATTGTTAAAGAGCCTATGGTGATCGGCCATGAATGTGCTGGGATCATTGATGAAGTTGGTCCCCAGGT
GAAGTCTCTGGTGCCAGGGGATCGAGTGGCATTGGAGCCTGGAATCAGTTGCTGGCGATGCCAACTCTGCAAAGAAGGCCGATACAATCTGT
GCCCTGAGATGAAATTTTTTGCTACACCTCCTGTTCATGGTTCTCTTGCCAATCAGGTTGTGCATCCTGCAGATCTATGCTTTAAACTGCCAGACA
ATGTGAGCTTGGAGGAAGGGGCAATGTGTGAGCCATTAAGTGTTGGTGTTCATGCTTGTCGCCGTGCTGATATAGGTCCAGAGTCGAATGTAT
TGGTCATGGGTGCAGGGCCAATTGGGCTTGTTACAATGCTTGCAGCTCGTGCTTTTGGAGCACCCAGAATTGTCATTGTGGATGTGGACGATTA
CCGGTTATCTGTTGCAAAGGATCTTGGTGCAGATGAGATTGTTAAAGTTTCGACAAATATTCAGGATGTAGCTGAAGAAGTGGTGCAGATACA
TAAAGCTATGGGAGCTAGAGTGGATGTGAGCTTTGATTGTGCCGGCTTTGATAAAACCATGTCAACCGCTCTAAGTGCCACTTCTACTGGTGGC
AAAGTTTGTCTTGTGGGGATGGGGCACAATGAAATGACGGTTCCGCTTACTCCTGCTGCTGCAAGGGAAGTTGATGTGGTTGGTGTGTTCCGG
TACAAGAACACATGGCCGATATGCATCGAGTTCCTGAGGAGTGTGAAGATCGACGTGAAGCCCCTGATAACACATAGGTTCGGCTTCTCCCAA
AGGGAGGTAGAAGAAGCCTTTGAAACCAGTGCTCGTGGTGGTACTGCCATTAAGGTCATGTTCAACCTCTAG

>GSVIVT01010644001

ATGGCAGCTTGGCTTCTGGGCATCAAGACCCTCAAGATTCAACCCTACATTCTCCCTTCTCTGGGCCCCCATGATGTTAAAGTTAGGATCAAAGC
TGTAGGGATATGTGGAAGTGATGTTCATCACTTCAAGACAATGAGGTGCGCAAATTTTATTGTGAAGAAGCCAATGGTGATAGGACATGAGTG
TGCTGGTATCATAGAAGAAGTTGGGAGTGAAGTGAAGAATCTTGTAGTAGGTGACCGGGTTGCTCTGGAGCCCGGTATTAGCTGTAACCGAT
GCAGTCTTTGCAGAAATGGTCAATACAATCTATGCAGAGAAATGAAGTTTTTTGGATCTCCTCCAACCAATGGATCTCTAGCTAACCAGGTGGT
CCATCCTTCAAATCTTTGTTTCAAGCTACCTGACAATGTGAGCTTGGAGGAAGGAGCAATGTGTGAGCCGCTCAGTGTCGGCATCCATGCTTGT
CGTCGTGCTAATGTTGGCCCTGAGACCAACGTACTGATCATGGGATCAGGCCCCATCGGCCTCGTTACAATGCTGGCTGCTCGTGCTTTTGGAG
CGCCGAGGATTGTCCTTGTGGATGTAGATGATCAGCGTCTAGCTATTGCAAAAGATCTTGGCGCAGACGACATTATCCGGGTTTCAACGAATAT
TCAGGATCTAGACGAAGAAGTGGCAAAAATACAAAGCACAATGGTTACTGGAGTTGATGTGAGCTTAGACTGTGTTGGCTTCAACAAAACCAT
GTCAACAGCTTTGAACGCGACTCGAGCAGGCGGCAAAGTTTGCCTTGTGGGTTTGGCCCAGAGTGAGATGACTGTTCCTCTCACTCCAGCTGCT
GCCAGGGAGGTCGATATCGTTGGCATATTCCGCTATAGGAACACATGGCCGCTCTGCCTTGAGTTTTTGAGGAGTGGCAAGATTGATGTTAAA
CCCCTGATAACCCACAGGTTTACTTTCTCACAGAAGGACGTGGAAGAAGCCTTTGAAACCAGTGCTCGTGGGGGTAATGCCATCAAGGTCATG
TTTAATCTCTGA

>GSVIVT01010646001

ATGGCAGCTTGGCTTCTGGGCATCAAGACCCTCAAGATTCAACCCTACATTCTCCCTTCTCTGGGCCCTTATGATGTCAAAGTTAGGATCAAAGC
TGTAGGGATATGTGGAAGTGATGTTCAT***CACTTC***AAGACAATGAGGTGCGCAAATTTCATTGTGAAGAAGCCAATGGTGATAGGGCATGAGT
GTGCTGGCATCATAGAAGAAGTTGGGAGTGAAGTGAAGAATCTTGTAGCAGGTGACCGGGTTGCTCTGGAGCCCGGTATTAGCTGCAACCGA
TGCAGTCTTTGCAGGAATGGTCAATACAATCTATGCAGAGAAATGAAGTTTTTT***GGATCT***CCTCCAACCAATGGTTCTCTAGCTAACCAGGTGG
TCCATCCTTCAAATCTTTGTTTCAAGCTACCTGACAATGTGAGCTTGGAGGAAGGAGCAATGTGTGAGCCGCTCAGTGTCGGCATCCATGCTTG
TCGCCGTGCTAATGTTGGCCCTGAGACCAACGTACTGATCATGGGATCAGGCCCCATCGGCCTTGTCACAATGCTGGCTGCTCGTGCTTTTGGA
GCGCCGAGGATTGTCCTTGTGGACGTAGATGATCAGCGACTAGCTATTGCAAAAGATCTTGGCGCAGACGACATTATCCGGGTTTCAACGAAT
ATTCAGGATCTAGATGAAGAAGTGGCAAAAATACAAAGCACAATGGTTACTGGAGTTGATGTGAGCTTTGATTGCGTCGGCTTCAACAAAACC
ATGTCAACAGCTTTGAACGCGACTCGAGCAGGCGGTAAAGTTTGCCTTGTGGGTTTGGCCCAGAGTGAGATGACTGTTCCTCTCACTCCAGCTG
CTGCCAGGGAGGTCGATATTGTTGGCATATTCCGCTATAGGAACACATGGCCGCTCTGCCTTGAGTTTTTGAGGAGTGGCAAGATTGATGTTAA
ACCCCTGATAACCCACAGGTTTACTTTCTCACAGAAGGATGTGGAGGAAGCCTTTGAAACCAGTGCTCGTGGGGGTAATGCTATCAAGGTCAT
GTTTAATCTCTAA

Solanum tuberosum

>PGSC0003DMT400081907

ATGGCTGCTTGGCTTCTTGGTGTTAACACCCTCAAGATTCAGCCTTTCAATCTCCCTGCTTTGGGACCCCATGATGTTAGAGTTAGGATGAAGGC
TGTCGGTATTTGTGGAAGTGATGTTCATTACCTCAAGACCATGAGGTGTGCGGATTTTGTGGTTAAAGAGCCAATGGTGATTGGGCATGAATG
TGCTGGGATCATAGAGGAAGTTGGCGGTGAAGTCAAGACATTGGTTCCTGGAGATCGTGTAGCGCTAGAGCCAGGAATTAGTTGCTGGAGAT
GTGATCTTTGCAAAGAAGGGCGATATAATCTCTGCCCCGAGATGAAGTTCTTCGCTACTCCCCCTGTTCATGGTTCTCTCGCAAATCAGGTAGTC
CATCCTGCTGATCTATGTTTCAAGCTCCCGGATAATATAAGTTTAGAGGAGGGAGCAATGTGTGAGCCACTTAGTGTTGGTGTTCATGCTTGTC
GGCGTGCAAATGTTGGTCCTGAGACAAACATTTTAGTGCTGGGAGCTGGACCAATCGGGCTTGTCACACTTCTCGCTGCTCGTGCTTTTGGTGC
CCCAAGAATTGTTATTGTGGACGTAGATGACTATCGTCTTTCAGTTGCAAAGAAGTTAGGAGCAGATGAGATCGTCAAGGTTTCAATCAATCTT
CAGGATGTAGCTACAGATATAGAAAACATTCAGAAAGCGATGGGAGGTGGAATTGATGTGAGTTTTGACTGTGCTGGCTTTAACAAAACTATG
TCGACCGCCCTTGGTGCAACTCGTCCAGGTGGCAAAGTTTGCTTGGTGGGAATGGGACATCATGAGATGACCGTTCCTCTCACTCCAGCTGCTG
CAAGGGAGGTCGACATCATCGGCATATTTCGCTACAAGAATACATGGCCATTGTGTCTTGAGTTCTTAAGAAGTGGAAAGATTGATGTGAAAC
CTATGATCACACACAGGTTTGGATTCTCTCAAAAAGAAGTTGAAGAAGCTTTTGAAACAAGTGCTCGTGGTGGTGATGCTATTAAAGTCATGTT
TAATCTGTAA

>PGSC0003DMT400065063

ATGGCAGATGAAGTAGAAGAAGAGAACTTGGCAGTTTGGCTTCTTGCTATCAAAACACTTCGGATTCAGCCCTTTAGTCTCCCATCTCTTGGTC
CTTATGATGTCAAGATTCGAATTAAAGCTGTTGGTATATGTGGAAGTGATATCCATCACTTTAAGAATATGAGAGTAGCTAATTTCGTGGTTAA
AAAGCCAATGGTACTTGGCCATGAGTGTGCTGGAATTGTAGAACAAGTTGGTAGCCAAGTGAAGTTTCTGATGATCGGTGATCGTGTAGCTTT
GGAACCTGGTATTAGTTGCAGGCAGTGTCAACTATGCAAAGACGGTCGCTACAATCTCTGCCGTAAAATGAAGTTCTTTGGCTCTCCTCCAACA
AATGGTGCTTTAGCCAACCAGGTGGTGCATCCTGCAGATCTATGTTTTAAGTTACCTGATAATGTAAGCTTGGAGGAAGGTGCAATATGTGAAC
CTTTGAGTGTTGGTGTCCATGCTTGCCGCCGTGCCAATGTAGGTCCTGACACCAAACTAATCATCATAGGTGCGGGACCTATTGGCCTGGTCAC
TATGCTGGCAGCTCGTGCTTTTGGATCACCCAAGATAGTCATCGTTGATGTGGATGATTGTCGTTTATCTTTCGCGAAAGACCTGGGTGCTGAT
GAGATCATTAAAGTTTCTTCCATTATGCAGGATGTGGAGGAGGAAGTGGTGCGAATACGTAATGCAATGGGTGGTCCAGTGGATTTGAGCTTT
GATTGTGTTGGTTTTAACAAGACTATGACAACGGCTCTACAAGCCACCCGCACTGGTGGTAAAGTATGCCTCGTTGGATTAGGCCAGAGTGAG
ATGACTCTCCCTCTTACTTCAGCTGCTGCAAGGGAGGTGGATGTCATAGGCATTTTCAGATACCGAAATACATGGCCGCTCTGCATTGAACTTCT
GAGAACTGGGGAAATAGATGTGAAGCCTCTCATAACCCACAGGTATAACTTTACTCAAGAGGGTGTGGATGAAGCATTTGAGACCAGTTCTCG
CGGCGGAAATGCCATCAAGGTCATGTTCAACCTG**TAG**

Theobroma cacao

>Tc03_g019270

ATGGTACCTAATGAATCGTTGCTACAGAGGACAATATGGGGTTTTCCCATTTGTATTATCCTATTTATGATTCGAGCATTTGAGGAACTAAATTA
TAACACGAGGGTAGACGTTGACCATCTCGAGACCGTTTGGAAAGACCAAGACGCAACTGGCTGTTTCGAGGAGGATCAAAACGTGGACTCAG
CTTATTTGAAGAGGATAAGACAGTGCGAAGTGCTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGGCAGAGACAAAGATATCAGAGATGG
CGCACGTTGCTCAAGCTAAGGCCGGCCAAGATGATGGGGAGAAGAATATGGCAGCTTGGCTTCTTGGAATCAAAACCCTTAAAATTCAACCTT
ACCTTCTTCCTCCTCTCGGTCCTCAGGATGTGAAAGTGCGTATAAAAGCTCTTGGCATATGCGGAAGTGATGTTCACCATTTCGAGACAATGAG
ATGTGCAAATTTTATTGTGAAAAAACCAATGGTCATTGGACATGAGTGCGCTGGGATCATCGAAGAAGTTGGGAGCGAAGTGAAGTCCCTGG
CTGTGGGAGATCGTGTGGCATTGGAGCCTGGCATTAGTTGCCAGAGATGCGGTATCTGCAGAGATGGACGCTACAATCTTTGTCCAGAAATGA
AGTTCTTTGGATCTCCTCCAACTAATGGCTCTCTTGCCAATAAGGTGGTGCATCCTGCAAACCTCTGTTTTAAACTACCTGACACTGTGAGCCTG
GAAGAAGGAGCAATGTGCGAACCCCTCAGTGTTGGAGTCCATGCCTGTCGCCGGGCAGGCATTTGTCCTGACACAAATGTGTTGATCATGGGA
GCAGGTCCCATAGGCCTTGTTACCTTATTGGCTGCTCGAGCATTCGGAGCCCCCAGAATTGTCATAGGTGATGTTGATGATTGCCGTCTATCTAT
TGCAAAGAATCTCGGTGCAGATGAGATTGTTCAAGTTTCAACAAATATTCAGGATGTGGGTGAAGAAGTAGTGAAGATACAAAATGCAATGA
GTTCACGCATTGATGTGAGTTTTGATTGTGTTGGTTTCAACAAAACCATGTCTACAGCTTTGAGAGCCACTGGAGCTGGTGGCAAAGTTTGCCT
TATTGGATTGGCGCAGAGCGAAATGACTATCCCCCTCACTCCGGCAGCAGTGAGGGAGATTGATGTGTTTGGCATATTCCGGTATAGAAATAC
GTGGCCACTCTGCATTGAATTCTTGAGGACTGGAAAGATTGATGTGAAGCCCCTTATAACTCATAGGTTTGGATTCTCTCAAAAGGGTATAGAA
GATGCTTTCCGAACCAGTGCTGCTGGTGGCAATGCCATTAAAGTCATGTTTAATCTG**TAA**

>Tc03_g019280

ATGGGCAAAGGAGGGAAATCTCATGAAGAAGCCAGTATTGGAGAAGAAGAAAACATGGCTGCTTGGCTTGTGGGTCTCAACACCCTCAAAAT
TCAACCATTCAAGCTCCCGCCTCTTGGACCCCGTGATGTCAGAGTTAGGATGAAAGCTGTTGGTATCTGTGGAAGTGATGTTCACTATCTCAAG
ACATTGAGGTTGGCAGATTTTGTGGTAAAAGAACCAATGGTGATTGGGCATGAGTGTGCTGGGATCATAGAGGAGGTTGGAGGTGAAGTCAA
GAATTTAGTTCCTGGAGACCGTGTGGCATTGGAACCGGGGATCAGTTGCTGGCGATGTGATCTTTGCAAGGAAGGTCGATACAATCTATGCCC
TGAAATGAAGTTTTTTGCCACACCACCAGTTCATGGTTCTCTTGCTAATCAGGTTGTGCATCCTGCAGATTTATGTTTCAAACTACCTGACAATGT
GAGCTTGGAAGAAGGGGCTATGTGTGAGCCCTTGAGTGTAGGGGTTCATGCTTGTCGCCGAGCTAATATTGGTCCAGGAAAAAATGTGTTGG
TCATGGGAGCAGGACCTATAGGGCTTGTTACGATGCTGGCAGCTCGTGCTTTTGGGGCACCTAGAATTGTTGTTGTAGATGTGGATGACAATC
GATTATCTGTTGCTAAGGATCTTGGTGCAGATGGAGTTGTTAAAGTTTCAACAAATATGCAGGAGTTTGGTGATTTTGACATTGCTCAACTAAA
GGATGTACCTGAAGAAGTTGAAAGAATATGTAAAGTGATGGGAGCAGGAGTGGATGTAAGCTTTGATTGTGCAGGCTTTAACAAAACGATGT
CAACTGCTTTGAGTGCCACTCGTGCTGGTGGCAAGGTTTGCCTTGTGGGAATGGGCCATTCTGAGATGACTGTCCCACTCACTCCAGCTGCTGC
AAGGGAGGTTGATATTATTGGAATCTTCCGGTATAAGAACACTTGGCCTTTGTGCCTTGAGTTTCTAAGGAGCGGTAAGATTGATGTGAAGCCC
CTCATAACCCACAGGTATGGGTTTTCCCAGAAAGAGGTGGAAGAAGCTTTTGAGACCAGTGCTCGTGGTGGCAATGCCATTAAGGTCATGTTT
AATCTG**TGA**

PEP sequences of target SDHs

Vitis vinifera

>GSVIVT01010642001

MGKGGMSQGGDGRGEGEENMAAWLLGVNNLKIQPFILPPLGPHDVRVRMKAVGICGSDVHYLKKLRCADFIVKEPMVIGHECAGIIDEVGPQVKS
LVPGDRVALEPGISCWRCQLCKEGRYNLCPEMKFFATPPVHGSLANQVVHPADLCFKLPDNVSLEEGAMCEPLSVGVHACRRADIGPESNVLVMGA
GPIGLVTMLAARAFGAPRIVIVDVDDYRLSVAKDLGADEIVKVSTNIQDVAEEVVQIHKAMGARVDVSFDCAGFDKTMSTALSATSTGGKVCLVGM
GHNEMTVPLTPAAAREVDVVGVFRYKNTWPICIEFLRSVKIDVKPLITHRFGFSQREVEEAFETSARGGTAIKVMFNL

>GSVIVT01010644001

MAAWLLGIKTLKIQPYILPSLGPHDVKVRIKAVGICGSDVHHFKTMRCANFIVKKPMVIGHECAGIIEEVGSEVKNLVVGDRVALEPGISCNRCSLCRN
GQYNLCREMKFFGSPPTNGSLANQVVHPSNLCFKLPDNVSLEEGAMCEPLSVGIHACRRANVGPETNVLIMGSGPIGLVTMLAARAFGAPRIVLVD
VDDQRLAIAKDLGADDIIRVSTNIQDLDEEVAKIQSTMVTGVDVSLDCVGFNKTMSTALNATRAGGKVCLVGLAQSEMTVPLTPAAAREVDIVGIFRY
RNTWPLCLEFLRSGKIDVKPLITHRFTFSQKDVEEAFETSARGGNAIKVMFNL

>GSVIVT01010646001

MAAWLLGIKTLKIQPYILPSLGPYDVKVRIKAVGICGSDVHHFKTMRCANFIVKKPMVIGHECAGIIEEVGSEVKNLVAGDRVALEPGISCNRCSLCRN
GQYNLCREMKFFGSPPTNGSLANQVVHPSNLCFKLPDNVSLEEGAMCEPLSVGIHACRRANVGPETNVLIMGSGPIGLVTMLAARAFGAPRIVLVD
VDDQRLAIAKDLGADDIIRVSTNIQDLDEEVAKIQSTMVTGVDVSFDCVGFNKTMSTALNATRAGGKVCLVGLAQSEMTVPLTPAAAREVDIVGIFR
YRNTWPLCLEFLRSGKIDVKPLITHRFTFSQKDVEEAFETSARGGNAIKVMFNL

Solanum tuberosum

>PGSC0003DMT400081907

MAAWLLGVNTLKIQPFNLPALGPHDVRVRMKAVGICGSDVHYLKTMRCADFVVKEPMVIGHECAGIIEEVGGEVKTLVPGDRVALEPGISCWRCDL
CKEGRYNLCPEMKFFATPPVHGSLANQVVHPADLCFKLPDNISLEEGAMCEPLSVGVHACRRANVGPETNILVLGAGPIGLVTLLAARAFGAPRIVIV
DVDDYRLSVAKKLGADEIVKVSINLQDVATDIENIQKAMGGGIDVSFDCAGFNKTMSTALGATRPGGKVCLVGMGHHEMTVPLTPAAAREVDIIGIF
RYKNTWPLCLEFLRSGKIDVKPMITHRFGFSQKEVEEAFETSARGGDAIKVMFNL

>PGSC0003DMT400065063

MADEVEEENLAVWLLAIKTLRIQPFSLPSLGPYDVKIRIKAVGICGSDIHHFKNMRVANFVVKKPMVLGHECAGIVEQVGSQVKFLMIGDRVALEPGI
SCRQCQLCKDGRYNLCRKMKFFGSPPTNGALANQVVHPADLCFKLPDNVSLEEGAICEPLSVGVHACRRANVGPDTKLIIIGAGPIGLVTMLAARAFG
SPKIVIVDVDDCRLSFAKDLGADEIIKVSSIMQDVEEEVVRIRNAMGGPVDLSFDCVGFNKTMTTALQATRTGGKVCLVGLGQSEMTLPLTSAAAREV
DVIGIFRYRNTWPLCIELLRTGEIDVKPLITHRYNFTQEGVDEAFETSSRGGNAIKVMFNL

Theobroma cacao

>Tc03_g019270

MAHVAQAKAGQDDGEKNMAAWLLGIKTLKIQPYLLPPLGPQDVKVRIKALGICGSDVHHFETMRCANFIVKKPMVIGHECAGIIEEVGSEVKYLAVG
DRVALEPGISCQRCGICRDGRYNLCPEMKFFGSPPTNGSLANKVVHPANLCFKLPDNVSLEEGAMCEPLSVGVHACRRAGICPDTNVLIMGAGPIGL
VTLLAARAFGAPRIVIGDVDDCRLSIAKNLGADEIVQVSTNIQDVGEEVVKIQNAMSSRIDVSFDCVGFNKTMSTALRATGAGGKVCLIGLAQSEMTIP
LTLAAVREIDVFGIFRYRNTWPLCIEFLRTGKIDVKPLITHRFGFSQKGIEDAFRTSAAGGNAIKVMFNL

>Tc03_g019280

MGKGGKSHEEASIGEEENMAAWLVGLNTLKIQPFKLPPLGPRDVRVRMKAVGICGSDVHFLKTLRLADFVVKEPMVIGHECAGIIEEVGGEVKNLVP
GDRVALEPGISCWRCDLCKEGRYNLCPEMKFFATPPVHGSLANQVVHPADLCFKLPDNVSLEEGAMCEPLSVGVHACRRANIGPGKNVLVMGAGP
IGLVTMLAARAFGAPRIVVVDVDDNRLSVAKDLGADGVVKVSTNMQDVPEEVERICKVMGAGVDVSFDCAGFNKTMSTALSATRAGGKVCLVGM
GHSEMTVPLTPAAAREVDIIGIFRYKNTWPLCLEFLRSGKIDVKPLITHRYGFSQKEVEEAFETSARGGNAIKVMFNL

2.2 Supplementary file S2: Codon-optimized CDS of Tco3_g019270, Tco3_g019280 and PGSC0003DMT400065063

Site-directed mutation forward primer, reverse primer, ***mutation sites***

>Tco3_g019270

CATATGGCACATGTTGCACAGGCAAAAGCAGGTCAGGATGATGGTGAAAAAAACATGGCAGCATGGCTGCTGGGTATTAAAACCCTGAAAATTCA
GCCGTATCTGCTGCCTCCGCTGGGTCCGCAGGATGTTAAAGTTCGTATTAAAGCACTGGGTATTTGCGGTAGTGATGTGCAT***CATTTT***GAAACCATG
CGTTGCGCCAACTTCATTGTTAAAAAACCGATGGTGATTGGCCATGAATGTGCCGGTATTATTGAAGAAGTTGGCAGCGAAGTTAAATATCTGGCA
GTTGGTGATCGTGTTGCACTGGAACCGGGTATTAGCTGTCAGCGTTGTGGTATTTGTCGTGATGGTCGTTATAATCTGTGTCCGGAAATGAAATTTT
TC***GGTAGC***CCTCCGACCAATGGTAGCCTGGCAAATAAAGTTGTTCATCCGGCAAATCTGTGTTTCAAACTGCCGGATAATGTTAGCCTGGAAGAGG
GTGCAATGTGTGAACCGCTGAGCGTTGGTGTTCATGCATGTCGTCGTGCAGGCATTTGTCCGGATACCAATGTTCTGATTATGGGTGCAGGTCCGA
TTGGTCTGGTTACCCTGCTGGCAGCACGTGCATTTGGTGCACCGCGTATTGTTATTGGTGATGTTGATGATTGTCGTCTGAGCATTGCAAAAAATCT
GGGTGCAGATGAAATTGTTCAGGTGAGCACCAATATTCAGGATGTTGGCGAAGAAGTGGTGAAAATCCAGAATGCAATGAGCAGCCGTATTGATG
TTAGCTTTGATTGCGTGGGTTTCAACAAAACCATGAGCACCGCACTGCGTGCAACCGGTGCCGGTGGTAAAGTTTGTCTGATTGGCCTGGCACAG
AGCGAAATGACCATTCCGCTGACCCTGGCAGCAGTTCGTGAAATTGATGTTTTTGGCATTTTTCGCTATCGCAATACCTGGCCTCTGTGTATTGAAT
TTCTGCGTACCGGTAAAATCGATGTTAAACCGCTGATTACCCATCGTTTTGGTTTTAGCCAGAAAGGTATTGAAGATGCATTTCGTACCAGCGCAGC
GGGTGGTAATGCAATTAAAGTTATGTTTAATCTGTAA**GGATCC**

>Tco3_g019280

CATATGGGTAAAGGTGGCAAAAGCCATGAAGAAGCAAGCATTGGCGAAGAAGAAAATATGGCAGCATGGCTGGTTGGTCTGAATACCCTGAAAA
TTCAGCCGTTTAAACTGCCTCCGCTGGGTCCGCGTGATGTTCGTGTTCGTATGAAAGCAGTTGGTATTTGTGGTTCCGATGTGCATTTTCTGAAAAC
CCTGCGTCTGGCAGATTTTGTTGTTAAAGAACCGATGGTGATTGGTCATGAATGTGCCGGTATTATTGAAGAGGTTGGTGGTGAAGTTAAAAATCT
GGTTCCGGGTGATCGTGTTGCACTGGAACCGGGTATTAGCTGTTGGCGTTGTGATCTGTGTAAAGAGGGTCGTTATAATCTGTGTCCGGAAATGAA
ATTTTTCGCAACCCCTCCGGTTCATGGTAGCCTGGCAAATCAGGTTGTTCATCCGGCAGATCTGTGTTTCAAACTGCCGGATAATGTTAGCCTGGAA
GAGGGTGCAATGTGTGAACCGCTGAGCGTTGGTGTTCATGCATGTCGTCGTGCAAATATTGGTCCGGGTAAAAATGTTCTGGTTATGGGTGCAGG
TCCGATTGGTCTGGTTACCATGCTGGCAGCACGTGCATTTGGTGCACCGCGTATTGTTGTTGTTGATGTGGATGATAATCGTCTGAGCGTGGCCAA
AGATCTGGGTGCAGATGGTGTTGTGAAAGTTAGCACCAATATGCAGGATGTTCCGGAAGAAGTTGAACGTATTTGTAAAGTGATGGGTGCCGGTG
TTGATGTTAGCTTTGATTGTGCAGGTTTCAACAAAACCATGAGCACCGCACTGAGCGCAACCCGTGCCGGTGGTAAAGTTTGTCTGGTTGGTATG
GGTCATAGCGAAATGACCGTTCCGCTGACACCGGCAGCAGCCCGTGAAGTTGATATTATTGGTATTTTCCGCTATAAAAACACCTGGCCTCTGTGTC
TGGAATTTCTGCGTAGCGGTAAAATTGATGTTAAACCGCTGATTACCCACCGCTATGGTTTTAGCCAGAAAGAAGTGGAAGAAGCATTTGAAACCA
GCGCACGTGGTGGTAATGCAATCAAAGTTATGTTTAATCTGTAA**GGATCC**

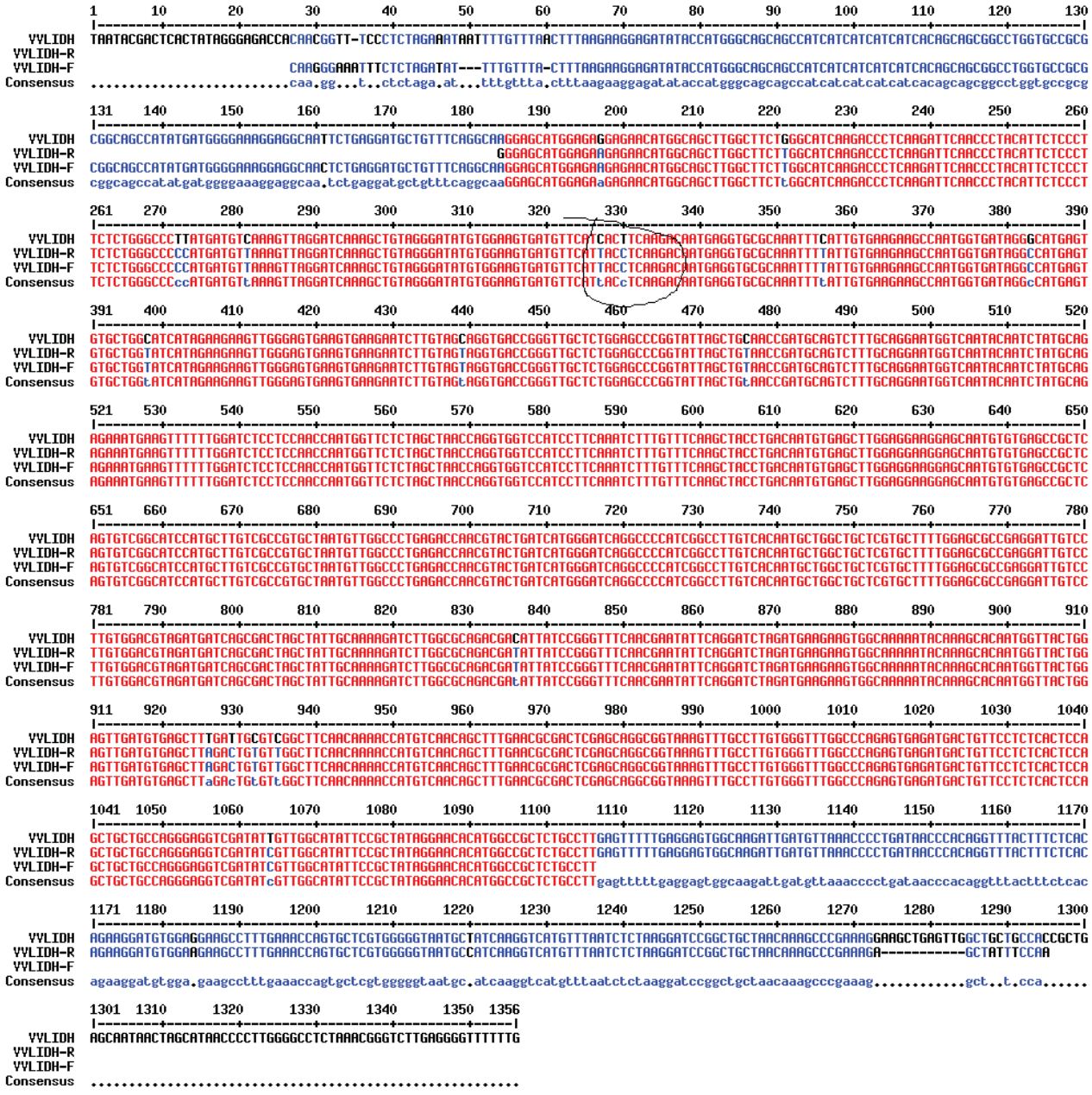
>PGSC0003DMT400065063

CATATGGCCGATGAAGTGGAAGAAGAAAATCTGGCAGTTTGGCTGCTGGCAATTAAAACCCTGCGTATTCAGCCGTTTAGCCTGCCGAGCCTGGG
TCCGTATGATGTTAAAATTCGTATTAAAGCCGTGGGTATTTGCGGCAGCGATATTCAT***CATTTC***AAAAATATGCGCGTGGCCAACTTCGTTGTTAAAA
AACCGATGGTTCTGGGTCATGAATGTGCAGGTATTGTTGAACAGGTTGGTAGCCAGGTTAAATTTCTGATGATTGGTGATCGTGTTGCACTGGAAC
CGGGTATTAGCTGTCGTCAGTGTCAGCTGTGTAAAGATGGTCGTTATAATCTGTGCCGCAAAATGAAATTTTTC***GGTAGC***CCTCCGACCAATGGTGC
ACTGGCAAATCAGGTTGTTCATCCGGCAGATCTGTGTTTTAAACTGCCGGATAATGTTAGCCTGGAAGAAGGTGCAATTTGCGAACCGCTGAGCG
TTGGTGTTCATGCATGTCGTCGTGCAAATGTTGGTCCGGATACCAAACTGATTATTATCGGTGCAGGTCCGATTGGTCTGGTTACCATGCTGGCAGC
ACGTGCATTTGGTAGCCCGAAAATTGTTATTGTTGATGTGGATGATTGCCGTCTGAGCTTTGCAAAAGATCTGGGTGCAGATGAAATTATCAAAGT
GAGCAGCATTATGCAGGATGTCGAAGAAGAAGTTGTTCGTATTCGTAATGCAATGGGTGGTCCGGTTGATCTGAGTTTTGATTGTGTTGGTTTCAA
CAAAACCATGACCACCGCACTGCAGGCAACCCGTACCGGTGGTAAAGTTTGTCTGGTTGGTCTGGGTCAGAGCGAAATGACCCTGCCGCTGACC

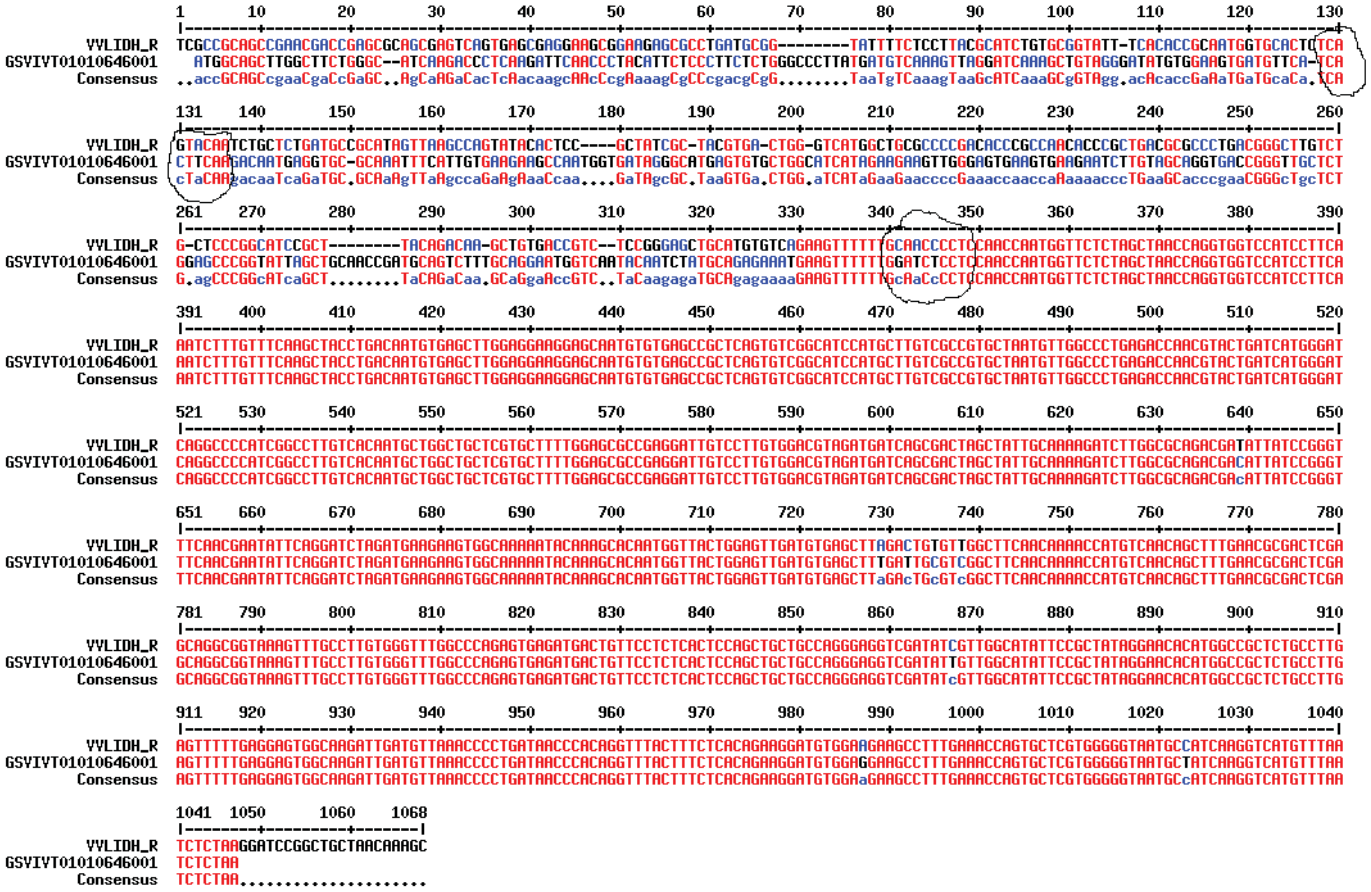
AGCGCAGCAGCCCGTGAAGTTGATGTTATTGGTATTTTTCGTTATCGCAATACCTGGCCTCTGTGTATTGAACTGCTGCGCACCGGTGAAATTGATG
TGAAACCGCTGATTACCCACCGTTATAACTTTACCCAAGAGGGTGTTGATGAAGCATTTGAAACCAGCAGCCGTGGTGGTAATGCCATTAAAGTTA
TGTTTAATCTGTAA**GGATCC**

2.3 Supplementary file S3: Sequencing results

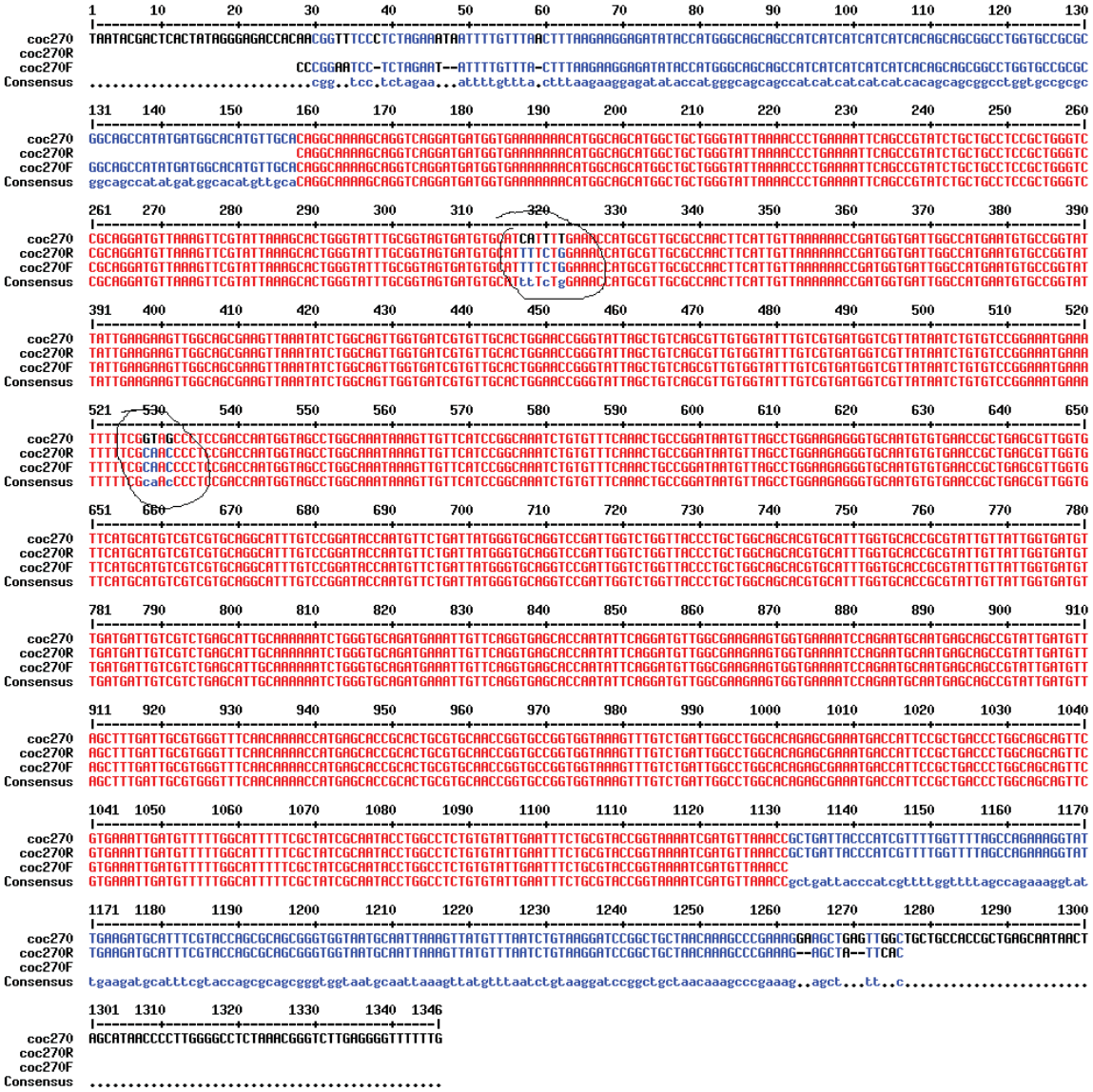
>GSVIVT01010646001_YL



>GSVIVT01010646001_YL_AT



>Tco3_g019270_YL_AT



>PtLIDH_YL_AT

Bibliography

(Chapter 1, 3, 4 & 5)

Adams, K. L. and J. F. Wendel (2005). "Polyploidy and genome evolution in plants." Current Opinion in Plant Biology **8**(2): 135-141.

Adams, P. D., P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallographica Section D-Biological Crystallography **66**: 213-221.

Amerine, M. A., H. W. Berg and W. V. Cruess (1972). The technology of wine making, Avi Pub. Co.

Ameyama, M. and O. Adachi (1982). "2-Keto-D-gluconate reductase from acetic-acid bacteria." Methods in Enzymology **89**: 203-210.

Antonyuk, S. V., R. W. Strange, M. J. Ellis, Y. Bessho, S. Kuramitsu, Y. Inoue, S. Yokoyama and S. S. Hasnain (2009). "Structure of D-lactate dehydrogenase from Aquifex aeolicus complexed with NAD(+) and lactic acid (or pyruvate)." Acta Crystallographica Section F-Structural Biology and Crystallization Communications **65**: 1209-1213.

Baack, R. D., J. Markwell, P. L. Herman and J. C. Osterman (2003). "Kinetic behavior of the Arabidopsis thaliana leaf formate dehydrogenase is thermally sensitive." Journal of Plant Physiology **160**(5): 445-450.

Bantog, N. A., K. Shiratake and S. Yamaki (1999). "Changes in sugar content and sorbitol- and sucrose-related enzyme activities during development of loquat (Eriobotrya japonica Lindl. cv. Mogi) fruit." Journal of the Japanese Society for Horticultural Science **68**(5): 942-948.

Bantog, N. A., K. Yamada, N. Niwa, K. Shiratake and S. Yamaki (2000). "Gene expression of NAD(+)-dependent sorbitol dehydrogenase and NADP(+)-dependent sorbitol-6-phosphate dehydrogenase during development of loquat (Eriobotrya japonica Lindl.) fruit." Journal of the Japanese Society for Horticultural Science **69**(3): 231-236.

Barch, W. E. (1933). "The oxidation of 5-keto-gluconic acid with nitric acid in the presence of vanadium." Journal of The American Chemistry Society **55**: 3653-3658.

Bartlett, G. J., C. T. Porter, N. Borkakoti and J. M. Thornton (2002). "Analysis of catalytic residues in enzyme active sites." Journal of Molecular Biology **324**(1): 105-121.

Battye, T. G. G., L. Kontogiannis, O. Johnson, H. R. Powell and A. G. W. Leslie (2011). "iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM." Acta Crystallographica Section D-Biological Crystallography **67**: 271-281.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne (2000). "The Protein Data Bank." Nucleic Acids Research **28**(1): 235-242.

Bernard, N., K. Johnsen, T. Ferain, D. Garmyn, P. Hols, J. J. Holbrook and J. Delcour (1994). "NAD-dependent D-2-hydroxyisocaproate dehydrogenase of Lactobacillus delbrueckii subsp bulgaricus - Gene cloning and enzyme characterization." European Journal of Biochemistry **224**(2): 439-446.

Beruter, J. (1985). "Sugar accumulation and changes in the activities of related enzymes during development of the apple fruit." Journal of Plant Physiology **121**(4): 331-341.

Booth, M. P. S., R. Conners, G. Rumsby and R. L. Brady (2006). "Structural basis of substrate specificity in human glyoxylate reductase/hydroxypyruvate reductase." Journal of Molecular Biology **360**(1): 178-189.

Boss, P. K. and M. R. Thomas (2002). "Association of dwarfism and floral induction with a grape 'green revolution' mutation." Nature **416**(6883): 847-850.

Bouckaert, R., J. Heled, D. Kuhnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut and A. J. Drummond (2014). "BEAST 2: A software platform for Bayesian evolutionary analysis." Plos Computational Biology **10**(4).

Boulton, R. (1980). "The general relationship between potassium, sodium and pH in grape juice and wine." American Journal of Enology and Viticulture **31(2)**:182-186.

Burbidge, C. A. (2011). Identification and characterization of the enzymes involved in the biosynthetic pathway of tartaric acid in Vitis vinifera. Doctor of Philosophy, The University of Flinders.

Castineira, A., R. M. Pena, C. Herrero and S. Garcia-Martin (2002). "Analysis of organic acids in wine by capillary electrophoresis with direct UV detection." Journal of Food Composition and Analysis **15**(3): 319-331.

Chaib, J., L. Torregrosa, D. Mackenzie, P. Corena, A. Bouquet and M. R. Thomas (2010). "The grape microvine - a model system for rapid forward and reverse genetics of grapevines." Plant Journal **62**(6): 1083-1092.

Chen, V. B., W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." Acta Crystallographica Section D-Biological Crystallography **66**: 12-21.

Conant, G. C. and K. H. Wolfe (2008). "Turning a hobby into a job: How duplicated genes find new functions." Nature Reviews Genetics **9**(12): 938-950.

Coombe, B. G. and M. G. McCarthy (2000). "Dynamics of grape berry growth and physiology of ripening." Australian Journal of Grape and Wine Research **6**(2): 131-135.

Dai, M. S., Z. B. Shi and C. J. Xu (2015). "Genome-wide analysis of sorbitol dehydrogenase (SDH) Genes and their differential expression in two sand pear (Pyrus pyrifolia) fruits." International Journal of Molecular Sciences **16**(6): 13065-13083.

Danielsson, O., S. Atrian, T. Luque, L. Hjelmqvist, R. Gonzalezduarte and H. Jornvall (1994). "Fundamental molecular differences between alcohol-dehydrogenase classes." Proceedings of the National Academy of Sciences of the United States of America **91**(11): 4980-4984.

Dawson, R. M. C. (1959). Data for biochemical research, Oxford, Clarendon Press.

De Bodt, S., S. Maere and Y. Van de Peer (2005). "Genome duplication and the origin of angiosperms." Trends in Ecology & Evolution **20**(11): 591-597.

de Sousa, S. M., M. D. Paniago, P. Arruda and J. A. Yunes (2008). "Sugar levels modulate sorbitol dehydrogenase expression in maize." Plant Mol Biol **68**(3): 203-213.

DeBolt, S. (2006). L-Tartaric acid synthesis in plants. Doctor of Philosophy, The University of Adelaide.

DeBolt, S., D. R. Cook and C. M. Ford (2006). "L-Tartaric acid synthesis from vitamin C in higher plants." Proceedings of the National Academy of Sciences of the United States of America **103**(14): 5608-5613.

DeBolt, S., J. Hardie, S. Tyerman and C. M. Ford (2004). "Composition and synthesis of raphide crystals and druse crystals in berries of Vitis vinifera L. cv. Cabernet Sauvignon: Ascorbic acid as precursor for both oxalic and tartaric acids as revealed by radiolabelling studies." Australian Journal of Grape and Wine Research **10**(2): 134-142.

DeBolt, S., V. Melino and C. M. Ford (2007). "Ascorbate as a biosynthetic precursor in plants." Annals of Botany **99**(1): 3-8.

Dengler, U., K. Niefind, M. Kiess and D. Schomburg (1997). "Crystal structure of a ternary complex of D-2-hydroxyisocaproate dehydrogenase from Lactobacillus casei, NAD(+) and 2-oxoisocaproate at 1.9 angstrom resolution." Journal of Molecular Biology **267**(3): 640-660.

Dey, S., G. A. Grant and J. C. Sacchettini (2005). "Crystal structure of Mycobacterium tuberculosis

D-3-phosphoglycerate dehydrogenase." Journal of Biological Chemistry **280**(15): 14892-14899.

Doehlert, D. C. (1987). "Ketose reductase-activity in developing maize endosperm." Plant Physiology **84**(3): 830-834.

Duester, G., J. Farres, M. R. Felder, R. S. Holmes, J. O. Hoog, X. Pares, B. V. Plapp, S. J. Yin and H. Jornvall (1999). "Recommended nomenclature for the vertebrate alcohol dehydrogenase gene family." Biochemical Pharmacology **58**(3): 389-395.

Emsley, P., B. Lohkamp, W. G. Scott and K. Cowtan (2010). "Features and development of Coot." Acta Crystallographica Section D-Biological Crystallography **66**: 486-501.

Flagel, L. E. and J. F. Wendel (2009). "Gene duplication and evolutionary novelty in plants." New Phytologist **183**(3): 557-564.

Force, A., M. Lynch, F. B. Pickett, A. Amores, Y. L. Yan and J. Postlethwait (1999). "Preservation of duplicate genes by complementary, degenerative mutations." Genetics **151**(4): 1531-1545.

Freddi, G., G. Allara and G. Candiani (1996). "Degumming of silk fabrics with tartaric acid." Journal of the Society of Dyers and Colourists **112**(7-8): 191-195.

Fuleki, T., E. Pelayo and R. B. Palabay (1995). "Carboxylic-acid composition of varietal juices produced from fresh and stored apples." J Agr Food Chem **43**(3): 598-607.

Galdon, B. R., D. R. Mesa, E. M. R. Rodriguez and C. D. Romero (2010). "Influence of the cultivar on the organic acid and sugar composition of potatoes." J Sci Food Agric **90**(13): 2301-2309.

Gerós, H., M. Chaves and S. Delrot (2012). The biochemistry of the grape berry, Bentham Science Publishers.

Goldberg, J. D., T. Yoshida and P. Brick (1994). "Crystal structure of a NAD-dependent D-glycerate dehydrogenase at 2.4 angstrom resolution." Journal of Molecular Biology **236**(4): 1123-1140.

Grant, G. A. (1989). "A new family of 2-hydroxyacid dehydrogenases." Biochemical and Biophysical Research Communications **165**(3): 1371-1374.

Green, M. A. and S. C. Fry (2005). "Vitamin C degradation in plant cells via enzymatic hydrolysis of 4-O-oxalyl-L-threonate." Nature **433**(7021): 83-87.

Guo, Z. X., T. F. Pan, K. T. Li, F. L. Zhong, L. Lin, D. M. Pan and L. X. Lu (2012). "Cloning of NAD-SDH cDNA from plum fruit and its expression and characterization." Plant Physiology and Biochemistry **57**: 175-180.

Hale, C. R. (1962). "Synthesis of organic acids in the fruit of the grape." Nature **195**: 917-918.

Han, X. D., X. H. Xiong, X. L. Hu, M. T. Li, W. C. Zhang and X. Q. Liu (2014). "Crystallization and structural analysis of 2-hydroxyacid dehydrogenase from Ketogulonicigenium vulgare." Biotechnology Letters **36**(2): 295-300.

Hancock, R. D. and R. Viola (2005). "Biosynthesis and catabolism of L-ascorbic acid in plants." Critical Reviews in Plant Sciences **24**(3): 167-188.

Hardie, W. J. (2000). "Grapevine biology and adaptation to viticulture." Australian Journal of Grape and Wine Research **6**: 74-81.

Hayward, S. and H. J. C. Berendsen (1998). "Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme." Proteins-Structure Function and Genetics **30**(2): 144-154.

Hilbert, B. J., B. L. Morris, K. C. Ellis, J. L. Paulsen, C. A. Schiffer, S. R. Grossman and W. E. Royer (2015). "Structure-guided design of a high affinity inhibitor to human CtBP." ACS Chemical Biology **10**(4): 1118-1127.

Hough, L. and J. K. Jones (1956). "The biosynthesis of the monosaccharides." Advances in Carbohydrate

Chemistry **48**(11): 185-262.

Hudina, M. and F. Stampar (2000). "Sugars and organic acids contents of European (Pyrus communis L.) and Asian (Pyrus serotina Rehd.) pear cultivars." Acta Alimentaria **29**(3): 217-230.

Hughes, A. L. (1994). "The evolution of functionally novel proteins after gene duplication." Proceedings of the Royal Society B-Biological Sciences **256**(1346): 119-124.

Hughes, A. L. (2002). "Adaptive evolution after gene duplication." Trends in Genetics **18**(9): 433-434.

Hurles, M. (2004). "Gene duplication: The genomic trade in spare parts." Plos Biology **2**(7): 900-904.

Iida, M., N. A. Bantog, K. Yamada, K. Shiratake and S. Yamaki (2004). "Sorbitol- and other sugar-induced expressions of the NAD+-dependent sorbitol dehydrogenase gene in Japanese pear fruit." Journal of the American Society for Horticultural Science **129**(6): 870-875.

Iland, P. G. and B. G. Coombe (1988). "Malate, tartrate, potassium, and sodium in flesh and skin of shiraz grapes during ripening- concentration and compartmentation." American Journal of Enology and Viticulture **39**(1): 71-76.

Isbell, H. S. and N. B. Holt (1945). "Oxidation of galacturonic acid and of 5-keto-gluconic acid in alkaline solution." Journal of Research of the National Bureau of Standards **35**(5): 433-438.

Ito, A., H. Hayama and Y. Kashimura (2005). "Partial cloning and expression analysis of genes encoding NAD(+)-dependent sorbitol dehydrogenase in pear bud during flower bud formation." Scientia Horticulturae **103**(4): 413-420.

Jackson, R. S. (2000). Wine Science: Principles, practice, perception, Elsevier Science.

Jaillon, O., J. M. Aury, B. Noel, A. Policriti, C. Clepet, A. Casagrande, N. Choisne, S. Aubourg, N. Vitulo, C. Jubin, A. Vezzi, F. Legeai, P. Hugueney, C. Dasilva, D. Horner, E. Mica, D. Jublot, J. Poulain, C. Bruyere, A. Billault, B. Segurens, M. Gouyvenoux, E. Ugarte, F. Cattonaro, V. Anthouard, V. Vico, C. Del Fabbro, M. Alaux, G. Di Gaspero, V. Dumas, N. Felice, S. Paillard, I. Juman, M. Moroldo, S. Scalabrin, A. Canaguier, I. Le Clainche, G. Malacrida, E. Durand, G. Pesole, V. Laucou, P. Chatelet, D. Merdinoglu, M. Delledonne, M. Pezzotti, A. Lecharny, C. Scarpelli, F. Artiguenave, M. E. Pe, G. Valle, M. Morgante, M. Caboche, A. F. Adam-Blondon, J. Weissenbach, F. Quetier, P. Wincker and F.-I. Public (2007). "The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla." Nature **449**(7161): 463-465.

Janiak, V., M. Petersen, M. Zentgraf, G. Klebe and A. Heine (2010). "Structure and substrate docking of a hydroxy(phenyl)pyruvate reductase from the higher plant Coleus blumei Benth." Acta Crystallographica Section D-Biological Crystallography **66**: 593-603.

Jeffery, J., J. Chesters, C. Mills, P. J. Sadler and H. Jornvall (1984). "Sorbitol dehydrogenase is a zinc enzyme." EMBO Journal **3**(2): 357-360.

Jeffery, J., L. Cummins, M. Carlquist and H. Jornvall (1981). "Properties of sorbitol dehydrogenase and characterization of a reactive cysteine residue reveal unexpected similarities to alcohol dehydrogenases." European Journal of Biochemistry **120**(2): 229-234.

Jia, Y., D. C. J. Wong, C. Sweetman, J. B. Bruning and C. M. Ford (2015). "New insights into the evolutionary history of plant sorbitol dehydrogenase." Bmc Plant Biology **15**.

Johansson, K., M. El-Ahmad, C. Kaiser, H. Jornvall, H. Eklund, J. O. Hoog and S. Ramaswamy (2001). "Crystal structure of sorbitol dehydrogenase." Chemico-Biological Interactions **130**(1-3): 351-358.

Jones, D. T., W. R. Taylor and J. M. Thornton (1992). "The rapid generation of mutation data matrices from protein sequences." Computer Applications in the Biosciences **8**(3): 275-282.

Jornvall, H., O. Danielsson, H. Eklund, L. Hjelmqvist, J. O. Hoog, X. Pares and J. Shafqat (1993). "Enzyme and isozyme developments within the Medium-Chain Alcohol-Dehydrogenase family."

Enzymology and Molecular Biology of Carbonyl Metabolism 4 **328**: 533-544.

Jornvall, H., E. Nordling and B. Persson (2003). "Multiplicity of eukaryotic ADH and other MDR forms." Chemico-Biological Interactions **143**: 255-261.

Jornvall, H., M. Persson and J. Jeffery (1981). "Alcohol and polyol dehydrogenases are both divided into 2 protein types, and structural-properties cross-relate the different enzyme-activities within each type." Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences **78**(7): 4226-4230.

Jose, S. (2008). Tartaric acid: A global strategic business report, Global Business Analysts, Inc.

Kassaian (2003). Tartaric acid, Wiley-VCH.

Kim, H. Y., J. C. Ahn, J. H. Choi, B. Hwang and D. W. Choi (2007). "Expression and cloning of the full-length cDNA for sorbitol-6-phosphate dehydrogenase and NAD-dependent sorbitol dehydrogenase from pear (Pyrus pyrifolia N.)." Scientia Horticulturae **112**(4): 406-412.

Kim, K. H., V. Janiak and M. Petersen (2004). "Purification, cloning and functional expression of hydroxyphenylpyruvate reductase involved in rosmarinic acid biosynthesis in cell cultures of Coleus blumei." Plant Molecular Biology **54**(3): 311-323.

Klasen, R., S. Bringermeyer and H. Sahm (1992). "Incapability of gluconobacter oxydans to produce tartaric acid." Biotechnology and Bioengineering **40**(1): 183-186.

Kleczkowski, L. A. and G. E. Edwards (1989). "Identification of Hydroxypyruvate and Glyoxylate Reductases in Maize Leaves." Plant Physiology **91**(1): 278-286.

Kleczkowski, L. A., G. E. Edwards, R. D. Blackwell, P. J. Lea and C. V. Givan (1990). "Enzymology of the reduction of hydroxypyruvate and glyoxylate in a mutant of barley lacking peroxisomal hydroxypyruvate reductase." Plant Physiology **94**(2): 819-825.

Kleczkowski, L. A. and D. D. Randall (1988). "Purification and characterization of a novel NADPH(NADH)-dependent hydroxypyruvate reductase from spinach Leaves - Comparison of immunological properties of leaf hydroxypyruvate reductases." Biochemical Journal **250**(1): 145-152.

Kotera, U., K. Yamada, K. Umehara and T. Kodama (1972). "Microbial formations of tartaric acid from glucose. 4. Isolation method of highly tartaric acid producing mutants of gluconobacter-suboxydans." Agricultural and Biological Chemistry **36**(8): 1307-1313.

Kumar, V., J. E. Carlson, K. A. Ohgi, T. A. Edwards, D. W. Rose, C. R. Escalante, M. G. Rosenfeld and A. K. Aggarwal (2002). "Transcription corepressor CtBP is an NAD(+)-regulated dehydrogenase." Molecular Cell **10**(4): 857-869.

Kuo, T. M., D. C. Doehlert and C. G. Crawford (1990). "Sugar metabolism in germinating soybean seeds - Evidence for the sorbitol pathway in soybean axes." Plant Physiol **93**(4): 1514-1520.

Lamzin, V. S., Z. Dauter, V. O. Popov, E. H. Harutyunyan and K. S. Wilson (1994). "High resolution structures of holo and apo formate dehydrogenase." Journal of Molecular Biology **236**(3): 759-785.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Lavery, R. and S. Sacquin-Mora (2007). "Protein mechanics: a route from structure to function." Journal of Biosciences **32**(5): 891-898.

Lichtenthaler, F. W. (2002). "Unsaturated O- and N-heterocycles from carbohydrate feedstocks." Accounts of Chemical Research **35**(9): 728-737.

Lindstad, R. I., L. F. Hermansen and J. S. McKinley-Mckee (1992). "The kinetic mechanism of sheep liver sorbitol dehydrogenase." European Journal of Biochemistry **210**(2): 641-647.

Lindstad, R. I., P. Koll and J. S. McKinley-McKee (1998). "Substrate specificity of sheep liver sorbitol dehydrogenase." J Biochemical **330**(Pt 1): 479-487.

Loescher, W. H. (1987). "Physiology and metabolism of sugar alcohols in higher-plants." Physiologia Plantarum **70**(3): 553-557.

Loescher, W. H., G. C. Marlow and R. A. Kennedy (1982). "Sorbitol metabolism and sink-source interconversions in developing apple leaves." Plant Physiology **70**(2): 335-339.

Loewus, F. A. (1957). "Observations on the incorpotation of C14 into tartaric acid." Plant physiology **33**(2): 155-156.

Loewus, F. A. (1999). "Biosynthesis and metabolism of ascorbic acid in plants and of analogs of ascorbic acid in fungi." Phytochemistry **52**(2): 193-210.

Loewus, F. A. and H. A. Stafford (1958). "Observations on the incorporation of C-14 into tartaric acid and the labeling pattern of D-glucose from an excised grape leaf administered L-ascorbic acid-6-C-14." Plant Physiology **33**(2): 155-156.

M. Francisca Aguayo, J. C. C., Matías Fuentealba, Rodrigo Muñoz, Claudia Stange, Ricardo Cabrera and Michael Handford (2015). "Polyol specificity of recombinant Arabidopsis thaliana sorbitol dehydrogenase studied by enzyme kinetics and in silico modeling." Frontiers in Plant Science.

Malipiero, U., H. P. Ruffner and D. M. Rast (1987). "Ascorbic to tartaric acid conversion in grapevines." Journal of Plant Physiology **129**(1–2): 33-40.

Mano, S., M. Hayashi, M. Kondo and M. Nishimura (1997). "Hydroxypyruvate reductase with a carboxy-terminal targeting signal to microbodies is expressed in Arabidopsis." Plant and Cell Physiology **38**(4): 449-455.

Manuel Salgado, J., N. Rodriguez, S. Cortes and J. Manuel Dominguez (2010). "Improving downstream processes to recover tartaric acid, tartrate and nutrients from vinasses and formulation of inexpensive fermentative broths for xylitol production." Journal of the Science of Food and Agriculture **90**(13): 2168-2177.

Mares, D. J. and H. Marschner (1980). "Assimilate conversion in potato-tubers in relation to starch deposition and cell-growth." Berichte Der Deutschen Botanischen Gesellschaft **93**(1): 299-313.

Maret, W. (1997). "Human sorbitol dehydrogenase - A secondary alcohol dehydrogenase with distinct pathophysiological roles - pH-dependent kinetic studies." Enzymology and Molecular Biology of Carbonyl Metabolism 6 **414**: 383-393.

Maria, F. A., A. Diego, M. Patricio, P. Roberto, M. Rodrigo, G. Marta, A. Teresa, C. Ricardo, S. Claudia and H. Michael (2013). "Sorbitol dehydrogenase is a cytosolic protein required for sorbitol metabolism in Arabidopsis thaliana." Plant Science **205-206**(1): 63-75.

Matías D Hartman, C. M. F., Claudia V Piattoni and Alberto A Iglesias (2014). "Glucitol dehydrogenase from peach (Prunus persica) fruits is regulated by thioredoxin." Plant Cell Physiol **55**(6): 1157-1168.

Mattick, L. R., R. A. Plane and L. D. Weirs (1980). "Lowering wine acidity with carbonates." American Journal of Enology and Viticulture **31**(4): 350-355.

Mccoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." Journal of Applied Crystallography **40**: 658-674.

Mdluli, K., M. P. S. Booth, R. L. Brady and G. Rumsby (2005). "A preliminary account of the properties of recombinant human glyoxylate reductase (GRHPR), LDHA and LDHB with glyoxylate, and their potential roles in its metabolism." Biochimica Et Biophysica Acta-Proteins and Proteomics **1753**(2): 209-216.

Melino, V. J. (2009). Ascorbate metabolism in grape berries during development. Doctor of Philosophy,

The University of Adelaide.

Miles;, C. M. and M. Wayne (2008). "Quantitative trait locus (QTL) analysis." Nature Education **1**(208).

Milsom, P. E. (1987). Organic acids by fermentation, especially citric acid. Food Biotechnology-1. R. D. King and P. S. J. Cheetham, Elsevier Applied Science**:** 273-307.

Murray, A. J. S., R. D. Blackwell and P. J. Lea (1989). "Metabolism of Hydroxypyruvate in a Mutant of Barley Lacking Nadh-Dependent Hydroxypyruvate Reductase, an Important Photorespiratory Enzyme-Activity." Plant Physiology **91**(1): 395-400.

Nardini, M., S. Spano, C. Cericola, A. Pesce, A. Massaro, E. Millo, A. Luini, D. Corda and M. Bolognesi (2003). "CtBP/BARS: a dual-function protein involved in transcription co-repression, and Golgi membrane fission." EMBO Journal **22**(12): 3122-3130.

Negm, F. B. and W. H. Loescher (1979). "Detection and characterization of sorbitol dehydrogenase from apple callus-tissue." Plant Physiology **64**(1): 69-73.

Ng, K., R. Q. Ye, X. C. Wu and S. L. Wong (1992). "Sorbitol dehydrogenase from Bacillus-subtilis - Purification, characterization, and gene cloning." Journal of Biological Chemistry **267**(35): 24989-24994.

Nordling, E., H. Jornvall and B. Persson (2002). "Medium-chain dehydrogenases/reductases (MDR) - Family characterizations including genome comparisons and active site modelling." European Journal of Biochemistry **269**(17): 4267-4276.

Nosarszewski, M., A. M. Clements, A. B. Downie and D. D. Archbold (2004). "Sorbitol dehydrogenase expression and activity during apple fruit set and early development." Physiologia Plantarum **121**(3): 391-398.

Nosarzewski, M. and D. D. Archbold (2007). "Tissue-specific expression of sorbitol dehydrogenase in apple fruit during early development." Journal of Experimental Botany **58**(7): 1863-1872.

Nosarzewski, M., A. B. Downie, B. Wu and D. D. Archbold (2012). "The role of sorbitol dehydrogenase in Arabidopsis thaliana." Functional Plant Biology **39**(6): 462-470.

Nour, V., I. Trandafir and M. E. Ionica (2010). "HPLC organic acid analysis in different citrus juices under reversed phase conditions." Notulae Botanicae Horti Agrobotanici Cluj-Napoca **38**(1): 44-48.

Ohshima, T., N. Nunoura-Kominato, T. Kudome and H. Sakuraba (2001). "A novel hyperthermophilic archaeal glyoxylate reductase from Thermococcus litoralis - Characterization, gene cloning, nucleotide sequence and expression in Escherichia coli." European Journal of Biochemistry **268**(17): 4740-4747.

Ohta, K., R. Moriguchi, K. Kanahama, S. Yamaki and Y. Kanayama (2005). "Molecular evidence of sorbitol dehydrogenase in tomato, a non-Rosaceae plant." Phytochemistry **66**(24): 2822-2828.

OIV (2014). Global economic viticulture data. J. M. Aurand, International Organization of Vine and Wine.

Orengo, C. A., A. E. Todd and J. M. Thornton (1999). "From protein structure to function." Current Opinion in Structural Biology **9**(3): 374-382.

Oura, Y., K. Yamada, K. Shiratake and S. Yamaki (2000). "Purification and characterization of a NAD(+)-dependent sorbitol dehydrogenase from Japanese pear fruit." Phytochemistry **54**(6): 567-572.

Park, S. W., K. J. Song, M. Y. Kim, J. H. Hwang, Y. U. Shin, W. C. Kim and W. I. Chung (2002). "Molecular cloning and characterization of four cDNAs encoding the isoforms of NAD-dependent sorbitol dehydrogenase from the Fuji apple." Plant Science **162**(4): 513-519.

Parnham, M. J. and K. Kesselring (1985). "Rosmarinic acid." Drugs of the Future **10**(9): 756-757.

Pauly, T. A., J. L. Ekstrom, D. A. Beebe, B. Chrunyk, D. Cunningham, M. Griffor, A. Kamath, S. E. Lee, R. Madura, D. McGuire, T. Subashi, D. Wasilko, P. Wafts, B. L. Mylari, P. J. Oates, P. D. Adams and V. L. Rath (2003). "X-ray crystallographic and kinetic studies of human sorbitol dehydrogenase." Structure

11(9): 1071-1085.

Persson, B., J. Hedlund and H. Jornvall (2008). "The MDR superfamily." Cellular and Molecular Life Sciences **65**(24): 3879-3894.

Petsko, G. A. and D. Ringe (2004). Protein Structure and Function, New Science Press.

Prave, P., U. Faust, W. Sittig and D. Sukatsch (1987). Fundamentals of biotechnology, VCH.

Razeto, A., S. Kochhar, H. Hottinger, M. Dauter, K. S. Wilson and V. S. Lamzin (2002). "Domain closure, substrate specificity and catalysis of D-Lactate dehydrogenase from Lactobacillus bulgaricus." Journal of Molecular Biology **318**(1): 109-119.

Rensing, S. A., D. Lang, A. D. Zimmer, A. Terry, A. Salamov, H. Shapiro, T. Nishiyama, P. F. Perroud, E. A. Lindquist, Y. Kamisugi, T. Tanahashi, K. Sakakibara, T. Fujita, K. Oishi, T. Shin-I, Y. Kuroki, A. Toyoda, Y. Suzuki, S. Hashimoto, K. Yamaguchi, S. Sugano, Y. Kohara, A. Fujiyama, A. Anterola, S. Aoki, N. Ashton, W. B. Barbazuk, E. Barker, J. L. Bennetzen, R. Blankenship, S. H. Cho, S. K. Dutcher, M. Estelle, J. A. Fawcett, H. Gundlach, K. Hanada, A. Heyl, K. A. Hicks, J. Hughes, M. Lohr, K. Mayer, A. Melkozernov, T. Murata, D. R. Nelson, B. Pils, M. Prigge, B. Reiss, T. Renner, S. Rombauts, P. J. Rushton, A. Sanderfoot, G. Schween, S. H. Shiu, K. Stueber, F. L. Theodoulou, H. Tu, Y. Van de Peer, P. J. Verrier, E. Waters, A. Wood, L. X. Yang, D. Cove, A. C. Cuming, M. Hasebe, S. Lucas, B. D. Mishler, R. Reski, I. V. Grigoriev, R. S. Quatrano and J. L. Boore (2008). "The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants." Science **319**(5859): 64-69.

Rivas, B., A. Torrado, A. Belen Moldes and J. Manuel Dominguez (2006). "Tartaric acid recovery from distilled lees and use of the residual solid as an economic nutrient for Lactobacillus." Journal of Agricultural and Food Chemistry **54**(20): 7904-7911.

Rosenberg, M., H. Mikova and L. Kristofikova (1999). "Production of L-tartaric acid by immobilized bacterial cells Nocardia tartaricans." Biotechnology Letters **21**(6): 491-495.

Ruffner, H. P. (1982). "Metabolism of tartaric and malic acids in Vitis-A review." Vitis **21**: 247-259.

Saito, K. (1984). "Synthesis of L-(+) -tartaric acid from 5-Ketio-D-gluconic acid in Pelargonium." Plant and Cell Physiology **25**(7): 1223-1232.

Saito, K. (1992). "Metabolism of L-threotetruronic acid by Pelargonium-crispum." Phytochemistry **31**(4): 1219-1222.

Saito, K. and Z. Kasai (1968). "Accumulation of tartaric acid in the ripening process of grape." Plant and Cell Physiology **9**(3): 529-537.

Saito, K. and Z. Kasai (1969). "Tartaric acid synthesis from L-Ascorbic acid-C-14 in grape berries." Phytochemistry **8**(11): 2177-2182.

Saito, K. and Z. Kasai (1982). "Conversion of L-ascorbic acid to L-idonic acid, L-idono gamma-lactone and 2-keto-L-idonic acid inslices of immature grapes." Plant and Cell Physiology **23**(3): 499-507.

Saito, K. and Z. Kasai (1984). "Synthesis of L-(+)-tartaric acid from L-ascorbic acid via 5-Keto-D-gluconic acid in grapes." Plant Physiology **76**(1): 170-174.

Saito, K. and F. A. Loewus (1989). "Formation of tartaric acid in Vitaceous plants - Relative contributions of L-ascorbic acid-inclusive and acid-noninclusive pathways." Plant Cell Physiol **30**(6): 905-910.

Saito, K. and F. A. Loewus (1989). "Occurrence of dextro-tartaric acid and its formation from D-gluconate or D-xylo-5-hexulosonate in bean leaf (Phaseolus vulgaris L.)." Plant Science **62(2)**: 175-180.

Saito, K. and F. A. Loewus (1989). "Occurrence of L(+)-Tartaric acid and its formation from D-Gluconate or D-Xylo-5-Hexulosonate in bean leaf (Phaseolus Vulgaris L)." Plant Science **62**(2): 175-180.

Saito, K., J. Ohmoto and N. Kuriha (1997). "Incorporation of O-18 into oxalic, L-threonic and L-tartaric acids during cleavage of L-ascorbic and 5-keto-D-gluconic acids in plants." Phytochemistry **44**(5): 805-809.

Saito, L. (1979). "The metabolism of L-[6-14C]ascorbic acid in detached grape leaves." Plant Cell Physiol **20**: 1481-1488.

Salusjarvi, T., M. Povelainen, N. Hvorslev, E. V. Eneyskaya, A. A. Kulminskaya, K. A. Shabalin, K. N. Neustroev, N. Kalkkinen and A. N. Miasnikov (2004). "Cloning of a gluconate/polyol dehydrogenase gene from Gluconobacter suboxydans IFO 12528, characterisation of the enzyme and its use for the production of 5-ketogluconate in a recombinant Escherichia coli strain." Applied Microbiology and Biotechnology **65**(3): 306-314.

Schuller, D. J., G. A. Grant and L. J. Banaszak (1995). "The Allosteric Ligand Site in the V-Max-Type Cooperative Enzyme Phosphoglycerate Dehydrogenase." Nature Structural Biology **2**(1): 69-76.

Sha, S. F., J. C. Li, J. Wu and S. L. Zhang (2011). "Characteristics of organic acids in the fruit of different pear species." African Journal of Agricultural Research **6**(10): 2403-2410.

Singh, B. (1976). "Synthesis of DL-2-amino-1-butanol." US Patent 3944617.

Smirnoff, N. and G. L. Wheeler (2000). "Ascorbic acid in plants: Biosynthesis and function." Critical Reviews in Biochemistry and Molecular Biology **35**(4): 291-314.

Solieri, L. and P. Giudici (2009). Vinegars of the world, Springer Science & Business Media.

Soltis, D. E., Senters, A. E., Zanis, M. J., Kim, S., Thompson, J. D., Soltis, P. S., De Craene, L. P. R., Endress, P. K. and Farris, J. S. (2003). "Gunnerales are sister to other core eudicots: implications for the evolution of pentamery." American Journal of Botany **90(3)**:461-470.

Stafford, H. A. (1959). "Distribution of tartaric acid in the leaves of certain angiosperms." American Journal of Botany **46**(5): 347-352.

Stoll, V. S., M. S. Kimber and E. F. Pai (1996). "Insights into substrate binding by D-2-ketoacid dehydrogenases from the structure of Lactobacillus pentosus D-lactate dehydrogenase." Structure **4**(4): 437-447.

Strommer, J. (2011). "The plant ADH gene family." Plant Journal **66**(1): 128-142.

Suarez, M. H., E. R. Rodriguez and C. D. Romero (2008). "Analysis of organic acid content in cultivars of tomato harvested in Tenerife." European Food Research and Technology **226**(3): 423-435.

Sun, Z., C. Ma, J. Zhou and S. Zhu (2013). "Cloning, expression, purification and assay of sorbitol dehydrogenase from "Feicheng" peach fruit (Prunus persica)." Brazilian Archives of Biology and Technology **56**(4): 531-539.

Sutsawat, D., K. Yamada, K. Shiratake, Y. Kanayama and S. Yamaki (2008). "Properties of sorbitol dehydrogenase in strawberry fruit and enhancement of the activity by fructose and auxin." Journal of the Japanese Society for Horticultural Science **77**(3): 318-323.

Sweetman, C., L. G. Deluc, G. R. Cramer, C. M. Ford and K. L. Soole (2009). "Regulation of malate metabolism in grape berry and other developing fruits." Phytochemistry **70**(11-12): 1329-1344.

Sweetman, C., D. C. J. Wong, C. M. Ford and D. P. Drew (2012). "Transcriptome analysis at four developmental stages of grape berry (Vitis vinifera cv. Shiraz) provides insights into regulated and coordinated gene expression." BMC Genomics **13**(1): 691.

Synoradzki, L., P. Ruskowski and U. Bernas (2005). "Tartaric acid and its O-acyl derivatives. Part 1. Synthesis of tartaric acid and O-acyl tartaric acids and anhydrides." Organic Preparations and Procedures International **37**(1): 37-63.

Szabo, E., A. Thelen and M. Petersen (1999). "Fungal elicitor preparations and methyl jasmonate

enhance rosmarinic acid accumulation in suspension cultures of Coleus blumei." Plant Cell Reports **18**(6): 485-489.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013). "MEGA6: Molecular evolutionary genetics analysis Version 6.0." Mol Biol Evol **30**(12): 2725-2729.

Tang, H. B., J. E. Bowers, X. Y. Wang and A. H. Paterson (2010). "Angiosperm genome comparisons reveal early polyploidy in the monocot lineage." Proceedings of the National Academy of Sciences of the United States of America **107**(1): 472-477.

Taylor, J. S. and J. Raes (2004). "Duplication and divergence: The evolution of new genes and old ideas." Annual Review of Genetics **38**: 615-643.

Thompson, J. R., J. K. Bell, J. Bratt, G. A. Grant and L. J. Banaszak (2005). "V-max regulation through domain and subunit changes. The active form of phosphoglycerate dehydrogenase." Biochemistry **44**(15): 5763-5773.

Timm, S., A. Florian, K. Jahnke, A. Nunes-Nesi, A. R. Fernie and H. Bauwe (2011). "The hydroxypyruvate-reducing system in arabidopsis: Multiple enzymes for the same End." Plant Physiology **155**(2): 694-705.

Timm, S., A. Nunes-Nesi, T. Pamik, K. Morgenthal, S. Wienkoop, O. Keerberg, W. Weckwerth, L. A. Kleczkowski, A. R. Fernie and H. Bauwe (2008). "A cytosolic pathway for the conversion of hydroxypyruvate to glycerate during photorespiration in arabidopsis." Plant Cell **20**(10): 2848-2859.

Tolbert, N. E., R. K. Yamazaki and A. Oeser (1970). "Localization and properties of hydroxypyruvate and glyoxylate reductases in spinach leaf particles." Journal of Biological Chemistry **245**(19): 5129-5136.

Truesdell, S. J., J. C. Sims, P. A. Boerman, J. L. Seymour and R. A. Lazarus (1991). "Pathways for netabolism of ketoaldonic acids in an Erwinia sp." Journal of Bacteriology **173**(21): 6651-6656.

Van de Peer, Y., J. A. Fawcett, S. Proost, L. Sterck and K. Vandepoele (2009). "The flowering world: a tale of duplications." Trends in Plant Science **14**(12): 680-688.

van der Pijl, L. (1982). Principles of dispersal in higher plants, Springer Berlin Heidelberg.

Velasco, R., A. Zharkikh, M. Troggio, D. A. Cartwright, A. Cestaro, D. Pruss, M. Pindo, L. M. FitzGerald, S. Vezzulli, J. Reid, G. Malacarne, D. Iliev, G. Coppola, B. Wardell, D. Micheletti, T. Macalma, M. Facci, J. T. Mitchell, M. Perazzolli, G. Eldredge, P. Gatto, R. Oyzerski, M. Moretto, N. Gutin, M. Stefanini, Y. Chen, C. Segala, C. Davenport, L. Dematte, A. Mraz, J. Battilana, K. Stormo, F. Costa, Q. Z. Tao, A. Si-Ammour, T. Harkins, A. Lackey, C. Perbost, B. Taillon, A. Stella, V. Solovyev, J. A. Fawcett, L. Sterck, K. Vandepoele, S. M. Grando, S. Toppo, C. Moser, J. Lanchbury, R. Bogden, M. Skolnick, V. Sgaramella, S. K. Bhatnagar, P. Fontana, A. Gutin, Y. Van de Peer, F. Salamini and R. Viola (2007). "A high quality draft consensus sequence of the genome of a heterozygous grapevine variety." Plos One **2**(12): e1326.

Wada, Y., S. Iwai, Y. Tamura, T. Ando, T. Shinoda, K. Arai and H. Taguchi (2008). "A new family of D-2-hydroxyacid dehydrogenases that comprises D-mandelate dehydrogenases and 2-ketopantoate reductases." Bioscience Biotechnology and Biochemistry **72**(4): 1087-1094.

Wagner, G. and F. Loewus (1973). "The biosynthesis of L-(+)-tartaric acid in Pelargoniumm crispum." Plant Physiology **52(6)**: 651-654.

Wagner, G., J. C. Yang and F. A. Loewus (1975). "Stereoisomeric characterization of tartaric acid produced during L-ascorbic metabolism in plants." Plant Physiology **55**(6): 1071-1073.

Wagner, L. (1974). "Ascorbic acid metabolism in vitaceae conversion to tartaric acid and hexoses." Plant physiology **54**(5): 784-787.

Wang, X. L., Y. H. Xu, C. C. Peng, R. C. Fan and X. Q. Gao (2009). "Ubiquitous distribution and different subcellular localization of sorbitol dehydrogenase in fruit and leaf of apple." Journal of Experimental

Botany **60**(3): 1025-1034.

Wen, Y. Q., J. M. Li, Z. Z. Zhang, Y. F. Zhang and Q. H. Pan (2010). "Antibody preparation, gene expression and subcellular localization of L-idonate dehydrogenase in grape berry." Bioscience Biotechnology and Biochemistry **74**(12): 2413-2417.

Willaert, R. and L. De Vuyst (2006). "Continuous production of L(+)-tartaric acid from cis-epoxysuccinate using a membrane recycle reactor." Applied Microbiology and Biotechnology **71**(2): 155-163.

Williams, M. and F. A. Loewus (1978). "Biosynthesis of L-(+)-tartaric acid from L-ascorbic-4-C-14 acid in grape and geranium " Plant Physiology **61**(4): 672-674.

Williams, M., K. Saito and F. A. Loewus (1979). "Ascorbic acid metabolism in geranium and grape." Phytochemistry **18**(6): 953-956.

Winn, M. D., C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallographica Section D-Biological Crystallography **67**: 235-242.

Worth, C. L., S. Gong and T. L. Blundell (2009). "Structural and functional constraints in the evolution of protein families." Nature Reviews Molecular Cell Biology **10**(10): 709-720.

Wu, B. H., S. H. Li, M. Nosarzewski and D. D. Archbold (2010). "Sorbitol dehydrogenase gene expression and enzyme activity in apple: tissue specificity during bud development and response to rootstock vigor and growth manipulation." Journal of the American Society for Horticultural Science **135**(4): 379-387.

Yamada, K., H. Mori and S. Yamaki (1999). "Gene expression of NAD - Dependent sorbitol dehydrogenase during fruit development of apple (Malus pumila Mill. var. domestica Schneid.)." Journal of the Japanese Society for Horticultural Science **68**(6): 1099-1103.

Yamada, K., N. Niwa, K. Shiratake and S. Yamaki (2001). "cDNA cloning of NAD-dependent sorbitol dehydrogenase from peach fruit and its expression during fruit development." Journal of Horticultural Science & Biotechnology **76**(5): 581-587.

Yamada, K., Y. Oura, H. Mori and S. Yamaki (1998). "Cloning of NAD-dependent sorbitol dehydrogenase from apple fruit and gene expression." Plant Cell Physiol **39**(12): 1375-1379.

Yamaguchi, H., Y. Kanayama, J. Soejima and S. Yamaki (1996). "Changes in the amounts of the NAD-dependent sorbitol dehydrogenase and its involvement in the development of apple fruit." Journal of the American Society for Horticultural Science **121**(5): 848-852.

Yamaguchi, H., Y. Kanayama and S. Yamaki (1994). "Purification and properties of NAD-dependent Sorbitol Dehydrogenase from apple fruit." Plant Cell Physiol **35**(6): 887-892.

Yamaki, S. a. I., K. (1986). " Roles of four sorbitol-related enzymes and invertase in the seasonal alteration of sugar metabolism in apple tissue." Journal of the American Society for Horticultural Science **111**: 134–137.

Yancey, P. H., M. E. Clark, S. C. Hand, R. D. Bowlus and G. N. Somero (1982). "Living with water-stress - Evolution of osmolyte systems." Science **217**(4566): 1214-1222.

Yang, J. C. and F. A. Loewus (1975). "Metabolic conversion of L-ascorbic-acid to oxalic-acid in oxalate-accumulating plants." Plant Physiology **56**(2): 283-285.

Yennawar, H., M. Moller, R. Gillilan and N. Yennawar (2011). "X-ray crystal structure and small-angle X-ray scattering of sheep liver sorbitol dehydrogenase." Acta Crystallographica Section D-Biological Crystallography **67**(Pt5): 440-446.

Yoshikawa, S., R. Arai, Y. Kinoshita, T. Uchikubo-Kamo, T. Wakamatsu, R. Akasaka, R. Masui, T. Terada, S. Kuramitsu, M. Shirouzu and S. Yokoyama (2007). "Structure of archaeal glyoxylate reductase from Pyrococcus horikoshii OT3 complexed with nicotinamide adenine dinucleotide phosphate." Acta Crystallographica Section D-Biological Crystallography **63**: 357-365.

Yum, D. Y., S. S. Bae and J. G. Pan (1998). "Purification and characterization of the 2-ketoaldonate reductase from Brevibacterium ketosoreductum ATCC21914." Bioscience Biotechnology and Biochemistry **62**(1): 154-156.

Yum, D. Y., B. Y. Lee, D. H. Hahm and J. G. Pan (1998). "The yiaE gene, located at 80.1 minutes on the Escherichia coli chromosome, encodes a 2-ketoaldonate reductase." Journal of Bacteriology **180**(22): 5984-5988.

Zhang, J. Z. (2003). "Evolution by gene duplication: an update." Trends in Ecology & Evolution **18**(6): 292-298.

Zhang, J. Z., Y. P. Zhang and H. F. Rosenberg (2002). "Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey." Nature Genetics **30**(4): 411-415.

Jia et al. BMC Plant Biology (2015) 15:101
DOI 10.1186/s12870-015-0478-5

BMC Plant Biology

RESEARCH ARTICLE Open Access

New insights into the evolutionary history of plant sorbitol dehydrogenase

Yong Jia[1], Darren CJ Wong[1,2], Crystal Sweetman[1,3], John B Bruning[4] and Christopher M Ford[1*]

Abstract

Background: Sorbitol dehydrogenase (SDH, EC 1.1.1.14) is the key enzyme involved in sorbitol metabolism in higher plants. SDH genes in some *Rosaceae* species could be divided into two groups. L-idonate-5-dehydrogenase (LIDH, EC 1.1.1.264) is involved in tartaric acid (TA) synthesis in *Vitis vinifera* and is highly homologous to plant SDHs. Despite efforts to understand the biological functions of plant SDH, the evolutionary history of plant SDH genes and their phylogenetic relationship with the *V. vinifera* LIDH gene have not been characterized.

Results: A total of 92 SDH genes were identified from 42 angiosperm species. SDH genes have been highly duplicated within the *Rosaceae* family while monocot, *Brassicaceae* and most *Asterid* species exhibit singleton SDH genes. Core Eudicot SDHs have diverged into two phylogenetic lineages, now classified as SDH Class I and SDH Class II. *V. vinifera* LIDH was identified as a Class II SDH. Tandem duplication played a dominant role in the expansion of plant SDH family and Class II SDH genes were positioned in tandem with Class I SDH genes in several plant genomes. Protein modelling analyses of *V. vinifera* SDHs revealed 19 putative active site residues, three of which exhibited amino acid substitutions between Class I and Class II SDHs and were influenced by positive natural selection in the SDH Class II lineage. Gene expression analyses also demonstrated a clear transcriptional divergence between Class I and Class II SDH genes in *V. vinifera* and *Citrus sinensis* (orange).

Conclusions: Phylogenetic, natural selection and synteny analyses provided strong support for the emergence of SDH Class II by positive natural selection after tandem duplication in the common ancestor of core Eudicot plants. The substitutions of three putative active site residues might be responsible for the unique enzyme activity of *V. vinifera* LIDH, which belongs to SDH Class II and represents a novel function of SDH in *V. vinifera* that may be true also of other Class II SDHs. Gene expression analyses also supported the divergence of SDH Class II at the expression level. This study will facilitate future research into understanding the biological functions of plant SDHs.

Keywords: Sorbitol dehydrogenase, L-idonate-5-dehydrogenase, Gene duplication, Functional divergence, Tartaric acid, Ascorbic acid, Grapevine

Background

Sorbitol dehydrogenase (SDH, EC 1.1.1.14) is commonly found in all kinds of life forms, including animals [1-4], yeasts [5], bacteria [6] and plants [7-13]. It represents the early divergence within the NAD (H)-dependent medium-chain dehydrogenase/reductase (MDR) superfamily (with a typical ~350-residue subunit), sharing a distant homology with alcohol dehydrogenase (ADH, EC 1.1.1.1) [14-17]. SDH catalyses the reversible oxidation of a range of related sugar alcohols into their corresponding ketoses [7,13,18-21], preferring polyols with a d-*cis*-2,4-dihydroxyl (2S,4R) configuration and a C1 hydroxyl group next to the oxidation site at C2, such as sorbitol, xylitol and ribitol (Additional file 1). It exhibits the highest activity on sorbitol while also being able to oxidize the other polyols at lower reaction rates [6,13,18,20]. The process of sorbitol oxidation by human SDH requires a catalytic zinc atom which is coordinated by the side chains of three amino acids (44C, 69H, 70E, numbering in human SDH) and one water molecular. NAD^+ binds to the protein first, followed by sorbitol. The backbone of sorbitol stacks against the nicotinamide ring while the C1 and C2 oxygen atoms are coordinated to the zinc. The water molecule coordinating the zinc atom acts a general base and abstracts

* Correspondence: christopher.ford@adelaide.edu.au
[1]School of Agriculture, Food and Wine, University of Adelaide, Adelaide 5005, Australia
Full list of author information is available at the end of the article

the proton of the C2 hydroxyl, which creates an electron flow to NAD^+, leading to the oxidation of sorbitol at C2 and the final production of NADH [22].

Plant SDH is the key enzyme in the sorbitol metabolism pathway [7,13,20,21,23] and has been associated with resistance to abiotic stresses such as drought and salinity. SDH activity regulates the levels of polyols [13,23], which act as important osmolytes during drought stress and recovery processes [24]. In *Rosaceae* species sorbitol occurs as the major photosynthate and phloem transported carbohydrate [25]. In these plants, which include apple [26-31], pear [32,33] and loquat [34,35], SDH plays a crucial role in the oxidation of sorbitol and its translocation to sink tissues such as developing fruits and young leaves. Gene transcript level and enzyme activity remain high during fruit development and maturation, dropping gradually in later stages, and contributing to the sugar accumulation in the ripening fruits [27-30,34-36]. The role of sink strength regulation for SDH is of particular research interest given the economic importance of these fruit species. Additionally, SDH has been shown to be involved in the sugar metabolism process during seed germination of some herbaceous plants including soybean [37] and maize [8,38].

Despite efforts to understand the physiological role of SDH in plants, little attention has been paid toward the evolutionary history of the plant SDH gene family. The distribution of the SDH genes in higher plants appears to be species-dependant. In particular, 9 paralogous SDH genes have been reported in apple [27] and 5 in Japanese pear [39]. In contrast, other plant genomes such as *A. thaliana* [23], tomato [11] and strawberry [12] contain only one SDH gene. Recent studies have indicated that there are two groups of SDH present in some *Rosaceae* plants. Park et al. [10] isolated four SDH isoforms (MdSDH1-4) from Fuji apple and found that MdSDH2-4 could be clearly distinguished from MdSDH1 based on the deduced amino acid sequence, showing 69–71% identity with MdSDH1 and 90–92% identity with each other. In addition, MdSDH2-4 were expressed only in sink tissues such as young leaves, stems, roots and maturing fruits while MdSDH1 was highly expressed in both sink and source organs [10]. Nosarzewski et al. [27] identified nine SDHs (SDH1-9) from the Borkh apple genome and showed that all isoforms except SDH1 (71–73% identity with SDH2-9) were highly homologous with an identity of 91–97%. Similar observations have been made with the SDH isoforms (PpySDH1-5) identified in pear whereby PpySDH5 differed from PpySDH1-4 at both the primary structure level and the gene transcriptional level [39]. Preliminary phylogenetic analyses have classified these homologous SDHs into two groups based on primary protein structures [10,29,33,40]. However, these studies focused on only one or just a few related *Rosaceae* species. No comprehensive phylogenetic analysis has been performed on SDH across a broad range of angiosperm species.

Gene duplication is widespread in plant genomes. Functional divergence after gene duplication is the major mechanism by which genes with novel function evolve; this phenomenon plays a key role in the evolution of phenotypic diversity [41-44]. The current understanding of gene evolution via duplication suggests that duplicated genes could arise through different mechanisms including unequal crossing over (resulting in tandem duplication), retrotransposition, segmental duplication and chromosomal (or whole genome) duplication [42,45]. Most duplicated genes are lost due to the accumulation of mutations that render them non-functional (pseudogenization) [42]. However, they can be retained under certain circumstances whereby the acquisition of beneficial mutations leads to novel function (neofunctionalization), which requires positive natural selection, or through adoption of part of the functions of the ancestral gene (sub-functionalization), which could occur by expression divergence or functional specialization of protein [41,42,46,47]. The latter usually involves a shift in the enzyme substrate specificity.

Protein structural analyses have shown that the LIDH of *V. vinifera*, which catalyses the inter-conversion of L-idonate and 5-keto-D-gluconate (5KGA) in the tartaric acid (TA) synthesis pathway [48], is highly homologous to plant SDHs, sharing ~77% amino acid sequence similarity with SDH from tomato (Gene ID: 778312) and *A. thaliana* (Gene ID: AT5G51970) [48]. The 366 amino acid LIDH (UniProt ID: Q1PSI9) contains an N-terminal GroES-like fold and a C-terminal Rossmann fold [48], characteristics of the ADH family [49], which has a distant homology to SDH [14-17]. However, unlike other plant SDHs, LIDH displays principal activity against L-idonate and has a low reaction rate with sorbitol [48]. The unique substrate specificity of LIDH was suggested to be due to small changes in amino acid sequence encoded by paralogous genes [48].

In this study, a comprehensive phylogenetic analysis of angiosperm SDHs was conducted using currently available genomic data. A computational approach was employed to characterise the natural selection pressure on plant SDH. The protein structures of the SDH homologues in *V. vinifera* were modelled based on human SDH (PDB:1PL8) to identify the putative active site residues of plant SDHs. Transcription and co-expression data of SDH genes were also extracted from recent publicly available microarray and co-expression databases and analysed. New insights into the evolution history of the plant SDH family and the evolutionary origin of *V. vinifera* LIDH will be discussed.

Results and discussion

Identification of sorbitol dehydrogenase (SDH) homologous genes in higher plants

A database homology search identified 92 SDH homologous genes from 42 species (Figure 1; See Additional file 2: Table S1 for identified gene IDs and Additional file 3 for gene sequences in corresponding species). At least one putative SDH gene was present in each plant genome studied, consistent with previous studies [17] that suggested the ubiquity of SDH and its functional importance across all life forms. However, the distribution of SDH homologous genes varied dramatically across species. Monocot species (n = 8) uniformly presented a single SDH gene, and this same observation was made with *Brassicaceae* plants (n = 7) from the Eudicot group. It was recently reported that there are 2 SDH genes in both rice (monocot) and *A.thaliana* (*Brassicaceae*) [50], however, in both cases these SDH genes were found to be alternative transcripts of a single gene. All except one species from the *Asterid* clade and the *Leguminosae* family had one SDH gene, the exceptions being *Solanum tuberosum* (potato) and *Glycine max* (soybean), respectively, which both had two copies. By contrast, numerous copies of SDH genes were found in *Rosaceae* species, which employ sorbitol as the major transported carbohydrate [25]. *Malus* × *domestica* (apple) contained 16 putative SDH genes, the highest number among all species investigated. A previous study [50] identified 17 SDH genes in the apple genome, however, the extra putative SDH (MDP0000506359) was only a partial gene (177 residues) and was excluded from the present study. In addition to apple, other *Rosaceae*

			Species Name	Class I	Class II	Total
Monocots			*Brachypodium distachyon*	0	0	1
			Oryza sativa	0	0	1
			Panicum virgatum	0	0	1
			Setaria italica	0	0	1
			Zea mays	0	0	1
			Sorghum bicolor	0	0	1
			Aegilops tauschii	0	0	1
			Hordeum vulgare	0	0	1
Lower Eudicots			*Aquilegia coerulea*	0	0	7
Core Eudicots	*Asterids*		*Solanum lycopersicum*	1	0	1
			Solanum tuberosum	1	1	2
			Capsicum annuum cv. CM334	1	0	1
	Rosids	*Vitaceae*	***Vitis vinifera***	1	2	3
		Brassicaceae	*Eucalyptus grandis*	1	1	2
			Citrus sinensis	2	1	3
			Theobroma cacao	1	1	2
			Cucumis sativus	2	0	2
			Gossypium raimondii	3	0	3
			Carica papaya	1	0	1
			Thellungiella halophila	1	0	1
			Brassica rapa Chiifu-401	1	0	1
			Brassica oleracea	1	0	1
			Capsella rubella	1	0	1
			Arabidopsis lyrata	1	0	1
			Arabidopsis thaliana	1	0	1
			Eutrema salsugineum	1	0	1
		Geraniaceae	*Pelargonium x hortorum*	1	1	2
		Rosaceae	*Fragaria vesca*	1	0	1
			Malus domestica	15	1	16
			Eriobotrya japonica	0	1	1
			Prunus persica	1	3	4
			Prunus mume	1	2	3
			Pyrus bretschneideri	4	1	5
		Leguminosae	*Glycine max*	2	0	2
			Cajanus cajan	1	0	1
			Phaseolus vulgaris	1	0	1
			Medicago truncatula	1	0	1
		Euphorbiaceae	*Populus trichocarpa*	1	1	2
			Linum usitatissimum	2	2	4
			Ricinus communis	1	0	1
			Jatropha curcas	2	1	3
			Manihot esculenta	1	1	2

Figure 1 Distribution of SDH homologous genes in higher plants. Closely related species were specified accordingly. The gene abundance heat map was based on the total copy number of SDH genes in each species. SDHs of *P. bretschneideri* [39] and *E. japonica* (loquat) [35] were obtained from literature; additional SDHs may be identified in these two species when complete genome information becomes available. The classification of SDH Class I and SDH Class II was based on the phylogenetic analysis carried out in the present study.

species such as *Prunus persica* (peach), *Prunus mume* (Chinese plum), *Eriobotrya japonica* (loquat) and *Pyrus bretschneideri* (pear) had 4, 3, 1 and 5 putative SDH genes respectively. It should be noted that the information of SDH numbers in loquat [35] and pear [39] was retrieved from earlier reports, and that more SDH genes may be found when complete genome data for these species become available. Although *Fragaria vesca* (strawberry) belongs to the *Rosaceae* family, only one SDH gene was present in this species. Unlike other *Rosaceae* fruit species, *F. vesca* utilizes sucrose instead of sorbitol as the main translocated carbohydrate [51]. According to a recent development in the evolution by duplication theory, a proper gene dosage should be kept to maintain a stoichiometric balance in macromolecular complexes such as functional proteins, thereby ensuring the normal functioning of a particular biological process [41,52]. Transportation and assimilation of sorbitol is a *Rosaceae*-specific metabolism. The retention of highly duplicated SDH genes in *Rosaceae* species suggests that a higher dosage of SDH transcription or enzyme activity is needed to facilitate sorbitol metabolism in these species.

Three putative SDH genes were identified in the *V. vinifera* genome. One (GSVIVT01010646001) corresponded to the previously characterized LIDH (Uniprot No. Q1PSI9) [48] while the other two shared 99% (GSVIVT01010644001) and 77% (GSVIVT01010642001) amino acid sequence identity with *V. vinifera* LIDH (Additional file 2: Table S4). Other important crops such as *C. sinensis* (orange), *Theobroma cacao* (cocoa), and *Pelargonium hortorum* (a geranium species) had 3, 2 and 2 SDH genes respectively. *P. hortorum* and *S. tuberosum* are of particular interest in this study because they have also been shown to accumulate significant levels of TA, like *V. vinifera* [53,54]. Another species that should be noted is *Aquilegia coerulea* (a flower native to the Rocky Mountains), which belongs to the Eudicot family but has been recognized as an evolutionary intermediate [55] between monocot and core Eudicot plants, and contained 7 SDH paralogues.

Phylogenetic analysis of plant sorbitol dehydrogenase families

To determine the evolutionary history of plant SDH family and the phylogenetic relationship between LIDH and SDH, a phylogeny of the SDH family was reconstructed. Consistent results were obtained using both Neighbour Joining (Figure 2A; Additional file 4) and Maximum Likelihood (Figure 2B) methods. As can be seen in the Maximum Likelihood tree (Figure 2B), the target proteins divided at the basal nodes into three major clusters, corresponding to the three life kingdoms: fungi, animal and plant (Bootstrap supports at 0.98, 1 and 1 respectively). The overall topology of the plant SDH clade was in agreement with the Phytozome species tree (http://www.phytozome.net/), indicating that the phylogeny results were reliable. Specifically, monocot plants (n = 8) formed a single clade with strong support (0.91), corresponding to the early split between monocot and dicot lineages. *A. coerulea* SDHs separated into a single group (0.91) which positioned itself between monocot and core Eudicot plants. The *Aquilegia* genus belongs to the Eudicot order *Ranunculales* which has been established as a sister clade to the rest of the core Eudicot [56-58] and agrees with the present phylogenetic analysis.

The core Eudicot SDHs split into two distinct lineages in the Maximum Likelihood tree (Figure 2B). The first lineage (classified as Class I) covered all core Eudicot species included in this study while the second (Class II) had a narrower coverage and was less expanded compared to SDH Class I. The divergence of core Eudicot SDHs into two lineages was in agreement with previous reports that SDHs from some *Rosaceae* species could be separated into two groups [10,29,33]. All *Rosaceae* plants (n = 5) investigated in this study except *F. vesca* (strawberry) had multiple copies of SDH genes that covered both SDH Class I and SDH Class II. However, within these species, the distribution of SDHs among the two SDH classes varied greatly. In particular, 15 out of the 16 SDHs from *M. domestica* and 4 out of the 5 SDHs from *P. bretschneideri* fell into SDH Class I while 3 out of the 4 SDHs from *P. persica* and 2 out of the 3 SDHs from *P. mume* belonged to SDH Class II. Other species retaining two classes of SDHs included *S. tuberosum*, *V. vinifera*, *Eucalyptus grandis*, *C. sinensis*, *T. cacao*, *P. hortorum*, *Populus trichocarpa*, *Linum usitatissimum*, *Jatropha curcas* and *Manihot esculenta*, from different orders or families. In contrast, *Brassicaceae* plants (n = 7), *Leguminosae* plants (n = 4) and *Asterid* plants (n = 2) except *S. tuberosum* contained either a single SDH or two SDHs that could only be classified into SDH Class I. Within both SDH Class I and Class II clades, *Rosaceae* SDHs (except *F. vesca*) formed separate phylogeny groups (Figure 2B), implying divergent molecular characteristics for SDHs from this family. Most recent phylogenetic analyses [59,60] have placed *Vitaceae* as a sister clade to the *Rosid* plants in the core Eudicot group. The presence of two classes of SDHs in both *V. vinifera* and *S. tuberosum* (*Asterid*s) indicated that the divergence between SDH Class I and Class II occurred before the species radiation of the core Eudicot plants. Moreover, although 7 SDH genes were retained in the genome of the evolutionarily intermediate species *A. coerulea*, none of them could be classified into SDH Class I or SDH Class II. Taken together, our results suggested that SDH Class I and Class II might have diverged during the common ancestor of core Eudicot plants

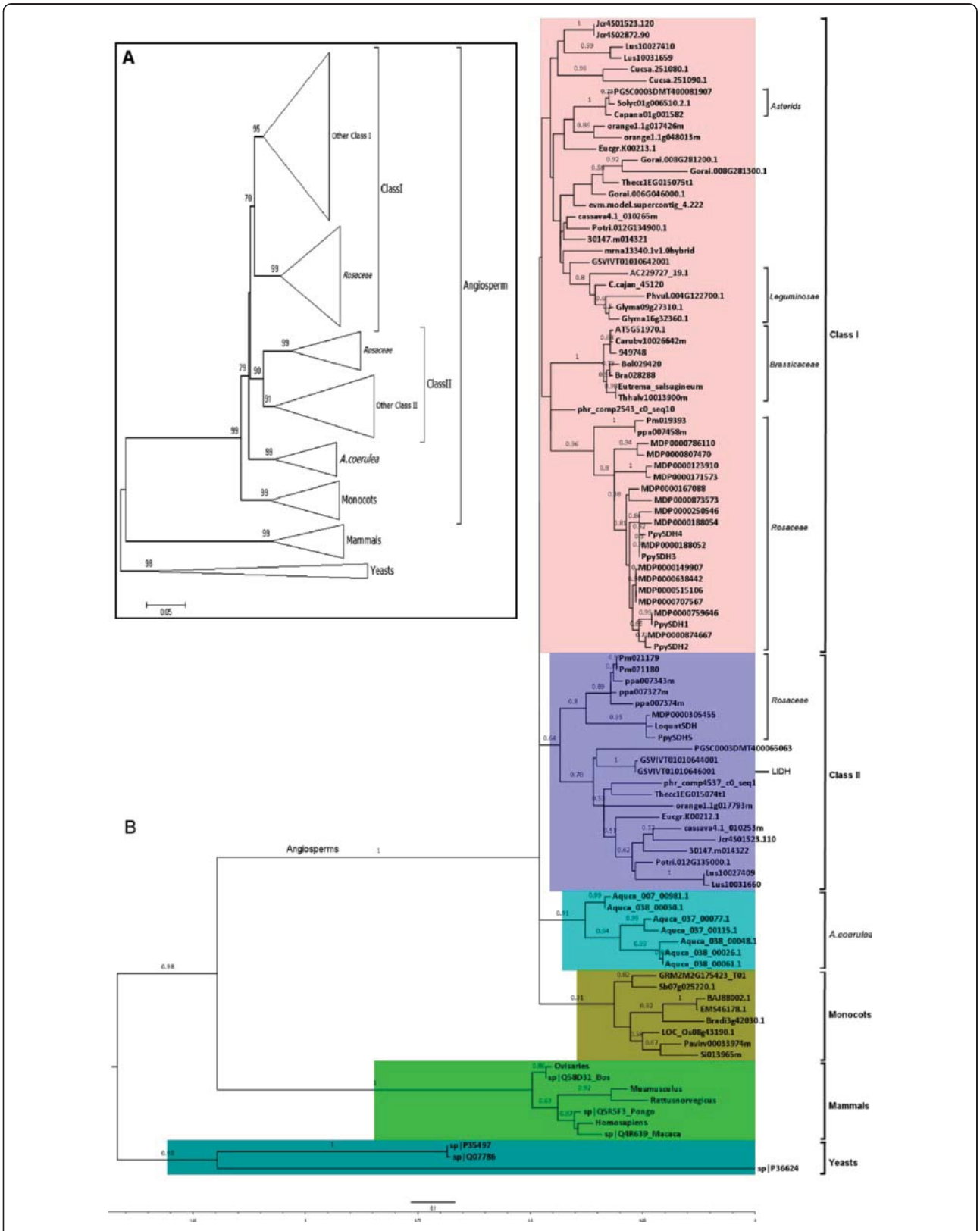


Figure 2 (See legend on next page.)

(See figure on previous page.)
Figure 2 Phylogenetic tree showing the evolutionary history of the angiosperm SDH family. **A**: A simplified schematic phylogeny of the SDH family inferred by MEGA 6.0 [97] software using the Neighbour Joining method. Values (as percentage, cutoff value 50) of Internal branch test (1000 replicates) supports are indicated above the corresponding branches. **B**: The Maximum Likelihood phylogeny of the SDH family developed by MEGA 6.0 [97] software using the selected best-fitting substitution model JTT + G [99]. 1000 times Bootstraping supports (cut off at 0.5) are displayed above corresponding branch. Closely related species are annotated accordingly. The *V. vinifera* LIDH (GSVIVT01010646001) is also marked.

but after the branching of the basal Eudicots such as *Ranunculales.* This corresponds to a period of about 125Mya ~ 115Mya [55,58].

In the Maximum Likelihood tree, the Class II clade was well-supported and separated from Class I with longer branch length in general (Figure 2B), suggesting a higher level of amino acid substitution within this clade. In addition, the topology of the Class II clade (except the *Rosaceae* group) was in good agreement with the species tree at Phytozome (http://www.phytozome.net/search.php), with *S. tuberosum* (*Asterids*) diverging first followed by *V. vinifera* and the rest of the rosid species. This indicates that the Class II SDHs have evolved vertically within respective species, which lends further support to the suggestion above that SDH Class I and Class II have existed during the common ancestry of core Eudicot plants. The backbone topology of the more inclusive Class I clade in the Maximum Likelihood tree was weakly supported (Bootstrap support under 0.5; Figure 2B), in contrast with the strong clustering support for this clade in the Neighbour Joining tree (Figure 2A; Additional file 4). The weak bootstrap support for the topology of SDH Class I may have resulted from a lack of amino acid substitution in this clade, as reflected by the short branch length (Figure 2B). The calculation of evolutionary distances for plant SDHs revealed a pair-wise distance under 0.3 in general (Additional file 2: Table S2), sequence alignment showed that Class I SDHs tend to be more conserved (average sequence pair-wise identity 83.4%; Table 1) than Class II (79%; Table 1), which means less amino acid substitution within the Class I clade. These results are consistent with the strong clustering support for the major sub-clades of the Class I branch in the Neighbour Joining tree (Figure 2A; Additional file 4).

In contrast to the ubiquity of Class I SDHs, the absence of Class II SDHs in some species may be due to gene loss after duplication, a common mechanism in gene evolution via duplication [42,61]. This also indicated that SDH Class II members may not be essential for the normal growth of plants, suggesting a divergent function for this class of SDH genes. Interestingly, the previously characterized *V. vinifera* LIDH (GSVIVT01010646001) [48] was grouped into SDH Class II, providing direct support that in at least one case SDH Class II may have acquired a novel function, in this instance its involvement in the synthesis of TA. While the identity of additional functions for Class II SDHs in other species is unknown, support for a role of some Class II SDHs in TA metabolism may be proposed. Only a few plant families, including *Vitaceae*, *Geraniaceae* and *Leguminosae* have been shown to accumulate significant levels of TA [54] and the present results showed that Class II SDHs were present in both *Vitaceae* and *Geraniaceae*. The absence of Class II SDHs in *Leguminosae* plants could be explained by the fact that the synthesis of TA in *Leguminosae* proceeds via a different pathway, which bypasses the interconversion of L-idonate and 5KGA (catalysed by LIDH) [62]. Recent studies have revealed that potato [53], citrus fruits [63] and pear [64,65] (all containing Class II SDHs) also produce TA, although to a lesser degree than *V. vinifera*. This is consistent with the potential correlation between Class II SDHs and TA synthesis. However, it has also been reported that TA is absent or found only in trace amount in apple [66], and no information is available about the occurrence of TA in

Table 1 Amino acid sequence identity between different SDH groups

Identity	Class I	Class II	*A. coerulea*	Monocot	Mammal	Yeast
Class I	83.4 (71-99.7)	75.2 (67-83)	78.5 (71-86)	77.5 (71-83)	48.0 (44-50)	40.9 (38-43)
Class II		79.0 (71-99)	73.2 (68-80)	71.0 (67-74)	46.4 (43-49)	39.3 (37-42)
A. coerulea			86.7 (83-99.7)	75.7 (72-79)	48.0 (47-50)	41.4 (40-43)
Monocot				88.4 (86-93)	47.4 (46-49)	41.5 (40-45)
Mammal					87.8 (82-99.8)	42.3 (39-44)
Yeast						65.5 (48-99.7)

SDH sequences were divided into six groups (Class I, Class II, *A. coerulea*, Monocot, Mammal and Yeast SDHs) according to the phylogenetic analysis carried out in the present study (Figure 2). The amino acid sequence identity (as percentage) was obtained using all-vs-all BLAST tool. The average pair-wise identity between each group is presented, followed by the identity range (in bracket).

peach even though three copies of Class II SDH genes were identified in this species (Figure 1). It is possible that Class II SDHs have evolved varied functions to meet the different environmental challenges faced by respective plants. In this context, it would also be valuable for future work to investigate the in-planta function of SDH and the occurrence of TA in the evolutionarily intermediate plant *A.coerulea*, for which 7 SDH paralogues were identified.

Sequence alignment and protein subdomain analysis

Sequence alignment and protein subdomain analyses were performed to investigate the molecular characteristics of plant SDHs. Results showed that plant SDHs shared an overall identity above 67% (Table 1), while having ca 48% and ca 41% identities with mammal and yeast SDHs respectively (Additional file 2: Table S4). Plant SDHs were clustered into four groups in the present phylogenetic analysis: monocot SDH, *A. coerulea* SDH, core Eudicot SDH Class I and SDH Class II. Protein BLAST results showed that Class I and Class II SDHs within the same species generally had an inter-class identity of around 70% and an intra-class identity above 90% (Additional file 2: Table S4). When compared with monocot and *A. coerulea* SDHs, Class I SDHs always demonstrated a significantly higher similarity than Class II SDHs (77.5% vs 71.0% and 78.5% vs 73.2% respectively; Table 1), suggesting that core Eudicot Class I SDHs have a closer distance to monocot and *A. coerulea* SDHs and that SDH Class II may have diverged from SDH Class I. In addition, Class I SDHs tend to be more homologous than Class II SDHs (83.4% vs 79.0%; Table 1). No significant difference between the two SDH classes was observed when compared to mammal or yeast SDHs (48.0% vs 46.4% and 40.9% vs 39.3% respectively; Table 1). Protein functional domain prediction identified two functional domains for plant SDHs: an N-terminal GroES-like fold and a C-terminal Rossmann fold (Figure 3; See Additional file 5 for the complete sequence alignment). Secondary structure analysis showed that these two domains tended to be highly conserved among all plant SDHs, and amino acid substitutions mainly occurred at boundary regions linking secondary structural elements such as alpha-helices and beta-sheets (Figure 3).

Gene duplication pattern characterization and synteny analysis

To characterise the expansion patterns of plant SDH gene family, nine species that were from different families and contained both classes of SDHs were selected for gene duplication and synteny analyses (*C. sinensis*, *E. grandis*, *P. mume*, *P. persica*, *Populus trichocarpa*, *M. domestica*, *S. tuberosum*, *T. cacao* and *V. vinifera*). As shown in Table 2 (See Additional file 6 for the original output data), tandem duplication contributed the most to the expansion of the core Eudicot SDH family, followed by WGD/Segmental duplication. Dispersed SDHs (MDP0000305455, MDP0000759646 and PGSC0003DMC400055323) and a single proximal SDH (MDP0000188054) were identified only in *M. domestica* and *S. tuberosum*. Based on phylogenetic classification in the present study, Class I and Class II SDH genes from *E. grandis*, *P. trichocarpa*, *T. cacao* and *V. vinifera* are located in a tandem manner in their corresponding chromosomes, which provides strong support that SDH Class I and SDH Class II are tandem duplications. A similar pattern was observed with *C. sinensis* whereby Cs9g16660.1 (SDH Class II) is separated by a single-gene insertion with the two Class I SDH genes (Cs9g16680.1, Cs9g16690.1; data not shown). This may be caused by gene insertion after tandem duplication. Class I and Class II SDH genes in the three *Rosaceae* species (*M. domestica*, *P. mume*, *P. persica*) and in *S. tuberosum* are separated either on the one chromosome or on separate chromosomes altogether, indicating a divergent evolutionary history for SDH genes in the *Rosaceae* family and in *S. tuberosum* compared to other plants. SDH genes on chromosome 1 (md1) and chromosome 7 (md7) in *M. domestica* were highly duplicated by tandem duplication (Table 2), in contrast to the other *Rosaceae* species (*P. mume*, *P. persica*). Notably, the Class I SDH gene from *S. tuberosum* (PGSC0003DMC400055323) and the Class II SDH gene from *M. domestica* (MDP0000305455) were identified as dispersed duplicates, which may underpin the divergent sorbitol metabolism profiles across these species.

To investigate the conservation of SDH genes across species, collinear SDH gene pairs were identified within and across species. SDH genes from the nine above-mentioned species were analysed. The single SDH gene (AT5G51970) from the model plant *A. thaliana* was also used as a reference for collinear block identification. As shown in Figure 4, all target plant genomes contained at least one SDH gene (corresponding to chromosome positions A, B, C, D, E, H, J, L, N, P and Q in Figure 4) with collinear SDH genes in all other nine species studied, indicating a conserved collinear SDH block. SDH genes at gene positions F, G, I, K and O, concerning only the *Rosaceae* species investigated, were collinear with SDH genes in only some of the species included in the present analysis. In particular, position F at chromosome 8 (pp8) of *P. persica* paired only with position I at chromosome 6 (Pm6) of *P. mume*. While position F was found collinear only with position I, position I had another collinear region at position O from *E. grandis*. Position G at chromosome 4 (pp4) of *P. persica* was

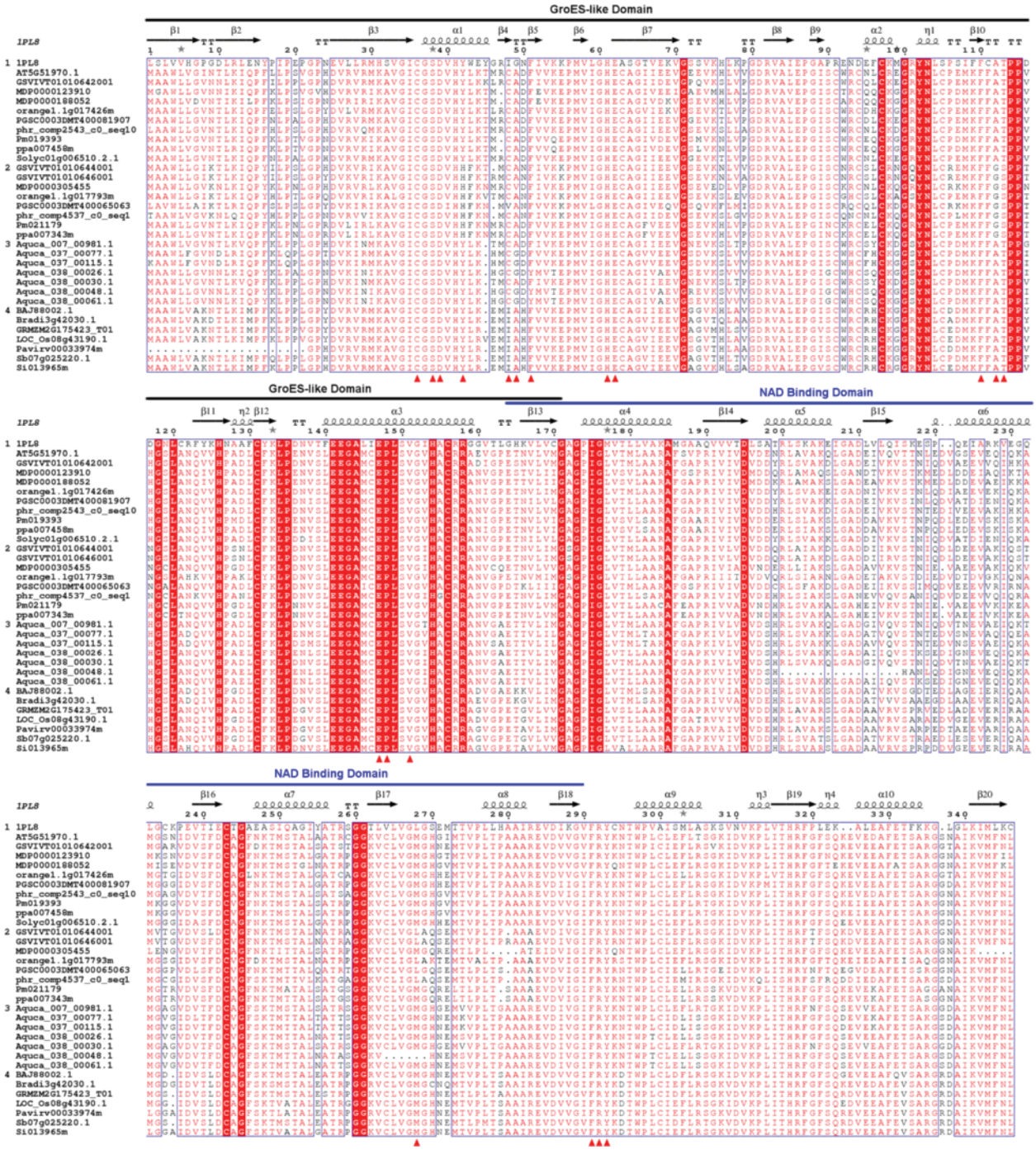


Figure 3 Multiple sequence alignment of plant SDH family. ESPript output was obtained with the sequence alignment of plant SDHs and human SDH. Secondary structures were inferred using human SDH (PDB: 1PL8) as a template, with springs representing helices and arrows representing beta-strands. Sequences were grouped into 1 (1PL8 and core Eudicot SDH Class I), 2 (core Eudicot SDH Class II), 3 (*A.coerlea* SDH) and 4 (monocot SDH). Amino acid site numbering above the alignment is according to LIDH (Q1PSI9) without the first 20 amino acids. Adjacent similarity amino acid sites were boxed in blue frame. Similarity calculations were based on the complete SDH alignments but only partial sequences for SDH Class I and SDH Class II were displayed. The active site residues identified in this study are marked with red triangles. Conserved domains are indicated above the alignment.

Table 2 Gene duplication patterns of plant SDH

Species	Chromosome ID	SDH gene ID	SDH class	Duplication pattern	Start position	End position
C. sinensis	cs9	Cs9g16680.1 (orange1.1g017426m)	I	Tandem	16143063	16147624
	cs9	Cs9g16690.1 (orange1.1g048013m)	I	Tandem	16150122	16154404
	cs9	Cs9g16660.1 (orange1.1g017793m)	II	WGD or Sgm	16135216	16138066
E. grandis	eg11	Eucgr. K00213.1	I	Tandem	2624187	2627945
	eg11	Eucgr.K00212.1	II	Tandem	2615486	2618589
M. domestica	md1	MDP0000786110	I	Tandem	25191824	25193641
	md1	MDP0000873573	I	Tandem	25182502	25183812
	md1	MDP0000707567	I	Tandem	25180931	25182241
	md1	MDP0000515106	I	Tandem	25177288	25178612
	md1	MDP0000250546	I	Tandem	25173127	25174375
	md1	MDP0000874667	I	Tandem	25157544	25158783
	md1	MDP0000638442	I	WGD or Sgm	25149134	25150444
	md1	MDP0000123910	I	WGD or Sgm	25087036	25088743
	md1	MDP0000305455	II	Dispersed	14150327	14159200
	md7	MDP0000188052	I	Tandem	23301490	23302735
	md7	MDP0000171573	I	WGD or Sgm	23281847	23283529
	md7	MDP0000188054	I	Proximal	23310942	23312187
	md7	MDP0000167088	I	Tandem	23405354	23406795
	md7	MDP0000807470	I	WGD or Sgm	23390960	23392683
	md14	MDP0000759646	I	Dispersed	24043122	24044360
P. mume	Pm5	Pm019393	I	WGD or Sgm	23673441	23675177
	Pm6	Pm021180	II	Tandem	7217228	7219256
	Pm6	Pm021179	II	Tandem	7217228	7225304
P. persica	pp2	ppa007458m|PACid:17644502	I	WGD or Sgm	24766424	24768515
	pp4	ppa007327m|PACid:17655491	II	WGD or Sgm	17729024	17731238
	pp8	ppa007343m|PACid:17644328	II	Tandem	15254677	15256888
	pp8	ppa007374m|PACid:17655656	II	Tandem	15249947	15251989
P .trichocarpa	pt12	POPTR_0012s13780	II	WGD or Sgm	13789342	13787442
	pt12	POPTR_0012s13790	I	WGD or Sgm	13790093	13792804
S. tuberosum	st01	PGSC0003DMC400055323	I	Dispersed	1594220	1598967
	st06	PGSC0003DMC400043871	II	WGD or Sgm	24156879	24158593
T. cacao	tc03	Tc03_g019280	I	WGD or Sgm	18300080	18303115
	tc03	Tc03_g019270	II	WGD or Sgm	18298897	18296706
V. vinifera	vv16	GSVIVT01010642001	I	WGD or Sgm	15653874	15651701
	vv16	GSVIVT01010646001	II	Tandem	15675560	15678887
	vv16	GSVIVT01010644001	II	Tandem	15666264	15664425

SDH gene duplication patterns were characterized by the *duplicate_gene_classifier* program in the MCScanX package. "WGD or Sgm" refers to Whole Genome Duplication or segmental duplication. "SDH Class" is defined according to the present phylogenetic analysis. Notably, MDP0000149907 from *M. domestica* could not be anchored in any chromosome and was therefore absent in this table.

only paired with positions A, E and K from *A. thaliana*, *P. trichocarpa* and *M. domestica* respectively. Some collinear SDH gene pairs, such as F-I, G-K and K-O, were restricted to *Rosaceae* species only, reflecting genetic features shared only by these plants. Notably, intra-species collinear SDH pairs were identified only within *M. domestica* but not in *P. mume*, *P. persica* and *S. tuberosum* although all of these species have SDH genes located on multiple chromosomes (Figure 4; See Additional file 2: Table S5 for identified collinear SDH gene pairs). This observation could be explained by the fact that the apple genome

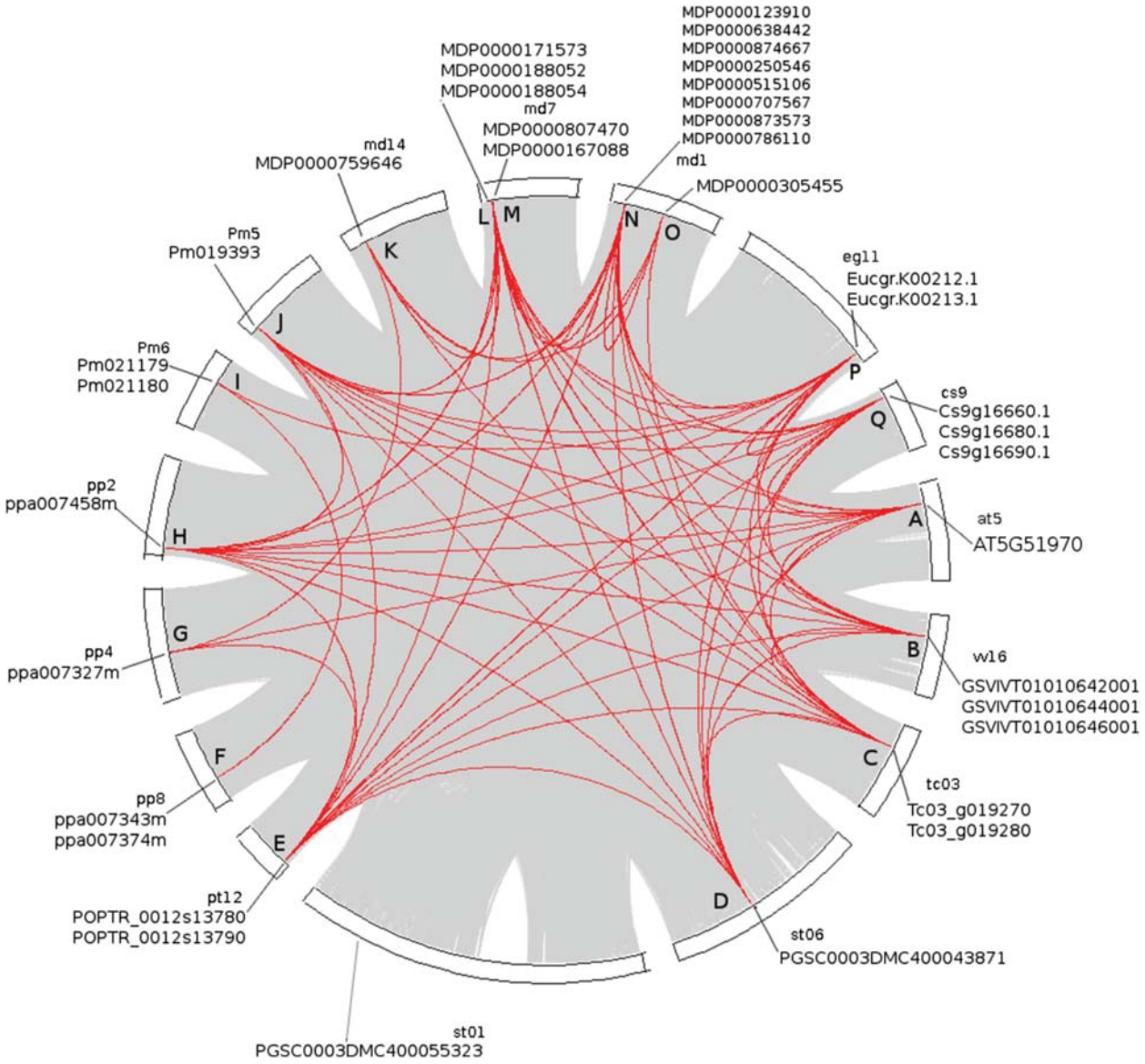


Figure 4 Identification of collinear gene pairs among plant SDH families. A circular plot of SDH gene family collinearity. Collinear SDH genes are linked by red curved lines. SDH genes located at each position in corresponding chromosomes are indicated. Family collinearity is shown in the genomic collinearity background. Only those chromosomes containing SDH genes are included.

underwent a recent (>50Mya) WGD, which doubled the chromosome number from nine to 17 in the *Pyreae* [50] while most other *Rosaceae* plants have a haploid chromosome number of 7, 8 or 9. *S. tuberosum* was unique among the species investigated in that it had a Class II SDH gene (PGSC0003DMC400043871) but no Class I SDH gene preserved in the collinear region (Figure 4). The Class I SDH gene (PGSC0003DMC400055323), which was identified as a dispersed duplication (Table 2), was the only SDH gene for which no collinear gene was identified in the present analysis. Since the Class II SDH homologue (LIDH) in *V. vinifera* has been shown to be involved in TA synthesis [48], it would be of great interest to investigate the potential role of SDHs in *S. tuberosum*, which has also been shown to accumulate a significant amount of TA [53]. Noteworthy, *S. lycopersicum*, another species from the *Solanale* order, accumulates no TA [67] and contains only a single SDH, which belongs to Class I (Figure 2B).

Natural selection analysis

Assessment of synonymous and non-synonymous substitution ratios is important to understand molecular

evolution at the amino acid level [68,69]. To examine the intensity of natural selection acting on the specific clade, the ratio (w) of non-synonymous substitution to synonymous substitution in the developed plant SDH phylogeny was investigated, whereby w<1, w=1 and w>1 indicated purifying selection, neutral evolution and positive selection respectively. Based on our phylogeny results, four branches ("monocot SDH", "*A. coerulea* SDH", "core Eudicot SDH Class I" and "core Eudicot SDH Class II") were specified for w assessments (w [mono], w [Aer], w [sdhC1] and w [sdhC2] respectively). Firstly, the branch-specific likelihood model [70] was applied to the SDH data. As can be seen in Table 3, Likelihood-ratio tests (LRT) showed that the two-ratio model and the four-ratio model fit the dataset significantly better ($2\Delta l = 12.6$ with $p = 0.0004$, $df = 1$ and $2\Delta l = 13.2$ with $p = 0.0042$, $df = 3$ respectively) than the one-ratio model. In contrast, the three-ratio model assumption lacked statistical support ($2\Delta l = 0.2$ with $p = 0.9048$, $df = 2$). Given that the two-ratio and four-ratio models assume unequal w ratios for the Class I and Class II branches while the three-ratio model specifies w(sdhC1)=w (sdhC2) (Table 3), the above calculation suggested that the w ratio for the core Eudicot SDH Class II was significantly different from that of Class I. Moreover, the four-ratio model, which assumes unequal w ratios for the monocot, *A.coerulea* and Class I branches (Table 3), was not significantly better ($2\Delta l = 0.6$ with $p = 0.7408$, $df = 2$) than the two-ratio model (assuming uniform ratio for these branches; Table 3). This indicated that the w ratios for monocot, *A. coerulea* and core Eudicot Class I branches had no significant difference. Notably, all branch-specific models tested demonstrated a low w value for the monocot, *A. coerulea* and Class I branches (w[mono]=w[Aer]=w[sdhC1]=0.10415 with the two-ratio model and w[mono]=0.10428, w[Aer]=0.09731, w [sdhC1]=0.0001with the four-ratio model), suggesting that plant SDHs have been under strong purifying selection. This agrees well with the suggestion that functional proteins are usually under strong structural and functional constraints [71]. It should be noted that w[sdhC2] were infinite in both multi-ratio models (w[sdhC2]=859 and 999 respectively). This is because an extremely low level of synonymous substitution or no synonymous substitution was detected in the SDH Class II clade. On the other hand, the number of non-synonymous substitutions in the core SDH Class II clade was estimated to be 12.7 and 12.8 respectively for the two-ratio model and the four-ratio model. In contrast, only 0.4 non-synonymous substitution was detected for the SDH Class I clade with the two-ratio model

Table 3 Natural selection tests of plant SDH

Model	np	l = ln L	Estimates of parameters	Positively selected sites
M0: one-ratio				
w(mono)=w (Aer)=w(sdhC1)=w(sdhC2)	1	-30147.4	w(mono)=w(Aer)=w(sdhC1)=w(sdhC2)=0.10492	Not Allowed (NA)
Branch-specific models				
w(mono)=w(Aer)=w(sdhC1)≠w(sdhC2) (two ratios)	2	-30141.1	w(mono)=w(Aer)=w(sdhC1)=0.10415, w(sdhC2)=859.33956	NA
w(mono)≠w(Aer)≠w(sdhC1)=w(sdhC2) (three ratios)	3	-30147.3	w(mono)=0.10510, w(Aer)=0.10821, w(sdhC1)=w(sdhC2)=0.06935	NA
w(mono)≠w(Aer)≠w(sdhC1)≠w(sdhC2) (four ratios)	4	-30140.8	w(mono)=0.10428, w(Aer)=0.09731, w(sdhC1)=0.0001, w(sdhC2)=999	NA
w(mono)=w(Aer)=w(sdhC1)≠w(sdhC2) (two ratios with w(sdhC2) fixed to 1)	1	-30141.4	w(mono)=w(Aer)=w(sdhC1)=0.10424 (w(sdhC2)=1)	NA
Site-specific models				
M1:Neutral (2 site classes)	2	-29650.0	p0=0.87775 (p1=1-p0=0.12225); w0=0.07628 (w1=1)	NA
M2:Selection (3 site classes)	3	-29650.0	p0=0.87775, p1=0.07499 (p2=1-p0-p1=0.04726); w0=0.07628 (w1=1), w2=1	None
Branch-site models (SDH Class II as foreground lineage)				
Model A Null (4 site classes)	3	-29643.2	p0=0.33951, p1=0.04783 (p2+p3=0.61266); w0=0.07544	NA
Model A (4 site classes)	4	-29640.9	p0=0.82864, p1=0.11666 (p2+p3=0.0547), w0=0.07544 (w1=1), w2=132.6226	Sites for foreground lineage: 42H,43F,112G, 113S,116T, 270Q ($p > 0.99$);

All calculations were implemented using codeml at PAML4.7. Different models were specified according to the software instruction. "np" refers to the number of parameters, "l = (ln L)" refers to the log value of the likelihood. The estimated parameters w and p refer to the K_a/K_s ratio and the percentage of the corresponding site classes respectively. In the one-ratio model M0 and the Branch-specific models, w(mono), w(Aer), w(sdhC1) and w(sdhC2) stand for the w ratios for the monocot, *A. coerulea*, SDH Class I and SDH Class II branches respectively. In the Site-specific models and the Branch-site models, w0, w1 and w2 represent the w ratios for the specific site classes in respective models (see the Methods section for more details). For the Branch-site models, the SDH Class II branch was specified as the foreground branch. Amino acid site numbering is according to LIDH (Uniprot No: Q1PSI9) without the first 20 amino acids.

(Additional file 7: branch-specific-two-ratio-output) and no non-synonymous substitution was detected with the four-ratio model (Additional file 7: branch-specific-four-ratio-output). These results provided clear evidence that positive selection had occurred in the lineage leading to core Eudicot SDH Class II. To test whether w[sdhC2] is significantly higher than 1, the log likelihood value (Table 3; Additional file 7: branch-specific-two-ratio-null-output) was calculated for the two-ratio model with w[sdhC2]=1 fixed. Results showed that this model was not significantly worse than the two-ratio model without the "w[sdhC2]=1" constraint ($2\Delta l = 0.6$ with $p = 0.4386$, $df = 1$), suggesting that w[sdhC2] was not significantly greater than 1 at the 5% significance level. This leads to the hypothesis that positive selection in SDH Class II might have only affected particular amino acid residues in the protein sequence, which is possible for a functional protein under strong structural and functional constraints [72]. To test this, Site-specific likelihood analysis was performed on the same data, which assumes variable selection pressures among amino acid sites but no variation among branches in the phylogeny. Results (Table 3: model M2) showed that the selection model (M2) fitted the dataset significantly better ($2\Delta l = 994.8$ with $p = 0.0001$, $df = 2$) than the one-ratio model but was not better ($2\Delta l = 0$ with $p = 1$, $df = 1$) than the neutral model (M1). These results indicated a significant variation of selection pressure among amino acid sites of plant SDH. However, the Selection model failed to detect any positively selected amino acid site at a significant level (Table 3; Additional file 7: site-specific-output), which suggested that no positively selected amino acid site could be identified across all branches. Therefore, we speculate that the positive selection might have only acted on a few amino acid sites in the core Eudicot SDH Class II clade.

In this context, a Branch-site model [73] that permits variable w ratios among both amino acid sites and branches was applied. Model A successfully identified the potential amino acid sites under positive selection in the SDH Class II branch (Table 3; Additional file 7: branch-site-modelA-output). Specifically, 42H, 43F, 112G, 113S, 116T and 270Q (numbering in LIDH (Q1PSI9) without the first 20 amino acids) were identified with Model A (Bayes Empirical Bayes analysis possibility >0.99; Additional file 7: branch-site-modelA-output). LRTs test showed that Model A fit the data significantly better ($2\Delta l = 18.2$ with $p = 0.0001$, $df = 2$) than the neutral model M1. The comparison ($2\Delta l = 4.6$ with $p = 0.0320$, $df = 1$) of Model A with its null hypothesis which assumes w2=1 (Additional file 7: branch-site-modelA-null-output) indicated that these amino acid sites had undergone positive selection in SDH Class II but not in the background branches. In addition, the Model A test demonstrated that 82.90% (model A: p0 = 0.82864; Table 3) of the amino acids of SDH were under strong purifying selection (model A: w0=0.07544; Table 3) and 11.7% were under neutral selection (model A: p1=0.11666, w1=1; Table 3) in all branches. No positive selection could be detected in the background branches (Additional file 7: branch-site-modelA-output). Taken together, these calculations demonstrated that plant SDHs were under strong purifying selection pressure and were highly conserved across all the plant species, and more importantly, that positive natural selection had occurred in the SDH Class II clade, affecting specific amino acids, namely 42H, 43F, 112G, 113S, 116T and 270Q.

Ancestral sequence reconstruction and evolution rate analysis

To characterize the evolutionary rates for different groups of plant SDHs, ancestral amino acid sequences for the developed SDH phylogeny were reconstructed. Results (Additional file 8: ancestral-sequence-construction-output) showed that 9 potential amino acid substitutions (Y42H, L43F, A112G, T113S, V116T, Q228K, H270Q, N271S, R283A; numbering in LIDH (Q1PSI9) without the first 20 amino acids) occurred in the branch leading to SDH Class II from the common ancestor of core Eudicot SDH. This finding corresponded well with the natural selection analysis, whereby six out of the nine amino acid sites were identified to be under positive selection (42H, 43F, 112G, 113S, 116T and 270Q; Table 3). In contrast, no substitution was detected in the branch leading to core Eudicot SDH Class I (Additional file 8: ancestral-sequence-construction-output and interpreted-ancestral-sequences.fasta). Relative rate tests (RRT) [74] using monocot SDH as the out-group showed that core Eudicot SDH Class II evolved significantly faster than core Eudicot SDH Class I (Additional file 9: ClassI-vs-ClassII.txt), indicating a relaxed selection pressure on SDH Class II. In contrast, *A. coerulea* SDH and core Eudicot Class I SDH demonstrated no significant difference (Additional file 9: Aer-vs-ClassI.txt).

Protein structure modelling analysis

To deduce the reaction mechanism and identify the potential active sites of plant SDHs, protein structure models of *V. vinifera* Class I SDH (Vv_SDH, UniProt No: D7TMY3) and Class II SDH (Vv_LIDH, UniProt No: Q1PSI9) were created based on human SDH (PDB: 1PL8; 46 ~ 47% identity with Vv_SDH and Vv_LIDH). Ligands including zinc, NAD^+, D-sorbitol and L-idonate were docked into the models (Additional file 10). Our models contain one zinc binding site, located in the active site. Some published SDH crystal structures (eg. PDB: 1E3J) contain a second, structural zinc-binding site distant from the active site catalytic zinc atom; this is

not however a universal feature of these enzymes. No function has been correlated with the second, structural zinc-binding site. The sequence of our homology models does not support a second, structural zinc-binding site, as the necessary side chains required for zinc coordination are absent. A ribbons diagram of the overall structure of the homology models can be seen in Figure 5A, with Vv_SDH and Vv_LIDH adopting a typical dehydrogenase fold with an NAD^+ binding site conforming to a Rossmann fold. The catalytic zinc ion in the active site was modelled coordinating to 36C, 61H and 62E (Figure 5C; numbering in LIDH (Q1PSI9) without the first 20 amino acids). All three of these residues together with 147E (corresponding to 155E in human SDH, mediating the water molecule coordinating the zinc atom [22]) are strictly conserved in plant SDHs (Figure 3). The 2′ and 3′ hydroxyls of the NAD^+ ribose in our model were poised to 195D (203D in human SDH), potentially forming hydrogen bonds (Additional file 10: Asp195-NAD.png). The preservation of 195D instead of 195A at this amino acid site has been shown to be the structural basis for the selection of NAD (H) over NADP (H) as co-enzyme [75]. This amino acid site is strictly conserved in all plant SDHs (Figure 3), implying that plant SDHs preferably utilize NAD (H). This suggestion is consistent with the lack of NADP-SDH activity for plant SDHs [7,10,11,13]. Previous characterizations of SDHs from Arabidopsis [13], tomato [11], apple [7,76] and pear [20]

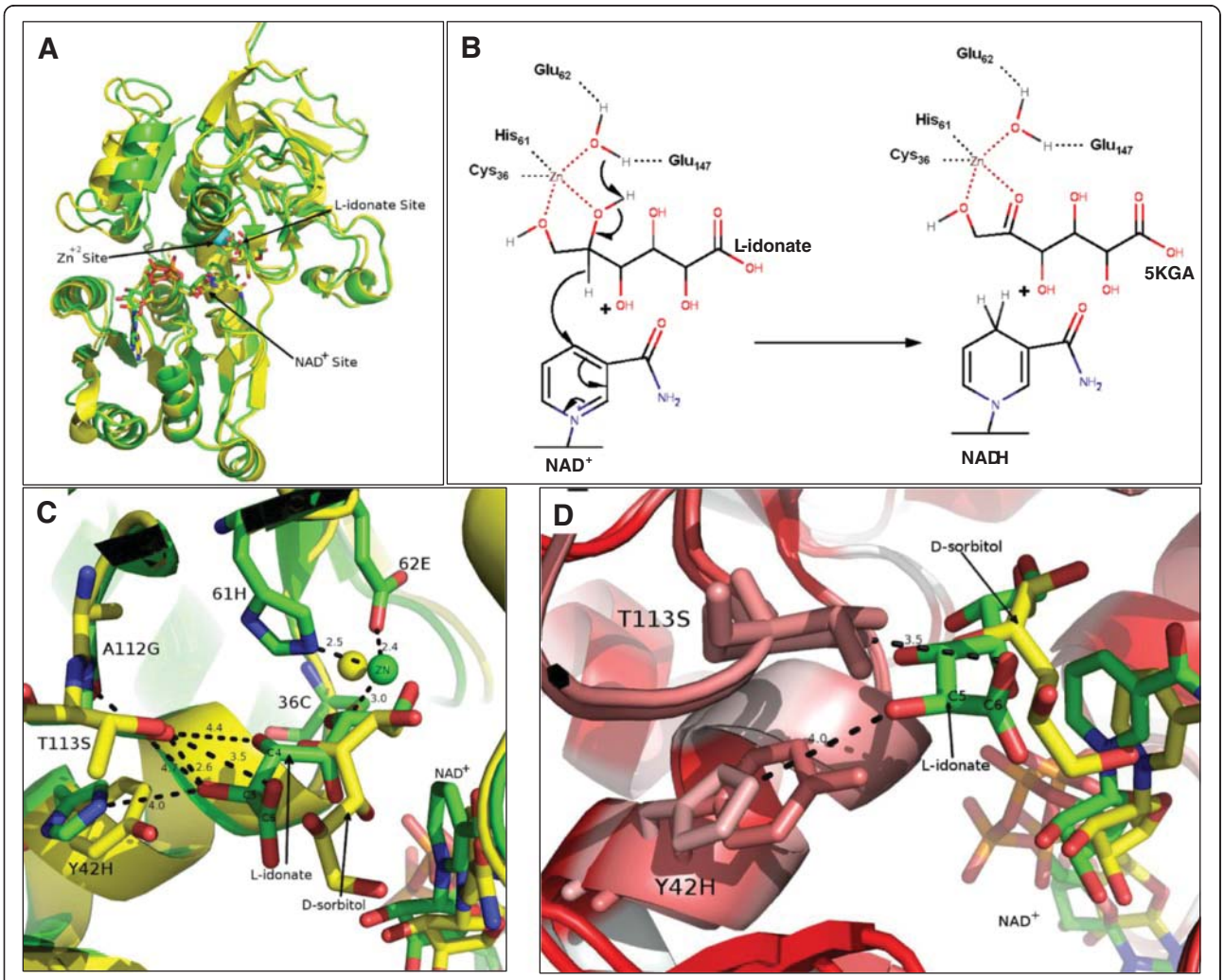


Figure 5 Homology models of Vv_LIDH and Vv_SDH and proposed reaction mechanisms. **A**. Structure superimposition of Vv_LIDH_idonate (green) and Vv_SDH_sorbitol (yellow) in Ribbon forms. **B**. The proposed reaction mechanism for Vv_LIDH on the oxidation of L-idonate into 5-keto-D-gluconate (5KGA). **C**. Superimposition of the active site residues of Vv_LIDH (green) and Vv_SDH (yellow). The distances (Å) between corresponding atoms are labelled. Target active site residues are shown in stick forms and labelled correspondingly. **D**. Hydrophobicity variance at Y42H between Vv_LIDH (green) and Vv_SDH (yellow) with red and white colours representing the highest hydrophobicity and the lowest hydrophobicity respectively. (All amino acid site numbering is according to LIDH (UniProt No: Q1PSI9) without the first 20 amino acids).

have suggested that plant SDHs exhibit highest activity for the oxidation of sorbitol, while also being able to oxidize other polyols such as xylitol and ribitol at lower reaction rates. However, the characterization of *V. vinifera* LIDH showed that this enzyme demonstrated the highest reaction rate on L-idonate but had a low reaction rate with sorbitol [48]. Upon docking of L-idonate, we found overall similar hydrogen bonding patterns with sorbitol as those proposed by Pauly et al. [22] and Yennawar et al. [77]. Earlier studies on enzyme substrate specificity also indicated that SDHs preferentially use substrates with a d-cis-2,4-dihydroxyl (2S,4R) configuration [6,13,18,20] (Additional file 1). L-idonate and D-sorbitol have the same molecular configuration from C1 to C4 and differ only at C5 (D and L chirality) and C6 (a hydroxyl group in sorbitol is replaced by a carboxyl group in L-idonic acid) (Additional file 1). Protein modelling analyses showed that L-idonate occupied a comparable position in the active site to sorbitol (Figure 5C). Therefore a similar reaction mechanism for L-idonate oxidation by *V. vinifera* LIDH is possible with D-sorbitol oxidation by human SDH [22]. The hydroxyl groups at C1 and C2 of L-idonate were modelled within interacting distance of the zinc atom in *V. vinifera* LIDH (Additional file 10: C1-C2-Zn.png), which may facilitate the proton transfer from C2 hydroxyl to NAD^+, ultimately resulting in an oxidized C2 with ketone and the production of NADH (Figure 5B). Previous work suggested that the preferential binding of L-idonate over sorbitol seen in *V.vinifera* LIDH may be attributed to amino acid substitution at the catalytic sites between paralogous proteins [48]. As a result, the catalytic site of plant SDHs was investigated based on our models of *V.vinifera* SDH homologs.

Nineteen putative active site residues (36C, 38S, 39D, 42H, 48C, 49A, 51F, 61H, 62E, 110F, 112G, 113S, 147E, 148P, 151V, 268L, 291F, 292R and 293Y; numbering in LIDH(Q1PSI9) without the first 20 amino acids) were identified either coordinating the zinc ion or forming potential non-covalent interactions with NAD(H) and L-idonate. Ten out of the 19 residues were considered strictly conserved throughout all plant SDH forms, and six additional residues are also largely conserved with variations in only a few SDH sequences (Figure 3). These observations revealed a potential structural basis for the preserved function of plant SDHs. Interestingly, three other residues were found to be uniformly exchanged (Y42H, A112G and T113S) between core Eudicot SDH Class I and Class II while monocot and *A. coerulea* SDHs resemble SDH Class I at these amino acid sites (Figure 3). A closer inspection of these residues showed that the oxygen atom of C5 hydroxyl of L-idonate was poised to potentially interact with both 42H and 113S within distances of 4 Å and 2.6 Å respectively (Figure 5C). Additionally, the oxygen atom of the C6 ketone group of L-idonate was within non-covalent interaction distance to 113S (3.5 Å; Figure 5C). Notably, the replacement of 42Y (hydrophobic aromatic side chain) with 42H (charged side chain) in LIDH has the potential to change the hydrophobicity in the substrate-binding pocket (Figure 5D), which may lead to the preferential binding of L-idonate over D-sorbitol. These observations potentially provided a structural explanation for the unique activity of *V. vinifera* LIDH compared to other plant SDHs. Previous studies have indicated that the chiral configuration at C5 is not a determining factor for SDH substrate specificity [18,20], however, our analysis suggested that the C5 hydroxyl group and the C6 ketone group of L-idonate potentially affect substrate binding affinity due to amino acid substitutions at 42H, 112G and 113S in Class II SDHs. A previously identified SDH from apple fruit [9] was found to be the single Class II SDH (MDP0000305455) in *M. domestica* in the present study. This SDH has a much lower affinity for sorbitol (K_m 247 mM [9]) compared to other SDHs purified (K_m 40.3 mM [76], 86.0 mM [7]) or cloned (K_m 83.0 mM [10]; SDH Class I) from apple species. While the kinetic differences were suggested to be due to protein configuration changes between the fusion protein and native protein [9], the present analysis indicated that they might have been be due also to amino acid substitutions at the catalytic site.

From an evolutionary point of view, amino acid changes leading to the shift of enzyme substrate specificity are usually derived from positive Darwinian selection after gene duplication [41,43]. Results from the natural selection analyses in the present study are consistent with this suggestion. The three amino acid sites (42H, 112G and 113S) displaying substitutions between SDH Class I and Class II are all under positive natural selection (Table 3). At the moment, the enzymatic characterization of plant SDH is still fragmentary; no information is available regarding plant SDH activity with L-idonate, except for the activity of *V. vinifera* LIDH [48]. Site mutation and enzymatic studies are currently underway in our laboratory to investigate this hypothesis.

Meta-analysis of sorbitol dehydrogenase related gene expression

In addition to changes in enzyme activity, gene evolution after duplication can also occur at the transcriptional level [42]. Expression division appears to be more common than structural evolution and often occurs rapidly after gene duplication [42,78,79]. To further characterize the evolutionary pattern of plant SDH genes and also to explore the role of SDH related genes during plant development, a survey of transcriptional data was undertaken. Based on the availability of microarray and RNA sequencing data and the presence of both classes of

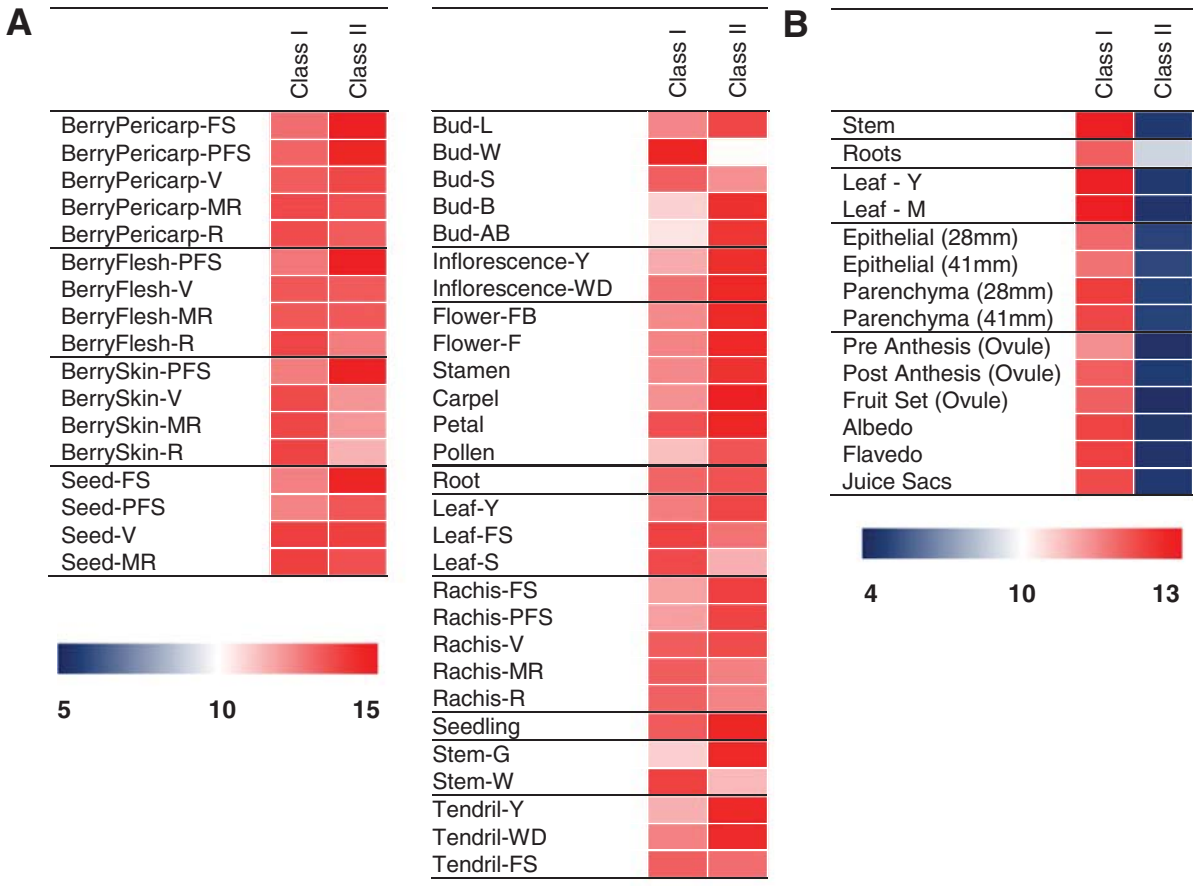
SDH in the genome, grapevine and citrus species were selected. In addition, the expression profile of the single Class I SDH (AT5G51970, Figure 2) in *A. thaliana* was used as a model reference [80]. This gene was highly expressed in cotyledons, leaves and late stages of seed development compared to organs such as flowers (stamen, petal, carpel) and shoots (inflorescence, vegetative, transition), where it was marginally expressed (data not shown).The results support a potential role for SDH Class I during seed germination in *A. thaliana* [23], soybean [37] and maize [8,38]. In grapevines, transcriptional patterns of VIT_16s0100g00290 (SDH Class II, LIDH) and VIT_16s0100g00300 (SDH Class I, SDH) were analysed using the normalised grapevine gene expression atlas of the 'Corvina' cultivar [81]. Notable differences in gene expression intensities and dynamics were observed between SDH Class I and Class II (Figure 6A; Additional file 11: Table S1). The transcript abundance of grapevine SDH Class I was highest in the ripening stages of berries (measured in pericarp, pulp, seeds and skins), resembling the expression profiles reported for Class I SDHs in apple [10,27,29]. In most cases, transcript abundance was lowest in young berry growth stages and increased gradually until harvest in berry tissues. Developmental up-regulation of SDH Class I transcripts in other cultivars such as 'Shiraz' [82] and 'Tempranillo' [83] during berry development under normal conditions was also evident. In addition, the latter work showed sorbitol is present in leaves and berries, and that the biochemical activity of SDH Class I, involving sorbitol oxidation, coincided with SDH class I transcripts levels in these berries during development [83]. Similarly, developmental increases of the grapevine SDH Class I transcript were observed in leaf, rachis, seed and tendrils. Interestingly, gene expression of grapevine SDH Class I was highly induced in winter buds and followed a gradual down-regulation during dormancy release. A similar gene expression and protein activity pattern reported in raspberry [84] and pear [39] respectively may reflect a response to the environment where dormancy periods encompasses dehydration and temperature (cold) stress, although developmental processes could take place concurrently. Taken together, this suggests an active role for SDH Class I in developmental processes through the coordinated regulation of transcript and protein activities in controlling the flux of sorbitol (and related polyols) in grapevines which may be critical in maintaining cell and tissue homeostasis in the mature tissues [83] where oxidative stress is inherent [85,86].

Expression profiles of SDH Class II were well represented in most grapevine organs with the highest expression in berries at fruit-set and in flower carpels. A striking developmental down-regulation of grapevine SDH Class II genes was evident in most grapevine organs, where expression levels in young tissues of berries (pericarp, flesh, skin and seed), buds, leaves, stems and tendrils were high and gradually decreased during development (Figure 6A). We have previously demonstrated in a cross-comparison study involving RNA-seq, microarray and qRT-PCR in young, early veraison, late veraison and ripening berries of grapevine [82] that SDH Class II genes were developmentally down-regulated consistently in all profiling platforms. This distinct expression coincides with the accumulation of TA biosynthesis in young/immature tissues [48,87].

In citrus, SDH Class I and SDH Class II genes were represented by probesets "Cit.9778.1.S1_s_at" and "Cit.9780.1.S1_s_at" respectively. Although gene expression studies encompassing developmental series in citrus are not as comprehensive compared to *A. thaliana* and grapevine, several striking observations could be inferred (Figure 6B; Additional file 11: Table S1). The citrus SDH Class I gene was highly expressed regardless of organ and tissue, including stems, roots, leaves, ovules and fruit tissues (albedo, flavedo, juice sacs), similar to that of grapevine SDH Class I. Interestingly, SDH Class II genes were expressed to a very low level (possibly in fact not at all) in the majority of organs, including fruit tissues, except for the root where expression was highest. It is speculated that this may reflect the trace amount of TA detected in fruits of sweet oranges and other citrus species [63]. Until now, no information, to our knowledge, has been reported on the function of citrus SDHs. Given the novel transcription profiles of one the two citrus Class II SDHs (specifically expressed in root tissues), and the presence of an additional Class II SDH (albeit this sequence was not represented in the array from which these data were analysed), these features may indicate a novel function of SDHs specific to root tissues of sweet oranges and therefore, deserve more attention in future research. In addition to *V. vinifera* and citrus, divergent transcription profiles have also been reported for SDHs from apple [10] and pear [39] where the single copy Class II SDH genes were shown to be under independent transcriptional regulation from other SDH genes. Taken together, divergent expression profiles for SDH Class I and SDH Class II appear to be true to all species where two classes are present, supporting a gene functional divergence at the expression level.

Gene co-expression mining in various plant species

Gene co-expression network analysis (GCA) is based on the principle that genes involved in similar and/or related biological processes may be expressed in a proportional manner, thereby providing a unique tool to understand gene function. Based on information availability, co-expressed gene lists of SDHs from *A. thaliana*,



C

GO	Description	Class I At	Class I Vv	Class I Pt	Class I Cs	Class I Os	Class II Vv	Class II Cs	Class II Pt	F
GO:0050896	response to stimulus									5
GO:0009628	response to abiotic stimulus									4
GO:0042221	response to chemical stimulus									4
GO:0010035	response to inorganic substance									3
GO:0010033	response to organic substance									3
GO:0009266	response to temperature stimulus									3
GO:0016054	organic acid catabolic process									3
GO:0006082	organic acid metabolic process									3
GO:0043436	oxoacid metabolic process									3
GO:0046395	carboxylic acid catabolic process									3
GO:0019752	carboxylic acid metabolic process									3
GO:0044281	small molecule metabolic process									3
GO:0044282	small molecule catabolic process									3
GO:0006573	valine metabolic process									3
GO:0009081	branched-chain amino acid metabolism									3
GO:0006814	sodium ion transport									2
GO:0015672	monovalent inorganic cation transport									2

1 3 5

Figure 6 (See legend on next page.)

(See figure on previous page.)
Figure 6 Transcript and gene co-expression profiles of SDH in different plants. **A**. Expression profiles for Class I and Class II SDH genes in various tissues and developmental stages of *V. vinifera*. Class I and II SDH genes were moderately to highly expressed in most tissues (Log2 intensity > 10; 50th percentile of all gene expression values, see Methods). The heatmap was adjusted to colour ranges between log2 intensity of 5 (blue), 10 (white) and 15 (red) to illustrate low, moderate and high expression when compared to all other genes respectively. **B**. Expression profiles for Class I and Class II SDH gene in citrus. The heatmap was adjusted to colour ranges between log2 intensity of 4 (blue), 10 (white) and 14 (red) to illustrate low, moderate and high expression when compared to all other genes respectively. **C**. Heatmap of selected enriched GO terms (−log10 (adj. *p*-value) for genes co-expressed with SDHs from *A. thaliana* (At), *V. vinifera* (Vv), *C. sinensis* (Cs), *P .trichocarpa* [84], *O. sativa* (Os) and associated frequencies in the plants tested. Light and dark orange denote enrichment scores between 1 and 3 respectively. Highly enriched scores (>5) are coloured in red. Grey colour denotes no significant enrichment.

rice, poplar, grapevine and citrus (Additional file 11: Table S2-S9) were retrieved from publicly available co-expression databases [88-90]. In *A. thaliana*, the SDH Class I homologue (At5g51970) was significantly co-expressed with 67 genes (33% of total genes in the list) involved in branched chain amino acid metabolism, 72 genes (36%) involved in response to various stimuli, 37 genes (19%) involved in protein import in the peroxisome and 17 genes (9%) involved in auxin metabolism (Additional file 11: Table S2). In grapevines, the SDH Class I homologue (VIT_16s0100g00300) was significantly co-expressed with genes involved in abiotic stress (21%), peptide metabolism (13%) and lipid metabolism (13%) (Additional file 11: Table S3; Additional file 12: Table S2–S3). The co-expression results presented here corroborated with recent findings that the importance of SDH Class I lies in regulating sorbitol levels via its biochemical activity and gene expression during various abiotic stresses [83]. More importantly, intracellular accumulation of sorbitol to high levels, accentuated under salt and osmotic stress, significantly reduced stress-induced biomass loss of grapevine berry cell suspensions which were likely the results of the polyol utilisation as an effective osmoprotectant and cellular homeostasis buffer [83]. Similar to its Arabidopsis counterpart (At5g51970), it is therefore likely that grapevine SDH Class I plays an important role in abiotic stress tolerance via the synergistic regulation of polyol transport and metabolism. The SDH Class II homologue (LIDH, VIT_16s0100g00290) was also significantly co-expressed with genes related to abiotic stress response (35%). Other genes related to hexose biosynthetic pathways and carbohydrate metabolism (25%), protein biogenesis and catabolism (8%) and malic acid transport (6%) were also evident in the list of co-expressed genes (Additional file 11: Table S4). GO terms associated with these genes were also enriched within the gene lists (FDR < 0.05). Interestingly, GO enrichment analysis of co-expressed genes showed that terms associated with "malate trans-membrane transport" and "response to abiotic stimulus" were highly enriched (FDR < 1.51E-04 and 3.5E−03 respectively) (Additional file 12: Table S2). Similarly to the grapevine SDH Class I gene, SDH Class II transcription was also stress responsive, being down-regulated during the heat stress recovery of grapevine leaves and up-regulated during exposure to UV-C light irradiation (Additional file 12: Table S3). Based on our coexpression analysis, we speculate that the involvement of Class II SDHs in abiotic stress responses is likely to occur via a separate mechanism from that of sorbitol metabolism, namely the ascorbate-glutathione cycle [91] and specifically in regulating the balance between the biosynthesis of ascorbate by the L-galactose pathway [92] and its catabolism. This is supported in part in grapevines in which a marked down-regulation of SDH Class II (LIDH) protein (impeding TA formation) and the up-regulation of proteins involved in L-galactose pathway (favouring Asc formation) in shoots of grapevines during drought stress were observed [93]. Therefore, the stress responsive nature of SDH Class II gene and enzyme could potentially function as an extra level of control (preventing loss of Asc to TA). The *C. sinensis* SDH Class II gene (Cit.9780.1.S1_at) was significantly co-expressed with genes involved in ion transport (11%), ubiquinone biosynthesis/oxidative phosphorylation (20%) and ribosome biogenesis (9%) (Additional file 11: Table S6). GO terms associated with these genes were highly enriched within the co-expressed gene lists (Additional file 12: Table S5). Unlike Class I SDHs, enriched GO terms associated with Class II SDH co-expressed genes were more specialised to each corresponding plant but shared a common set of co-expressed genes related to transporters (Additional file 11: Table S7; Additional file 12: Table S6). In rice, the top 200 genes co-expressed with SDH (Os08g0545200) were primarily enriched for genes involved in stress response (31%), carboxylic acid biosynthesis (16%), plastid organisation (11%), protein transport (10%) and starch metabolism (5%) (Additional file 11: Table S5; Additional file 12: Table S4).

Enriched GO parent terms such as "response to stimulus" and descendent terms "response to abiotic stimulus", were frequently enriched in SDH Class I co-expressed lists and slightly in SDH Class II containing plant species (Figure 6C; Additional file 12: Table S1-S9). These observations agreed with previous

reports that SDHs (Class I) in *A. thaliana* [13,23] and grapevine [83] play an active role during drought stress and recovery processes and also suggest some shared functions related to stress tolerance between the two classes of SDH, even though to a conservative degree and potentially involving a separate mechanistic route. Therefore, enriched GO parent terms associated with "organic acid metabolic process" and "branched-chain amino acid metabolism" were demonstrated to be more relevant to SDH Class I co-expressed genes but not to SDH Class II (Figure 6C). This is not surprising as response to various stresses involves the coordinated regulation of amino acid and polyol accumulation [94]. On the other hand, co-expression analysis showed that plant SDH Class II could be tightly linked to mechanisms related to transport and compartmentation of cations and solutes (Figure 6C). In membrane transport and compartmentation systems involving pumps, carriers and ion channels are also pivotal for ion homeostasis and equivocally involved in a wide range of stress conditions [95]. In addition, divergent co-expression profiles across species have also been observed for both classes of SDH. In general, monocot rice SDH-related genes have more common co-expression responses with core Eudicot SDH Class I than with SDH Class II, corresponding with the finding that monocot SDH has a closer relationship with core Eudicot SDH Class I than SDH Class II at the enzyme structural level.

Conclusions

SDH is the key enzyme involved in sorbitol metabolism in higher plants. The results of the present study demonstrated that core Eudicot SDHs have evolved into two distinct lineages: SDH Class I and SDH Class II. Class I SDH genes were present in all core Eudicot species investigated in this study and appear to be essential for the normal growth of plants. Class II SDH genes were found to be absent in *Brassicaceae*, *Leguminosae*, most *Asterid*s (except *S. tuberosum*) and some other plants. The previously characterized LIDH involved in TA synthesis in *V. vinifera* has now been identified as a Class II SDH and represents a novel function of SDH genes in *V. vinifera*. The role of LIDH in TA synthesis may be relevant to the function of Class II SDHs in other species. Phylogeny, natural selection and genomic structure analyses supported the emergence of SDH Class II as a result of positive natural selection after tandem duplication, which might occur in the common ancestor of core Eudicot plants. Furthermore, positive natural selection has only acted on specific amino acid sites in the SDH Class II lineage. Protein modelling analyses revealed substitutions of three putative active site residues for Class I and Class II SDHs, which may be responsible for the unique enzyme activity of *V. vinifera* LIDH. Gene expression analysis demonstrated a clear transcriptional divergence between SDH Class I and Class II in several plants and supports the divergence of Class II SDHs at the expression level as well. Future work should be dedicated to uncovering the enzymatic activities and roles of Class II SDH gene products in plant metabolism.

Methods

Identification of sorbitol dehydrogenase homologous genes in higher plants

To identify homologous SDHs in angiosperm plants, the amino acid sequence of *A. thaliana* SDH (accession no. At5g51970) was used as a query to BLAST against the genomes of angiosperm species at Phytozome (http://www.phytozome.net/), with the exception of *M. domestica* for which genome dataset at Plant Genome Duplication Database (PGDD, http://chibba.agtec.uga.edu/duplication/) was used instead. To increase dataset coverage, the genomes of 8 recently sequenced species including *Cajanus cajan*, *Jatropha curcas*, *Capsicum annuum*, *Brassica oleracea*, *Eutrema saisugineum*, *P. mume*, *Hordeum vulgare* and *Aegilops tauschii* were also queried using the corresponding genome databases. BLAST hits with an expectancy value (E value) of zero were selected as SDH homologs were subjected to another round of BLAST searches within the genomes from which they were identified. Only the primary transcript was chosen when alternative transcripts occurred. In addition, five partial SDH protein sequences of *P. bretschneideri* [39] and one SDH sequence of *Eriobotrya japonica* [35] were obtained from literature searches. Homologous SDHs of *P. hortorum* were provided by the *P. hortorum* genome sequencing project author (Prof. Robert K. Jansen, The University of Texas at Austin).

Phylogenetic analysis of sorbitol dehydrogenase

The Uniprot database was queried for previously identified MDR mammal SDHs and yeast SDHs. Only reviewed entries were selected and used as the outgroup in this phylogenetic analysis. Multiple sequence alignments of 102 sequences (92 plant SDHs, 7 mammal SDHs and 3 yeast SDHs) were carried out using ClustalW2 [96]. The evolutionary distances of target SDHs (pairwise p-distance) were estimated using MEGA6 software [97]. The Neighbour Joining tree was inferred by MEGA6 software [97] using the p-distance [98] substitution model, the certainty at each node was assessed by the Interior-branch Test method (1000 times iteration). Maximum likelihood trees were estimated by MEGA6 software [97] using the JTT+GAMMA substitution model [99], the best fitting model as determined by the "Find Best DNA/Protein Models" function in MEGA6. Bootstrap supports for Maximum likelihood trees were calculated from 1000 replicates. For both Neighbour Joining and Maximum

likelihood methods, the Gaps/Missing Data Treatment parameter was set as Complete-Deletion to eliminate the effects of gaps and insertions. The developed phylogenetic trees were rooted on the yeast SDHs and annotated using the FigTree version 1.4.2 software (http://tree.bio.ed.ac.uk/software/figtree/).

Sequence alignment and protein subdomain analysis

Preliminary sequence identity of SDHs was obtained by local all-vs-all BLAST using NCBI-BLAST-2.2.29 tool [100] downloaded from ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/. The BLAST results were sorted according to respective phylogeny groups. Average pair-wise sequence identities were calculated using Microsoft Excel software based on the BLAST results. Protein functional domains were predicted using InterPro (http://www.ebi.ac.uk/interpro/). Secondary structure analysis was implemented with ESPript3.0 tool (http://espript.ibcp.fr/ESPript/ESPript/) using human SDH (PDB: 1PL8) as a template. All residue numberings in the present study are according to LIDH (Q1PSI9) without the first 20 amino acids (unless otherwise declared) which was predicted to be a mitochondria-targeting signal sequence (data not shown; alignment corresponding to this region was highly divergent).

Gene duplication pattern characterization and synteny analysis

The MCScanX package [101] from http://chibba.pgml.uga.edu/mcscan2/ was employed to investigate gene duplication patterns of plant SDHs. In order to elaborate on the origin of the core Eudicot Class II SDHs, plant genomes containing SDHs from both Class I and Class II were selected. These were further refined to genomes for which predicted genes have been mapped into corresponding chromosome locations. *A.thaliana* was included as a reference for inter-species collinear block analysis. Amino acid sequence files and gene position files were downloaded either from PGDD or from Phytozome databases and were further modified to suit the requirements of the MCScanX software. BLAST tool NCBI-BLAST-2.2.29 [100] was used for intra and inter species genome comparisons. The E-value threshold was set at 10^{-5} for all analyses. For gene duplication pattern identification, self-genome all-vs-all BLAST was performed. The *duplicate_gene_class ifier* program from the MCScanX package was applied to each dataset. For collinear SDH gene pair identification, amino acid sequences and genetic position information of chromosomes containing SDHs were extracted from each species, then combined to perform the multi-species MCScanX analysis. The SDH gene family file was created manually by including all the SDHs identified from the selected species. The *family_circle_plotter.java* tool at MCScanX package was used to display the results.

Natural selection analysis

Natural selective pressure on plant SDH was examined by measuring the ratio of non-synonymous to synonymous substitutions (dN/dS=w). Codon-based maximum-likelihood estimates of w was performed using codeml in PAML4.7 [73]. Multiple-alignment of conserved domain sequences (CDS) for those identified plant SDHs was carried out using ClustalW2 [96]. Significant insertions and gaps were removed manually. To facilitate the input data requirements of codeml, an additional Maximum Likelihood tree was constructed using a smaller dataset where SDHs with no CDS sequence available were removed. The sub-tree covering the plant SDHs was used in codeml. Branch pattern specification was implemented using Treeview1.6.6 (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). Four target clades were specified based on the present phylogenetic analysis: monocot SDH, *A. coerulea* SDH, core Eudicot SDH Class I and core Eudicot SDH Class II. The w values for these clades were represented as w[mono], w[Aer], w[sdhC1] and w[sdhC2] respectively. Nested likelihood ratio tests(LRTs) were performed to assess the significance of the model under different hypothesises: (w[mono]≠w[Aer]≠w[sdhC1]=w[sdhC2], w[mono]=w[Aer]≠w[sdhC1]≠w[sdhC2], w[mono]≠w[Aer]≠w[sdhC1]≠w[sdhC2], w[mono]=w[Aer]=w[sdhC1]≠w[sdhC2], w[mono]=w[Aer]=w[sdhC1]≠w[sdhC2] with w[sdhC2]=1). The corresponding p values were calculated using the online tool at http://graphpad.com/quickcalcs/PValue1.cfm. In the Site-specific model M1, two site classes were specified: highly conserved sites (w0) and neutral sites (w1=1). For the Site-specific model M2, there were three site classes: highly conserved sites (w0), neutral sites (w1=1) and positively selected sites (w2). For w assessments with the Branch-site models, core Eudicot SDH Class II was specified as the foreground group. In the Branch-site model A, four site classes were specified. The first two classes have w ratios of w0 and w1 respectively, corresponding to highly conserved sites and neutral sites across all lineages. In the other two site classes, the background lineages have w0 or w1 while the foreground lineages have w2.

Ancestral sequence reconstruction and evolution rate analyses

The ancestral sequence (amino acid) reconstruction for the internal nodes of the obtained plant SDH phylogeny was carried out using codeml in PAML4.7 [73]. The Empirical_Frequency model, which allowed the estimates of the stationary frequencies based on user dataset, was performed on the plant SDHs. Ancestral amino

acid sequences for nodes representing monocot SDH, *A. coerulea* SDH, core Eudicot SDH Class I and core Eudicot SDH Class II were used for Tajima's RRT analysis [74] using MEGA6.0 software [97].

Protein structure modelling analysis

SDH homology modelling was carried out using ICM Pro (Molsoft LLC, La Jolla, CA, USA). Models of *V. vinifera* LIDH (Uniprot ID: Q1PSI9; accession no: GSVIVT01010646001) and *V. vinifera* SDH (Uniprot ID: D7TMY3; accession no: GSVIVT01010642001) structures were generated with the human SDH (PDB:1PL8) as a template. Given that no plant SDH structures exist in the protein data bank we chose the model with the highest identity as performed within the Molsoft software package. Ligands including the zinc atom, NAD^+, D-sorbitol and L-idonate were docked into the models using the Molsoft Monte Carlo method [102]. Residues within 5 Å to the ligands were inspected for enzyme-ligand interaction potential. All molecular visualizations were obtained using the PyMOL graphic tool (The PyMOL molecular graphics system, Version 1.3r1. Schrodinger, LLC). The deduced reaction mechanism of *V. vinifera* LIDH on the oxidation of L-idonate was created using the Marvin online tool (http://www.chemaxon.com/marvin/sketch/index.php). Protein hydrophobicity profiles were implemented in PyMOL using the Color_h script (http://www.pymolwiki.org/index.php/Color_h), based on the hydrophobicity scale defined at http://us.expasy.org/tools/pscale/Hphob.Eisenberg.html. All residue numberings are according to LIDH (Q1PSI9) without the first 20 amino acids.

Meta-analysis of developmental gene expression

Identification of corresponding probesets in the microarray platforms of *A. thaliana*, rice, poplar, grapevine and citrus were performed using the BLAST software (NCBI-BLAST-2.2.29+) [100], and grapevine Class I (VIT_16s0100g00290) and Class II (VIT_16s0100g00290) SDH sequences with default settings. The top hits for each corresponding probeset in the microarray platform of each species were selected for downstream analysis (Additional file 11). Normalised gene expression atlases encompassing transcriptional data during growth and development of *A. thaliana*, grapevine and citrus were retrieved from the Botany Array Resource (BAR) [80], *Vitis* co-expression database (VTCdb) [88] and Network inference of citrus co-expression (NiCCE) [89] webservers, respectively. Only experimental conditions relating to tissue/organ development and probesets intensities (normalised) corresponding to Class I and Class II SDHs were retained. Normalised log2 intensities were deemed highly, well and lowly/not expressed when the intensities of total background distribution > 95th, at the 50th and < 20th percentile respectively.

Gene co-expression mining in various plant species

Information on co-expressed genes with Class I and Class II SDHs in plants such as *A. thaliana*, poplar and rice (version 7.1) [90], grapevine (version 2.1) [88] and citrus [89] were retrieved from the various plant gene co-expression webservers. The top 200 co-expressed genes (unless otherwise specified) for each SDH class in each species were empirically chosen as a cut-off for significant co-expression, and to provide comparisons of enriched gene ontology (GO) terms within the co-expressed gene lists from each species. Enrichment of GO terms (i.e. biological processes, BP; molecular function, MF; cellular component, CC) were evaluated by hypergeometric distribution, adjusted by false discovery rate (FDR) for multiple hypothesis correction and using the 'gProfileR' package [103] in R (http://www.r-project.org) which interfaces g:profiler webserver (http://gprofiler.at.mt.ut.ee/gprofiler/). The 'ordered query' option was enabled to perform incremental enrichment analysis, which prioritises highly co-expressed genes and results in better functional GO term associations. GO terms were considered to be significantly enriched when FDR < 0.05 and > 2 genes were annotated with the same GO term. Enriched GO terms from the SDH co-expressed gene lists across tested plants (*A. thaliana*, poplar, rice, grapevine and citrus), were considered 'commonly occurring' when more than 3 counts were present for each enriched GO term.

Availability of supporting data

All relevant supporting data can be found within the additional files accompanying this article. Phylogenetic data supporting the results of this article are available in the TreeBASE repository at http://purl.org/phylo/treebase/phylows/study/TB2:S17300.

Additional files

Additional file 1: Displays the molecular structures of SDH substrates.

Additional file 2: Table S1. Contains SDH gene IDs from corresponding species and organisms. **Table S2.** Contains pairwise p-distance values of SDH sequences. **Table S3.** Contains information on sequence renaming. **Table S4.** Contains the all-vs-all BLAST results of SDH amino acid sequences. Table S5 contain the identified collinear SDH gene pairs.

Additional file 3: Contains the original amino acid sequences of the identified plant, mammal and yeast SDHs.

Additional file 4: Displays the complete Neighbour Joining tree for Figure 2A.

Additional file 5: Displays complete sequence alignment for Figure 3.

Additional file 6: Contains gene duplication pattern information. Tables "cs", "eg", "md", "pm", "pp", "pt", "st", "tc", "vv" refer to *C. sinensis*, *E. grandis*, *M. domestica*, *P. mume*, *P. persica*, *P. trichocarpa*, *S. tuberosum*, *T. cacao* and *V. vinifera* respectively.

Additional file 7: Contains input and output data for natural selection modelling analyses. "-output" files are codeml outputs and are recommended to be viewed using Microsoft WordPad. ".phy" is phylogenetic tree file and can be viewed using Treeview software. ".ctl" is a control file and can be viewed using any text viewer. "sdh-pep2.fas" sequence file was produced from Additional file 3 by manually removing the significant gaps, insertions; sequences with no CDS sequence available were also removed. "sdh-cds2.fas" is the corresponding CDS sequences for "sdh-pep2.fas". "sdh-pep2.nwk" is the phylogenetic tree produced from "sdh-cds2.fas" and can be viewed using any phylogenetic tree viewer software. Sequence IDs are represented by numbers for software input convenience (see Additional file 2: Table S3 for sequence ID renaming information). Amino acid site numbering is according to LIDH (Uniprot No: Q1PSI9) without the first 20 amino acids.

Additional file 8: Contains input and output data for the reconstruction of ancestral SDH sequences. "sdh-pep.fas" contains amino acid sequences for the plant SDH sub-branch. The "ancestral-sequence-construction_output" file is codeml output and can be viewed using any text viewer. Ancestral sequences for corresponding branches were extracted and put in the "interpreted-ancestral-sequence.fas" file for readers' convenience.

Additional file 9: Contains the Tajima's RRT test outputs.

Additional file 10: Contains the modelled structures files of Vv_LIDH and Vv_SDH and additional illustration figures. "Asp195_NAD.png" displays the interaction of Asp195 with the hydroxyl groups at C1 and C2 of L-idonate. "LIDH-hydrophobicity.png" and "SDH-hydrophobicity.png" display the overall hydrophobicity profiles of Vv_LIDH and Vv_SDH respectively. Amino acid site numbering is according to LIDH (Uniprot No: Q1PSI9) without the first 20 amino acids.

Additional file 11: Contains a Microsoft Excel spread sheet with detailed results of transcript and gene co-expression analysis of Class I and Class II SDH in plants. Table S1 contains gene expression profile of Class I and Class II SDH profile in various tissues of (A) grapevine and (B) sweet oranges. Table S2 – S9 contains lists of all significantly co-expressed genes and respective rank, function description, and co-expression metric with class I and II SDH in *A. thaliana* (Table S2), grapevine (Table S3 and S4), rice (Table S5), sweet orange (Table S6 and S7) and poplar (Table S8 and S9).

Additional file 12: Contains a Microsoft Excel spread sheet with detailed results of functional (GO) enrichment analysis of significantly co-expressesed genes of class I and II SDH in plants. Table S1 – S8 contains outputs of GO enrichment analysis containing enriched GO ID, description, adjusted p-value, and lists of genes having the enriched GO term for A. thaliana (Table S1), grapevine (Table S2 and S3), rice (Table S4), sweet orange (Table S5 and S6) and poplar (Table S7 and S8). Table S9 contains a summary of common enriched GO ID/term identified among the co-expressed genes with SDHs in the aforementioned plants tested.

Abbreviations

SDH: Sorbitol dehydrogenase; LIDH: L-idonate-5-dehydrogenase; TA: Tartaric acid; MDR: Medium-chain dehydrogenase/reductase; ADH: Alcohol dehydrogenase; 5KGA: 5-keto-D-gluconate; Mya: Million years ago; WGD: Whole genome duplication; RRT: Relative rate tests; GCA: Gene co-expression network analysis; PGDD: Plant Genome Duplication Database; CDS: Conserved domain sequences; LRTs: Likelihood ratio tests; BAR: Botany Array Resource; VTCdb: *Vitis* co-expression database; NiCCE: Network inference of citrus co-expression; GO: Gene ontology; BP: Biological processes; Asc: Ascorbate; MF: Molecular function; CC: Cellular component; FDR: False discovery rate.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YJ conceived the research. YJ and DCJW did sequence retrieval, curation and gene duplication characterization. YJ performed phylogenetic, synteny, natural selection modeling and ancestral sequence analyses and drafted the manuscript. JBB and YJ carried out protein modeling analyses. DCJW performed the transcript expression and gene co-expression analysis. CS and DCJW assisted with the drafting of the manuscript. CMF and JBB supervised the project. All authors have read and approved the final manuscript.

Acknowledgements

We acknowledge the related research groups for making the genomic information and microarray data available to the public. We are very grateful to Dr Anthony Borneman and Dr Julian Schwerdt for their valuable suggestions about the phylogenetic analyses. We thank the anonymous referees for their constructive comments and suggestions. This work was part-supported by Australia's grape growers and winemakers through the Grape and Wine Research and Development Corporation with matching funds from the Australian Government (project UA 10/01). YJ is supported by a postgraduate scholarship from China Scholarship Council.

Author details

[1]School of Agriculture, Food and Wine, University of Adelaide, Adelaide 5005, Australia. [2]Present address: Wine Research Center, Faculty of Land and Food Systems, University of British Columbia, Vancouver V6T 1Z4BC, Canada. [3]Present address: School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide 5001, Australia. [4]School of Biological Sciences, University of Adelaide, Adelaide 5005, Australia.

Received: 14 November 2014 Accepted: 23 March 2015
Published online: 12 April 2015

References

1. Iwata T, Hoog JO, Reddy VN, Carper D. Cloning of the human Sorbitol Dehydrogenase gene. Invest Ophth Vis Sci. 1993;34(4):712–2.
2. Karlsson C, Jornvall H, Hoog JO. Sorbitol Dehydrogenase - cDNA coding for the rat enzyme - variations within the Alcohol-Dehydrogenase family independent of quaternary structure and metal content. Eur J Biochem. 1991;198(3):761–5.
3. Wang T, Hou M, Zhao N, Chen Y, Lv Y, Li Z, et al. Cloning and expression of the sorbitol dehydrogenase gene during embryonic development and temperature stress in *Artemia sinica*. Gene. 2013;521(2):296–302.
4. Niimi T, Yamashita O, Yaginuma T. A cold-inducible Bombyx gene encoding a protein similar to mammalian Sorbitol Dehydrogenase - yolk nuclei-dependent gene-expression in diapause eggs. Eur J Biochem. 1993;213(3):1125–31.
5. Sarthy AV, Schopp C, Idler KB. Cloning and sequence determination of the gene encoding Sorbitol Dehydrogenase from *Saccharomyces cerevisiae*. Gene. 1994;140(1):121–6.
6. Ng K, Ye RQ, Wu XC, Wong SL. Sorbitol Dehydrogenase from *Bacillus subtilis* - purification, characterization, and gene cloning. J Biol Chem. 1992;267(35):24989–94.
7. Negm FB, Loescher WH. Detection and characterization of Sorbitol Dehydrogenase from apple callus-tissue. Plant Physiol. 1979;64(1):69–73.
8. Doehlert DC. Ketose reductase-activity in developing maize endosperm. Plant Physiol. 1987;84(3):830–4.
9. Yamada K, Oura Y, Mori H, Yamaki S. Cloning of NAD-dependent sorbitol dehydrogenase from apple fruit and gene expression. Plant Cell Physiol. 1998;39(12):1375–9.
10. Park SW, Song KJ, Kim MY, Hwang JH, Shin YU, Kim WC, et al. Molecular cloning and characterization of four cDNAs encoding the isoforms of NAD-dependent sorbitol dehydrogenase from the Fuji apple. Plant Sci. 2002;162(4):513–9.
11. Ohta K, Moriguchi R, Kanahama K, Yamaki S, Kanayama Y. Molecular evidence of sorbitol dehydrogenase in tomato, a non-Rosaceae plant. Phytochemistry. 2005;66(24):2822–8.
12. Sutsawat D, Yamada K, Shiratake K, Kanayama Y, Yamaki S. Properties of sorbitol dehydrogenase in strawberry fruit and enhancement of the activity by fructose and auxin. J Jpn Soc Hortic Sci. 2008;77(3):318–23.
13. Aquayo MF, Ampuero D, Mandujano P, Parada R, Muñoz R, Gallart M, et al. Sorbitol dehydrogenase is a cytosolic protein required for sorbitol metabolism in *Arabidopsis thaliana*. Plant Sci. 2013;205–206(1):63–75.
14. Persson B, Zigler JS, Jornvall H. A super-family of medium-chain dehydrogenases/reductases (MDR) - Sub-lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductases, enoyl reductases, Vat-1 and other proteins. Eur J Biochem. 1994;226(1):15–22.

15. Persson B, Hedlund J, Jornvall H. The MDR superfamily. Cell Mol Life Sci. 2008;65(24):3879–94.
16. Jornvall H, Persson M, Jeffery J. Alcohol and polyol dehydrogenases are both divided into 2 protein types, and structural-properties cross-relate the different enzyme-activities within each type. P Natl Acad Sci-Biol. 1981;78(7):4226–30.
17. Nordling E, Jornvall H, Persson B. Medium-chain dehydrogenases/reductases (MDR) - family characterizations including genome comparisons and active site modelling. Eur J Biochem. 2002;269(17):4267–76.
18. Lindstad RI, Koll P, McKinley-McKee JS. Substrate specificity of sheep liver sorbitol dehydrogenase. J Biochemical. 1998;330(Pt 1):479–87.
19. Lindstad RI, Hermansen LF, McKinley-Mckee JS. The kinetic mechanism of sheep liver sorbitol dehydrogenase. Eur J Biochem. 1992;210(2):641–7.
20. Oura Y, Yamada K, Shiratake K, Yamaki S. Purification and characterization of a NAD(+)-dependent sorbitol dehydrogenase from Japanese pear fruit. Phytochemistry. 2000;54(6):567–72.
21. Guo ZX, Pan TF, Li KT, Zhong FL, Lin L, Pan DM, et al. Cloning of NAD-SDH cDNA from plum fruit and its expression and characterization. Plant Physiol Biochem. 2012;57:175–80.
22. Pauly TA, Ekstrom JL, Beebe DA, Chrunyk B, Cunningham D, Griffor M, et al. X-ray crystallographic and kinetic studies of human sorbitol dehydrogenase. Structure. 2003;11(9):1071–85.
23. Nosarzewski M, Downie AB, Wu B, Archbold DD. The role of sorbitol dehydrogenase in *Arabidopsis thaliana*. Funct Plant Biol. 2012;39(6):462–70.
24. Yancey PH, Clark ME, Hand SC, Bowlus RD, Somero GN. Living with water-stress - evolution of osmolyte systems. Science. 1982;217(4566):1214–22.
25. Loescher WH. Physiology and metabolism of sugar alcohols in higher-plants. Physiol Plantarum. 1987;70(3):553–7.
26. Loescher WH, Marlow GC, Kennedy RA. Sorbitol metabolism and sink-source interconversions in developing apple leaves. Plant Physiol. 1982;70(2):335–9.
27. Nosarszewski M, Clements AM, Downie AB, Archbold DD. Sorbitol dehydrogenase expression and activity during apple fruit set and early development. Physiol Plantarum. 2004;121(3):391–8.
28. Nosarzewski M, Archbold DD. Tissue-specific expression of sorbitol dehydrogenase in apple fruit during early development. J Exp Bot. 2007;58(7):1863–72.
29. Wang XL, Xu YH, Peng CC, Fan RC, Gao XQ. Ubiquitous distribution and different subcellular localization of sorbitol dehydrogenase in fruit and leaf of apple. J Exp Bot. 2009;60(3):1025–34.
30. Yamaguchi H, Kanayama Y, Soejima J, Yamaki S. Changes in the amounts of the NAD-dependent sorbitol dehydrogenase and its involvement in the development of apple fruit. J Am Soc Hortic Sci. 1996;121(5):848–52.
31. Wu BH, Li SH, Nosarzewski M, Archbold DD. Sorbitol dehydrogenase gene expression and enzyme activity in apple: tissue specificity during bud development and response to rootstock vigor and growth manipulation. J Am Soc Hortic Sci. 2010;135(4):379–87.
32. Iida M, Bantog NA, Yamada K, Shiratake K, Yamaki S. Sorbitol- and other sugar-induced expressions of the NAD+–dependent sorbitol dehydrogenase gene in Japanese pear fruit. J Am Soc Hortic Sci. 2004;129(6):870–5.
33. Kim HY, Ahn JC, Choi JH, Hwang B, Choi DW. Expression and cloning of the full-length cDNA for sorbitol-6-phosphate dehydrogenase and NAD-dependent sorbitol dehydrogenase from pear (*Pyrus pyrifolia N.*). Sci Hortic. 2007;11(4):406–12.
34. Bantog NA, Shiratake K, Yamaki S. Changes in sugar content and sorbitol- and sucrose-related enzyme activities during development of loquat (*Eriobotrya japonica Lindl. cv. Mogi*) fruit. J Jpn Soc Hortic Sci. 1999;68(5):942–8.
35. Bantog NA, Yamada K, Niwa N, Shiratake K, Yamaki S. Gene expression of NAD(+)-dependent sorbitol dehydrogenase and NADP(+)-dependent sorbitol-6-phosphate dehydrogenase during development of loquat (*Eriobotrya japonica Lindl.*) fruit. J Jpn Soc Hortic Sci. 2000;69(3):231–6.
36. Beruter J. Sugar accumulation and changes in the activities of related enzymes during development of the apple fruit. J Plant Physiol. 1985;121(4):331–41.
37. Kuo TM, Doehlert DC, Crawford CG. Sugar metabolism in germinating soybean seeds - evidence for the sorbitol pathway in soybean axes. Plant Physiol. 1990;93(4):1514–20.
38. de Sousa SM, Paniago MD, Arruda P, Yunes JA. Sugar levels modulate sorbitol dehydrogenase expression in maize. Plant Mol Biol. 2008;68(3):203–13.
39. Ito A, Hayama H, Kashimura Y. Partial cloning and expression analysis of genes encoding NAD(+)-dependent sorbitol dehydrogenase in pear bud during flower bud formation. Sci Hortic. 2005;103(4):413–20.
40. Hartman MD, Figueroa CM, Piattoni CV, Iglesias AA. Glucitol Dehydrogenase from peach (*Prunus persica*) fruits is regulated by thioredoxin h. Plant Cell Physiol. 2014;55(6):1157–68.
41. Flagel LE, Wendel JF. Gene duplication and evolutionary novelty in plants. New Phytol. 2009;183(3):557–64.
42. Zhang JZ. Evolution by gene duplication: an update. Trends Ecol Evol. 2003;18(6):292–8.
43. Hughes AL. The evolution of functionally novel proteins after gene duplication. P Roy Soc B-Biol Sci. 1994;256(1346):119–24.
44. Hughes AL. Adaptive evolution after gene duplication. Trends Genet. 2002;18(9):433–4.
45. Hurles M. Gene duplication: the genomic trade in spare parts. Plos Biol. 2004;2(7):900–4.
46. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999;151(4):1531–45.
47. Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet. 2008;9(12):938–50.
48. DeBolt S, Cook DR, Ford CM. L-Tartaric acid synthesis from vitamin C in higher plants. P Natl Acad Sci USA. 2006;103(14):5608–13.
49. Strommer J. The plant ADH gene family. Plant J. 2011;66(1):128–42.
50. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, et al. The genome of the domesticated apple (*Malus x domestica Borkh.*). Nat Genet. 2010;42(10):833–9.
51. Forney CF, Breen PJ. Growth of strawberry fruit and sugar uptake of fruit disks at different inflorescence positions. Sci Hortic. 1985;27(1–2):55–62.
52. Veitia RA, Bottani S, Birchler JA. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. Trends Genet. 2008;24(8):390–7.
53. Galdon BR, Mesa DR, Rodriguez EMR, Romero CD. Influence of the cultivar on the organic acid and sugar composition of potatoes. J Sci Food Agric. 2010;90(13):2301–9.
54. Stafford HA. Distribution of tartaric acid in the leaves of certain angiosperms. Am J Bot. 1959;46(5):347–52.
55. Kramer EM. *Aquilegia*: a new model for plant development, ecology, and evolution. Annu Rev Plant Biol. 2009;60:261–77.
56. Worberg A, Quandt D, Barniske AM, Lohne C, Hilu KW, Borsch T. Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. Org Divers Evol. 2007;7(1):55–77.
57. Hoot SB, Magallon S, Crane PR. Phylogeny of basal eudicots based on three molecular data sets: atpB, rbcL, and 18S nuclear ribosomal DNA sequences. Ann Mo Bot Gard. 1999;86(1):1–32.
58. Moore MJ, Bell CD, Soltis PS, Soltis DE. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. P Natl Acad Sci USA. 2007;104(49):19363–8.
59. Wang HC, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, et al. Rosid radiation and the rapid rise of angiosperm-dominated forests. P Natl Acad Sci USA. 2009;106(10):3853–8.
60. Bremer B, Bremer K, Chase MW, Fay MF, Reveal JL, Soltis DE, et al. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc. 2009;161(2):105–21.
61. Roulin A, Auer PL, Libault M, Schlueter J, Farmer A, May G, et al. The fate of duplicated genes in a polyploid plant genome. Plant J. 2013;73(1):143–53.
62. Saito K, Loewus FA. Formation of tartaric acid in Vitaceous plants - relative contributions of L-ascorbic acid-inclusive and acid-noninclusive pathways. Plant Cell Physiol. 1989;30(6):905–10.
63. Nour V, Trandafir I, Ionica ME. HPLC organic acid analysis in different citrus juices under reversed phase conditions. Not Bot Horti Agrobo. 2010;38(1):44–8.
64. Hudina M, Stampar F. Sugars and organic acids contents of European (*Pyrus communis L.*) and Asian (*Pyrus serotina Rehd.*) pear cultivars. Acta Aliment Hung. 2000;29(3):217–30.
65. Sha SF, Li JC, Wu J, Zhang SL. Characteristics of organic acids in the fruit of different pear species. Afr J Agr Res. 2011;6(10):2403–10.
66. Fuleki T, Pelayo E, Palabay RB. Carboxylic-acid composition of varietal juices produced from fresh and stored apples. J Agr Food Chem. 1995;43(3):598–607.
67. Suarez MH, Rodriguez ER, Romero CD. Analysis of organic acid content in cultivars of tomato harvested in Tenerife. Eur Food Res Technol. 2008;226(3):423–35.
68. Ina Y. Pattern of synonymous and nonsynonymous substitutions: an indicator of mechanisms of molecular evolution. J Genet. 1996;75(1):91–115.

69. Kimura M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Letters to Nature. 1977;267(5608):275–6.
70. Yang ZH. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 1997;13(5):555–6.
71. Worth CL, Gong S, Blundell TL. Structural and functional constraints in the evolution of protein families. Nat Rev Mol Cell Bio. 2009;10(10):709–20.
72. Yang ZH, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol. 2002;19(6):908–17.
73. Yang ZH. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
74. Tajima F. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics. 1993;135(2):599–607.
75. Baker PJ, Britton KL, Rice DW, Rob A, Stillman TJ. Structural consequences of sequence patterns in the fingerprint region of the nucleotide binding fold - implications for nucleotide specificity. J Mol Biol. 1992;228(2):662–71.
76. Yamaguchi H, Kanayama Y, Yamaki S. Purification and properties of NAD-dependent Sorbitol Dehydrogenase from apple fruit. Plant Cell Physiol. 1994;35(6):887–92.
77. Yennawar H, Moller M, Gillilan R, Yennawar N. X-ray crystal structure and small-angle X-ray scattering of sheep liver sorbitol dehydrogenase. Acta Crystallogr D. 2011;67(Pt5):440–6.
78. Gu ZL, Nicolae D, Lu HHS, Li WH. Rapid divergence in expression between duplicate genes inferred from microarray data. Trends Genet. 2002;18(12):609–13.
79. Wagner A. Decoupled evolution of coding region and mRNA expression patterns after gene duplication: implications for the neutralist-selectionist debate. P Natl Acad Sci USA. 2000;97(12):6579–84.
80. Toufighi K, Brady SM, Austin R, Ly E, Provart NJ. The botany array resource: e-northerns, expression angling, and promoter analyses. Plant J. 2005;43(1):153–63.
81. Fasoli M, Dal Santo S, Zenoni S, Tornielli GB, Farina L, Zamboni A, et al. The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. Plant Cell. 2012;24(9):3489–505.
82. Sweetman C, Wong DCJ, Ford CM, Drew DP. Transcriptome analysis at four developmental stages of grape berry (*Vitis vinifera cv. Shiraz*) provides insights into regulated and coordinated gene expression. BMC Genomics. 2012;13(1):691.
83. Conde A, Regalado A, Rodrigues D, Costa JM, Blumwald E, Chaves MM, et al. Polyols in grape berry: transport and metabolic adjustments as a physiological strategy for water-deficit stress tolerance in grapevine. J Exp Bot. 2015;66(3):889–906.
84. Mazzitelli L, Hancock RD, Haupt S, Walker PG, Pont SDA, McNicol J, et al. Co-ordinated gene expression during phases of dormancy release in raspberry (*Rubus idaeus L.*) buds. J Exp Bot. 2007;58(5):1035–45.
85. Pilati S, Perazzolli M, Malossini A, Cestaro A, Dematte L, Fontana P, et al. Genome-wide transcriptional analysis of grapevine berry ripening reveals a set of genes similarly modulated during three seasons and the occurrence of an oxidative burst at veraison. BMC Genomics. 2007;8(1):428.
86. Fortes AM, Agudelo-Romero P, Silva MS, Ali K, Sousa L, Maltese F, et al. Transcript and metabolite analysis in *Trincadeira* cultivar reveals novel information regarding the dynamics of grape ripening. BMC Plant Biol. 2011;11(1):149.
87. Melino VJ, Soole KL, Ford CM. A method for determination of fruit-derived ascorbic, tartaric, oxalic and malic acids, and its application to the study of ascorbic acid catabolism in grapevines. Aust J Grape Wine Res. 2009;15(3):293–302.
88. Wong DCJ, Sweetman C, Drew DP, Ford CM. VTCdb: a gene co-expression database for the crop species *Vitis vinifera* (grapevine). BMC Genomics. 2013;14(1):17.
89. Wong DCJ, Sweetman C, Ford CM. Annotation of gene function in citrus using gene expression information and co-expression networks. BMC Plant Biol. 2014;14(1):17.
90. Obayashi T, Okamura Y, Ito S, Tadaka S, Aoki Y, Shirota M, et al. ATTED-II in 2014: evaluation of gene coexpression in agriculturally important plants. Plant Cell Physiol. 2014;55(1):e6.
91. Foyer CH, Noctor G. Ascorbate and glutathione: the heart of the redox hub. Plant Physiol. 2011;155(1):2–18.
92. Wheeler GL, Jones MA, Smirnoff N. The biosynthetic pathway of vitamin C in higher plants. Nature. 1998;393(6683):365–9.
93. Cramer GR, Van Sluyter SC, Hopper DW, Pascovici D, Keighley T, Haynes PA. Proteomic analysis indicates massive changes in metabolism prior to the inhibition of growth and photosynthesis of grapevine (*Vitis vinifera L.*) in response to water deficit. BMC Plant Biol. 2013;13(1):49.
94. Krasensky J, Jonak C. Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. J Exp Bot. 2012;63(4):1593–608.
95. Conde A, Chaves MM, Geros H. Membrane transport, sensing and signaling in plant adaptation to environmental stress. Plant Cell Physiol. 2011;52(9):1583–602.
96. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23(21):2947–8.
97. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis Version 6.0. Mol Biol Evol. 2013;30(12):2725–9.
98. Nei M, Kumar S. Molecular Evolution and Phylogenetics. New York: Oxford University Press; 2000.
99. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 1992;8(3):275–82.
100. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
101. Wang YP, Tang HB, DeBarry JD, Tan X, Li JP, Wang XY, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.
102. Abagyan R, Totrov M. Biased probability Monte-Carlo conformational searches and electrostatic calculations for peptides and proteins. J Mol Biol. 1994;235(3):983–1002.
103. Reimand J, Arak T, Vilo J. g: profiler-a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res. 2011;39(suppl2):W307–15.